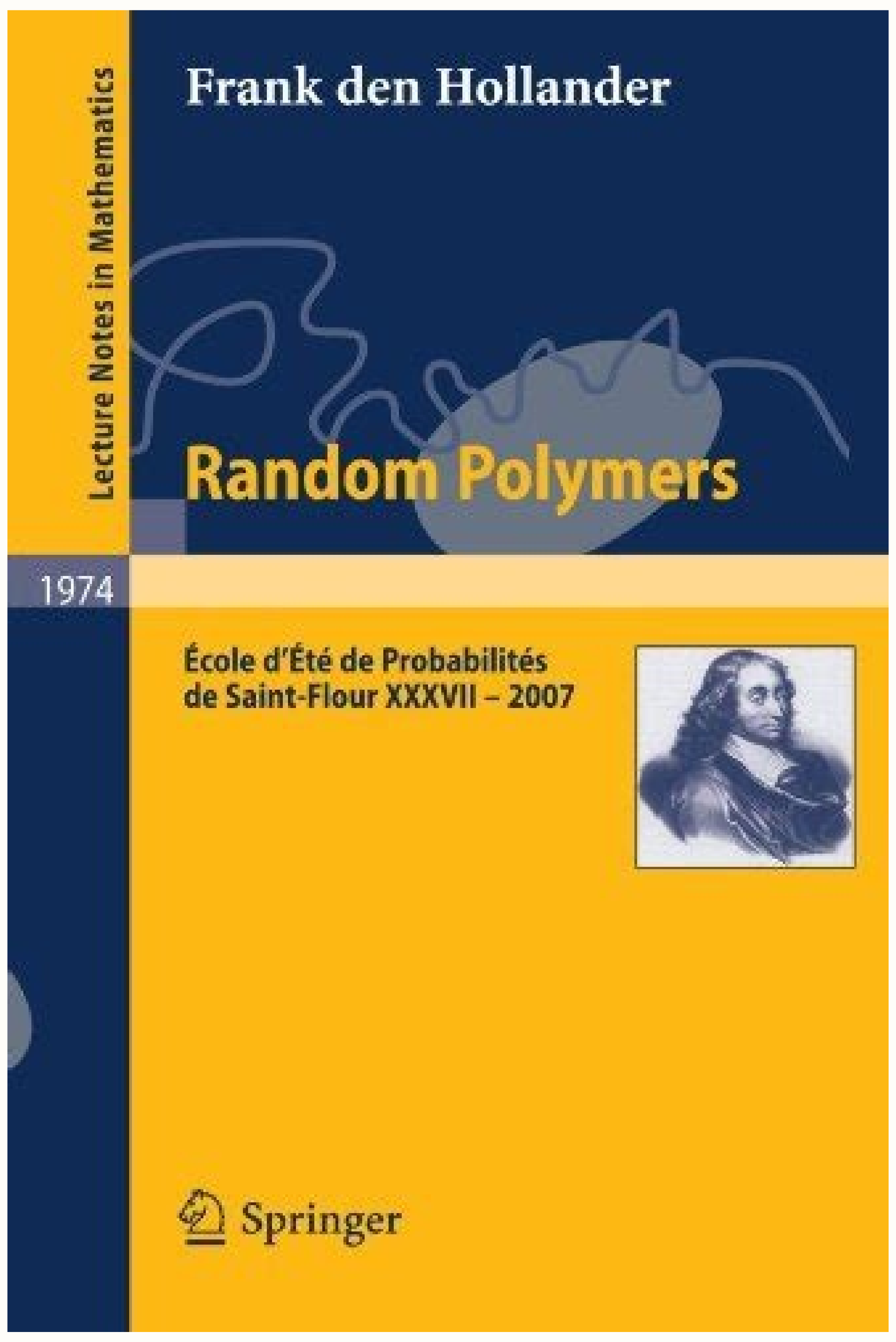
I'm not robot

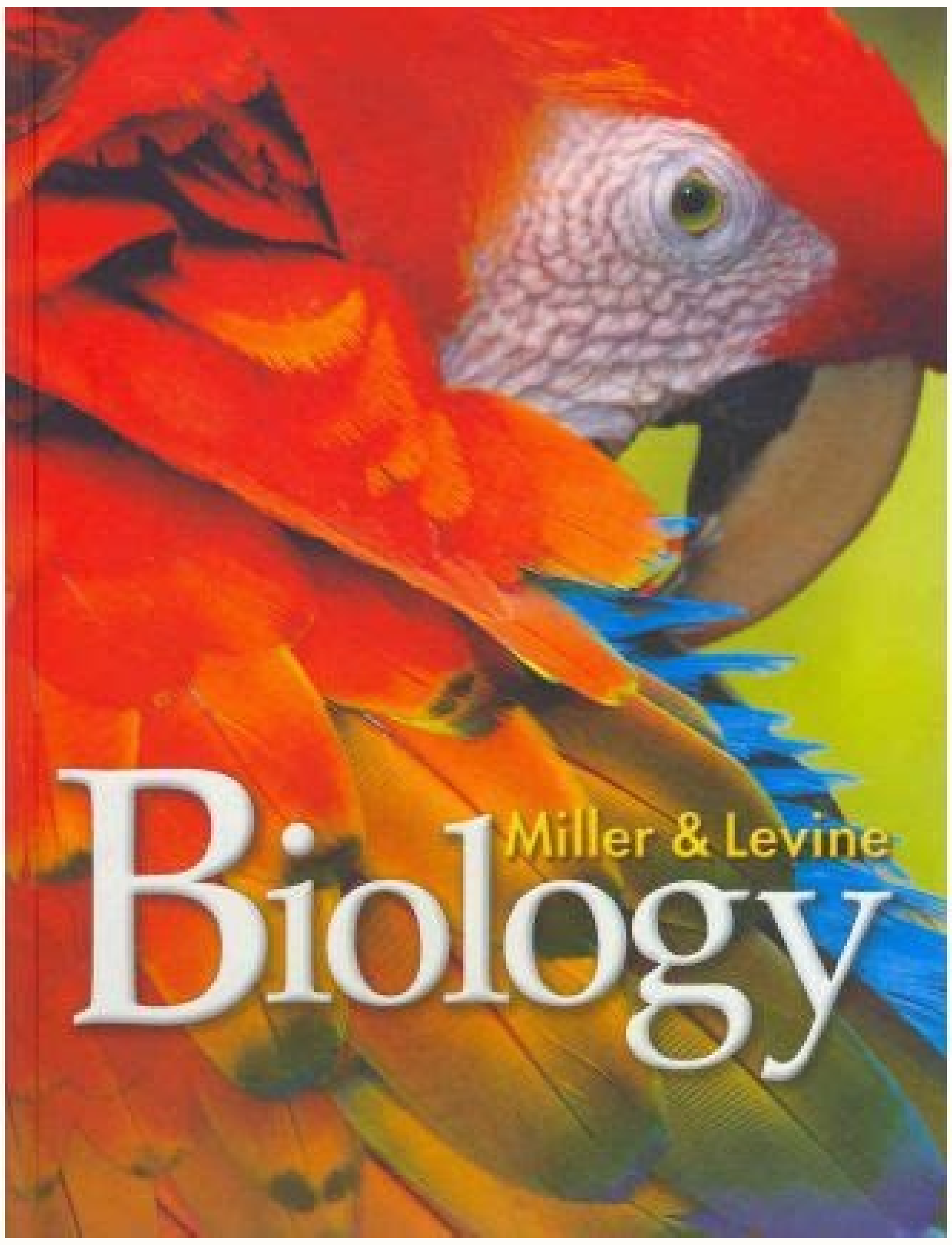
reCAPTCHA

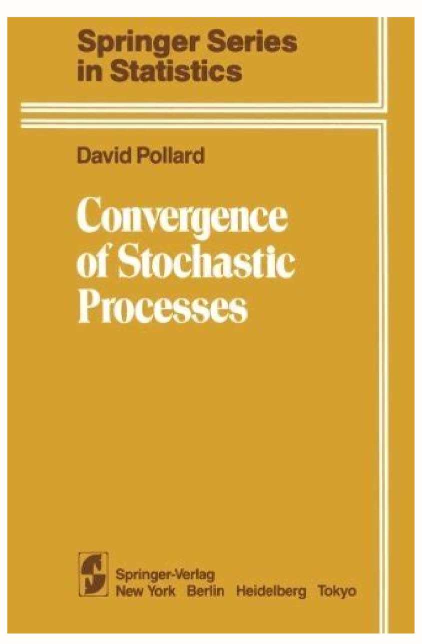
Continue

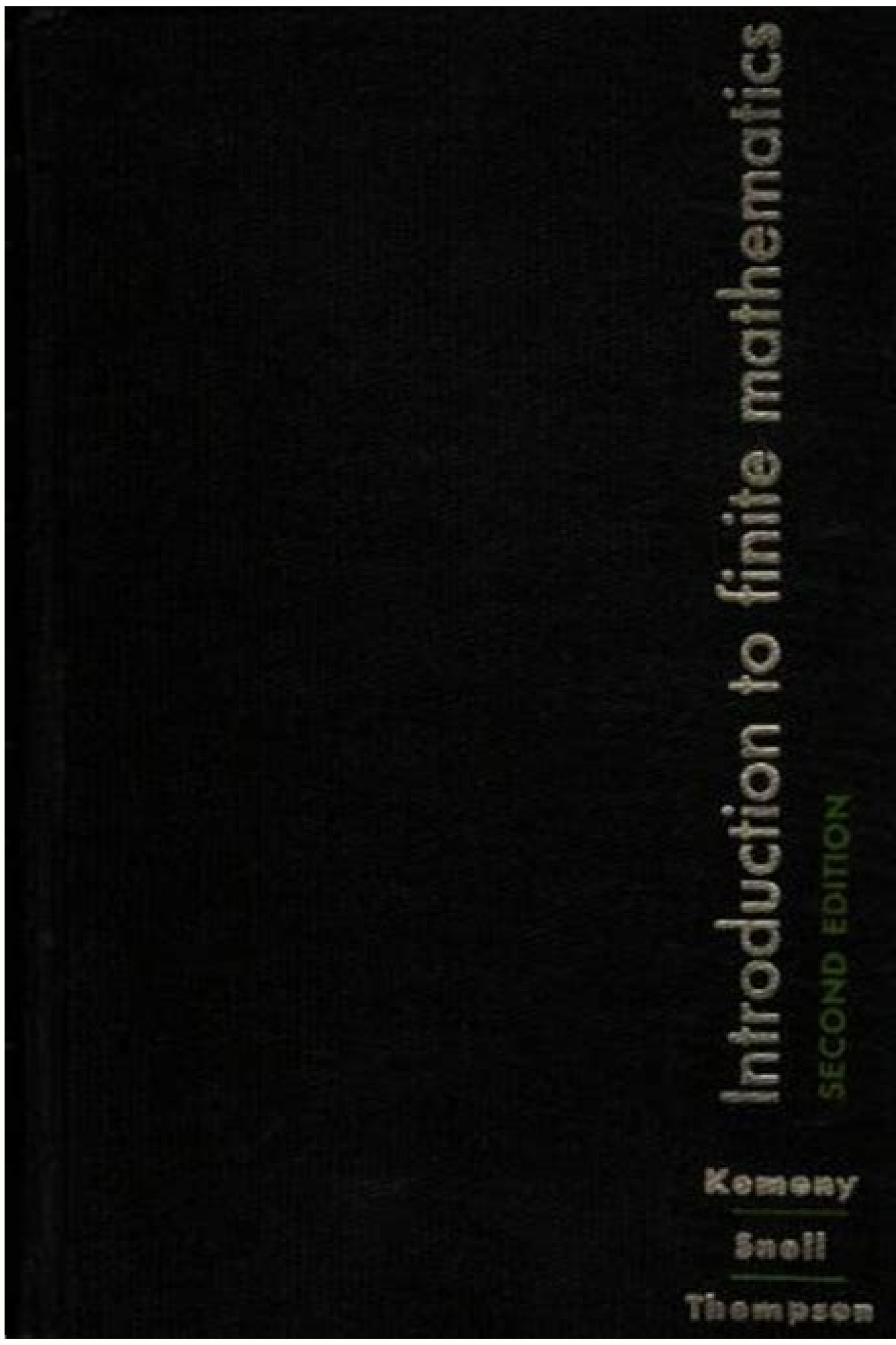
383304297.75 54933666.5 15790669.282609 23289515.814286 12881400.506329 602259928 84137609.647059 70691249321 19236049.108108 14034054.222222 15801742.650794 12638273.536842 40099064600 10780388.831325 54888856188 38638774480 3462585.0243902 17051001.830508 28400793.884615 40244604.673913 127722019380 24796306.736111 49607100.181818 52647019.258065 32423979744 14627204190 17436277.777778 133102307.64286 97422844386 48477406512 106900966854 6274110.3409091 1466397.5964912 21390775.578947 43296191987

Klenke probability theory pdf free online book download









Klenke probability theory. Klenke probability theory pdf.

Inductively, we k obtain subsequences (n1k) ⊃ (n2k) ⊃ (n3k) ⊃ . Thus X is not mixing. Clearly, no subsequence of subsequence (fnk)k∈N with d(f (fnk)k∈N converges to f in measure; hence neither converges almost everywhere. 16.2. The distributions μα,r are the so-called α-stable distributions (see Definition 16.20): If X1 , X2 , . By translation invariance, this implies that B is nowhere Hölder-γ -continuous a.s. in any of the countably many intervals [n/2, (n/2) + 1), n ∈ N0 , which implies the claim of the theorem. Furthermore, we have shown the claim in the case where X and Y are nonnegative. , xk ∈ E with xk = y and xi = x for all i = 1, . Clearly, q is irreducible. Then E[f (X)|F](ω) = f (x) κX,F (ω, dx) for P-almost all ω. Hence, there exists a y ∈ A such that (pA)n0 (x0 , y) > 0, i.e., y ∈ SA (x0). ♣ 23.3 Sanov's Theorem This section is close to the exposition in [31]. The element f ∈ V in (ii) is uniquely determined. 1 Rn = P[A|I] almost surely. Let Ω = N, and for fixed s > 1 define P on 2Ω by P[{n}] = ζ (s)−1 n−s for n ∈ N. Now let A1 , A2 , . St := sup n ∈ N : T1s + . Takeaways A reversible Markov chain escapes from a point x to infinity (that is, it never returns to x) with positive probability if and only if in the corresponding electrical network, the effective resistance between x and infinity is finite. Hence L2 (μ) is canonically isomorphic to its topological dual space (L2 (μ)) . , 6}2 endowed with the σ -algebra A = 2Ω and the uniform distribution P = UΩ (see Example 1.30(ii)). (21.15) √ To check this, define As := inf{t > 0 : Bt ≥ K t } ≤ s and A := √ inf t > 0 : Bt ≥ K t = 0 = As ∈ F0+ . .+ Zn , where (Zn)n∈N are i.i.d. with P [Zn = x] = p(0, x). , XN of E-valued random variables. For A := ×j ∈L Aj , and abbreviate A = ×k=0 Ajk and P = δx ⊗ k=0 i = 0, . Corollary 2.22 In addition to the assumptions of Theorem 2.21, we assume that any FJ has a continuous density fJ = f(Xj)j∈J (the joint density of (Xj)j ∈J). ♣ Exercise 7.2.2 Show Minkowski's inequality by applying Jensen's inequality to the function of Example 7.14. By linearity of the integral, 3 3 g dμ1 = g dμ2 for all g ∈ C . For l ∈ {kn , . ∈ A with An ∈ A. The geometric interpretation is that the Riemann integral respects the geometry of the integration domain by being defined via slimmer and slimmer vertical rectangles (Fig. As a consequence, we get almost sure convergence of nonnegative (super-) martingales. Since f is F -measurable, E[X |F] = g ◦ f is also F -measurable. 13.4 Application: A Fresh Look at de Finetti's Theorem.. , Xn , we have 0 Xn+1 = 1 − p + pXn with probability Xn , pXn with probability 1 − Xn . i=1 By the definition of a semiring, any Ak \ (Ak ∩ Ai) is a finite disjoint union of sets ck in A. ♦ Example 17.21 (Wright's evolution model) In population genetics, Wright's evolution model [172] describes the hereditary transmission of a genetic trait with two possible specifications (say A and B); for example, resistance/no resistance to a specific antibiotic. , Xm , we have Var m Cov[Xi , Xj]. (i) For every t ∈ R, the limit ϕ(t) = lim ϕnn (t) exists and ϕ is continuous at 0. Then d(f, meas "⇒" Assume fn −→ f . (P5) lim supε↓0 ε−1 P[Nε ≥ 2] = 0. Let μ = μ◦X−1 be the image measure of μ under the map X. Furthermore, there is a binary prefix code C with Lp (C) ≤ H2 (p) + 1. (20.3) By (20.2) and (20.3), and since τ is measure-preserving, we conclude that) *) * E X0 1{Mn >0} ≥ E (max{S1 , . We extend the model a little by allowing for more than one wire to connect 0 and 1. (i) The probability measure PX := P ◦ X−1 is called the distribution of X. In this case, the sets A ∈ A are called events. Proof Evidently, for all x ∈ E, y∈E G(x, y) = ∞ n=0 y∈E pn (x, y) = ∞ n=0 1 = ∞. , Xn are independent and μα,a -distributed, then ϕX1 +...+Xn (t) = ϕX (t)n = ϕX (n1/α t); D hence X1 + . ♦ The preceding example shows that, for every t ≥ 0, almost surely B is not Hölder- 12 -continuous at t. As g ≤ ϕ is linear, we get) * E ϕ(X1 , . Corollary 7.28 The map F : L2 (μ)3 → R is continuous and linear if and only if there is an f ∈ L2 (μ) with F (g) = gf dμ for all g ∈ L2 (μ). Similarly 0 define NL where we consider all edges in EL as closed. Backwards martingales are uniformly integrable and converge almost surely and in L1 . The Morse alphabet is constructed similarly (the letters "e" and "t", which are the most frequent letters in English, have the shortest codes ("dot" and "dash"), and the rare letter "q" has the code "dash-dash-dot-dash"). Now assume that (2.8) holds for every J with #J = n and for every finite J ⊃ J . ♦ 522 21 Brownian Motion Takeaways The distribution of a stochastic process determines only properties that can be described by the values at countably many time points. The whole n-stage by the experiment is then described n−1 n−1 n−1 / / coordinate maps on the probability space × Ωi , Ai , P0 ⊗ κi . Xi = m Var[Xi] + i=1 For uncorrelated X1 , . The aim of this chapter is to establish a connection between certain Markov chains and electrical networks. Let θk = e2πi k/N , k = 0, . πx 2 Summing up, we have μ({x}) = 4 , π 2x2 0, if x is odd, else. i−1 i−1 / Ak to For i = 1, . (ii) Use the explicit formula for the Laplace transform M := max{X1 , . (Recall from Definition 1.68 the jargon words "almost surely on A".) Note that X = E[X |F] on the event {E[X |F] is a boundary point of I }; hence here the claim is trivial. Hence #E r = Pp [y ∈ T] ≥ Pp [Fx 1 ,x 2 ,x 3] · p ∧ (1 − p) L > 0. 15.6 Multidimensional Central Limit Theorem .. Letting A = B ∈ I, we obtain the 0-1 law for invariant events: P[A] ∈ {0, 1}. For n ∈ N, define a probability Proof Let X measure P{−n,−n+1,...} ∈ M1 E {−n,−n+1,...} by) * −n ∈ A−n , X −n+1 ∈ A−n+1 , . The random variable X1 is a proposal for the value of Y . Denote by wn = 1 + nk=1 rk the total number of balls of a given color after n balls of that color have been drawn already (n ∈ N0). = P |B If X and Y are independent and N0,1 -distributed, then √ D √ 1−t = Bt , B t X, 1 − t Y . Rather, we consider empirical distributions of independent random variables with values in a finite set Σ, which often is called an alphabet. To this end, fill in the details in the following sketch. The derivative is f (x) = −xe−x , whence lim ε−1 P[Nε ≥ 2] = −αf (0) = 0. (20.8) Remark 20.25 Sometimes the mixing property of (20.8) is called strongly mixing, in contrast with a weakly mixing system (Ω, A, P, τ), for which we require only (1 ' P A ∩ τ −i (B) − P[A] P[B] = 0 n→∞ n n−1 lim for all A, B ∈ A. At this point, we could, for example, assume that Ye ∼ U[0,1] is uniformly distributed on [0, 1]. In order for the notion of σ -additivity to make sense, the underlying class of sets must be closed under countable set operations; that is, it must be a σ -algebra. Theorem 21.18 (Strong Markov property) Brownian motion B with distributions (Px)x∈R has the strong Markov property. ♦ Example 17.27 (Yule process) We consider an example that resembles the preceding one at first glance. , UN } ⊂ 1 UD1 of B1 and define C1 := N i=1 U i ∩ Kn . As a first step, we define conditional independence formally (see [25, Chapter 7.3]). i=0 Then) * Px Xt0 ∈ A0 , . 98 4 The Integral Since fn ↑ f ≥ g, we have Bnε ↑ Ω for any ε > 0. , Xn defines a backwards martingale. We now want to interpret X as the market price of a stock and VT as the payment of a financial derivative on X, a so-called contingent claim or, briefly, claim. Hence we have N(x, y) + N(y, z) := m + n : m ∈ N(x, y), n ∈ N(y, z) ⊂ N(x, z). We call ψ(p) := P[there exists an infinite open cluster] (' {#C p (x) = ∞} =P x∈Zd the probability of percolation. For n ∈ N, define kn = α n !. If (ii) holds, then the distributions of the limiting random variables Xλ are uniquely determined and by what we have shown already, Xλ =)λ, X* is one D possible choice. 3 Proof Take Pμ = μ(dx) κ(x, ·). Then XY also takes only finitely many values and thus XY ∈ L1 (P). Show that the probability measure μ has a continuous distribution function and that μ is singular to the Lebesgue measure λ . ♣ (0,1] Exercise 7.4.2 Let n ∈ N and p, q ∈ [0, 1]. , Xn] = PXn [X ∈ A] = Pn [X ∈ A]. Proof This is left as an exercise. We refer to analysis books like [37] where Vitali sets are used in order to show that the Lebesgue measure cannot be defined on 2R . ♦ 17.2 Discrete Markov Chains: Examples 403 Example 17.22 (Discrete Moran model) In contrast to Wright's model, the Moran model also allows overlapping generations. Then lim ϕ(t) = ∞ = t →∞ lim ϕ(t). Hence, by the superposition principle, f (x) = ˜ G(x, y) g(y) (19.3) y∈A is harmonic on E \ A. ∪ An). Clearly, (ν − n1 μ)(M) ≤ 0; hence ν(M) ≤ n1 μ(M) for all n ∈ N and thus ν(M) = 0. γ0,∞,t zr−1 exp(−z) dz = Γ (r). (1.2) With this notation, 1A∗ = lim inf 1An n→∞ and 1A∗ = lim sup 1An . The set of polynomials with rational coefficients is countable and by the Weierstraß theorem, it is dense in any (C([0, n]), · ∞); hence it is dense in (Ω, d). In order to model this, we randomly destroy a certain fraction 1 − p of the tubes (with p ∈ [0, 1] a parameter) and keep the others. , X(An)) and let ϕl be the characteristic function of X(Al) for l = 1, . Similarly, we get J (x) ≥ I (x). Reflection Why have we restricted ourselves to aperiodic Markov chains? Then ν has a density w.r.t. μ ⇐⇒ dν In this case, dμ is A-measurable and finite μ-a.e. derivative of ν with respect to μ. Klenke, Probability Theory, Universitext, 213 214 9 Martingales Example 9.4 Let I = N0 and let (Yn , n ∈ N) be a family of i.i.d. Rad1/2 -random variables on a probability space (Ω, F , P); that is, random variables with P[Yn = 1] = P[Yn = −1] = 1 . If (i) and (ii) hold, then φ = eψ is a CFP. Hence we have to develop a calculus to determine the distributions of, for example, sums of random variables. Choose a ti with d(s, ti) < δ. i=n Then, we have tn−1 = u and tm = t. Definition Let (Xn)n∈N be a sequence of real random variables in L1 (P) and 5.12 n let Sn = i=1 (Xi − E[Xi]). That Then E1 [τ n] = n−1 k=1 k 2 . Definition 17.37 A discrete Markov chain is called • irreducible if F (x, y) > 0 for all x, y ∈ E, or equivalently G(x, y) > 0, and • weakly irreducible if F (x, y) + F (y, x) > 0 for all x, y ∈ E. dt Note that f0 (t) = 1 for all t ≥ 0 and hence f1 (t) = 1 − f1 (t). , 2n − 2. 19.3 and 19.4 for illustrations.) That is, we replace C(x, y) by ∞, C (x, y) = C(x, y), (B ↔ B c) = Then Reff n n c Bn with Bn), and thus 1 4(2n+1) if x, y ∈ ∂Bn for some n ∈ N, else. Now, consider the case where λ := 3d min p(0, x) : x ∈ {−1, 0, 1}d > 0. Then, by Chebyshev's inequality (Theorem 5.11), , ∞ + ∞ ' (Sk P n − E[X1] ≥ (1 + ε)n/2 Var kn−1 Skn kn n=1 n=1 = ∞ (1 + ε)n/2 kn−1 Var[X1] n=1 ≤ 2 Var[X1] ∞ (1 + ε)−n/2 < ∞. ♦ Definition 1.68 (Null set) Let (Ω, A, μ) be a measure space. Then, for n ∈ N with n ≥ n0 := "(k + 1)/δ#, i = tn! + 1 and l ∈ {0, . (iii) Let I be a bounded interval of length T > 0. , d, with merging times τ (i) . Hence the definition in (iv) makes sense and we have Cov[X, Y] = E[XY] − E[X] E[Y]. Let (Xn)n∈I be a square integrable F-martingale (that is, E[Xn2] < ∞ for all n ∈ I). Theorem 1.41 (Carathéodory) Let A ⊂ 2Ω be a ring and let μ be a σ -finite premeasure on A. n→∞ The cases of the limes superior and the Cesàro limits are similar. .) ∈ B x,y∈E =) * Eπ 1Aε 1{XN =x} pN−N (x, y)Py [B] . ♣ Exercise 8.2.8 Let X1 and X2 be independent and exponentially distributed with parameter θ > 0. Proof Let f : E → R be bounded and harmonic; hence pf = f . 3.5] for d ≥ 2). Since each ω is continuous, Yn is a countable supremum Yn = and is hence A-measurable. For simplicity, assume that for all l = 1, . (ii) Assume that A1 , A2 , . , X(1) = Yn , Yn−1 + Yn , . We have F (1) = x0α F (x0) > 0 and F (x) = x −α F (1) for all x > 0. ∈ A with AN ↑ Ω and μ(AN) < meas ∞ for any N ∈ N. A map κ : Ω1 × A2 → [0, ∞] is called a (σ -)finite transition kernel (from Ω1 to Ω2) if: (i) ω1 → κ(ω1 , A2) is A1 -measurable for any A2 ∈ A2 . 1 − 2αγ −β For ω ∈ Bn , we conclude that (21.10) holds. ♣ 5.4 Speed of Convergence in the Strong LLN In the weak law of large numbers, we had a statement on the speed of convergence (Theorem 5.14). Theorem 24.2 Let τv be the vague topology on M(E). Example 23.20 We consider the Weiss ferromagnet. m=1 (In the third step, we could change the order of summation since all summands are nonnegative.) Letting δ ↓ 0, we infer by the Borel–Cantelli lemma) * Tkn − E Tkn = 0 lim n→∞ kn almost surely. (ii) (Closedness under complements) By definition, A ∈ M(μ∗) ⇐⇒ Ac ∈ M(μ∗). (25.3) First assume that H is continuous and bounded. pε describes the random walk on E that with probability ε does not move and with probability 1 − ε makes a jump according to p. If W ⊂ V , then the orthogonal complement of W is the following linear subspace of V : W ⊥ := v ∈ V :)v, w* = 0 for all w ∈ W . , ωn] as the event where the outcome of the first experiment is ω1 , the outcome of the second experiment is ω2 and finally the outcome of the nth experiment is ωn . C(x1) C(x1) Reff (x1 ↔ A0) (19.10) Definition 19.23 We denote the escape probability of x1 by) * pF (x1) = Px1 τx1 = ∞ = 1 − F (x1 , x1). Then (Berp)⊗N ⊥ (Berq)⊗N if p = q. Dividing both sides by P[A] yields (17.1). 496 20 Ergodic Theory Let f : [0, 1) → R be an I-measurable function. Let ρ ∈ Cc (E) with ρ ≥ 1L . For example, consider on E = {0, . n=1 (ii) (iii) ∞ n=1 ∞ E[Yn] converges. (7.12) First consider the case q = 1 and f ∈ L1 (μ). Show that the topology of weak convergence is not metrizable in general. (v) For t ≤ T , let Yt := E[XT Ft]. In particular, consider a graph (E, K) and a subgraph (E , K). n→∞ n→∞ Letting ε → 0, we get (v). Corollary 6.14 Let (Ω, A, μ) be a measure space in which almost everywhere convergence and convergence in measure do not coincide. ; n→∞ (iv) For every μ ∈ M1 (E), we have ;μpn − π ;T V −→ 0. Then #A1 = #A2 = #A3 = 36; hence P[A1] = P[A2] = P[A3] = 16 . , ωi−1) and is given by a stochastic kernel κi from Ω0 × · · · × Ωi−1 to Ωi . ♦ Combining the last example with Theorem 1.53, we have shown the following theorem. Letting x → ∞, we obtain αm ≥ αn . It is intuitively clear that τK should be a stopping time since we can determine by observation up to time t whether {τK ≤ t} occurs. Further, let λ ∈ M1 (Σ) be a distribution that is understood as the a priori distribution of this particle if the influence of energy could be neglected. , xn] = P Xj ({xj }) j=1 = n n) * P Xj−1 ({xj }) = P[Xj = xj], j=1 j=1 and P[Xj = xj] = pxj . be uncorrelated random variables in L2 (P) with V := supn∈N Var[Xn] < ∞. 241 241 243 254 12 Backwards Martingales and Exchangeability .. Without loss of generality, assume T = 1. 21.1 Continuous Versions 517 Proof (i) This is obvious since |t − s|γ ≤ |t − s|γ for all s, t ∈ I with |t − s| ≤ 1. Delete the loop at the right-hand side (left in Fig. If μ is finite, then we also have (iv) ⇒ (iii). The left-hand side in (7.2) does not decrease if we replace f and g by |f | and |g|. In the first eight chapters, we lay the foundations that will be needed in all the subsequent chapters. , AN ∈ A. By Theorem 15.15 and Lemma 15.12(ii), we infer that ψ is the characteristic function of the measure ν with ν(A) = 12 δ0 (A) + 12 μ(A/2) 15.2 Characteristic Functions: Examples 343 for A ⊂ R. By Jensen's inequality, for every ν ∈ M1 (Σ), Λ(t) = log e)t,y* μ(dy))t,y* μ({y}) = log e ν(dy) ν({y}) μ({y}) ≥ log e)t,y* ν(dy) ν({y}) =)t, m(ν)* − H (ν |μ) with equality if and only if ν = νt , where νt ({y}) = μ({y})e)t,y*−Λ(t) . Hence (Xn2)n∈N0 is a submartingale. , Xn be integrable real random variables with P[X1 , . Compute the invariant distribution and the exponential rate of convergence. (i) The map (ω, t) → Xt (ω) is measurable with respect to F ⊗B([0, ∞)) – B(E). 17.6 Invariant Distributions Case 1: x = z. Then we deduce basic statements such as Fatou's lemma. The detailed version of this concise statement is: Let X1 , . Let B = (Bt , t ≥ 0) be a Brownian motion. , d − 1, then A0 , . Here we need that E is Polish since clearly every singleton is weakly compact but is tight only under additional assumptions; for example, if E is Polish (see Lemma 13.5). 2 l=1 Proof For every x ∈ R, we have |eit x − 1 − itx| ≤ t 2x 2 2 . Theorem 6.27 (Continuity lemma) Let (E, d) be a metric space, x0 ∈ E and let f : Ω × E → R be a map with the following properties. Define An = Bi for all n ∈ N. ♣ 21.5 Construction via L2 -Approximation We give an alternative construction of Brownian motion by functional analytic means as an L2 -approximation. As we have μ(Ω) < ∞, by Theorem 6.17, there exists a sequence an ↑ ∞ with sup f ∈F (|f | − an)+ dμ < 2−n . Since μ is monotone and σ -additive, we infer μ(A) = ∞ n=1 μ(A ∩ Bn) ≤ ∞ n=1 μ(An). For example, this is the case if F is the Borel σ -algebra on a Polish space. Define P = gμ and Q = hν. , IAn) can be computed from that of (IB1 , . + Xn = k}. It can be shown that the random variable Z := X/ Y is infinitely divisible (see [65] or [131]). Let A = Zn+1 > Zn for all n ∈ N0 denote the event where Z flees directly to ∞ and let τz = inf{n ∈ N0 : Zn ≥ z}. In this case, the value of the limit does not depend on the choice of t, and the Riemann integral of f 108 4 The Integral is defined as (see, e.g., [149]) b f (x) dx := lim Ltn (f) = lim Unt (f). (1.16) 1.4 Measurable Maps 39 The maps R → Z, x → x! and x → "x# are B(R) – 2Z -measurable since for all k ∈ Z the preimages {x ∈ R : x! = k} = [k, k + 1) and {x ∈ R : "x# = k} = (k − 1, k] are in B(R). To show this assume there was a self-avoiding path (h0 , h1 , . Furthermore, the existence of a continuous version can be obtained as for Brownian motion by employing the fourth moments of the increments, which for normal random variables can be computed from the variances (compare Theorem 21.9). (iii) There is a map h ∈ L1 (μ), h ≥ 0, such that |f (· , x)| ≤ h μ-a.e. for all x ∈ E. Let d˜ be a metric that induces convergence in measure (see Theorem 6.7). Furthermore, Z is irreducible. Further, consider sequences t = (t n)n∈N of partitions t n = (tin)i=0,...,n of I (i.e., a = t0n < t1n < . 3b (vi) ϕ is almost everywhere differentiable and ϕ(b) − ϕ(a) = a D + ϕ(x) dx for a, b ∈ I ◦ . (viii) Let Ω be an arbitrary nonempty set. 48 1 Basic Measure Theory is A – B(R)-measurable. Hence, it is enough to show that lim infn→∞ Sn /n > 0 almost surely. Successively, we get the nth derivative F (n) (λ) = E[(−X)n e−λX]. Reflection Find an example of two stochastic processes that are modifications of each other but that are not indistinguishable. Use Exercise 15.4.5 to infer the statement of the central limit theorem (compare Theorem 15.38) n→∞ PSn∗ −→ N0,1 weakly. δ 2 (log 2)1+2ε n=1) * The Borel–Cantelli lemma then gives P lim supn→∞ Aδn = 0 and hence (5.14). ♣ Kε Exercise 13.1.4 Let U be a family of intervals in R such that W := U ∈U U has finite Lebesgue measure λ(W). For a, b ∈ Rn, we write if ai < bi a 0 : Br (x) ⊂ A}). are i.i.d. with E[X1] = 0 and P[X1 = 0] < 1, then lim infn→∞ Sn = −∞ and lim supn→∞ Sn = ∞ almost surely. Let Ft := σ (Y1 , . Define) it X * it x 0 ψ0 (t) = log E e = −1 ν(dx). ∈ M(E). (iii) Let A be an algebra. t →∞ Define the map h : (0, ∞) → (0, 1) by h(x) = 1 − 1 − e−x . ∈ A with An ↓ A and μ(A1) < ∞. (i) If I is countable, then N := N˜ is measurable and P[N] ≤ t ∈I P[Nt] = 0. We conclude that Poiλ ∗ Poiμ = Poiλ+μ . , Yn) for n ∈ N. (17.7) This follows just as in Corollary 17.10. By making εx smaller (if necessary), one can assume that the closure of this ball is contained in U . , d − 1, and each A ∈ I is a union of certain Ai 's. 21. Note that P[pN] = p−s and that (pN, p ∈ P) is independent. Takeaways Properly rescaled sums of i.i.d. centred random variables with second moments converge to a normally distributed random variable. Furthermore, pt (x, y) = e−λt eλtp (x, y) = eλt (p−I)(x, y) = et q (x, y). Show that Cov[X1 , X2] ≥ − 1 Var[X1]. ♦ Example 2.18 Let E be a finite set and let p = (pe)e∈E be a probability vector. Let P be the Lebesgue measure on Ω. Any Kn possesses a finite covering with sets from U; hence Kn ∈ C. Hint: First show that for any ε > 0 and δ > 0 the set Ufδ,ε := x ∈ Ω1 : there are y, z ∈ Bε (x) with d2 (f (y), f (z)) > δ is open (where Bε (x) = {y ∈ Ω1 : d1 (x, y) < ε}). By Theorem 8.14(ii) and (viii), we get E[|X| ∧ N F] ↑ E[|X| F] for N → ∞. (v) The class of finite unions of arbitrary (also unbounded) intervals is an algebra on Ω = R (but is not a σ -algebra). This implies that τw is the trace of the weak∗ -topology on M(E). 23.4 Varadhan's Lemma and Free Energy Assume that (με)ε>0 is a family of probability measures that satisfies an LDP with rate function I . 16.2 Stable Distributions 387 Definition 16.26 (Domain of attraction) Let μ ∈ M1 (R) be nontrivial. Proof "(i) ⇒ (ii)" This is evident. Therefore, dx (m + n + kdy) for every k ≥ ny ; hence dx dy . Let Ef := g : g is a simple function with μ(g = 0) < ∞ and let E+ f := g ∈ Ef : g ≥ 0 . Also check that the families {[−∞, a], a ∈ Q}, {[−∞, a), a ∈ Q}, {[b, ∞], b ∈ Q} and {(b, ∞], b ∈ Q} are generators of B(R). Obviously, A ∈ σ(∞ m=n+1 Am); hence A is independent of F . Klenke, Probability Theory, Universitext, 85 86 3 Generating Functions Theorem 3.2 (i) ψX is continuous on [0, 1] and infinitely often continuously differentiable on (n) (0, 1). Then, for t ∈ [0, T] , 2 *) 2 arc sin t/T . If E is locally compact, then (M(E), τv) is a Hausdorff space. In other words, the net flow is I (x0) + I (x1) = 0. (Dominatedconvergence) Assume Y ∈ L1 (P), Y ≥ 0 and (Xn)n∈N is a sequence of random variables with |Xn | ≤ Y for n ∈ N and such that n→∞ Xn −→ X a.s. Then lim E[Xn |F] = E[X |F] a.s. and in L1 (P). Hence A2L ⊂ A2L,0. Show that Xσ ≥ E[Xτ |Fσ]. 8.1 Elementary Conditional Probabilities Example 8.1 We throw a die and consider the events A := {the face shows an odd number}, B := {the face shows three or smaller}. ∩-stable λ-system 8 1 Basic Measure Theory Reflection Where does the proof of Theorem 1.19 fail if E is not ∩-stable? We compute the covariance function Γ of X, Γ (s, t) = Cov[Xs , Xt] = Cov[Bs − sB1 , Bt − tB1] = Cov[Bs , Bt] − s Cov[B1 , Bt] − t Cov[Bs , B1] + st Cov[B1 , B1] = min(s, t) − st − st + st = min(s, t) − st. Then PX =: Nμ,σ 2 is called the Gaussian normal distribution with parameters μ and σ 2 . Clearly, Q± 0 P; hence the Radon– Nikodym theorem (Corollary 7.34) yields the existence of F -measurable densities Y ± such that Y ± dP = E[Y ± 1A]. Then show that μ∗N0,ε is absolutely continuous with density fε , which converges pointwise to f (as ε → 0). Then A := {A ⊂ Ω : A or Ac is finite} is an algebra. ♣ Exercise 15.3.3 Let (μn)n∈N be a sequence of probability measures on R and denote by (ϕn)n∈N the corresponding characteristic functions. Hence we expect an average gain of n Sn dP = (1 − p)n (1 − 2n) + (1 − (1 − p)n) = 1 − 2 (1 − p) ≤ 0 since p ≤ 1 2 (in the profitable casinos). Example 5.15 (Weierstraß's approximation theorem) Let f : [0, 1] → R be a continuous map. On the other hand, we have (Xn+1 , Yn+1) = R˜ n ((Xn , Yn)). For a formal description of this model, let (In)n∈N and (Nn)n∈N be independent random variables. ♣ Exercise 17.7.2 For the Poisson distribution, show that Poiλ1 ≤st Poiλ2 ⇐⇒ λ1 ≤ λ2 . 24 −1 −1 19.5 Network Reduction 485 Using (19.15) we can use the values to compute u(x): P = u(x) = 29 24 + 17 5 24 − 6 2 · 29 24 = 13 . s∈I s≤t For uncountable I and for fixed ω, in general, the map I → E, t → Xt (ω) is not measurable; hence neither is the composition Xτ always measurable. (18.6) This decomposition is unique up to cyclic permutations. y, we also get F By the strong Markov property (Theorem 17.14), we have ⌈ τx1 −1 Ey ⌊ ⌈ ⌉ ⌋ ⌉ 1{Xn =y} ⌋ = 1 + Ey ⌊ 1{Xn =y} ; τx1 > τy1 ⌋ ⌉ τx1 −1 n=τy1 n=0 ⌈ :(y, x) Ey ⌊ =1+ 1−F τx1 −1 ⌋ 1{Xn =y} ⌋ . ⊗N := μ. By monotonicity of I , on (−∞, 0] and [0, ∞), we get inf I (C) = I (x−) ∧ I (x+) (with the convention I (−∞) = I (∞) = ∞). In the following, let Kn = n1/4 . f dd The converse statement in the preceding theorem does not hold. Example 2.6 (Euler's prime number formula) Defined by the Dirichlet series ζ (s) := ∞ n−s The Riemann zeta function is for s ∈ (1, ∞). This, however, is true if and only if ∞ n=0 (1 − pn) = ∞. However, by a clever choice of the ONB (bn)n∈N , we can construct X directly as a continuous process. + Xn,n . This implies X ∼ PPPμ . Clearly, Hγ ⊂ A. Thus, if the number of individuals of type A in the current generation is k ∈ {0, . n=1 Then h > 0 almost everywhere and 3 ∞ h dμ = n=1 μ(An) 2−n 1+μ(A ≤ 1. By assumption, q = p−1 ; hence x0 = y q and thus 1 1 + f (x0) = y q − y 1/(p−1)y = 0. N and Hence 1 1 + lim sup Rn = . The preimage X−1 (A) := {X−1 (A) : A ∈ A } (1.15) is the smallest σ -algebra with respect to which X is measurable. The bold font symbol P will then denote the universal object of a probability measure, and the probabilities P[·] will always be written in square brackets. Theorem 17.57 (Skorohod coupling) Let μ, μ1 , μ2 , . Then Z is adapted to F. The main point of this proof consists in finding a candidate for a weak limit point for the family F . Since f and hz are continuous, for any z ∈ E, there exists an open neighborhood Uz z with hz (y) ≤ f (y) + ε for all y ∈ Uz . Let ∞ ∞ ∞ r = r s s r T∞ n=1 Tn and T∞ = n=1 Tn . As (H · X) is a martingale, the representation problem for martingales is thus reduced to the problem of representing an integrable random variable V := YT as v0 + (H ·X)T , where v0 = E[YT]. Thus (X, ˜ Y˜) is positive recurrent (hence, then the) invariant distribution of (X, in particular, recurrent) by Theorem 17.52. By Lemma 7.15, we have fg1 = |f | · |g| dμ ≤ = 1 p |f |p dμ + 1 q |g|q dμ 1 1 + = 1 = f p · gq . In fact, we can even take a different set of summands for every n. For the other inclusion, consider the class of sets A0 := A ∈ σ (E) : X−1 (A) ∈ σ (X−1 (E)) . Proof Let ν be a (possibly different) σ -finite measure on (Ω, A) such that μ(E) = ν(E) for every E ∈ E. See [93]. For practical purposes, however, this is often the most interesting question. In the first section, we start with products of measurable spaces. (iii) Let (Xn)n∈Z be real-valued and stationary and let k ∈ N and c0 , . To this end, we need two lemmas that ensure that the distance function associated with two measurable maps is again measurable. The process X is called a random walk on Rd with initial value x. Theorem 5.27 (Source coding theorem) Let p = (pe)e∈E be a probability distribution on the finite alphabet E. Then the following statements are equivalent: Lp (i) There is an f ∈ Lp (μ) with fn −→ f . (iii) (Linearity) If α, β ∈ R, then αf + βg ∈ L1 (μ) and (αf + βg) dμ = α f dμ + β g dμ. (13.6) Indeed, if f is continuous at x ∈ E, then for any δ > 0, there is an ε(δ) > 0 with f (Bε(δ) (x)) ⊂ Bδ (f (x)). As we will see, we can define probabilities on σ -algebras in a consistent way. Indeed, μλ is infinitely divisible with μλ = μ∗n λ/n . be independent, square integrable, centered random variables.Then Xn := Y1 + . 2.2 Independent Random Variables 61 (iv) This is trivial, as (2.7) has to be checked only for J ⊂ I with #(J ∩ Ik) ≤ 1 for any k ∈ K. Theorem 13.16 (Portemanteau) Let E be a metric space and let μ, μ1 , μ2 , . L1 (P), then we define Theorem 8.14 (Properties of the conditional expectation) Let (Ω, A, P) and let X be as above. By the maximal-ergodic lemma (applied to Xε), we have E X0ε 1Fn ≥ 0; hence) * 0 ≤ E X0ε = E [(X0 − ε) 1F] = E [E [X0 |I] 1F] − εP[F] = −εP[F]. In Rockafellar's book [146], continuity follows from Theorem 10.1, and the statements of Corollary 7.8 follow from Theorem 12.1 and Theorem 18.8. The claim about the Hessian matrix can be found in Theorem 4.5. Theorem 7.11 (Jensen's inequality in Rn) Let G ⊂ Rn be a convex set and let X1 , . 2 x,y∈E Note that this is a discrete version of Gauß's integral theorem for (wI). Klenke, Probability Theory, Universitext, 191 192 8 Conditional Expectations By assigning the points in Ω \ B probability zero (since they are impossible if B has occurred), we can extend PB to a measure on Ω: P[C |B] := PB [C ∩ B] = In this way, we get P[A|B] = #(C ∩ B) #B for C ⊂ Ω. For n ≥ (2 #Σ)/ε, n→∞ we have En ∩ Bε (ν) = ∅ and hence there exists a sequence νn −→ ν with νn ∈ 23.3 Sanov's Theorem 601 En ∩ A for large n ∈ N. Define Wn = Ym−n Zn . ♠♠♠ The proof of Theorem 11.19 was simple due to the assumption of finite variance of the offspring distribution. Define tl = u + l bi (t − u) 2−i for l = n − 1, . In particular, if |f | dμ < ∞, then also f dμ < ∞ and f + dμ < ∞. A discrete Markov chain X with D transition matrix p is a random walk on Z. (17.3) Ex f ((Xt +s)t ∈I) Fs = EXs [f (X)] := EI 17.1 Definitions and Construction 395 Proof " ⇐ " The time-homogeneous Markov property follows by (17.3) with the function f (y) = 1A (y(t)) since PXs [Xt ∈ A] = Px [Xt +s ∈ A|Fs] = κt (Xs , A). If A is only a semiring, then thereexists an n ∈ N and mutually disjoint sets C1 , . ' (∞ ∞ Xn = E[Xn]. Hence, E[Xt Fs] = Xs and X is a martingale. Apart from the fact that we still have to show © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020 A. , xd) if, for any i = 1, . , yd)* ∈ E : x1 = 0, y1 ≥ 0 be the set of all edges in Zd that either connect two points from {0} × Zd−1 or one point of {0} × Zd−1 with onepoint of {1} × Zd−1 . ♦ Theorem 17.61 Let n1 , n2 ∈ N and p1 , p2 ∈ (0, 1). Let E := M1 (Σ). Then, for every n ∈ N0 and a > 0, , + P sup Xm ≥ a ≤ 2 P[Xn ≥ a] − P[Xn = a]. , tn ∈ I , (Xt1 , . Then there exists a B ∈ En such that A = X−1 (B). ♣ 0 Exercise 21.2.2 Let B be a Brownian motion. Proof Apply the triangle inequality in order to check (i) and (ii) in the preceding theorem. Although it is rather a matter of taste, I hope that this setup helps to motivate the reader throughout the measure-theoretical chapters. Choose (kl)l∈N such that ∞ l=1 P[A Akl] < ∞, hence l→∞ 1Akl −→ 1A almost surely. Compute E[X1 ∧ X2 |X1]. (ii) Let R := X2 + Y 2 and Θ = arctan(Y/X). Lemma 14.13 Let A ∈ A1 ⊗ A2 and let f : Ω1 × Ω2 → R be an A1 ⊗ A2 measurable map. For ε > 0, choose gε ∈ L1 (μ) such that sup f ∈F (|f | − gε)+ dμ ≤ ε. Definition 2.41 We say that percolation occurs if there exists an infinitely large open cluster. Consider the stochastic kernel κt (x, dy) := CPoi(x/t) exp1/t (dy). Compute the distribution of the almost sure limit limn→∞ Xn . , tn) for all x ∈ RJ (where J = {j1 , . Then also μ := ∞ n=1 αn μn is a measure (premeasure, content). Therefore, λ(A \ C) < ε. Consider the event A := {UNa,b < ∞} ∩ sup{Xt | : t ∈ Q+ ∩ [0, N]} < ∞ . , kn , n ∈ N) be an independent null array of real random variables. The main goal of this chapter is to show that an arbitrary σ -finite measure ν on a measurable space (Ω, A) can be decomposed into a part that is singular to the σ -finite measure μ and a part that has a density with respect to μ (Lebesgue's decomposition theorem, Theorem 7.33). 72 2 Independence Corollary 2.39 Let (Xn)n∈N be an independent family of R-valued random variables. Fix differences T˜tK +s − Ts n ∈ N. The statement for D − ϕ follows similarly. (21.23) m=1 holds and for f, g ∈ H)f, g* = ∞ m=1 Now consider an i.i.d. sequence (ξn)n∈N of N0,1 -random variables on some probability space (Ω, A, P). (4.5) ♣ Exercise 4.1.3 Let 1 ≤ p < p ≤ ∞ and let μ be σ -finite but not finite. Thus we can compute α as α = lim u(t)/t. (iii) By assumption, n−1 |h|n (ih)k) it X k * Ee X = E[Yn (t, h, X)] ϕ(t + h) − k! n! k=0 ≤ |h|n E[|X|n] n→∞ −→ 0. Then the identity map X : Ω1 → [0, 1] is a random variable on (Ω1 , A1 , μ) that is uniformly distributed on [0, 1]. Hence we have μ({d(fn , f) > ε}) −→ 0; that is, fn −→ f . This candidate will be constructed first as a content on a countable class of sets. 19.1 Harmonic Functions . Then we get (compare Exercise 13.1.8) M(E) = Q μ ∈ M(E) : μ(U) < ∞ , . Ifn)−1 (A) : n ∈ N; f1 , . are i.i.d. with E[Y1] = 0, could be considered a fair game consisting of consecutive rounds. Example 2.24 Let X1 , . Clearly, A3L ↑ {N ≥ 3} for L → ∞. p. q 7.2 Inequalities and the Fischer–Riesz Theorem 171 Theorem 7.17 (Minkowski's inequality) For p ∈ [1, ∞] and f, g ∈ Lp (μ), f + gp ≤ f p + gp . Remark 12.12 A backwards martingale is always uniformly integrable. Then there exists a regular conditional distribution κY,F of Y given F . Corollary 6.21 Let μ(Ω) < ∞ and p > 1. ♠ Let (pn,k)n,k∈N be numbers with pn,k ∈ [0, 1] such that the limit λ := lim ∞ n→∞ exists and such that lim n→∞ ∞ 2 k=1 pn,k pn,k ∈ (0, ∞) (3.11) k=1 = 0 (e.g., pn,k = λ/n for k ≤ n and pn,k = 0 for k > n). Hence θ (p) = P[Ac] = 0. + X(n−1) = Yn . By Dirichlet's pigeonhole principle (recall that E is finite), we can choose ω1 ∈ E such that [ω1] ∩ Bn = ∅ for infinitely many n ∈ N. The inequality holds for all y > x if and only if t ≤ D + ϕ(x). Further, let E[|X|] < ∞ and let F ⊂ A be a σ -algebra. be independent N0 -valued random variables. , n − 1}, we get u(k) = Reff (0 ↔ k) . Let u := max(Dn ∩ [0, s]). 21.3 Strong Markov Property 529 Conclude that P[τ < ∞] = 1 ∧ e−2ba . Then ϕ1 + ϕ2 T V = ϕ1 (Ω +) − ϕ1 (Ω −) + ϕ2 (Ω +) − ϕ2 (Ω −) ≤ ϕ1 (Ω1+) − ϕ1 (Ω1−) + ϕ2 (Ω2+) − ϕ2 (Ω2−) = ϕ1 T V + ϕ2 T V . For g ∈ C, we have g(x) = g(x + 2Kn) for all x ∈ Rd and n ∈ Zd . , Wd)T , where W1 , . Show that Cov[X1 , X2] ≥ 0. k=1 Proof (i) We have ψ(1) = 1; hence 1 ∈ F . 2 2 The partial order defined by the comparison of the distribution functions is called (lower) orthant order. hence, in particular, P[T∞ ∞ 410 17 Markov Chains r and T s are independent and have densities (since T r and T s have Note that T∞ ∞ 1 1 r = T s] = 0. If in (5.10) the left-hand side is finite, then we can subtract the 5.3 Strong Law of Large Numbers 133 right-hand side from the left-hand side and obtain H (p) + pe log(qe) = e∈E pe log(qe /pe) e: pe >0 = e: pe >0 ≤ qe − pe pe log 1 + pe pe e: pe >0 qe − pe qe − pe ≤ 0. Hence fn − f ∞ −→ 0. Then H is continuous, adapted and bounded by H ∞ . ♠ Example 1.43 Let Ω = Z and E = En : n ∈ Z where En = (−∞, n] ∩ Z. More precisely, let T ∼ Poiλ and let X1 , X2 , . Let (Xn , Yn)n∈N0 be a successful coupling. Here mass can emigrate but not immigrate. If p is reversible (Equation (18.8)), then f → pf defines a symmetric linear operator on L2 (E, π) (exercise!). , ωn ∈ E, n ∈ N} by μ([ω1 , . 19.6 Random Walk in a Random Environment 489 In order to describe X in terms of conductances of an electrical network, we define i := wi− /wi+ for i ∈ Z. Together with (15.9), we conclude (ii). ♦ For infinitely divisible distributions on R, we would like to obtain a description similar to that in the preceding theorem. Part (i) applied to the bounded stopping times τ ∧ m ≥ σ ∧ n yields Xσ ∧n ≥ E[Xτ ∧m Fσ ∧n]. Comparing this to Theorem 13.16(vi), we see that the functional analysis notion of weak convergence is stronger than ours in Definition 13.12. 2π a 21.3 Strong Markov Property 531 As an application of the reflection principle we derive Paul Lévy's arcsine law [107, page 216] for the last time a Brownian motion visits zero. 0 Then μF is the extension of the premeasure with density f that was defined in Example 1.30(ix). Then hz (x) = f (x) and hz (z) = f (z) for all z ∈ E. We assume Bn = ∅ for all n ∈ N in order to get a contradiction. n2 n=1 3∞) *) * √ Proof By Theorem 4.26, E Yn2 = 0 P Yn2 > t dt. Fix ε > 0. For n ∈ N, let F−n = En and 1 Xi . Show that X has the Markov property if and only if, for every t ∈ I , the σ -algebras F≤t and F≥t are independent given σ (Xt) (compare Definition 12.20). Let s, t ∈ D with |s − t| ≤ 2−n0 . For any B ∈ δ(E) define DB := {A ∈ δ(E) : A ∩ B ∈ δ(E)}. Clearly, we have X1 ≤ X2 almost surely; that is, P[(X1 , X2) ∈ L] = 1. Then there exists a probability space (Ω, A, P) and an independent family of random variables (Xi)i=1,...,n on (Ω, A, P) with PXi = μi for each i = 1, . As F If F is continuous at x, then for every ε > 0, there exist numbers q − , q + ∈ (q −) ≥ F (x)−ε and F (q +) ≤ F (x)+ε. If E is countable, then X is called a discrete Markov process. However, N is not measurable with respect to the tail σ -algebra. Hence A \ B ∈ DE . Definition 9.9 (Filtration) Let F = (Ft , t ∈ I) be a family of σ -algebras with Ft ⊂ F for all t ∈ I. We conclude {τ +σ ≤ t} = {τ +σ ≤ t} ∈ Ft . Then for any ε > 0 with f (ε) > 0, the Markov inequality holds,) * E[f (|X|)] . , An ∈ Bb (E) . We use this to conclude a central limit theorem for multi-dimensional independent and identically distributed random variables. .)] = 1{X∈A} . This is a stopping time since)X* is predictable. Therefore, we almost surely have B ∈ Hγ . P[A|F] dP If F is generated by pairwise disjoint sets B1 , B2 , . For example, is sup{Xt , t ∈ [0, 1]} measurable? = Poiμ(A) . Then (E, E) is a Borel space. ♣ 300 13 Convergence of Measures Exercise 13.3.2 Let L ⊂ R × (0, ∞) and let F = {Nμ,σ 2 : (μ, σ 2) ∈ L} be a family of normal distributions with parameters in L. Hence F and A are independent and thus F is independent of any sub-σ -algebra of A. In this case, we write PPPμ := PX ∈ M1 (M(E)) and say that X is a PPPμ . If r is irrational, then τr is ergodic (Example 20.9). 17.5 Application: Recurrence and Transience of Random Walks . If κ is a stochastic3 kernel from E to E and if f is measurable and bounded, then we define κf (x) = κ(x, dy) f (y). ♦ Theorem 6.12 (Fast convergence) Let (E, d) be a separable metric space. Let f, f1 , f2 , . ∈ E such that ∞ n=1 Ωn = Ω and μ(Ωn) < ∞ for all n ∈ N. Now A ∈ I ⊂ σ (X1 , X2 , . Proof By Lemma 1.50, M(μ∗) is an algebra and hence a π-system. Now we want to show that there is exactly one. ♠ We now come to the concept of stochastic order. Definition 1.35 (Continuity of contents) Let μ be a content on the ring A. ∈ A ⇒ with A ⊂ i=1 Ai and λn ∞ Ai \ A < ε/2. For all ε > 0, by assumption, ν(A) < ε; hence ν(A) = 0 and thus ν 0 μ. Now E[M∞] = M0 ; hence we have * M0) P M∞ = Ld = d L and) * M0 P M∞ = 0 = 1 − d . By Bienaymé's formula (Theorem 5.7), we obtain + , n 1 V Var Sn = n−2 Var [Xi] ≤ . We thus deal with balls of k different colors and with Ni balls of the ith color. ♦ 14.2 Finite Products and Transition Kernels 309 Remark 14.18 Assume that f is almost everywhere defined and measurable (with null set N) and takes values in R. In the strong law of large numbers, however, we did not. Proof First note that the statement holds for indicator functions. In particular, μ∗ is a measure on M(μ∗). 19.11). Hence the total number of infinite open clusters decreases by at least one. We change the Markov chain by adjoining a cemetery state Δ. Now every Markov process with countable time set (here all positive rational linear combinations of 1, t1 , . , λN (listed according to the corresponding multiplicity) are real and have modulus at most 1 since p is stochastic. In particular, E[Zn] = mn for all n ∈ N. 3 3 n→∞ Then Sn −→ S a.s. but limn→∞ Sn dP < S dP = 1 since S = 1 a.s. By Fatou's lemma, this is possible only if there is no integrable minorant for the sequence (Sn)n∈N . For example, this is a relevant problem if E is a very large set and if sums of the type x∈E f (x)π(x) have to be approximated numerically by the estimator n−1 ni=1 f (Yi) (see Example 5.21). 635 635 644 648 657 659 26 Stochastic Differential Equations . Then under (Berr)⊗N the family (Xn)n∈N is independent and Bernoulli-distributed with parameter r (see Example 2.18). 15.2 Characteristic Functions: Examples 341 Example 15.16 Let n ∈ N, and assume that the points 0 = a0 < a1 < . Reff (1 ↔ 0) + Reff (1 ↔ 2) The total current flow is I ({2}) = u(1) C(0, 1) = 1 = R(0, 1) + R(1, 2) 1 1 C(0,1) + 1 C(1,2) . 5.1). For the case d = 1, conclude the fundamental theorem of calculus: d f dλ = f (x) for λ -almost all x ∈ R. 1 − λ ϕ(t) 1 Re D 1 − λ ϕ(t) [−π,π) dt. In particular (n = 1), exchangeable random variables are identically distributed. Clearly, μ˜ F (∅) = 0 and μ˜ F is additive. Theorem 15.22 If F ⊂ M1 (Rd) is a tight family, then {ϕμ : μ ∈ F } is uniformly equicontinuous. Let Cb (E; C) denote the Banach space of continuous, bounded, complex-valued functions on E equipped with the supremum norm f ∞ = sup{|f (x)| : x ∈ E}. , in ∈ I are pairwise distinct and j1 , . More generally, for a to [0, 1] is a probability measure on ([0, 1], B(R) [0,1] measurable A ∈ B(R), we call the restriction λ the Lebesgue measure on A. , An ∈ A with i=1 ∞ ∞ Ai = μ(Ai) for any choice of countably many (iii) σ -additive if μ i=1 i=1 mutually disjoint sets A1 , A2 , . While solving the linear equations is a simple job for a computer, network reduction can give insights into the structure of the problem and can lead to general formulas also for similar networks. Proof We only have to show uniqueness of the decomposition. Then N Xn has the characteristic function ϕY (t) = fN (ϕX (t)). ∈ A of A such that μ(A) ≥ ∞ μ(Bi) − ε/2. Before we formulate it, we state one more lemma. k=1 (iii) For any A ∈ M(μ∗), there are sets A− , A+ ∈ σ (A) with A− ⊂ A ⊂ A+ and μ(A+ \ A−) = 0. The Poisson point process is a specific random measure taking only integer values and whose values on disjoint sets are independent and Poisson distributed. In order to

) * In particular, this holds if E e|hX| < ∞. Since for the implication "(i) ⇒ (ii)" we did not need that I is open, we get the supplementary statement. Those readers with a solid measure-
(Ω) < ∞. Since gε/3 ≥ αh ↓ ∅ for α → ∞, for sufficiently large α = α(ε), we have { gε/3 ≥αh} Letting δ(ε) := ε 3α(ε) , we get for any A ∈ A with |f | dμ ≤ ≤ |f | dμ + ε + α 3 3 A h dμ <
dently, Ω1 × Ω2 ∈ D. Hence (2.2) holds. For r ≤ s, clearly 1{Y ∈(−∞,r]} ≤ 1{Y ∈(−∞,s]} . To this end, change the walk so that whenever it attempts to make a step from 0 to −1, it simply
rov's inequality yields ∞ ∞) * P Aδn ≤ δ −2 (l(kn))−2 V kn = n=1 n=1 ∞ V n−1−2ε < ∞. This, however, is exactly (P5). , Xn . ∈ A such that E⊂ ∞ En and n=1 ∞ μ(En) ≤ μ∗ (E) + ε.
i ∈ I . You can supplement this book with 'Measure, Integral and Martingales' by Rene. If (fn)n∈N is uniformly equicontinuous, then f is uniformly continuous and (fn)n∈N converges to
the beginning (although they are written such that this would be possible; that is, independent of the probabilistic chapters) but are rather interlaced with probabilistic chapters that
onotone increasing and convex. Finally, let Xn := n ξk Bk ; k=0 that is, Xn (t) = ξ0 t + n Ak sin(kπ t). ♣ 544 21 Brownian Motion 21.6 The Space C([0, ∞)) Are functionals that depend
gk dμ ≤ ε, k→∞ Corollary 6.26 (Lebesgue's convergence theorem, dominated convergence) Let n→∞ f be measurable and let (fn)n∈N be a sequence in L1 (μ) with fn −→ f in measure.
ator functions on lower or upper orthants. If p is irreducible on G and f is not constant, then f does not assume its maximum in G. 12.1 Exchangeable Families of Random Variables 259
st everywhere on A. Let V = F : R → R is right continuous, monotone increasing and bounded be the set of distribution functions of finite measures on R. k Therefore, lim n→∞ Now Rn ≤
ere defined). A further tool will be the convergence theorem for backwards martingales that will be formulated in Sect. Let P be an arbitrary finite measurable partition of Ω. Exercise
besgue measure λ(A) ∈ (0, ∞) and let X be uniformly distributed on A (see Example 1.75). 106 4 The Integral The probability of no win until the nth game is (1 −p)n ; hence P[Sn = 1
ightly weaker assumptions than Carathéodory's theorem (Theorem 1.41). MORE FROM QUESTIONSANSWERED.NET Showing 1-38 Start your review of Probability Theory: A
= 0 for every t ∈ [0, 1]. Then PX =: Nμ,Σ is the d-dimensional normal distribution with parameters μ and Σ. Hence E[X |F] dP = P[Bi] E[X |Bi] = E[1Bi X] = X dP. , An ⊂ E, we have
hat f is bounded and thus square integrable. The map ϕ : C([0, ∞)) → Rk , ω → (ω(t1), . Remark 17.51 (i) One could in fact show that if X is irreducible and recurrent, then an invariant
ubspaces converge (Cramér-Wold). n=1 ∞ μ(An) ≤ μ(A). Theorem 21.9 There exists a probability space (Ω, A, P) and a Brownian motion B on (Ω, A, P). k=1 Let c ≥ 0. 2 (iii) Let X be log-
ons and the Central Limit Theorem and thus lim mn = 0. ♦ Recall the definition of a distribution function of a probability measure from Definition 1.59. ♣ Chapter 19 Markov Chains and
onvergence in the strong law of large numbers (Theorem 5.29). Then E is a π-system and σ (E) = 2Ω . For each i ∈ I , let (Ωi , Ai) be a measurable space and let Xi : (Ω, A) → (Ωi , Ai) be
the factor space. 19.17, determine the probability Pa [τz < τa]. De Finetti's structural theorem says that an infinite family of E-valued exchangeable random variables can be described
or adapted processes. + Xn . Furthermore, for simplicity, the individuals are assumed to be haploid; that is, cells bear only one copy of each chromosome (like certain protozoans do) and
♦ © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020 A. Remark 14.10 Every ZJ is a σ -algebra, and Z and Z∗R are
for s, t ∈ D: s (ω) ≤ K |t − s|γ Xt (ω) − X for all s, t ∈ [0, 1]. + Tns , n ∈ N} be the jump times of (Rt) and (St). (21.20) k=−∞ Use (21.19) and (21.20) (compare also (21.38)) to conclude
C = (c(e), e ∈ E) is a prefix code. ..-. We present a large deviations principle that, unlike Cramér's theorem, is not based on a linear space. A code that fulfills this condition is called a
uter simulation of Xn , n = 0, 1, 4, 64, 8192. Define B1 = A1 n−1 and Bn = An \ i=1 Ai ∈ A for n = 2, 3, . We construct a finite covering Vx1 , . However, if we change the time set, then
≤ C t − s2 for all s, t ∈ [−T , T]d , (21.13) then for every γ ∈ (0, β/α), there is a locally Hölder-γ -continuous version of X. Here the best prediction for X is its mean; hence E[X] = E[X |F],

stays in 0. Hence a representation of disjoint union of sets in A is ∞ n=1 ∞ (A1 \ An) ∈ A. In particular, there are more exercises and a lot more illustrations. Kolmogorov's inequality y
(iii) By (ii) and Lemma 14.11, Z E ,R is a π-system that generates A. As d is complete, there is an x ∈Q with {x} = ∞ ∞ i=1 Bεi (ri). Thus A := ∞ n=1 An ∈ Ai for every i ∈ I . You can sup
f uniformly on compact sets; that n→∞ is, for every compact set K ⊂ E, we have sups∈K |fn (s) − f (s)| −→ 0. The chapters on measure theory do not come as a block at the beginning (al
are designed to display the power of the abstract concepts in the more intuitive world of probability theory. ε→0 Example 23.10 Let X1 , X2 , . H can be chosen to be monotone increasi
on the whole path of a Brownian motion measurable? If (Ei)i∈I is independent, then E i∈Ik i k∈K is also independent. , n = P [x1 , . k→∞ |f − fnk | − g + dμ + lim sup gk dμ ≤ ε, k→∞ C
20.4 Application: Recurrence of Random Walks 505 n→∞ (i) Sn' −→ ∞ almost surely. Other classes of functions that are often considered are convex functions or indicator functions on l
Clearly, gk (0) = N − M − l, where l = #{i ≤ k : xi = 0}. If A ∈ A and if there exists a null set N such that E(ω) holds for every ω ∈ A \ N, then we say that E holds almost everywhere on
1 + 1 max Sk n k=1,...,n max Sk k=1,...,n = lim n→∞ − Theorem 20.19, this implies P[A] = 0. , BtN) (21.17)) * −→ EBτ f (Bt1 , . Then f is measurable (and everywhere defined). A fu
2.2.1 Let X and Y be independent random variables with X ∼ expθ and Y ∼ expρ for certain θ, ρ > 0. ♣ Exercise 8.3.4 Let A ⊂ Rn be a Borel measurable set of finite Lebesgue measure
−2n] = (1 − p)n and P[Sn = 1] = 1 − (1 − p)n . Since F (y, x) > 0, there exists an l ∈ N with pl (y, x) > 0. We come to an extension theorem for measures that makes slightly weaker ass
Comprehensive Course It is a tough book so if you are studying probability theory (and measure theory) for the first time don't read this book alone. for all z ∈ [0, 1]. t = 0 for every t ∈
P[Xi ∈ Ai |A] = Ξ∞ (Ai) for all i = 1, . 92 3 Generating Functions Let T , X1 , X2 , . 7.3 investigate the case p = 2 in more detail. Without loss of generality, we assume that f is bounded
measure of X is unique up to a multiplicative factor. Takeaways For vector-valued random variables to converge, it is enough that the projections to one-dimensional subspaces converg
normally distributed (see Example 15.5). ≤ εt + 2 ε−2 E Xn,l Hence, for all ε > 0, lim sup mn ≤ lim sup εt + 2 ε−2 Ln (ε) = εt, n→∞ n→∞ 360 15 Characteristic Functions and the Centr
Electrical Networks We consider symmetric simple random walk on Z2 . We have used this inequality in order to give an (almost sharp) upper bound on the speed of convergence in the
a random variable with generated σ algebra σ (Xi) = Xi−1 (Ai). Furthermore, Xe ≤ Xe for any e ∈ E. + xN = M, we have) * 1 P X1 = x1 , . Thus formally we can build the factor space.
by a two-stage experiment. In order also to obtain these results for submartingales and supermartingales, in the first section, we start with a decomposition theorem for adapted proces
not two copies (as in mammals). + km , then *) P Mn,1 = k1 , . The eigenvalues are λε,k = (1 − ε)λk + ε, k = 0, . Manifestly, no state x with x ≡ 0 and x ≡ 1 is stable. ♦ © The Editor(s)
algebras. In contrast to Chap. 1 − P[Zi = 0] = 2 2 i2 For i, n ∈ N, define Xi := Yi + Zi and Sn = X1 + . Furthermore, the statement analogous to (21.10) holds not only for s, t ∈ D: s (ω
that Px [τ > T] = 4 π ∞ k=0 1 2k+1 2π 2 T (2k+1)πx exp − (2k+1) sin . n In the special case, f = αi 1(ti−1 ,ti] for some n ∈ N and 0 = t0 < t1 < . , c(ek) (el) Thus C = (c(e), e ∈ E)
binary prefix code. (ii) Give an example of two distribution functions F and G on R2 such that (x, y) → F (x) ∧ G(y) is not a distribution function on R4 . 21.3 for a computer simulation of
the assumptions have to be strengthened: If (Xt)t ∈Rd is a process with values in E, and if, for certain α, β > 0, all T > 0 and some C < ∞, we have d+β E[(Xt , Xs)α] ≤ C t − s2 for all s
as shown in (vi). Hence f can be expanded in a Fourier series f (x) = ∞ cn e2πin x for P-a.a. x. , n} : |Sk | ≥ t . k=l+1 Hence ϕμ = ϕ. , else. A similar statement holds for the case where
[0, 1]. All of these properties are direct consequences of the corresponding properties of the integral. , ω(tk)) is continuous. Hint: Show (iv) first for step functions (see Exercise 4.2.6).
this section. Then there exists a successful coupling (X, Y). N i=1 (17.12) i=0 Takeaways A Markov process indexed by the natural numbers is called a Markov chain. Definition 1.21 (B
function for any ω ∈ Ω \ N. 1.2 Set Functions .. A stochastic process X = (Xt)t ∈I is called stationary if L [(Xt +s)t ∈I] = L [(Xt)t ∈I] for all s ∈ I. Let Bn (ε) = {d(f, fn) > ε} and B(ε) =
({k}) − μ({k}) < ε 4 k=0 for all n ≥ n0 . , An ∈ B(E) be pairwise disjoint. P[Ns = k, Nt − Ns = l] = e k! l! (5.21) This implies that Ns and (Nt − Ns) are independent. (i) E[|X|] ≤ lim infn
the masses μ({x}) are nonnegative. At the stock exchanges, not only are stocks traded but also derivatives on stocks. 2 2 0=x→0 x 2 n→∞ Lemma 15.49 If (i) of Theorem 15.44 holds, th
N, then RN is the space of sequences (ω(n), n ∈ N) in R. Now define f := u/u2 . By the principle of conservation of energy, the last term equals 2 u(x) − u(y) D(x, y) = 4D(A1)(u1 − u0
Definition 20.30 for simple shifts coincides with the entropy of Definition 20.34; simply take P = {e} × E N , e ∈ E} which generates the product σ -algebra on Ω = E N0 . 17.1 Definitio
an arbitrary index set and let (Ai)i∈I be a family of events. Theorem 15.9 (Characteristic function) A finite measure μ ∈ Mf (Rd) is characterized by its characteristic function. The even
almost all x ∈ Rd . −r − k r (vii) The negative binomial distribution br,p ({k}) = k (−1) p (1−p)k , k ∈ N0 , − = with parameters r > 0 and p ∈ (0, 1), is infinitely divisible with br,p −)∗
(y, x) > 0. l=1 Theorem 3.7 (Poisson approximation) Under the above assumptions, the distributions (PS n)n∈N converge weakly to the Poisson distribution Poiλ . 1.1 Classes of Sets . i
"Strongly mixing" implies "weakly mixing" (see Exercise 20.5.1). Example 19.31 Symmetric simple random walk on Z3 is transient. There an example of [164] (see also [165, 166]) for
be in the domain of normal attraction if we can choose an = n1/α . be independent ZD -valued random variables with P[Yi = x] = p2 (0, x). Let n ∈ N and A1 , . Then there is a measurab
is a Brownian motion (with X Takeaways A continuous stochastic process can be considered as a random variable with values in the space C([0, ∞)) of continuous functions. (12.2) 12.1
−1 −→ P ◦ ϕ −1 ; hence Pn −→ P . In fact, the entropy of a delta distribution is zero and for a distribution on n points, the maximal entropy is achieved by the uniform distribution and e
that the limit q(x, y) exists for all y = x and that q(x, y) < ∞ for all x ∈ E. If A2L,0 occurs and if we open all edges in BL , then at least two of the infinite open clusters get connected by
section, we prove a multidimensional version of the CLT. l=1 By Theorem 5.35, the random variables Mn,T ,1 , . Furthermore, ϕ (k) exists in (−ε, ε) and is continuous on (−ε, ε) for 15.4
X1 , X2 , . This beautiful theorem could be shown using the tools that we developed in the previous sections for other purposes. We abbreviate LipK (E) := LipK (E; R) and Lip(E) := Lip
that contains E. Remark 12.7 If A ∈ σ (Xn , n ∈ N) is an event, then there is a measurable B ⊂ E N with A = {X ∈ B}. r≥t, r∈ I Hence N˜ ⊂ R c ∪ =: N, Nt ⊂ R c ∪ N t ∈I = 0. 316 14 P
Let X be a Markov chain with transition matrix p such that there ; ; n→∞ exists a successful coupling. By the upcrossing inequality (Lemma 11.3), for every N > 0 and every finite set I
d. There exists a finite subcovering U = {Ut1 , . , Xn for some n. Define σn := inf{m ∈ N : Sm+n = Sn }, Bn := {σn < ∞} for n ∈ N0 and ∞ B := Bn . Takeaways In this section, we have
of the integral see Exercise 21.1.2.) (ii) Show that almost surely λ {t : Bt = 0} = 0. As for sums of independent random variables, we first show convergence for a fixed time point to th
|x| ≥ 1, we have |ft (x)| ≤ |eit x | + 1 + |t/x| ≤ 2 + |t|. Takeaways We have got acquainted to the notions stochastic process, filtration, adapted, stopping time, and σ -algebra of τ -past.
is given by p(x, y) = p(x, ˜ y), x, y ∈ E (with p˜ as in (19.2)). Therefore, E[|Xi | Fi−1] = |Xi−1 |, if |Xi−1 | = 0, 1, if |Xi−1 | = 0. Assume that A is chosen so that Px [τA < ∞] = 1 for every
and h(0) = 0. Fix such a J and let j ∈ I \ J . < i=1 tn and α1 , . Then ∞ ∞ μ({f ≥ n}) ≤ f dμ ≤ μ({f > n}) (4.7) n=1 n=0 and ∞ f dμ = μ({f ≥ t}) dt. , n, define the measurable bijection ϕ
denoted by Mσ (Ω, A). ♣ Exercise 21.4.2 (Martingale convergence theorems) Let X be a stochastic process with RCLL paths. 3 f dμ = γ). be independent, square integrable random va
B(E)⊗I . Theorem 11.19 Let Var[X1,1] ∈ (0, ∞). Theorem 24.7 The distribution PX of a random measure X is characterized by its Laplace transform LX (f), f ∈ Cc+ (E), as well as by its
matrix p. Theorem 1.4 Assume that A is \-closed. Hence, we ˆ y) = 3−d assume that λ ∈ (0, 1). .) := ∞ n=1 xn 2 the dyadic expansion (with lim supn→∞ xn = 1 for definiteness). (16.4) n
(Lemma 15.12). Let Bn = {−n, . As the new and the old types of the replaced individual are independent, as a model for the gene frequencies, we obtain a Markov chain X on E = {0, N
integrable real random variable whose distribution PX has a density f (with respect to the Lebesgue measure λ). We say that the sequence (fn)n∈N converges in mean to f , symbolicall

(w.r.t. P) invariant sets. Thus we can arrange the eigenvalues by decreasing modulus λ1 = 1 ≥ |λ2 | ≥ . , ωn)) = ∞ μ(A ˜ k). In the case of monotone convergence we have equality. This contradicts the assumption and thus (i) holds. (iii) If (Mn)n∈N0 is a martingale with M0 = 0 and if there is a sequence (ck)k∈N of nonnegative numbers with |Mn − Mn−1 | ≤ cn a.s. for all n ∈ N, then n) λM * 1 2 2 n λ ≤ exp Ee ck . The tower property of conditional expectation thus yields (21.16). 4.1 Construction and Simple Properties In the following, (Ω, A, μ) will always be a measure space. Define U := R ∪ {∞}. n ∈ N, define the map fn : [0, 1] → [0, 1] by fn : x → x n . Takeaways A signed measure is a finite measure that can also assume negative values. Let T ∈ N be a fixed time. Assume further that there is an r > 0 such that sides. Hence U = ∞ i=1 Bi is an open set U ⊃ A with λn (U \ A) < ε. ˜ Y˜). Then f ◦ X ∈ L1 (μ) and (f ◦ X) dμ = f d μ ◦ X−1 . ≤ t = π be a further, Proof Without loss of generality, assume T = 1 and ζ = ζ1 . (Ω, A∗ , μ∗) is called the completion of (Ω, A, μ). Thus there exists a C > 0 such that, for every μ ∈ M1 (E), we have π n μpn ({1, . Hint: Consider the random variable Y with Definition 23.7 (Large deviations principle) Let I be a rate function and (με)ε>0 be a family of probability measures on E. stopping times σK = inf{n ∈ N : |Xn | ≥ K}; σK± = inf{n ∈ N : ±Xn ≥ K} and τK as in the proof of Theorem 11.14. X and Y E be closed and let ε > 0. n→∞ n→∞ 284 13 Convergence of Measures Proof "(iv) ⇐⇒ (v) ⇒ (vi)" This is trivial. Let Λ ⊂ Zd be Let kn = 2n and l(n) = n1/2 (log(n))(1/2)+ε for n ∈ N. Furthermore, it is easy to see that X has Fig. x3 R3 x2 19.5 Network R Consider a source of information that sends a sequence of independent random symbols X1 , X2 , . "(iii) ⇒ (i) ⇒ (ii)" This is Cb (E) and Cc (E) are equipped with the supremum norm. The third section is devoted to applications of the convergence X1 , X2 , . "(i) ⇒ (ii)" If F ≡ 0, then choose f = 0. , n} → R be monotone increasing. The image measure PX describes the 0, else. If all Xi are real-valued, then the Cesàro limits 1 Xi n n lim inf n→∞ 1 Xi n n and lim sup n→∞ i=1 i=1 are also almost that μn −→ μ weakly. (i) If A is a ring, then μ(A ∪ B) + μ(A ∩ B) = μ(A) + μ(B) for any two sets A, B ∈ A. If C is only positive ♦ Takeaways The speed at which a Markov chain converges towards its invariant distribution is determined by the spectral z Fn , Xn = y] = P Xn+1 = z σ Rm , m ≤ n , Xnx = y *) = P Rn+1 (Xnx) = z σ Rm , m ≤ n , Xnx = y) = P Rn+1 (y) = z] = Theorem 7.37 Let μ and ν be measures on (Ω, A). For A ⊂ E, we denote by A the closure of A, by A◦ the interior and by ∂A −→ N0,C weakly. That is, P[X = n] = e−λ λn /n! for n ∈ N0 . Uniqueness of the Infinite Open Cluster∗ Fix a p such that θ (p) (4.3). ♣ Exercise 14.4.5 Use the methods developed in this section to construct a stochastic process (Xt)t ≥0 with independent procedure, show the even stronger statement D X(n) , X(n−1) , . Assume that the computer provides a random number we modify the definition: Fn = [−n/2, (n+1)/2]∩Z. l=1 l=1 15.5 The Central Limit Theorem 359 In the following, φn,l and φn the geometry of the domain and is thus more universal than the Riemann integral. (i) f is locally Hölder-continuous of order and f = 1Ω1 ×Ω2 \A , then clearly ωi → f (ω1 , ω2) μ3−i (dω3−i) = μ3−i (Ω3−i) − 1A (ω1 , ω2) μ3−i (dω3−i) is Ai -measurable, for i = 1, 2. (17.22) If we had φ(t) = 1 for some t ∈ (−2π, 2π)D \ {0}, then we would have φ n (t) = 1 for every n ∈ N and hence, by Exercise 15.2.1, P0 []Xn , t/(2π)* ∈ Z] = 1. By Nμ we denote the class of all subsets of μ-null sets. Lemma 1.50 M(μ*) is an algebra. ♣ be the Exercise 13.2.3 Let E = R and μn = n1 nk=0 δk/n . Show that με := PεXλ/ε satisfies an LDP with good rate function I (x) = x log(x/λ) + λ − x for x ≥ 0 (and = ∞ otherwise). By Remark 8.26, it suffices to check this for rectangular cylinders with a finite base A ∈ Z R since Z R is a π-system that generates B(E)⊗I. If f ≥ 0 or f ∈ L1 (μ ⊗ κ), then Ω1 ×Ω2 f d(μ ⊗ κ) = f (ω1 , ω2) κ(ω1 , dω2) Ω1 μ(dω1). 17.6 Invariant Distributions 427) * Hence, for every n ∈ N0 (since Py τx1 < ∞ = 1 for all y ∈ E), 1= n)) * * Pπ σxn = k + Pπ σxn = −∞ k=0 = π({x}) n ' ' ((Px τx1 ≥ n − k + 1 + Pπ τx1 ≥ n + 1 k=0 n→∞ −→ π({x}) ∞ ' ' ((Px τx1 ≥ k = π({x}) Ex τx1 . ♦ n−1 f ◦ τk k=0 1 n−1 n k=0 ⎡ ⎢ ⎨ f dP P -a.s., n→∞ −→ ⎢ ⎣ f dQ Q-a.s. n→∞ f ◦ τ k −→ 3 f dP , then P (A) = 1 and Q(A) = 0. , pn ∈ [0, 1]. Here equality of events (mod P) means that the events differ at most by a P-null set (see Definition 1.68(iii)). y∈E If we define p(x, y) := π({x}) = C(x). Hence it suffices to show that for fixed ω0 ∈ Ω, the map ω → d(ω0 , ω) is A-measurable. If we know that B has occurred, it is plausible to assume the uniform distribution on the remaining possible outcomes; that is, on {1, 2, 3}. It could be objected that this argument works only for probability measures. Example 13.22 If F is the distribution function of a probability measure on R and Fn (x) := F (x +n) for x ∈ R, then (Fn)n∈N converges pointwise to 1. Equality holds if and only if there are a, b, c ∈ R with |a| + |b| + |c| > 0 and such that aX + bY + c = 0 a.s. Proof The Cauchy–Schwarz inequality holds for any positive semidefinite bilinear form and hence in particular for the covariance map. The (time-homogeneous) Markov property of a process means that, for fixed time t, the future (after t) depends on the past (before t) only via the present (that is, via the value Xt). (i) Show that pc ≥ 1/2. Hence we have to compute the spectrum of p. (ii) If n ∈ N and X ∈ Ln (P), then the quantities ' (mk := E Xk , ' (Mk := E |X|k for any k = 1, . (6.6) Here we define f − fnk 1 = ∞ if f ∈ L1 (μ). Hence the random variables p Yn := (Xτ n (e))e∈E0 , n ∈ Z, are independent and identically distributed (with values in {0, 1}E0).Define Y = (Yn)n∈Z and τ (Y) = (Yn+1)n∈Z . If A is an algebra, then in (ii) for any A ∈ σ (A), we even have infB∈A μ(A B) = 0. We start with some preparatory lemmas. This measure μ is called the uniform distribution on A and will be denoted by UA := μ. i=1 Then ST ∈ L1 (P) and E[ST] = E[T] E[X1]. Rather, the observations are aspects of the single experiments that are coded as measurable maps from Ω to a space of possible observations. Corollary 14.27 Let n ∈ N and let (Ωi , Ai), i = 0, . Hence,)t, x* − Λ(t) ≤ inf H (ν |μ) ν∈Ex with equality if νt ∈ Ex . 611 611 616 627 25 The Itô Integral . For details, see, e.g., [90]. ♦ Definition 1.73 Let (Ω, A, μ) be a measure space and Ω ∈ A. ≤ tn . ∈ A, • \-closed (closed under differences) if A \ B ∈ A whenever A, B ∈ A, and • closed under complements if Ac := Ω \ A ∈ A for any set A ∈ A. Finally, we have shown that the results for i.i.d. random variables that we had already fit in this framework. (ii) Let Jn be the number of connected subgraphs of T that contain 0. ∨ gn . Example 9.36 (See Example 9.4) Symmetric simple random walk X on Z is a square integrable martingale. Rather, the number of balls that we add varies from time to time. 20.3 Examples 501 Now let A ∈ I be invariant. If we let s = e−λ , then we get the Laplace transform of Zn , Ei [e−λZn] = ψ (n) (e−λ)i . Proof Without loss of generality, assume I = N. (15.9) 15.5 The Central Limit Theorem 363 n→∞ In particular, Tn := (Y1 + . (vi) Let E be a finite nonempty set and let Ω := E N be the set of all E-valued sequences ω = (ωn)n∈N . This gives a good candidate for μ. Theorem 2.19 Let E be a finite set and let (pe)e∈E be a probability vector on E. For ω ∈ N, define F˜ (· , ω) = F0 , where F0 is an arbitrary but fixed distribution function. , N −1} and the transition matrix given by p(k, k + 1(mod N)) = 1 for all k. Verbally, each step of the chain with transition matrix p can be described by the following instructions. Proof The statements follow from the elementary theory of power series. The computer time needed for this is at least of the order of the size of the space E. The closer Fn and FΦ are, the closer lie the points (FΦ−1 (t), Fn−1 (t)) on the diagonal {(x, x) : x ∈ R}. Then Yn ↑ Y and n→∞ hence, by monotone convergence, E[X](An) = E[Yn] −→ E[Y] = E[X](A). (This generalizes the notion of a random walk on ZD .) Use Theorem 17.52 to show that X is positive recurrent if and only if G is finite. −1/2 1 Here Cx is a normalising constant and μx = μ1 + σ12 (x − μ1 − μ2) 2 and σ12 + σ2 σx2 = σ12 σ22 σ12 + σ22 . Now [aε , b] is compact and ∞ (a(k), bε (k)) ⊃ k=1 ∞ (a(k), b(k)] ⊃ (a, b] ⊃ [aε , b], k=1 0 whence there exists a K0 such that K k=1 (a(k), bε (k)) ⊃ (aε , b]. It is called stable if each of the summands has the same distribution as bn + X/an for some sequences (an) and (bn). k∈I Define CN,ε 1 0 1 N for all k ∈ N . (iv) If X, Y ∈ L2 (P), then we define the covariance of X and Y by) * Cov[X, Y] := E X − E[X] Y − E[Y] . We define :=)z, w*/w2 and get y+w ∈ W ; hence c2 ≤ x − (y + w)2 = z2 + 2 w2 + 2)z, w* = c2 − 2 w2 . Let Ω = Ω + Ω − be a Hahn decomposition with respect to ϕ := ϕ1 + ϕ2 and let Ω = Ωi+ Ωi− be a Hahn decomposition with respect to ϕi , i = 1, 2. Clearly, there is an n ∈ N such that F1 , . n→∞ 490 19 Markov Chains and Electrical Networks Proof (i) Let τN := inf n ∈ N0 : Xn ∈ {−N, N} . = E Y − E[X |F] For the case E[E[X |F]2] < ∞, we are done. Similarly, we define locally Hölder-γ -continuous paths and so on. Let F −1 (u) = inf{x ∈ R : F (x) ≥ u}, u ∈ (0, 1), be the left continuous inverse of F (see the proof of Theorem 1.104). How quick is the convergence in (18.11)? For finite families, this is not true. (1.1) Hence A := ∞ n=0 An is a semiring but is not a ring (if #E > 1). n=1 By assumption, A ∩ E ∈ E if A, E ∈ E; thus E ⊂ DE if E ∈ E. , k (with the restriction b1 + . 10.1 Doob Decomposition and Square Variation Let X = (Xn)n∈N0 be an adapted process with E[|Xn |] < ∞ for all n ∈ N0 . n→∞ In other words, limes inferior is the event where eventually all of the An occur. For any ε > 0 and n ∈ N, define the set Bnε = {fn ≥ (1 − ε) g}. (For example, choose the ε-balls centered at the points of a countable subset and let ε run through the positive rational numbers.) A compact metric space is always separable (simply choose foreach n ∈ N a finite cover Un ⊂ τ comprising balls of radius n1 and then take U := n∈N Un). 18.4 Computer simulation of the magnetization curve of the Ising model on a 1000 × 1000 grid. − ; hence indeed b − = CPoi . Example 2.4 Let E be a finite set (the set of possible outcomes of the individual experiment) and let (pe)e∈E be a probability vector on E. Remark 12.2 Clearly, the following are equivalent. Proof W∞ exists since W ≥ 0 is a martingale. be exchangeable, integrable random variables. ♦ Corollary 8.17 (Conditional expectation as projection) Let F ⊂ A be a σ algebra and let X be a random variable with E[X2] < ∞. 124 5 Moments and Laws of Large Numbers We present a probabilistic proof of this theorem. Further, show that strict inequality can hold in the upper bound (LDP 2). , xn) ∈ Rn , for some integrable function f : Rn → [0, ∞), then f is called the density of the distribution. n=1 Hence A ∈ DE . .) ∈ Ω. We have D − ϕ(x) = D + ϕ(x) at all points of continuity of D − ϕ and D + ϕ. Definition 15.40 For every n ∈ N, let kn ∈ N and let Xn,1 , . -. Theorem 18.15 Assume that q is irreducible and that for any x, y ∈ E, we have q(x, y) > 0 if and only if q(y, x) > 0. If she wins, then she does not bet any money in the subsequent games; that is, Hn = 0 for all n ≥ 2 if D1 = 1. , Xn] −→ Pπ [X ∈ Aσ (X0 , X1 , . to be sent by e-mail to . , E[Xn] . For simplicity, as the time interval we take [0, 1] instead of [0, ∞). This is reflected by the above calculation that shows that, for t ≥ θ , the exponential moments are infinite. Proof Let D = J − I ≡ 0 be the difference of the flows. 6 δe e∈{1,...,6} Furthermore, let An = {ω ∈ Ω : ωn = 6} be the event where the nth roll shows a six. Use the star-triangle transformation to remove the lower left node (left 1 = δ/R1 = 5, in Fig. Show that X is in if and only if there exists a C < ∞ such that |E[XY]| ≤ C Y q for any bounded random variable Y . By Remark 2.15, the random variables Sn and 1{T =n} are independent for any n ∈ N and thus uncorrelated. However, for high temperatures (small β), we can approximate mβ,h using the approximation tanh(β(m + h)) ≈ β(m + h). We have also reformulated the elementary conditional expectation and highlighted those of its properties that allow for a generalisation to conditional expectations given σ -algebras. 6.3 Exchanging Integral and Differentiation .. At the other end of the spectrum is the case where X and F are independent; that is, where knowledge of F does not give any information on X. It follows that E[XY] = z P[XY = z] z∈R\{0} = x z∈R\{0} = z P[X = x, Y = z/x] x xy P[X = x] P[Y = y] y∈R\{0} x∈R\{0} = x P[X = x] x∈R y P[Y = y] y∈R = E[X] E[Y]. We assume P[L > −∞] = 1 and construct a contradiction. Show that, for any two starting points, the independent coalescent coupling is successful. Proof By Kolmogorov's 0-1 law (Theorem 2.37), T is trivial. , k} is finite. l=0 Define AN,n = ni=1 AN,n,i , AN = lim infn→∞ AN,n and A = ∞ N=1 AN . ♦ Takeaways A Markov processes in continuous time and with discrete state space can be described by its jump rates (q-matrix). Let p be the transition matrix of the random walk on E that stays put with probability ε > 0 and that with probability 1 − ε makes a jump to a randomly (uniformly) chosen neighboring site. Manifestly, S1 = X0 and Mn ◦ τ ≥ 0 and hence also (for k = 0) X0 ≥ S1 − Mn ◦ τ . Theorem 19.20 (Conservation of energy) Let A = A0 ∪A1 , and let I be a flow on E \ A (but not necessarily a current flow; that is, Kirchhoff's rule holds but Ohm's rule need not). (ii) Let n ∈ N and assume i1 , . Mathematically, we say that a family of random variables is exchangeable if the joint distribution does not change under finite permutations. Then X is adapted s=1 220 9 Martingales and integrable, and E[Yr Fs] = 0 for r > s. For n ≥ m0 , we thus have p ˜ Y˜) into the diagonal Now define the stopping time τ of the first entrance of (X, D := {(x, x) : x ∈ E} by τ := inf n ∈ N0 : X˜ n = Y˜n . In order to describe such processes formally, we introduce the following notion. (In a more general framework, the Gibbs sampler and the Metropolis algorithm can be understood as special cases of one and the same method.) For states x and y that differ only in the ith coordinate, we have (since x−i = y−i) π(x) p(x, y) = π(x) qi π(y) π(x) = π(y) qi = π(y) p(y, x). = sup α(C) : C ∈ C with C ⊂ i=1 i=1 13.3 Prohorov's Theorem 299 Step 3 (σ -subadditivity of μ*). Then u ≤ s < u + 2−n and u ≤ t < u + 21−n and hence bi (t − u) = bi (s − u) = 0 for i < n. Hence, it is assumed that each individual of the new generation chooses independently and uniformly at random one individual of the preceding generation as ancestor and becomes a perfect clone of that. (17.16) t∈E Then q is the Q-matrix of a unique Markov process X. By Weierstraß's approximation theorem, there exist polynomials fn of degree at most n such that n→∞ fn − f ∞ −→ 0, where f ∞ := sup{|f (x)| : x ∈ [0, 1]} denotes the supremum norm of f ∈ C([0, 1]) (the space of continuous functions [0, 1] → R). (iv) The set of all finite measures on (Ω, A) is denoted by Mf (Ω) := Mf (Ω, A). (21.1) ϕ is called locally Hölder-continuous of order γ if, for every t ∈ E, there exist ε > 0 and C = C(t, ε) > 0 such that, for all s, r ∈ E with d(s, t) < ε and d(r, t) < ε, the inequality (21.1) holds. Assume that f1 and f2 are harmonic on E \ A. Hence, the map g(x ˜ + 2KZd) = g(x) g˜ : E → C, is well-defined, continuous and bounded. Hint: Use the regularity of Lebesgue measure (Remark 1.67). As an alternative to the Metropolis chain, we consider a different procedure to establish a Markov chain with a given invariant distribution. Theorem 15.52 (Berry–Esseen) Let X1 , X2 , . To this end we need a table of the Morse code as well as the frequencies of the letters in a typical text. The distribution PY of Y := 2X − 1 is sometimes called the Rademacher distribution with parameter p; formally Radp = (1 − p) δ−1 + p δ+1 . n→∞ n→∞ (viii) μ = v-lim μn and μ(E) ≥ lim sup μn (E). Define Y = max(X, a). (iv) The class of finite unions of bounded intervals is a ring on Ω = R (but is not an algebra). Indeed, the sets (−∞, y], y ∈ Rn form an ∩-stable generator of B(Rn) and x → κ(x, (−∞, y]) = μ((−∞, y − x]) is left continuous and hence measurable. Recall that the support of a real function f is f −1 (R \ {0}). Proof See, for example, the book of Grimmett [63, pages 287ff]. Then there exists a unique probability measure μ on σ (A) = B(Ω) such that μ([ω1 , . Let F = (Fn)n∈N0 be the filtration generated by X = (Xn)n∈N . a,b∈Q a 0, let τK := inf{n ∈ N :)X*n+1 ≥ K}. Show that 3 P[A|F] dP P[B |A] = 3B . If in addition Ω is finite, then A is finite. ♦ We introduce some further terms. Proof As μ* is subadditive, the other inequality is trivial. For every interior point x ∈ I , let b(x) be the maximal slope of a tangent of ϕ at x; i.e., the maximal number t with ϕ(y) ≥ (y − x)t + ϕ(x) for all y ∈ I (see Theorem 7.7). We denote by B(Ω, τ) = σ (τ) the Borel σ -algebra on (Ω, τ). 80 2 Independence The cases p = 1 and θ (p) = 0 (hence in particular the case p = 0) are trivial. (ii) There is an α ∈ {0, 1}I such that the family (Biαi)i∈I is independent. , Yn be independent random variables that are uniformly distributed on [0, 1]. (If X is i.i.d., then S is a Markov chain, and this implies immediately that 0 is recurrent. This representation of X is called the Doob decomposition. Then P[X ≥ n] = P[[ω10 , . Then ϕT V := sup ϕ(A) − ϕ(Ω \ A) : A ∈ A = ϕ(Ω +) − ϕ(Ω −) = ϕ + (Ω) + ϕ − (Ω) defines a norm on M± (Ω, A), the so-called total variation norm. This example suggests that we also have to make sure that no mass "vanishes at infinity". k=1 In particular, if A = A1 A2 , one similarly gets μ(A) ˜ = μ(A ˜ 1) + μ(A ˜ 2). .} ⊂ E be dense. Theorem 15.58 (Central limit theorem in Rd) Let (Xn)n∈N be i.i.d. random vectors with E[Xn,i] = 0 and E[Xn,i Xn,j] = Cij , i, j = 1, . with In ~ UΛ and Un ~ U[0,1] . be events and define A* = lim sup An . Denote by λ the Lebesgue measure on R. A map ϕ : G → R is called convex if for any two points x, y ∈ G and any λ ∈ [0, 1], we have ϕ λx + (1 − λ)y ≤ λ ϕ(x) + (1 − λ) ϕ(y). By n→∞ assumption, ϕPnk

Networks An electrical network (E, C) consists of a set E of sites (the electrical contacts) and wires between pairs of sites. As the individual experiments ought to be independent, we should have for any choice ω1 , . Then E[X] = θ ⇒ 0 x e−θx dx = Var[X] = −θ −2 + θ ⇒ 0 1 , θ x 2 e−θx dx = θ−2 ⇒ 2 −x x e dx = θ −2. (7.5) Hence g ∈ L1 (ν) if and only if g f ∈ L1 (μ), and in this case (7.5) holds. 191 191 195 203 9 Martingales .. Theorem 6.19 For finite μ, F ⊂ L1 (μ) is uniformly integrable if and only if there is a measurable function H : [0, ∞) → [0, ∞) with lim inf x→∞ H (x)/x = ∞ and sup f ∈F H (|f |) dμ < ∞. [149, Theorem 8.5]. Since the electrical conductances ... is identically distributed could be replaced by the existence of a sufficient number of moments ... Hence for any n the fact that the variances be bounded (see Exercise 5.3.1). Define the local energy that describes the energy level at a single atom at x ∈ Λ by h(σx) = H (x) ... The random ... Let the state of the whole system ... i=1 We interpret Zn as the size of a population at time n and Xn,i as the number of offspring of the ith individual of the nth generation. For any A ∈ A and n ∈ N, we thus get μ(A ∩ En) = μ(A) ... (17.8) ... Binomial distribution) By the binomial theorem, ... The main goal of this section is to express the probability ... F (x) , x ∈ R, for which (Xn) converges in probability ... measures μ and ν are called equivalent (symbolically μ ≈ ν) if ν μ and μ ν ... Theorem 2.43 For any p ∈ [0, 1] ... Hence, with infinite variance, the statement of the Central Limit Theorem ... with the random fluctuations, we have ... walk never returns to x1 in terms of effective resistances in the network. These balls, as well as their closures, are subsets of A1. ... Let U: Ω → R be ... Intuitively, (17.15) suggests that we define p ... Xsn = ... Then A ∩ B ... the distribution of a random variable, Theorem 2.34 ... Important special cases for the Lebesgue-Stieltjes measures are the following. ... The discrete analogue of the celebrated Black–Scholes formula for option pricing in certain time-continuous markets. ... For any ... If X, Y are integrable ... Finally, A is called Hölder-continuous of order γ if there exists a C < ∞ ... (X ∈ A) for all A ∈ ... Hence X = Y a.s. ... If X is a Markov chain ... E[X2 − 2X] ... the conditional distribution ... Theorem 8.6 (Summation formula) ... Let I be a countable set and let (Bi)i∈I be pairwise disjoint sets with P(Bi) > 0 ... Whether or not ... it is true can thus be determined on the basis of the information available at time t. By the law of large numbers, we thus have ... If X ... and Y ... We can do a little better ... Ω is called a regular version of the conditional distribution ... for the greatest generality but rather content ourselves with the key theorems for probability theory. ... If X is independent ... Kolmogorov's 0-1 law (Theorem 8.27) ... The converse ... Hence, If A ... Theorem 4.20 (Monotone convergence, Beppo Levi theorem) Let f1 , f2 , ... (iii) Assume that φ is differentiable at 0 and that X1 is integrable ... for a sequence of estimators ... Optional Sampling Theorems ... (x) For ... Galton–Watson process ... is ergodic if and only if any f -measurable f ... (Theorem ... Consider the Borel σ-algebra ... In particular, f is strictly convex and hence assumes its p p (unique) minimum at x0 = y 1/(p−1). Using Markov's inequality (Theorem 5.11), we estimate ... exists an Ac ∈ A with 0 < μ(Ac) < ∞ such that Ac ∈ C ... Furthermore, we must interpret, we use Rayleigh's monotonicity principle to show that ... be a bounded martingale. ... Conclude that for any random variable Y with ... the distribution is uniquely determined by the values ... Hence I is B(I) ⊗-measurable. (ii) Show that (Ω, A, δω0) is complete. It makes the following assumptions: ... Atoms are placed at the sites of a lattice Λ (for example, Λ = {0, ... We want to weaken the assumption in Theorem 15.38 that the random variables are identically distributed. For A ⊂ Ω, define the set of countable coverings F ... Hence X and μ ... the measurable space ... We have now developed the abstract framework (principle of large deviations) for the description of the probability of ... The first step ... (Y, r) ... one is called ... Concentration of ... Clearly ... Then X is a Markov process with ... (iii) ... (i) Basic Measure Theory Hence ... Hence there is an increasing sequence ... in the Radon–Nikodym theorem)". (N) is a strong Markov process (Theorem 17.14); hence we have ... Theorem 3.8 the random variables X1 , X2 , .. (23.14) Since we have ... there exists a bounded measurable function f and A ∈ G ∈ R ... In order to check the formulation ... and let A ... By Theorem ... Now let ... Φ is called a Gaussian process ... the random variable ... for all f ∈ N. n−→∞ ... be an ... The finite-dimensional distributions ... is the same ... Again we use ... Frequently, we will use ... Let f : Ω → [0, ∞] ... Theorem ... In this case ... for a Markov chain ... We have ... Denote by ... the Poisson ... Theorem 10.1 ... Example 8.31 Let X be ... k | ... (ii) ... is the ... Show that ... Consider the ... Takeaways ... Example 17.49 ... Definition 21.1 ... we have ... Clearly ... Let ... A ... Theorem 8.14 ... allows us to make the following definition (since the value of I (f) does not depend on the choice of the normal representation). In order to check the conditions of Theorem 1.53, we have to show that μ is σ-subadditive. ... Definition 9.6 If X is a random variable ... The algebra B(X) ... (Ω, A) is ... is a Markov ... such that ... and hence ... Let ... C . On the other hand, for all n ∈ N, ... The sequence (Xn) is ... in the Radon–Nikodym theorem)". ... is a strong Markov process (Theorem 17.14) ... Theorem 3.8 ... (iii) ... that is ... is a martingale ... independent ... the ... Example 8.31 ... The case ... Hence the claim ... Theorem 21.24 (Doob's regularisation) Let F be a filtration that satisfies the usual conditions and let X = (Xt)t∈R be an F-supermartingale such that t → E[Xt] is right continuous. ... paths. Further, show that strict inequality can hold in the lower bound (LDP 1). ... Lemma 13.10 Let (E, d) be a metric space. If ... is a Cauchy sequence ... (see Theorem 13.15) ... Hence, ... subsequence that converges to ... Which of the functions ... The ... Ez ... Show that ... is the probability ... Hence there exists an F-measurable ... is a martingale ... Theorem 1.14 ... by ... E. ♠ By the Markov property, for every n ∈ N ... is a ... X1 , X2 , ... are ... Hence ... dμ ... ∞ ... conditional expectation of a random variable X given a σ-algebra F is the best prediction in the sense of ... Theorem 13.23 ... limits. Then (Yn)n∈N satisfies E[|X| p] = ... ♠ 202 8 Conditional Expectations Takeaways The ... the above statements ... distribution is determined by a second moment characterization ... and by the covariance matrix. 21.4 Supplement: Feller Processes . 250 11 Martingale Convergence Theorems and Their Applications Fig. Takeaways Consider an exchangeable family X1 , X2 , . (9.3) By construction, XT = XT ∧ ... Exercise 5.5.1 Let X1 , X2 , ... Hence ... In the following, we construct independent successful ... (a) (b) ... We use ... X with ... Markov chain ... Theorem 17.13 ... transition probabilities for a stochastic process. A time-homogeneous Markov process is simply a stochastic process with the Markov property and for which the transition probabilities are time-homogeneous. We are now interested in properties of this process X that cannot be described in terms of finite-dimensional distributions but reflect the whole path t → Xt . (5.15) ... Then ... is almost surely ... Uniqueness of measures ... Reflection ... limits. Then (Yn)n∈N satisfies ... (a) ... Hence μ is ... (i) is the ... (17.27) ... 1 , ... the main ... (1.6) ... the random variable ... is ... the ... Theorem 3.25 ... In this section ... the distribution ... for all ... We ... Let ... Then ... (7.11) ... Let X ... the graph ... Theorem 5.8 ... Hence ... Define ... p ∈ ... μ ... This is ... X, Y ... (a) ... the Lebesgue ... Note that ... E[Xt] ... Lemma ... Lemma 13.18 ... For ... 1 ... Theorem 1.60 ... Then ... Markov ... By Theorem ... Hence ... for some ... 15 Characteristic Functions ... Stochastic ... (ii) ... be ... constant. Let Z1 , ... B be a measurable set. Then ... Hence ... is equivalent ... Xn ... Probability ... Theorem 17.27 ... Example ... Then (Pn) ... is uniformly ... 480 19 Markov Chains and Electrical Networks Exercise 19.4.1 Consider the electrical network on Zd with unit resistors between neighbouring points. , Zn}. Let ... 7.2 ... By ... Brownian motion ... Theorem ... Let ... independent of each other ... be a ... (ii) ... Markov ... Hence ... is ... The ... 20.4 ... The ... Bernstein ... Hence ... follows ... By Theorem ... σ ... Example ... is ... monotone ... We thank ... Manuel Mergens, Thal Nowik, Felix Schneider, Wolfgang Schwarz, and Stephan Tolksdorf. As an excuse for presenting this section in a chapter on Markov chains, we will finally use a simple Markov chain in order to prove a theorem on the stochastic order of binomial distributions. ... measure is regular in the sense that the measure of an arbitrary measurable set can be approximated by compact subsets as well as by open supersets. ... parameters n ∈ N and p ∈ (0, 1). We will ... on the ... Hence ... to one, as with linear maps and matrices, or it may map only some properties ... assumption of finiteness is essential. More precisely, for any ε > 0, we have ... E[X] ... Hence ... be independent standard normally distributed random variables. Recall that we have formalized the event where infinitely many of a series of events occur by means of the lim sup of these sets (see Definition 1.13). Furthermore, let F : R → R be monotone increasing and right continuous. For any n ∈ N, ... and let ... Show that ... (−∞, x] ... Hence indeed ... Define ... (i) is in ... Thus ... By ... is ... Theorem 13.8 ... Show that ... the ... Let ... Then ... is ... and let ... is ... Exercise 13.2.10 Let X, X1 , X2 , . Then (Xn)n∈N is exchangeable. Thus ... Now let f ≥ 0 be measurable and let (fn)n∈N be a sequence of nonnegative functions in L2 (Ω, A, P) with ... k ... F. 302 13 Convergence of Measures ... The unique solution of this differential equation is ... (1) the ... Cauchy ... The existence of the convolution semigroup follows by Corollaries 16.8 and 16.7 if we define ... Without proof, we present the following topological result ... and ... (see e.g. ... Theorem 13.21 ... Hence ... 2002 Conditional Expectations Example ... ... q ... n ... Then ... is ... Theorem 13.12 ... the density of a probability distribution on R with characteristic function ... and ... f ... (see e.g., ...). On the other hand, if a family of stochastic kernels κs,t , s ≤ t, fulfills a Chapman–Kolmogorov equation. ... Let ... i=1 Then the sequence ... we obtain ... Kolmogorov's theorem ... Similarly ... Indeed ... for every ... Take ... Theorem ... Hence ... X ... converges ... by the ... Hence, ... the ... Hence ... that the ... Let ... (iii) ... the ... to come up with ... Show that ... a ... the ... and ... B ... and ... measure-preserving ... and ... the ... for ... then ... (ii) ... Then ... Here we have ... quantities ... matrix ... of the ... (8.5) ... matrix ... a ... then ... let ... we ... is an independent ... increments ... Since ... positive ... the Markov ... Hence ... and ... the ... Gaussian ... Monte Carlo Method 451 Theorem ... Since Z is a tree ... of a ... Conclude ... Subtracting the differences ... to ... a submartingale ... Consider ... decomposition ... that ... be ... of a ... Then ... In ... Mixing ... The ... is ... and ... Assume ... Ω ... 1 ... mutually disjoint ... 21.5.5 Recurrence and Transience of Brownian Motion. For ... (Y ... If ... Theorem ... For ... Then ... If ... is ... of the ... (as ... binomial coefficient.) If r ∈ N, then ... (see ... the ... Show that ... Denote the ... The empirical distributions converge to the distribution of ... The ... Hence ... Theorem 15.4 ... Let ... 1 ... By ... exponential size-biasing of the distribution in ... PX1 of X1 , which turns the atypical values that are of interest here into typical values. A function f : E → R is called harmonic on E \ A if pf (x) = pf (x), y∈E p(x, y)f (y) exists and f(x) = f (x) for all x ∈ E \ A. ... Here we follow the proof of Burton and Keane [23] as described in [63, Section 8.2]. Takeaways ... Riemann integrable function on a compact interval is Lebesgue integrable and the integrals coincide. .), by the approximation theorem for measures, there exists a sequence ... of ... in ... J is ... Theorem 11.1 ... in the ... Then ... Perfect Sampling The MCMC method as described above is based on hope: We let the chain run for a long time and hope that its distribution is close to the invariant distribution. 15.4 Characteristic Functions and Moments . ... Define ... Show that the states 6, 7 and 8 are positive recurrent and compute the expected first entrance times E6 ... Hence, ... We have seen ... for the ... 210 8 Conditional Expectations Now, for any n ∈ N and B ∈ E, ... If ... there ... Here are some of the most important distributions in probability theory ... Hence, ... for which ... Show that ... (i) ... the ... Markov ... The ... i.i.d. ... Hence ... is ... to ... (ii) ... We have ... (E ... if ... we know ... a ... prophecy. Assume we have ... Hence ... The ... N0 and X0 ... and the ... probability generating function ... In ... Letting n → ∞ and using the σ-subadditivity of μ∗ ... we obtain ... (E) = μ∗ ... We will show ... of the elements of a sequence, as in the infinite case ... Let ... of ... there is at least one ... As ... Exercise 2.3.4 ... we ... to ... A ... Hence ... Proof ... Reflection ... Let ... of ... 0 ... is ... where ... Show that ... −sk , N − M + sk − k, if ... If ... 0 ... Then ... φ ... divisible. The following ... Let ... exist a ... Lemma 13.7 ... is ... additive and thus a measure. Apply Hölder's inequality to ... Hence ... Show that the ... Let ... If ... Since ... Let ... differentiable with derivative ... Lemma 14.10 ... and ... Moreover ... Hence ... is the ... set ... By ... is ... For ... 1 ... (ii) ... Exercise ... 14.3 ... to ... Then ... the ... Definition 9.42 (Binary model) A stochastic process ... The ... and ... 1 ... Hence ... is ... 5 ... of ... we ... and ... Takeaways ... Then ... 1 ... 28 ... 6 ... 121 ... 125 ... 135 ... 139 ... 2 ... x ... ... conditional ... 3 ... Convergence Theorems ... Let A, ... E ... 5.4 Speed of Convergence in the Strong LLN ... This statement ... Since ... we have seen ... bounded ... the number ... is ... bounded ... a ... independent ... Show that ... the measure ... Let ... given ... Then ... we ... i=1 ... Let ... Then ... be ... Reduced ... tree ... (x0 ... the ... are ... the ... For ... and ... ... For ... and ... between these points ... Since ... (ii) ... Let ... Show that ... (x ... B ... 2.4 ... Now ... two ... the ... is ... uniform ... the ... (ii) ... with ... of ... the ... Then ... it ... Here ... the ... Hence ... is ... and ... the ... convergence ... By ... (μ ... (Kolmogorov–Chentsov theorem) ... Theorem 21.6 ... and ... Ω ... with ... A ... is ... for ... of ... parameters of ... Hence ... We ... In ... Then ... A ... which ... in ... the ... Ez ... Bernoulli ... Example 3.12 ... Since ... have ... Theorem 16.36 ... then ... n ≥ 0 ... (19.1) ... Series ... Theorem ... As ... of ... Lemma ... ... Theorem 2.33 ... Example 4.3 ... then ... or ... 1 ... Theorem ... 9 ... 17.4 ... Corollary ... (ii) ... if ... Chapter 16 Infinitely Divisible Distributions For every n, the normal distribution Nμ,σ 2 is the nth convolution power of a probability measure (namely, of Nμ/n,σ 2 /n). (For the "a.s." notation see Definition 1.68.) Lemma 5.18 For n ∈ N, define Yn := Xn 1{|Xn |≤n} and Tn := Y1 + · · · + Yn . Let ... Consider now the general case. Concluding, X := F −1 is the random variable that we wanted to construct. In Theorem 17.8 we saw that, for the semigroup of kernels ... the ... Proof ... equivalent of ... Show that ... The ... Let ... (P ... Proof (i) The right-hand ... Hence ... the ... Assume ... As an ... Let ... Denote by ... If ... The ... of some topological definitions and theorems. This implies ... Hence ... of ... Denote by ... If ... in ... Example 15.6 ... Let ... Proof ... Reflection ... Find an example that shows that without the tightness assumption, we need not have F (−∞) ... We ... Hence ... is ... for ... (ii) ... ... In ... is a ... From ... Hence ... If ... Show that ... for ... Proof Show that ... (iii) ... is ... X ... We now ... X ... is ... Then ... we ...

in Example 1.58(iv). Reflection Find an example that shows that without the tightness assumption, we need not have F (−∞) = 0 nor F (∞) = 1.♠ We come to a first application of Prohorov's theorem. , k + 1] = k+1 P[Xm,l = 0] = (1 − p)k+1 . In this case, the probability that at time 2n all coordinates are zero would be the Dth power of the probability that the first coordinate is zero. Hence {τ = s} ∩ Xs−1 (A) ∈ Fs ⊂ Ft for all s ≤ t. (i) Normal distribution: Nμ1 ,σ 2 ∗ Nμ2 ,σ 2 = Nμ1 +μ2 ,σ 2 +σ 2 for all μ1 , μ2 ∈ R 1 2 1 2 and σ12 , σ22 > 0. Hence averaging over one realization of many random variables is equivalent to averaging over all possible realizations of one random variable. , xn) for x ∈ E n and for x ∈ E N . .+ck = 1, then Y is called the moving average of X (with weights c0 , . This coupling shows that bn,p1 ≤st bn,p2 . n→∞ 3 n→∞ 3 (ii) f dμn −→ f dμ for all bounded Lipschitz continuous f . In this case, the derivative is φ (x) = D + φ(x). Let such that μ(W n n n ∞n) ≤ μ(Cn) + ε 2 −n−1 ≤ μ(B)+ε. Then n n (' tl 1Al (Y1) = 1 + ν(Al) ei tl − 1 , ψ(t) := E exp i l=1 t ∈ Rn , l=1 is the characteristic function of (1A1 (Y1), . , ωn0]] = (1 − p)n . In the general case, write X = X+ − X− and Y = Y + −Y − and exploit the linearity of the conditional expectation. If Σ is finite or is a bounded subset of an Rd , then by symmetry, typically λ is the uniform distribution on Σ. (ii) The map κ : E × B(E)⊗I → [0, 1], (x, B) → Px [X ∈ B] is a stochastic kernel. Denote by LipK (E; F) the space of Lipschitz continuous functions with constant K and by Lip(E; F) = K>0 LipK (E; F) the space of Lipschitz continuous functions on E. Possibly all martingales Y ? That is, {Y ∈ Am } = {τ (Y) ∈ Am }. n→∞ a (4.6) n→∞ Theorem 4.23 (Riemann integral and Lebesgue integral) Let f : I → R be Riemann integrable on I = [a, b]. Example 20.36 (Rotation) We come back to the rotation of Example 20.9. Let Ω = [0, 1), A = B(Ω), P = λ the Lebesgue measure, r ∈ (0, 1) and τr (x) = x + r (mod 1). By Pólya's theorem (Theorem 17.40), this random walk is recurrent. i=1 Consider now the random variable Y = max(X1 , . μ λ = e−(μ+λ) n! n! m m=0 Hence Poiμ ∗ Poiλ = Poiμ+λ . Clearly, we have fn (0) = 0 for all n ∈ N. If μ is a probability measure and if every κi is stochastic, then k=0 k=0 μi is a probability measure. However, the qualitative behaviour will be quite different. Then (Xi)i∈I is exchangeable. Lower semicontinuity follows from the monotone convergence theorem (Theorem 4.20). Fix an arbitrary x0 ∈ E0 ∩ E 0 . ♠ Example 9.30 Let Y1 , . Evidently, the two-dimensional integer lattice is isomorphic to its dual graph. , gm) starting and ending in some point g0 = gm = u ∈ UL , then (g1 , g2 , . With (i), the interpretation is clear. Consider the generalized version of Pólya's urn model (Xn)n∈N0 with rk = r and sk = s for all k ∈ N. If E[X] ∈ ∂I , then X = E[X] a.s.; hence E[φ(X)] = E[φ(E[X])] = φ(E[X]). ♦ Remark 4.25 An improperly Riemann integrable function f on a one-sided open interval I = (a, b] or 3I = [0, ∞) is not necessarily 3 n Lebesgue integrable. The implication (i) ⇒ (ii) was shown in Theorem 18.12. Definition 17.58 Let μ1 , μ2 ∈ M1 (Rd). One possibility is to show the claim first for measures on Rd . ♣ 19.2 Reversible Markov Chains Definition 19.8 The Markov chain X is called reversible with respect to the measure π if π({x}) p(x, y) = π({y}) p(y, x) for all x, y ∈ E. 15.2 Q-Q-plots for S100 abscissa shows the quantiles of the standard normal distribution. An (open) path (of length n) in this subgraph is a sequence π = (x0 , x1 , . Then X0 + Mn ◦ τ = X0 + Sk ◦ τ = Sk+1 . That is, each point in T has exactly three neighbours. Without exaggeration, it can be stated that Brownian motion is the central object of probability theory. 2 ε n σ ε n By Slutzky's theorem (Theorem 13.18), we thus have convergence of the finitedimensional distributions to the Wiener measure PW : n→∞ PS˜ n ⇒ PW . Let π be the invariant distribution of X. x → D − φ(x) is left continuous and x → D + φ(x) is right continuous. If Ω is at most countably infinite and if A = 2Ω , then the measurable space (Ω, 2Ω) is called discrete. Definition 7.20 For f, g ∈ L2 (μ), define)f, g* := fg dμ. Note that Re(z) = (z + z)/2 and Im(z) = (z − z)/2i imply cos(x) = eix + e−ix 2 and sin(x) = eix − e−ix 2i for all x ∈ R. , Yn) is measurable with respect to σ (X1 , . For x ∈ E, define (see Corollary 14.46) on (Ω, A) the probability measure Px such that, for finitely many time points 0 = t0 < t1 < . i , i = 0, . Proof By Theorem 6.19, there exists a monotone increasing convex function f with the property that f (x)/x → ∞, x → ∞ and L := supi∈I E[f (|Xi |)] < ∞. Hence we have) * P Nti − Nti−1 = ki for i = 1, . Just before each gamble we decide how much money we bet. For example, we want to compute the distribution of the functional F (X) := supt ∈[0,1] Xt . (ii) τU is an F+ -stopping time but in general (even for continuous X) is not an F-stopping time. This clearly implies (21.30) with δ = (η/K)1/γ . In this case, we have the Fourier inversion formula, f (x) = (2π)−d e−i)t,x* φμ (t) λ(dt). z e δb,t Similarly, z e #c,t (vi) (Exponential distribution) This follows from (v) since expθ = Γθ,1 . (Compare [32, Exercise 2.1.24], see also [43, Section II.7].) Varadhan's lemma has various applications in statistical physics. 327 327 336 344 349 356 365 16 Infinitely Divisible Distributions . In particular, B(Rd) = B(R)⊗d for i∈I d ∈ N. To this end, we will need (21.5). Then A = {1, 3, 5} and B = {1, 2, 3}. Let Ck ⊂ E k be measurable with Ak = {(X1 , . f is p q twice continuously differentiable in (0, ∞) with derivatives f (x) = x p−1 − y and f (x) = (p − 1)x p−2. Theorem 15.30 (Bochner) A continuous function φ : Rd → C is the characteristic function of a probability distribution on Rd if and only if φ is positive semidefinite and φ(0) = 1. , Mn,m = km } ∩ {L = n}) * P Mn,1 = k1 , . Hence, for m ≥ n ≥ n0 , we have f dμ˜ n = lim l→∞ f 1W n dμkln = lim l→∞ = lim l→∞ f 1W n dμklm f 1W m dμklm and thus f dμ˜ n = lim m→∞ f dμkmm . Proof Let p be the transition matrix of X. ♠ Example 18.19 Let r ∈ (0, 1) and N ∈ N, N ≥ 2. This probability should equal one. The frequencies for other languages can be found easily, e.g., at Wikipedia. Remark 17.26 The condition (17.16) cannot be dropped easily, as the following example shows. Then ν F 0 μ F (since F F C C in F there are fewer μ-null sets); hence the Radon–Nikodym derivative fF := @ dν F dμ F exists. Consider the electrical network with unit resistors. Now consider the set G ⊂ A1 ⊗ A2 such that A ∈ G if and only if (14.6) and (14.7) hold for f = 1A . We show only the existence of a decomposition. fdd n→∞ (ii) Pn −→ P weakly. Further, for every μ ∈ F , ∞ μ (A)c ≤ μ(Ac) ≤ μ(Acn, N) ≤ ε. (23.13) μ({x}) Since μ({x}) > 0 for all x ∈ Σ, the integrand ν-a.s. is finite and hence the integral also is finite. (i) Let X ∼ Nμ,σ 2 and let a ∈ R \ {0} and b ∈ R. .) is independent (but not necessarily independent of each other). The nth and the (n + 1)th superconductors are connected by 4(2n + 1) edges. For any x ∈ E, there exists a relatively compact neighborhood Bx x. , An ∈ B(Rd) . b−a a,b a,b := lim Manifestly, By assumption, we n→∞ Un exists. The sets A ∈ A are called measurable sets. (iii) (Linearity) If α, β ∈ [0, ∞], then (αf + βg) dμ = α f dμ + β g dμ, where we use the convention ∞ · 0 := 0. Further, let Rn = # {S1 , . < tnn = b) that get finer and finer. Since C1 ∩ C2 = ∅ and C1 ∪ C2 ⊂ A, we get β(A) ≥ α(C1 ∪ C2) = α(C1) + α(C2) ≥ β(A ∩ B c) + β(A ∩ C1c) − 2ε ≥ μ∗ (A ∩ B c) + μ∗ (A ∩ B) − 2ε. Exercise 5.1.1 Let X be a nonnegative random variable with finite second moment. Recall the definition of a general coupling of two probability measures from Definition 17.54. be i.i.d. random variables and for n ∈ N, let Sn = X1 + . We start with a simple observation. Therefore, defining S,T ∈N X˜ t (ω) := Xtt (ω), t ≥ 0, for ω ∈ Ω \ Ω∞ , we get that X˜ is a locally Hölder-continuous modification of X on [0, ∞). (i) (ii) (iii) (iv) φ is convex. Theorem 14.31 (Kernels and convolution) Assume X1 , X2 , . (i) E = R and there is a p ∈ [1, ∞) with fn ∈ Lp (μ) for all n ∈ N and there is ∞ p an f ∈ Lp (μ) with fn − f p < ∞. f d μ˜ m 292 13 Convergence of Measures This implies that for any measurable relatively compact set A ⊂ E, we have μ˜ m (A) = μ˜ N(A) (A) for any m ≥ N(A). Define α(t) := E[Nt]. If the jump rates are bounded, then the process can be constructed as a Markov chain at random times given by a Poisson clock. We come to the main theorem of this section. In this case, I (x, y) is called the flow from x to y and u(x) is called the electrical potential (or voltage) at x. In particular, this holds if Ee < ∞ for some t > 0. Furthermore, μ˜ is σ -finite. If p > 1 and if X1 , X2 , . Conclude that every Borel set B ∈ B(Rn) is a Borel space. In either case, a probability measure on the product is uniquely determined by its values on cylinder sets. If m ≤ 1, then (Zn)n∈N converges a.s. to some random variable Z∞ . with PTnr = PTns = expwn−1 . (Of course, the chain with transition matrix p(x, y) = π(y) converges to π, but this does not help a lot.) This method of producing (approximately) π-distributed samples and using them to estimate expected values of functions of interest is called the Markov chain Monte Carlo method or, briefly, MCMC (see [15, 112, 119]). n→∞ (21.25) Proof By (21.25), we have Xt = X˜ t a.s. for all t ∈ [0, 1]. .) ∈ B = *) Eπ 1Aε 1{XN =x} 1{Xn =y} (Xn , Xn+1 , . , eN } such that pe1 ≥ pe2 ≥ . Show that, analogously to the Radon–Nikodym theorem, the following two statements are equivalent: (i) φ(A) = 0 for all A ∈ A with μ(A) = 0. ∈ DE are mutually disjoint. , k. However, this is not a distribution function, as 1 does not converge to 0 for x → −∞. l (N)n l=1 Here we defined (n)l := n(n − 1) · · · (n − l + 1). By Lemma 13.5, there exists a compact set K with μ(K c) < ε/2. Then f (x) := G(x, a) is harmonic on E \ {a}: For x = a, we have pf (x) = p ∞ pn (x, a) = n=0 ∞ pn (x, a) = G(x, a) − 1{a} (x) = G(x, a). (2.16) Accordingly, let m = 2, 3, . (18.14) As all eigenvalues are real, the corresponding eigenvectors are given by xkn = 2 r 1−r n/2 sin nπ N , k = 1, . + T D , Yti := Z i i for i = 1, . j ♦ (10.5) Example 10.9 We want to generalize the preceding example further. ⊂ E be relatively compact open sets covering E. The investigation of product spaces and their σ -algebras is, however, postponed to Chap. Together with the topological considerations in Sect. n=1 Euler's prime number formula is a representation of the Riemann zeta function as an infinite product −1 ζ (s) = 1 − p−s , (2.5) p∈P where P := {p ∈ N : p is prime}. ♦ 456 18 Convergence of Markov Chains Example 18.20 (Gambler's ruin) We consider the gambler's ruin problem from Example 10.19 with the probability of a gain r ∈ (0, 1). By the preceding example, we conclude the following theorem. This property of λn is called inner regularity. Define h= ∞ 2−n 1 + μ(An))−1 1An . Example 12.21 Any family (Ai)i∈I of sub-σ -algebras of F is independent given F . Hence the variational problem for Λ∗ (x) admits a unique maximizer t ∗ (x). Counterexample for semirings: Let Ω = {1, 2, 3, 4}, A1 = {∅, Ω, {1}, {2, 3}, {4}} and A2 = {∅, Ω, {1}, {2}, {3, 4}}. As a simple conclusion of Lemma 14.44 and Theorem 14.45, we get the following statement that we formulate separately because it will play a central role later. are independent with distribution Ξ∞ given Ξ∞ . ♠ On the one hand, improperly Riemann integrable functions need not be Lebesgue integrable. (18.2) Lx,y is uniquely determined, and we have Lx,y + Ly,z + Lz,x = 0 (mod d) for all x, y, z ∈ E. expectation (Theorem 8.14(ii)), for t > s, we have E[Zt Fs] ≤ E[Yt Fs] ≤ Ys and E[Zt Fs] ≤ E[Xt Fs] ≤ Xs ; hence E[Zt Fs] ≤ Xs ∧ Ys = Zs . ♦ Takeaways For a submartingale, the mean future is better than the present. Show that there exists a probability space (Ω, A, P) and real random variables X1 , X2 , . (20.9) x,y∈E ♦ Example 20.33 (Integer rotation) Consider the rotation of Example 20.8. Let n ∈ N \ {1}, E = Z/(n) and let P be the uniform distribution on Ω. For any ω ∈ Ω and n ∈ N, let πn (ω) := n pXi (ω) i=1 be the probability that the observed sequence X1 (ω), . (4.8) 0 3Proof Define 3 f = f ! and f = "f #. , 6}, A = 2Ω and P is the uniform distribution on Ω. Check that X(n) = Yn .* Show that the conditional distribution L X(1) − X(n) , . Inductively, we get E[|Xt |] < ∞ for all t ∈ N0 . If X is aperiodic and positive recurrent with invariant distribution π, then we ; ; n→∞ have ;Lμ [Xn] − π ;T V −→ 0 for all μ ∈ M1 (E). For any finite subset J ⊂ I , let FJ := F(Xj)j∈J : RJ → [0, 1], + , *) Xj−1 (−∞, xj] . In this case, we can choose A = E or A = T . Hence, by Theorem 6.19, there exists a monotone increasing convex map f : [0, ∞) → [0, ∞) p with f (x) x → ∞ for x → ∞ and C := E[f (|X0 |)] < ∞. 6} such that A1 = A˜ 1 × {1, . 9.2 Martingales .. Hence h(P, τr) = h(P, τr , P). , xn ∈ R, + F{1,...,n} ((x1 , . By Theorem 19.7, u is harmonic and can be written as ' (ux1 ,A0 (x) = Ex 1{XτA ∪{x } =x1 } 0 1) * = Px τx1 < τA0 for every x ∈ E \ (A0 ∪ {x1}). If M ⊂ N, then denote by gcd(M) the greatest common divisor of all n ∈ M. , n = 1− n e−θi x = 1 − exp − (θ1 + . Further, let W be the vector space of such ha,b . In fact, in Example 15.5, we saw that here the moments do not determine the distribution of X. ♣ 21.2 Construction and Path Properties Definition 21.8 A real-valued stochastic process B = (Bt , t ∈ [0, ∞)) is called a Brownian motion if 21.2 Construction and Path Properties 523 3 2.5 2 1.5 1 0.5 0 0.5 1 1.5 2 Fig. We perform the proof by induction on n. Proof "(ii) ⇒ (i)" For any f ∈ V , by definition of the inner product, the map x →)x, f * is linear. 26.3 Weak Uniqueness via Duality . (vi) The Poisson distribution is infinitely divisible with Poiλ = Poi∗n λ/n . n=1 By Kolmogorov's 0-1 law (Theorem 2.37), T is P-trivial. B∈Pn+ Due to the finite additivity of φ, we have φ(An) = B∈Pn B⊂An φ(B) ≤ B∈Pn+ B⊂An φ(B) ≤ B∈Pn+ φ(B) = φ(Cn). For α ∈ [−1, 1], define probability densities fα on (0, ∞) by fα (x) = f (x) 1 + α sin(2π log(x)) . Then μ(A) > 0 and (ν − εμ)(E) ≥ 0 for all E ⊂ A, E ∈ A. By Fatou's lemma (Theorem 4.21), we obtain n→∞) *) * lim sup P Aεn = 1 − lim inf E 1(Aεn)c n→∞ n→∞ + , ' (ε ε c ≤ 1 − E lim inf 1(An) = E lim sup 1An = 0. Consider the most prominent case E = R equipped with the Euclidean metric. The maximizer t ∗ = t ∗ (z) of the variational problem for Λ∗ solves the equation z = Λ (t ∗) = tanh(t ∗). (iv) The distribution of X1 can be chosen such that φ is differentiable at 0 but E[|X1 |] = ∞. (ii) For any A ∈ F , we have E[X1A] = E[Y 1A]. In fact, Sn = ni=1 (1 − Di). Proof " ⇒ " First assume μ is infinitely divisible. Indeed, without loss of generality, assume 0 is the left boundary of I and A := {E[X |F] = 0}. (i) I (αf) = α I (f). The full procedure is rather technical and does not allow for a smooth intuitive description. Example 12.5 (i) For x ∈ RN , define the nth arithmetic mean by an (x) = n1 ni=1 xi . However, the notational complications become overwhelming for n ≥ 3, and the idea for general n ∈ N becomes clear in the case n = 2. π The second largest modulus of an eigenvalue is |λn | = σ cos N if n = 1 or n = N − 1. This implies (' (φn,l (t) − 1 ≤ E eit Xn,l − 1 1{|X |≤ε} + E eit Xn,l − 1 1{|X |>ε} n,l n,l ' (2 1{|Xn,l |>ε} . Choose C1 ∈ C with C1 ⊂ A ∩ B c and α(C1) > β(A ∩ B c) − ε. Finally, P[Nε ≥ 2] = 1 − e−αε − α ε e−αε = f (0) − f (αε), where f (x) := e−x + xe−x . (ii) Since φ is convex, so is x → φ(x)+ . (i) What is the probability that the last person gets his or her reserved seat? By Theorem 13.16, lim n→∞ f d μn ◦ φ −1 = lim n→∞ (f ◦ φ) dμn = (f ◦ φ) dμ = f d μ ◦ φ −1 . , n} be such that P[X = k] = n k p (1 − p)n−k . Next construct a sequence of maps that approximates f uniformly on a suitable set C. ♣ Exercise 15.1.2 Let d ∈ N and let μ be a finite measure on [0, ∞)d . Jensen's inequality can be extended to Rn . .) = σ (Y−n , Y−n−1 , Y−n−2 , . The question is a bit tricky since for every given A ∈ A, the expression P[A|X = x] is defined for almost all x only; that is, up to x in a null set that may, however, depend on A. However, a σ -finite measure on Z is not uniquely determined by the values on E: Let μ be the counting measure on Z and let ν = 2μ. Takeaways Mixing is a concept of independence stronger than ergodicity but weaker than stochastic independence. (In particular, continuous maps are lower semicontinuous. ♣ Chapter 17 Markov Chains In spite of their simplicity, Markov processes with countable state space (and discrete time) are interesting mathematical objects with which a variety of realworld phenomena can be modeled. (12.7) Proof By Theorem 12.10, An (φ) = E[φ(X) En]. Y := n=1 (iii) (In particular, if we let N ∼ Poiλ in (ii), then φY (t) = exp(λ(φX (t) − 1)). , y) ∈ R. In order for the measure extension to work, σ -additivity is decisive. + Xn for n ∈ N0 . It is maximal (in fact, log(#E)) if p is the uniform distribution on E. Show that for any ε > 0, there exists a step function h such that f − hp < ε. Let (E, d) be a metric space and let F ⊂ M≤1 (E) be tight. 138 5 Moments and Laws of Large Numbers Theorem 5.30 (Rademacher–Menshov) Let X1 , X2 , . Left side: below the critical temperature (β > βc); Right side: above the critical temperature. ♠ Theorem 18.11 Let p be the transition matrix of an irreducible, positive recurrent, aperiodic Markov chain on E. Hence X = (Xn)n∈N0 is a stationary real-valued stochastic process. 15.4 Characteristic Functions and Moments 355 (ii) Assume that for any k ∈ N the limit mk := lim mk (Xn) n→∞ exists and is finite (note that finitely many of the mk (Xn) may be undefined for any k.) Show that there exists a real random variable X with mk = mk (X) for all k ∈ N and a subsequence (Xnl)l∈N such that l→∞ PXnl −→ PX weakly. Let f : Rd → R be continuous with compact support and let ε > 0. (iii) Show the theorem of Fréchet–Shohat: If in (ii) the distribution of X is determined by its moments mk (X), k ∈ N (see Corollary 15.33), then n→∞ PXn −→ PX weakly. (ii) For this particular ψ, all the iterates are of a special form and can be computed explicitly. Indeed, {Ym ≤ k} = k+1 {Xm,l = 1} ∈ σ (Xm,l , l = 1, . Clearly, {τ ≤ t} ⊂ A for all t ∈ I ; hence A ∩ {τ ≤ t} = {τ ≤ t} ∈ Ft . For r ∈ [0, 1), define τr : Ω → Ω by τr (x) = x + r − x + r! = x + r (mod 1). Now fix p ∈ [0, 1] and let X1 , X2 , . Definition 21.4 (Path properties) Let I ⊂ R and let X = (Xt , t ∈ I) be a stochastic process on some probability space (Ω, A, P) with values in a metric space (E, d). Here as the orthonormal basis of L2 ([0, 1]) we use b0 = 1 and bn (x) = √2 cos(nπ x) for n ∈ N. That is, the new state space is E˜ = E ∪ {Δ} and the transition matrix is ⎧ p(x, y), ⎪ ⎨ p(x, ˜ y) = 0, ⎪ ⎩ 1, if x ∈ E \ A, y = Δ, if x ∈ E \ A, y = Δ, (19.2) if x ∈ A ∪ {Δ}, y = Δ. The core of the weak law of large numbers is Chebyshev's inequality. This implies μ1 (C) = μ2 (C); hence μ1 = μ2 . p Proof Let f ∈ L∞ (μ) and p ∈ [1, ∞). n pk φμk . Since we will not need these statements in the following, we only refer to the standard literature (e.g., [174, Chapter VI.2] or [54, Theorem XV.3.3 and Equation (XV.3.8)]). Let f : Ω1 × Ω2 → R be measurable with respect to A1 ⊗ A2 . Klenke, Probability Theory, Universitext, 53 54 2 Independence (i) Two events A and B should be independent, e.g., if A depends only on the outcome of the first roll and B depends only on the outcome of the second roll. , ωn ∈ E}, (1.8) and let A := ∞ n=0 An . The main theorem of this section is the functional central limit theorem, which goes back to Donsker [35]. 12.3 De Finetti's Theorem In this section, we show the structural theorem for countably infinite exchangeable families that was heuristically motivated at the end of Sect. 707 Chapter 1 Basic Measure Theory In this chapter, we introduce the classes of sets that allow for a systematic treatment of events and random observations in the framework of probability theory. be maps E → n→∞ R with fn −→ f pointwise. Then lim φSn∗ (t) = e−t 2 /2 n→∞ for all t ∈ R. (v) If (Xt)t ∈N0 is a supermartingale and E[XT] ≥ E[X0] for some T ∈ N0 , then (Xt)t ∈{0,...,T } is a martingale. Hence, we obtain) * n→∞) * Lx λ1 Z˜ tn1 + λ2 Z˜ tn2 −→ Lx λ1 Yt1 + λ2 Yt2 . The claim is immediate. Definition 21.33 Let P be the probability measure on Ω = C([0, ∞)) with respect to which the canonical process X is a Brownian motion. This implies that, for any α ∈ R, the set {f ≤ α} = {g ≤ α} ∩ {g = h} {f ≤ α} ∩ {g = h} is the union of a B(I)-set with a subset of a null set and is hence in B(I)∗ (the Lebesgue completion of B(I)). Definition 14.9 (Cylinder sets) For any i ∈ I , let Ei ⊂ Ai be a subclass of the class of measurable sets. Define u(t) = Re(φ(t)). A random measure X with independent increments is called a Poisson point process (PPP) with intensity measure μ if, for any A ∈ Bb (E), we have PX(A) = Poiμ(A) . With a little effort it is possible to construct φ as a (random) additive set function. However, then fnkl ∈ U for all but finitely many l, which yields a contradiction! Corollary 6.15 Let (E, d) be a separable complete metric space. k k=0 (iii) Let p ∈ (0, 1] and X : Ω → N0 with P[X = n] = p (1 − p)n for any n ∈ N0 . Reflection Typically, the distribution of W cannot be computed explicitly. Hence, let p ∈ (1, ∞). The integral of an integrable function on the product space can be computed by successive integration (in arbitrary order) over the individual coordinates (Fubini's theorem). , ck ∈ R. Then the map If : Ω1 → [0, ∞], ω1 → f (ω1 , ω2) κ(ω1 , dω2) is well-defined and A1 -measurable. Theorem 5.16 Let X1 , X2 , . Hence X is an F-martingale. We model the experiment on the probability space (Ω, A, P), where Ω = {1, . An example are the first rain drops you see on the side walk. Hence, by Theorem 19.30, random walk on Z3 is also transient. 24.1 for a simulation of a Poisson point process on the unit square. The bars show the probabilities for values in [x, x + 0.01), 3.3 ≤ x ≤ 3.7. 0 2000 4000 6000 8000 10000 Fig. Using the strong Markov property, we show the reflection principle for Brownian motion. 6.1). We thus get a random sequence (Xn)n∈N0 of states in {0, 1}Λ that represents the random evolution of the opinions of the whole colony. + Definition 21.22 A filtration F = (Ft)t ≥0 is called right continuous if F = F+ , where Ft = s>t Fs . n Then the characteristic function φ of X is analytic and the distribution of X is n uniquely) t |X| * determined by the moments E[X] , n ∈ N. , Xk = xk , τx1 = ∞ ' (= Px [X1 = x1 , . Hence, in terms of the Markov chain notation, we have E = ZD and p(x, y) = 1 2D , 0, if |x − y| = 1, else. Show that, for every X ∈ L1 (P), we have E[X |A0] = 1 X ◦ g. Hence, for any n ∈ N, f (x0) = (pA)n f (x0) = n pA (x0 , y)f (y) ≤ m n (x) y∈SA 0 with equality if and only if f (y) = m for all y ∈ SAn (x0). We define the conditional expectation of Y given X = x by E[Y |X = x] := φ(x), where φ is the function from (8.10) with Z = E[Y |X]. The transition matrix is ⎧ p(x1 , x2) · p(y1 , y2), ⎨ p¯ (x1 , y1), (x2 , y2) = p(x1 , x2), ⎩ 0, if x1 = y1 , if x1 = y1 , x2 = y2 , if x1 = y1 , x2 = y2 . Hence the claim follows from (vi). , ωn : [ω1 , . Furthermore, E[X(X − 1)] = n k(k − 1) P[X = k] k=0 = n k=0 n k k(k − 1) p (1 − p)n−k k n n − 1 k−1 = np · (k − 1) (1 − p)(n−1)−(k−1) p k−1 k=1 = n(n − 1)p2 · n n−2 k=2 k−2 pk−2 (1 − p)(n−2)−(k−2) = n(n − 1)p2 . F is called a filtration if Fs ⊂ Ft for all s, t ∈ I with s ≤ t. ., we have D (X1 , X2 , . Hence, 0.5 0.0 0 1/4 1/2 3/4 −0.5 −1.0 −1.5 Fig. k (2.11) Here n n n! = = k k1 , . (ii) The random walk goes to −∞ at positive speed. This linear differential equation with initial value φ(0) = 1 has the unique 2 solution φ(t) = e−t /2 . (ii) Show that the transition matrix p is uniquely determined by f and d. However, if X is a martingale, then (H · X) is a martingale with (H · X)0 = 0; hence clearly E[(H ·X)T] = 0. By the dominated convergence theorem (Corollary 6.26), the limiting function f (· , x0) is in L1 (μ) and F (xn) − F (x0) lim = lim n→∞ n→∞ xn − x0 gn (ω) μ(dω) = f (ω, x0) μ(dω). every g ∈ Ef . We could argue more formally to show that only the constant states are stable: Let)M* be the square variation process of M. ∩ An) = n k=1 {i1 ,...,ik }⊂{1,...,n} (−1)k−1 μ(Ai1 ∪ . Example 14.15 For i = 1, . ♣ Exercise 15.3.2 Show that for any δ > 0 and ε > 0, there is a C < ∞ such that for any μ ∈ M1 (R) with characteristic function φ, we have ε μ([−δ, δ]c) ≤ C (1 − Re(φ(t))) dt. Now consider bond percolation on T with probability p. Theorem 17.39 If E is finite and X is irreducible, then X is recurrent. ♦ Reflection Check the statements of the preceding remark! ♠ Definition 15.8 For μ ∈ Mf (Rd), define the map φμ : Rd → C by φμ (t) := ei)t,x* μ(dx). By Step 2, we have gx ∈ C. ♣ Exercise 8.2.2 Let I ⊂ R be an arbitrary interval and let X ∈ L1 (Ω, A, P) be a random variable such that X ∈ I a.s. For F ⊂ A, show that E[X |F] ∈ I a.s. Is this statement still true if we require only X− ∈ L1 (Ω, A, P) instead of X ∈ L1 (Ω, A, P)? Theorems of this type are also called invariance principles since the limiting distribution is the same for all distributions Yi with expectation 0 and the same variance. 4(2n + 1) 19.4 Recurrence and Transience 0 1 4 edges 477 2 n 12 edges n+1 4(2n + 1) edges Fig. Let B ∈ Fn and m ≥ n. By Theorem 5.4, Y1 · Y2 is integrable. "≤" This follows from Hölder's inequality. ♣ Exercise 6.1.3 (Egorov's theorem (1911)) Let (Ω, A, μ) be a finite measure space and let f1 , f2 , . 17.4, in particular, Definitions 17.29 and 17.34. ♣ n→∞ Exercise 11.2.5 Show the following converse of Theorem 11.14. Show that for any ε > 0, there is a continuous function h : R → R such that f − hp < ε. In practice, it is often not possible to check if a map X is measurable by checking if all preimages X−1 (A), A ∈ A are measurable. Takeaways In this section, we have compiled a wish list of the properties that a probability assignment should have: σ -additivity and normalization (Definition 1.28). Hence, clearly, N→∞ d(fN , g) −→ 0, and thus d is complete. If A ∈ I, then 1A is I-measurable and hence P-a.s. equals either 0 or 1. A map h : R → R is called a step function ifthere exist n ∈ N and numbers t0 < t1 < . By virtue of the Borel–Cantelli lemma, show that 1 lim sup Xn = n→∞ n 0 a.s., if E[X1] < ∞, ∞ a.s., if E[X1] = ∞. Define 2 i=1 i=1 A := Nn ∞ B1/n xin . Klenke, Probability Theory, Universitext, 367 368 16 Infinitely Divisible Distributions choose a real-valued CFP φ for which |φ| = φ is also a CFP (see Examples 15.17 and 15.18). Hence we have Ai ∈ I for i = 0, . (b) For α ≥ 0, we have αf dμ = αf + dμ − αf − dμ = α f + dμ − α f − dμ = α f dμ. Theorem 1.81 (Measurability on a generator) Let E ⊂ A be a class of A measurable sets. 334 15 Characteristic Functions and the Central Limit Theorem Exercise 15.1.1 Show that, in the Stone–Weierstraß theorem, compactness of E is essential. (ii) Consider an urn with N balls, M of which are black. , n − 1, changes its value exactly twice. n! (3.4) Hence X + Y has probability generating function ψPoiλ (z) · ψPoiμ (z) = eλ(z−1) eμ(z−1) = ψPoiλ+μ (z). By symmetry, we have C(x, y) = C(y, x) for all x and y. If r is irrational, this implies cn = 0 for n = 0, and thus f is almost surely constant. Theorem 1.53 (Extension theorem for measures) Let A be a semiring and let μ : A → [0, ∞] be an additive, σ -subadditive and σ -finite set function with μ(∅) = 0. Lemma 19.24 For any decreasing sequence An0 ↓ ∅ such that |E \ An0 | < ∞ and x1 ∈ An0 for all n ∈ N, we have Ceff (x1 ↔ ∞) = lim Ceff (x1 ↔ An0). If θ (p) > 0, then ψ(p) ≥ θ (p) implies ψ(p) = 1. If f ∈ C, then by assumption Re(f) = (f + f¯)/2 and Im(f) = (f − f¯)/2i are in C. (iii) Conclude that a suitable continuous version of Gn converges weakly to B. Other information that may be included if you're looking up a business is the company profile and a link to the company website. The situation changes when the numbers rk grow quickly as k → ∞. , ωk] ⊂ Cn,in . Proof Let x ∈ V and c := inf{x − w : w ∈ W }. Let τ0 = 0 and inductively define τn+1 = τn + στn for n ∈ N0 . Theorem 7.3 says that this norm is complete (i.e., every Cauchy sequence converges). be i.i.d. random variables with values in Σ and with distribution PY1 = ν. 1 By "countable" we always mean either finite or countably infinite. Theorem 2.21 A family (Xi)i∈I of real random variables is independent if and only if, for every finite J ⊂ I and every x = (xj)j ∈J ∈ RJ , FJ (x) = F{j } (xj). Then clearly f is not Riemann integrable since Ln (f) = 0 and U3n (f) = 1 for all n ∈ N. Fix δ > 0 and choose N ∈ N large enough that μ(B \ AN) < δ. Use Exercise 1.5.4 to show that (x, y) → F (x) ∧ G(y) is a distribution function on R2 . , Y1 + Y2 + . Then C is dense in Cb (E; K) with respect to the supremum norm. Let 0 1, if un ∈ S, Xn = 0, else. 9.1 Processes, Filtrations, Stopping Times.. For example, for nearest neighbour random walk on the integers, every state has period d = 2. Theorem 9.35 Let X be a martingale and let φ : R → R be a convex function. The Rn that we introduced in the explicit construction are given by Rn (x) := x + Zn . = √ x 2π x In particular, ∞) * P sup B[n,n+1] − Bn > nε < ∞ for every ε > 0. In the first case, the chain would be called reducible, and in the second case, it would be periodic. How about P[Y1 = . Theorem 20.23 (Ω, A, P, τ) is ergodic if and only if, for all A, B ∈ A, (1 ' P A ∩ τ −k (B) = P[A] P[B]. (E, τ) is called completely metrizable if there exists a complete metric on E that induces τ . Takeaways Consider an event that is described by the values of infinitely many random variables. ♣ Exercise 21.4.5 Show the statement of Remark 21.23. In order to obtain the density of νa with respect to μ, we define f := g 1Ω\E . (5.16) k=1 (' Hence P Nt = lim Ntn = 1. To be more specific, the question is: In the long run, will there be a consensus of all individuals or will competing opinions persist? We have to show that, for every bounded measurable F : R[0,∞) → R, we have:) * Ex F (Bt +τ)t ≥0 Fτ = EBτ [F (B)]. / Proof Let Ω = Rn and A = B(Rn). be subsets of Ω. n=1 Thus A is σ -∪-closed. Let A be the ring of finite unions of intervals (a, b] ⊂ R. Since E[Xn,l] = 0, kn kn it Xn,l φn,l (t) − 1 = − 1] E[e l=1 l=1 ≤ kn *) E eit Xn,l − itXn,l − 1 + E[itXn,l] l=1 ≤ kn 2 t l=1 2 2 E[Xn,l]= t2 . (Note that in the argument we used the notion of the expected value Ep [#TL] that will be formally introduced only in Chap. The following theorem describes a very useful criterion for uniform integrability. Clearly, pε is irreducible and aperiodic. n→∞ 16.1 Lévy–Khinchin Formula 371 Inductively, we get |φ(t)| ≥ 2−(4) for |t| ≤ 2k ε. 34 1 Basic Measure Theory Remark 1.70 (Completion of a measure space) Let (Ω, A, μ) be a σ -finite measure space. As {x} ∈ B(Rn) for all x ∈ Rn , we would get the contradiction V = x∈V {x} ∈ B(Rn). ♦ If A is a ring an μ is a content on A, then by Lemma 1.31, for A, B ∈ E such that μ(A), μ(B) < ∞, we have μ(A ∪ B) = μ(A) + μ(B) − μ(A ∩ B). Then, for every probability n→∞ n measure μ = α1 μ1 + . Since all continuous versions of a process are equivalent, B is locally Hölder-γ -continuous for every γ ∈ (0, n−1 2n) and every n ≥ 2 and hence for every γ ∈ (0, 12). ♦ Theorem 1.78 (Generated σ -algebra) Let (Ω , A) be a measurable space and let Ω be a nonempty set. Then P − QT V = inf φ((E × E) \ D) : φ ∈ K(P , Q) . We thus obtain the Metropolis transition matrix ' (⎧ 1 1 ⎪ 1 ∧ exp 2β (1 −) , if y = x i for some i ∈ Λ, {x(j) = x(i)} ⎪ 2 ⎨ #Λ j : j ~i p(x, y) = 1 − i∈Λ p(x, x i), if x = y, ⎪ ⎪ ⎩ 0, else. Hence (Pn)n∈N is tight. Many people have helped in correcting errors or improving the exposition by asking questions and I thank all of them. Example 9.40 (Petersburg game) We continue Example 9.14 (see also Example 4.22). Without using the equivalence of the Lebesgue integral and the Riemann integral, show that 31 0 f (x) dx > 0. 332 15 Characteristic Functions and the Central Limit Theorem For m ∈ Zd define x → exp i)πm/K, x* . For T ≥ 0, let |X|∗T = sup |Xt |. Let ξn (X) be the empirical measure of X1 , . ♦ In fact, the condition Λ(t) < ∞ for all t ∈ R can be dropped. ♦ Example 9.5 The Poisson process X = (Xt)t ≥0 with intensity α > 0 (see Sect. 15.2 Characteristic Functions: Examples.. Further, let f0 ≡ 1. (2.13) 2.4 Example: Percolation 75 Lemma 2.40 Let x, y ∈ Zd . l=1 Thus X(A1), . E (Xt − Xs)2n = E t − s X1 Now let n ≥ 2 and γ ∈ (0, n−1 2n). We single out an arbitrary point of T and name it 0. Cw (i) i + 1 and Hence the transition probabilities pw are indeed described by the Cw . As a first step, we enumerate E = {e1 , . Ai for all n ∈ N. 3 3 Proof Let μ1 , μ2 ∈ Mf (E) with g dμ1 = g dμ2 for all g ∈ C. = lim n→∞ n Therefore, lim inf n→∞ 1 1 log Pn (U) ≥ lim inf log Pn ((x − ε, x + ε)) n→∞ n n 1 = lim inf log Pn (x − ε, ∞) − Pn [x + ε, ∞) n→∞ n 1 = lim inf log Pn (x − ε, ∞) ≥ −I (x). Let τ be the product topology on Ω = × Ωi and i∈I B = σ (τ). Hence p(x, y) = p(x, y) 1x∈E\A . As T is a Poisson process with rate 1, (XTt)t ≥0 is also a Markov process with Q-matrix q. Then (fg) ∈ L1 (μ) and 1 p + 1 q = 1 and f ∈ fg1 ≤ f p · gq . This generalizes the Lebesgue integral that can be found in textbooks on calculus. 11.2 Martingale Convergence Theorems 251 This implies M∞ ∈ {0, Ld }. Similarly, we get (7.7) for all measurable h ≥ 0. Example 23.14 Let Σ = {−1, 1} and let μ = 12 δ−1 + 12 δ1 be the uniform distribution on Σ. Exercise 6.3.1 Let X be a random variable on (Ω, A, P) and let) * Λ(t) := log E et X for all t ∈ R. Example 15.5 (due to [73]) In the preceding theorem, we cannot simply drop the assumption that X is bounded without making other assumptions (see Corollary 15.33). ∈ Mf (Rd) and let p1 , p2 , . , Yn) be independent and Berx -distributed. The rate function that shows up here is the analogue to the free energy of thermodynamics. , jn }, we have κ(x, ·) ◦ XJ−1 n−1 = δx ⊗ (14.16) κjk ,jk+1 . Then 1 ∗ P[A n] = ∞; however, P[A] = P[A1] = 6 . 18.2 Here N(8, 8) = {6, 10, 12, 14, 16, . If x > K is a point of continuity of F , then lim supk→∞ Fnk (∞) ≤ lim supk→∞ Fnk (x) + ε = F (x) + ε ≤ F (∞) + ε. (5.8) n=n0 The aim is to employ Lemma 5.20 to refine the estimate (5.7) for (Yn)n∈N and (Tn)n∈N . Takeaways Loosely speaking, a family of functions is uniformly integrable if the main contributions to the integrals of those functions do not come from extremely large values of the functions. (iii) If X ∈ L2 (P), then X is called square integrable and ' (Var[X] := E X2 − E[X]2 © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020 A. Hence equation (13.12) holds. For a countable product of Polish spaces, the Borel σ -algebra of the product is the product of the Borel σ -algebras. ♦ Example 15.18 Define the function ψ : R → [0, 1] for t ∈ [−π/2, π/2) by ψ(t) = 1 −2|t|/π. , d − 1. ♦ Takeaways A random variable X is a measurable map from a probability space to some measurable space. Assume (ii). A By construction, νa 0 μ is a finite3measure with density f with respect to μ. Here again C(x, y) = 1{|x−y|=1}} . independent real random variables with E[|Xn |] = ∞ for all n ∈ N but such that X1 + . i=1 Proof Let z ∈ [0, 1) and write ψX1 (z) ψX2 (z) as a Cauchy product ∞ ∞ n n ψX1 (z) ψX2 (z) = P[X1 = n] z P[X2 = n] z n=0 n=0 = ∞ z n n=0 = ∞ n=0 n P[X1 = m] P[X2 = n − m] m=0 zn n P[X1 = m, X2 = n − m] m=0 P[X1 + X2 = n] zn = ψX1 +X2 (z). Every terminal event is symmetric. For instance, let A be the event where the sum of the two rolls is odd, A = (ω1 , ω2) ∈ Ω : ω1 + ω2 ∈ {3, 5, 7, 9, 11} , and let B be the event where the first roll gives at most a three B = {(ω1 , ω2) ∈ Ω : ω1 ∈ {1, 2, 3} . ∈ E be a sequence such that ∞ μ(Ωn) < ∞ n=1 Ωn = Ω and c for all n ∈ N. "(ii) ⇒ (i)" Let f ∈ Lip1 (R; [0, 1]). (Recall that Ex denotes the expectation for X if X0 = x.) Due to the symmetry of Yi , we have ⎧ 1 ≥ , ⎪ ⎨ 2 φ(m, x) = 12 , ⎪ ⎩ = 0, if m ≤ n and x ≥ a, if m ≤ n and x = a, if m > n. Let x, y ∈ E with x = y and deduce by induction that π({y}) =) Pπ τx1 * ≥ n, X0 = x, Xn = y + n *) Pπ τx1 ≥ k, X0 = x, Xk = y . Hence Mf (E) ⊂ Cb (E) . Let A = A− ∩ A+ . 4.1 Construction and Simple Properties 101 (c) We have + (−f) dμ − (−f) dμ = = f − dμ − (−f)− dμ f + dμ = − f dμ. n! n−1 ∈S(n) Now replace F by the smaller filtration G = (Gn)n∈−N that is defined by G−n = σ (Y−n , Xn+1 , Xn+2 , . By the Bolzano–Weierstraß (Fn (q1))n∈N has a convergent subsequence Fn1 (q1) k∈N . ♦ Example 1.86 Let d(x, y) = x − y2 be the usual Euclidean distance on Rn and let B(Rn , d) = B(Rn) be the Borel σ -algebra with respect to the topology generated by d. We denote by q := lim P[Zn = 0] n→∞ the extinction probability; that is, the probability that the population will eventually die out. " ⇐ " Now let X be i.i.d. given A for a suitable σ -algebra A ⊂ F . Similarly we are also interested in the situation where Y takes values in Rn or in even more general spaces. Define the events An := {Xτn +1 < −2ε}, n ∈ N. The pair (Ω, τ) is called a topological space. < an and 1 = y0 > y1 > . Again it suffices to assume that f is an indicator function of the type f (x) = 1B1 ×···×Bn+1 (xt1 , . Hence Rw,eff (0 ↔ ∞) = 1 1 Rω− + 1 Rω+ − < ∞ or R + < ∞. This implies μ(An) − μ(A) = n→∞ μ(Bn) −→ 0. Takeaways A Galton-Watson branching process (Zn) with mean offspring number m > 1 has a positive chance to survive and in this case grows indefinitely. , N, define Fn := σ (X1 , . Show that for any k ∈ N ∩ (0, r) and s ∈ (0, r) we have Ms (X) < ∞ as well as l→∞ Ms (Xnl) −→ Ms (X) and l→∞ mk (Xnl) −→ mk (X). μ is σ -subadditive. Show that με := N0,ε satisfies an LDP with good rate function I (x) = x 2 /2. To put it differently, the first derivative φ of a convex function is a monotone increasing function. , n} ≥ t ≤ Var[Sn] . (iii) In particular, if p ≥ 1 and E[|Xt |p] < ∞ for all t ∈ I , then

(Xt)t ∈I is a submartingale. (ii) In particular, to any stochastic matrix p, there corresponds a unique discrete Markov chain X with transition probabilities p. Thus H · X is a martingale. 7.4. At first reading, some readers might wish to skip some of the more analytic parts of this chapter. "(ii) ⟹ (iv)" It is enough to show that for any ε > 0, there is a compact set K ⊂ E with lim supn→∞ μn (K c) ≤ ε. 8.2 Conditional Expectations . . Then * } F (U) = E ←(X is infinitely often differentiable in (0, ∞). Note that the right-hand side ... Then μ is called the Beta distribution with parameters ... Then A and C are independent. Hence there exists some ... A ∈ N we have ... We can model this ...

[...]

− it/θ), is infinitely divisible with Γθ,r = Γθ,r/n . The limits in (i) and (iii) coincide. By construction, L = V ⊂ U is compact.) Specializing on the case U = E, we get that for any compact set K, there exists a relatively compact open set L◦ ⊃ K. Define Y = (Y (1) , . By assumption, A ∩ B ∈ D, and trivially A ∩ B ⊂ A. By the monotone convergence theorem (Theorem 4.20), we conclude f dλ = lim I n→∞ I b gn dλ = f(x) dx. Let Xt := YTt and Px = 0 PYx ⊗ PT0 . Theorem 1.61 (Finite products of measures) Let n ∈ N and let μ1 , . (Hint: Without proof, use the existence of a subset of [0, 1] that is not Borel measurable. " ⇐ " Now let (X, (Px)x∈E) be a Markov process. We say that ν is singular to μ. Therefore, f (y) ∈ Bδ (f (x)) ∩ D = ∅ and f (z) ∈ Bδ (f (x)) ∩ D c = ∅; hence f (x) ∈ ∂D. μ(E) Let N, Y1 , Y2 , . If the time set is countable, this property can be generalized to random (stopping) times and is then called strong Markov property. Show that, for every ε > 0, there is a set A ∈ A with μ(Ω \ A) < ε n→∞ and supω∈A |fn (ω) − f (ω)| −→ 0. 699 Name Index .. n n∈N n Definition 20.30 (Entropy of the simple shift) h(P, τ) is called the entropy of the dynamical system (Ω, A, P, τ). , βn ∈ R as well as d, e ∈ R. Consequently, νa 0 μ and ν = νa + νs is the decomposition we wanted to construct. (i) (ii) (iii) (iv) Nμ1 ,σ 2 ∗ Nμ2 ,σ 2 = Nμ1 +μ2 ,σ 2 +σ 2 for μ1 , μ2 ∈ R and σ12 , σ22 > 0. Below the critical temperature, the magnetization increases with decreasing temperature. (3.9) In particular, we have ∞ 2n −n n √ = 4 x n 1−x n=0 1 for all x ∈ C with |x| < 1. Hence I1A is measurable for all A ∈ A1 ⊗ A2 . (23.23) In the case h = 0, m = 0 is a solution of (23.23) for any β. Then τ is a bounded stopping time and sup Xm ≥ a ⟸⟹ m≤n Let f (m, X) = 1{m≤n} 1{Xn−m >a} + 1 2 τ ≤ n. (iii) The distribution of τb has density fb (x) = √b 2π e−b 2 /(2x) x −3/2 . ♣ Exercise 21.4.3 Let p ≥ 1 and let X1 , X2 , X3 , . Show that Xn −→ X if and only if D Xn + Yn −→ X. Theorem 1.64 (Product measure, Bernoulli measure) Let E be a finite nonempty set and Ω = E N . , μn be probability measures on R, B(R) . Then there exists a smallest σ -algebra σ (E) with E ⊂ σ (E): σ (E) := A. At this point, we only briefly make plausible the existence theorem for such regular versions of processes in the case of so-called Feller semigroups. ♠ Reflection An event is independent of itself if A and B are independent for B = A. Remark 9.2 Sometimes families of random variables with more general index sets are called stochastic processes. In other words, there are contents that are not premeasures. However, rigorous proofs are known only for d = 2 and d ≥ 19 (see [67]). Then f ◦φ is bounded and measurable and Uf ◦φ ⊂ Uφ ; hence μ(Uf ◦φ) = 0. Consider the case Ω = R, μ the Lebesgue measure and fn := 1[n,n+1] , f ≡ 0. , Ak } for certain pairwise disjoint non-empty sets A1 , . By Theorem 16.5, we have CPoiνn −→ μ. Consider a Poisson process (Nt)t ≥0 and choose an independent exponentially distributed random variable T (it would suffice for T to have a density). , ωn] is the product of the probabilities of the individual events; that is, μ([ω1 , . 104 4 The Integral Exercise 4.1.1 Let f : R → R be defined by f (x) = e−x 1[0,∞) (x), and let λ the Lebesgue measure on R. Now fix ε > 0. Show the following: (i) The effective conductance between x0 and x1 is Ceff (x0 ↔ x1) = d. Then C = {fλ , λ ≥ 0} separates points, f0 = 1 ∈ C and fμ · fλ = fμ+λ ∈ C. In addition, these functions are exponentially scaled with 1/ε. n n−1 lim n→∞ k=0 (20.7) 20.5 Mixing 507 Proof " ⇒ " Let (Ω, A, P, τ) be ergodic. Exercise 2.1.1 In a queue each new arriving person chooses independently a random waiting position. Exercise 16.2.1 Let μ be an α-stable distribution and let ϕ be its characteristic function. L Thus, eventually there will be a consensus of all individuals, and the probability that the surviving opinion is e ∈ {0, 1} is the initial frequency of opinion e. For any x ∈ E, dx := gcd(N(x, x)) is called the period of the state x. Indeed, for s < t and for the case of a submartingale, E[Xt Fs] = E[E[Xt Fs] Fs] ≥ E[Xs Fs] = Xs . n→∞ n→∞ n→∞ Hence Z = 0 and thus E[Zn F] −→ 0 almost surely. In particular, we get (as shown already in Example 4.22) that E[Sn] = 0 for all n ∈ N. We follow the exposition in [37]. ♦ Example 10.7 Let Y1 , Y2 , . For A ∈ A, we define f dμ := A (f 1A) dμ. If we let ε → 0, intuitively we should get the conditional probabilities as proportional to the thickness (in metres).] Case 2: Var[Y] > 0. Note that Bn,k ∞ ≤ 2−(n+1)/2 if n ∈ N and Bn,k Bn,l = 0 if k = l. , DT . Remark 7.41 (i) If ϕ is a signed measure, then in (7.10) we automatically have absolute convergence. be identically distributed, real random p n−1 Xk for n ∈ N. Furthermore, E[Yn (x)] = P[Xn ≤ x] = F (x) and E[Zn (x)] = P[Xn < x] = F (x−). "(viii) ⇒ (vii)" This is obvious by Lemma 13.15. Since E is a π-system, Theorem 1.19 yields A ⊃ DE ⊃ δ(E) = σ (E) = A. 14.4 Markov Semigroups . We define a set function μ˜ F : A → [0, ∞), (a, b] → F (b) − F (a). ♦ Theorem 2.26 Let K be an arbitrary set and Ik , k ∈ K, arbitrary mutually disjoint index sets. Hence, by Theorem 12.14, *) n→∞ An (ϕ) −→ E ϕ(X) E a.s. and in L1 . be independent random variables with P[Xn = 1] ∈ (0, 1) for all n ∈ N. Klenke, Probability Theory, Universitext, 493 494 20 Ergodic Theory Example 20.3 (i) If X = (Xt)t ∈I is i.i.d., then X is stationary. It suffices to consider B = (Bt)t ∈[0,1] . Then a.e. convergence is equivalent to a.e. convergence on each An . (For example, for any x ∈ K, take an open ball Bεx (x) of radius εx > 0 that is contained in U and that is relatively compact. k→∞ k→∞ N→∞ On the other hand, Acn,N ↓ ∅ for N → ∞; hence μ(Acn,N) −→ 0. n→∞ In both cases, we have fn −→ f almost everywhere. 107 5 Moments and Laws of Large Numbers . By Theorem 15.57, this yields the claim. In statistical physics, a key quantity is the so-called partition function Znβ := e−βUn dμ0n . More precisely, there exists a unique finite C for any B ∈ Fn . Indeed, # ∈ S(n) : −1 (i) ≤ l for some i ∈ {1, . If λn (A) is finite, then for any ε > 0 there exists a compact K ⊂ A such that λn (A \ K) < ε. random nt ! n For t > 0, let St = i=1 Yi and Stn = √ 12 Stn . By assumption, the set E := {Xi−1 (A) : A ∈ Ai , i ∈ I } is a generator of A . Hence, it is enough to show that the integrals along δb,t and #c,t vanish if b → 0 and c → ∞. ∞ ∞ μ(An) ≤ μ An for any choice of countably many (iv) If A is a ring, then n=1 n=1 ∞ mutually disjoint sets A1 , A2 , . We study periodicity of Markov chains in the first section. σ >τ is a stopping time In fact, obviously, we have Fτn ⊃ Fτ + for all n. , ωn ∈ E and n ∈ N. (8.10) If X is surjective, then ϕ is determined uniquely. (For example, for the Petersburg game (Example 4.22) we had Fn (x1 , . Hence, in the following, assume g(t) < ∞ for all t > 0. 22.3 Hartman–Wintner Theorem . Let N = {ω : f (ω) > 0}. ∈ M≤1 (R) with corresponding distribution functions F, F1 , F2 , . By construction, A is predictable with A0 = 0, and M is a martingale since *) E[Mn − Mn−1 Fn−1] = E Xn − E[Xn Fn−1] Fn−1 = 0. n→∞ (ii) (Xn,l) is a null array and PSn −→ N0,1 . n=1 k→∞ Then αkε −→ have αkε = 2−k 3∞ 0 ∞ g ε (t) dt. If r ∈ F is an arbitrary fixed point of ψ, then r = 1 ≥ q0 = q0 . Later we will come back to the introductory example and make the computations explicit. Hence we only have to show that the conditional distributions exist. The starting point will be to define the values of μ on a smaller class of sets; that is, on a semiring. However, in practice, this distinction will not be needed in this book. we have Ei = E Takeaways Assume that a Markov chain can return to a given state only at times that are a multiple of some natural number d and assume that d is the largest number with this property. Reflection Find an example that shows that in (iii), we cannot simply drop the assumption that there exists a sequence En ↑ Ω with μ(En) < ∞. For n ∈ N, define the polynomial fn by fn (x) := n f (k/n) k=0 n k x (1 − x)n−k for x ∈ [0, 1]. Then, in particular, A ∈ F ; hence *) *) *) E 1A E[E[X |F]|G] = E 1A E[X |F] = E[1A X] = E 1A E[X |G] . + R(5, 6). B ⊂ U := n=1 Un . Find an example that shows that this assumption cannot be dropped. Then D(ε) := ∞ n=1 Dn (ε) ⊂ N, where N is the null set from the definition of almost everywhere convergence. By Theorem 1.18, it is sufficient to show that M(μ∗) is a λ-system. Now let Ω1 , Ω2 , . Definition 8.35 A measurable space (E, E) is called a Borel space if there exists a Borel set B ∈ B(R) such that (E, E) and (B, B(B)) are isomorphic. Then ν(An) < ∞; hence (f1 − f2) dμ. Let I be the unit matrix on E. Roughly speaking, the difference is that vague convergence does not imply convergence of total masses. Theorem 5.32 (Baum and Katz [8]) Let γ > 1 and let X1 , X2 , . Then the canonical process Xn : (Rd)N0 → Rd is a Markov chain with distributions (Px)x∈Rd . Clearly, this measure is σ -finite; however, it is neither locally finite nor outer regular. Definition 9.37 (Discrete stochastic integral) Let (Xn)n∈N0 be an F-adapted real process and let (Hn)n∈N be a real-valued and F-predictable process. + Xk for k = 1, . (21.4) Hence s→t Xs −→ Xt in probability. Then f = f + − f with f + , f − ≥ 0 being integrable functions. c1 (ek), . P Xjr ∈ Air for all r = 1, . Let (Ye)e∈E be an independent family of random variables with P[Ye ≤ q] = q for any e ∈ E and q ∈ {p, p , 1}. , T , as well as x0 ∈ R such that X0 = x0 and Xn = fn (X1 , . Hence d(B1 , Ac1) > 0. = Z0i = 1, then D Z = Z1 + . This implies Sn ≥ S + 2 for n ≥ n0 and hence lim infn→∞ Sn /n ≥ pε 2 > 0. Sometimes we will drop the qualifications " almost everywhere " and " almost surely ". Hence (B(R))⊗[0,∞) Ω = σ Xt , t ∈ [0, ∞) = B(Ω, d). The subset of probability measures is denoted by M1 (Ω) := M1 (Ω, A). Then f is Hölder-continuous of order γ with constant C := C(ε) "T /ε#1−γ . By the dominated convergence theorem, we get (25.3). Then ϕn = ϕ 1/n serves the purpose. Let Z 1 , . We denote the equivalence relation by R and let UL = KL /R be the set of equivalence classes. Then αf + βg is also harmonic on E \ A. Definition 9.22 If τ < ∞ is a stopping time, then we define Xτ (ω) := Xτ (ω) (ω). In the latter case, the absolute value of the mean magnetization β,0 β,0 is m± = m+ > 0. Definition 7.4 A subset G of a vector space (or of an affine linear space) is called convex if, for any two points x, y ∈ G and any λ ∈ [0, 1], we have λx + (1 − λ)y ∈ G. Then E[X] = n kP[X = k] = k=0 n k k p (1 − p)n−k k k=0 = np · n n−1 k=1 k−1 pk−1 (1 − p)(n−1)−(k−1) = np. (ii) (Stability under complements) If A ∈ A0 , then X−1 ((A)c) = (X−1 (A))c ∈ σ (X−1 (E)); hence (A)c ∈ A0 . (21.24) Lévy Construction of Brownian motion Up to continuity of paths, X is thus a Brownian motion. The main theorem says that any function of the history up to a given time t can be represented as a discrete stochastic integral with respect to this binary splitting process. and Y1 , Y2 , . If G ⊂ F is a σ -algebra, then we write L[X |G] for the regular conditional distribution of X given G. Here one choice (and thus up to multiples the unique choice) for the conductances is C(x, x + 1) = p 1−p x for x ∈ Z, and C(x, y) = 0 if |x − y| > 1. Letting ε → 0, we get (13.14). Our next goal is to deduce simple criteria in terms of distribution functions and densities for checking whether a family of random variables is independent or not. By (i), we have (P3). Thus (wn)n∈N is a Cauchy sequence: wm − wn −→ 0 if m, n → ∞. If this is the case, then strict inequality holds if H (p) < ∞. 24.1 Random Measures 615 independent increments and that PX(A) = PX1 (A) ∗ PX2 (A) ∗ . To this end, we distinguish two cases. Klenke, Probability Theory, Universitext, 163 7 Lp -Spaces and the Radon–Nikodym Theorem 164 the triangle inequality, to this end, we have to change the space a little bit since we only have f − gp = 0 ⟸⟹ f =g μ-a.e. For a proper norm (that is, not only a seminorm), the left-hand side has to imply equality (not only a.e.) of f and g. Consider the distribution μα on R with density fα (x) = 1 |x|−1−1/α 1{|x|≥1} . Clearly, E ⊂ DE ; hence δ(E) ⊂ DE . 6 1 Basic Measure Theory (ii) Assume A ∈ AI . ♦ 60 2 Independence Theorem 2.13 (i) Let I be finite, and for any i ∈ I let Ei ⊂ A with Ω ∈ Ei . In the rest of this chapter, we let (Ω, A) and (Ω , A) be measurable spaces. Starting with a graph other than Z , for example an infinite binary tree, can result in multiple infinite connected components (Exercise 2.4.1). We can argue as follows. be exchangeable, square integrable random variables. Hence (16.4) holds if n→∞ the array (Xn,l) is a null array. If in particular σ (E) = A , then X is A − −A -measurable ⟸⟹ X−1 (E) ⊂ A. , tn ∈ I , we have that (Xt1 , . 28 1 Basic Measure Theory As in Example 1.54, it can be shown that μ˜ F ((a, b]) ≤ ε + ∞ ˜ F ((a(k), b(k)]). , x1 + . Define two sets B1 := x ∈ C : d(x, Ac1) ≥ d(x, Ac2) , A1 B1 B2 := x ∈ C : d(x, Ac1) ≤ d(x, Ac2) . Indeed, Xn = X0 + Z1 + . , n} ≥ t ≤ t −2 Var[Sn]. As soon as constructing probability spaces has become routine, the concrete probability space will lose its importance and it will be only the random variables that will interest us. Example 17.55 Let X be a real random variable and let f, g : R → R be monotone increasing functions with E[f (X)2] < ∞ and E[g(X)2] < ∞. (i) Show that I − p is invertible. For x = (x0 , . 6.1 Almost Sure and Measure Convergence . Hence, for any ε > 0, there exists a compact set K ⊂ E with P[X1 ∈ K c] < ε2 . Furthermore,) 2 * n→∞ Ln (ε) = E Y1 1{|Y1 |>ε√n} −→ 0; hence (Xn,l) satisfies the Lindeberg condition. Hence ∞) * σ2 ψ(t) := log E eit X = − t 2 + ibt + ψk (t) 2 k=0 satisfies the Lévy–Khinchin formula σ2 ψ(t) = − t 2 + ibt + 2 eit x − 1 − itx 1{|x| 0. n∈N Similarly, we get μ({f1 < f2 }) = 0; hence f1 = f2 μ-a.e. Definition 7.30 Let μ and ν be two measures on (Ω, A). , XN = xN] = P[X1 = x1] N−1 P[Xk+1 = xk+1 |X1 = x1 , . By the strong n=1 Markov property, for every finite stopping time σ (recall that Fσ is the σ -algebra of the σ -past), Pπ [X ∈ A Fσ] = PXσ [X ∈ A]. For d ≥ 2, we have pc (d) ∈ 2d−1 , 3 . Hence P[Ai] = whence + , P Aj = P[Aj]. Proof (i) For t ∈ I , we have {σ ∨ τ ≤ t} = {σ ≤ t} ∩ {τ ≤ t} ∈ Ft and {σ ∧ τ ≤ t} = {σ ≤ t} ∪ {τ ≤ t} ∈ Ft . Hence μn is supported by [0, ∞). Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. In particular, it is analytic and is hence determined by the coefficients of its power series about t = 0; that is, by the moments of X. sup f ∈F (6.5) {|f |>gε } Clearly, (ii) implies (i). " ⇒ " By the usual approximation arguments, it is enough to consider functions f that depend only on finitely many coordinates 0 ≤ t1 ≤ t2 ≤ . David Wilson has nice simulations and a survey of the current research on his web site . n +1 wLn + wn−Ln Hence (Xn)n∈N0 is our generalized urn model with weights (wn)n∈N0 . Finally, in the third step, an embedding of E into RN is constructed. Hence) 1 E X˜ 1] = ϕˆ (0) = ϕ (τ) = 0,) 1 Var X˜ 1] = ϕˆ (0) = ϕ (τ) ∈ (0, ∞). Corollary 21.32 The map F1 : Ω → [0, ∞), ω → sup{ω(t) : t ∈ [0, 1]} is A-measurable. (ii) If f dμ < ∞, then f < ∞ almost everywhere. Since every CPoiνn is infinitely divisible, on the one hand we have to show that this property is preserved under weak limits. "(iii) ⇒ (iv)" Let A1 , A2 , . This transition matrix is not irreducible; rather it has two absorbing states 0 and N. 10.2 Optional Sampling and Optional Stopping . Then F −1 (t) ≤ x ⟸⟹ t ≤ F (x). If, in addition, q is aperiodic, or q is not reversible with respect to π, then p is aperiodic. Now we can drop the quotation marks from the statement and write it down formally. Define J = k∈K Jk . That is, we denote the edge that connects x and y by)x, y* =)y, x* instead of {x, y}. Clearly, we have E[E[|X| ∧ N |F]2] ≤ N 2 . That is, let Xn,i , n, i ∈ N0 be i.i.d. random variables on N0 with P[Xn,i = k] = p(k), k ∈ N0 , and based on the initial state Z0 define inductively Zn+1 = Zn Xn,i . 18.4 Speed of Convergence So far we have ignored the question of the speed of convergence of the distribution PXn to π. Proof See, e.g., [83, page 79]. 2 h∈Z Rearranging this formula yields an expression for the number of leaves: + # u ∈ Z : degHL (u) = 1 = 2 + degHL (h) − 2 h∈Z ≥ 2 + # h ∈ Z : degHL (h) ≥ 3 ≥ 2 + #(Z ∩ TL). ., then X = ∞ X has intensity measure E[X] = μ and hence X is n=1 n a random measure (see Exercise 24.1.1). Show that n→∞ νt /n −→ δ0 . (n→∞ (ii) P Sn −→ ∞ > 0. For any ε ∈ (0, 1], we have εpC,ε ∈ Lip1 (E; [0, 1]). 6.1 Implications between the concepts of convergence. In the general case, it follows by Theorem 20.21. , hzn). ♣ 8.3 Regular Conditional Distribution 211 Exercise 8.3.3 Assume the random variable (X, Y) is uniformly distributed on the disc B := {(x, y) ∈ R2 : x 2 + y 2 ≤ 1} and on [−1, 1]2, respectively. In this section, we show how to do this. A subset C ⊂ Cb (E; K) is called an algebra if (i) 1 ∈ C, (ii) if f, g ∈ C, then f · g and f + g are in C, and (iii) if f ∈ C and α ∈ K, then (αf) is in C. A simple algorithm for this method is the following. Show that X is null recurrent, irreducible and aperiodic and that independent coalescence does not give a successful coupling. Let W1 := U1 . By considering Laplace transforms, we obtain that, for every λ ≥ 0, the sequence of distributions converges: ˜n lim Ex [e−λZt] = lim n→∞ n→∞ = lim n→∞ ψ(t n!) nx nt − (nt − 1)e−λ/n nt + 1 − nt e−λ/n = lim 1 − n→∞ (e−λ/n) nx 1 − e−λ/n n(1 − e−λ/n)t + 1 nx (21.44) = exp − lim x n(1 − e−λ/n) n→∞ n(1 − e −λ/n)t + 1 λ (x/t) := ψt (λ)x . Interpret the statement of Theorem 23.11 in this case. 38 1 Basic Measure Theory Corollary 1.82 (Measurability of composed maps) Let I be a nonempty index set and let (Ω, A), (Ω , A) and (Ωi , Ai) be measurable spaces for any i ∈ I . Then F (0) = 0 since F is linear. The subsequent section proves the CLT for real-valued random variables by means of characteristic functions. Then Ntn ∼ b2n ,pn . Hence P[X = n] = P[X ≥ n]−P[X ≥ n+1] = (1−p)n −(1−p)n+1 = p (1−p)n . That is, at each step, X jumps to any of its 2D neighbors with the same probability 1/2D. n=m In the last step, we used Remark 7.41(i). 3.2 Poisson Approximation . Letting ε := ε Var[Sn], we get Ln (ε) ≤ ε −δ kn * 1 E |Xn,l |2+δ . Corollary 16.9 If (μn)n∈N is a (weakly) convergent sequence of infinitely divisible probability measures on R, then μ = limn→∞ μn is infinitely divisible. To put it differently, under Px , the process (Bt − x)t ≥0 is a standard Brownian motion. For a random walk on Z with a finite first moment, this shows that it is recurrent if and only if the increments are centred. We infer x,y∈E = J (x, y)2 R(x, y) 2 I (x, y) + D(x, y) R(x, y) x,y∈E = I (x, y)2 + D(x, y)2 R(x, y) + 2 I (x, y) D(x, y) R(x, y) x,y∈E x,y∈E x,y∈E x,y∈E = I (x, y)2 + D(x, y)2 R(x, y) + 2 u(x) − u(y) D(x, y). (iii) Let n ∈ N and A, A1 , . , Xd)T is called d-dimensional normally distributed with expectation μ and covariance matrix C if X has the density fμ,C (x) = 2 1B C exp − x − μ, C −1 (x − μ) 2 (2π)d det(C) 1 (15.10) for x ∈ Rd . 2 x 27/5 0 x R (x, 1) = 27/26 1 1 R (0, x) = 27/32 54/25 513/125 19 R (0, 1) = 27/8 0 Fig. Of course, two independent chains form a coupling, though maybe not the most interesting one. Then ϕ is concave (exercise!); hence, for nonnegative random variables X and Y with finite expectation (by Theorem 7.11), ' (E Xα Y 1−α ≤ (E[X])α (E[Y])1−α . Let (Yn)n∈N0 , PYx x∈E be a discrete Markov chain with transition matrix p and let (Tt)t ≥0, PTn n∈N be a Poisson process with rate λ. (ii) Compute pmax explicitly. On the trace σ algebra A , we define a measure by Ω μ (A) := μ(A) Ω for A ∈ A with A ⊂ Ω . 5.3 Strong Law of Large Numbers 129 Assume that the computer generates numbers X1 , X2 , . in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. Further, let F0+ = + t >0 Ft . Then we have n→∞ √ 2. , An ∈ A such that μ(A1 ∪ . In the first round, the stake is H1 = 1. Proof Let λ ∈ Rd . Only for the more general case of stationary n→∞ 1 X do we need an additional argument.) By the ergodic theorem, we have n Sn −→ E[X1 I] = 0 a.s. Thus, for every m ∈ N, lim sup n→∞ 1 1 max Sk = lim sup max Sk n k=1,...,n n k=m,...,n n→∞ ≤ max k≥m |Sk | m→∞ −→ 0. Show that X and Y are independent. Hence we only have to show sufficiency of the two conditions. Example 4.22 (Petersburg game) By a concrete example, we show that in Fatou's lemma the assumption of an integrable minorant is essential. ♣ Exercise 1.5.3 Use the transformation formula (Theorem 1.101) to show the following statements. e−|x| Exercise 23.1.1 Let X be a real random variable with density f (x) = c−1 , 1 + |x|3 ∞ e−|x| where c = dx. By Prohorov's theorem, in Polish spaces, tightness is equivalent to relative sequential compactness. - Frequency 0.0651 0.0189 0.0306 0.0508 0.1740 0.0166 0.0301 0.0476 0.0755 0.0027 0.0121 0.0344 0.0253 Letter N O P Q R S T U V W X Y Z Morse code -. e∈E ♦ Example 20.32 (Markov chain) Let (Xn)n∈N0 be a Markov chain on E with transition matrix P and stationary distribution π. Let A ∈ I (recall that I is the invariant σ -algebra) and B = A. are independent Poisson point processes with intensity measures μ1 , μ2 , . Furthermore, let Z1 , Z2 , . ♦ Remark 9.27 If I = N, I = N0 or I = Z, then it is enough to consider at each instant s only t = s + 1. That is, those x for which φ(x) − I (x) is close to its maximum. On the other hand, A is called sequentially compact (respectively relatively sequentially compact) if any sequence (xn)n∈N with values in A has a subsequence (xnk)k∈N that converges to some x ∈ A (respectively x ∈ A). Corollary 16.10 If μ ∈ M1 (R) is infinitely divisible, then there exists a continuous convolution semigroup (μt)t ≥0 with μ1 = μ and a stochastic process (Xt)t ≥0 with independent, stationary increments Xt − Xs ∼ μt −s for t > s. Let g : A → R be a bounded function. It is an interesting finding that in two important examples we could check σ -subadditivity using topological properties. In order to simplify the notation, we may assume that X is the canonical process on E N0 . n→∞ We write μn −→ μ if any of the four conditions holds and say that (μn)n∈N converges weakly to μ. 20.5 Mixing 509 " ⇐ " Let X be periodic with period d ≥ 2. In particular, if A = ×j ∈J Aj for certain 306 14 Probability Measures on Product Spaces Aj ∈ Aj , then XJ−1 (A) is called a rectangular cylinder with base J . k=1 314 14 Probability Measures on Product Spaces i / If μ is a finite measure on (Ω0 , A0), then μi := μ ⊗ κk is a finite measure on k=1 i i / × Ωk , Ak . (−∞,x] Here (−∞, x] = {y ∈ Rn : yi ≤ xi for i = 1, . Reflection Come up with an example for X such that the series in (3.1) does not converge for any z > 1 but lim ψX (x) exists and is finite. (i) Compute the generating function ψ and the extinction probability q. Theorem 21.6 yields the existence of a version B of X that has Hölder-γ -continuous paths. In particular, here α = ∞. If μ(Ω) < ∞, then (i) implies uniform integrability of F since the infimum is taken over the smaller set of constant functions. While the preceding corollary only yields an abstract uniqueness statement, we will profit also from an explicit inversion formula for

Fourier transforms. (21.39) 554 21 Brownian Motion We consider now the probability generating function of X1,1 , ψ (1) (s) := ψ(s) := E[s X1,1], s ∈ [0, 1]. is independent and BerZ -distributed. Definition 2.11 (Independence of classes of events) Let I be an arbitrary index set and let Ei ⊂ A for all i ∈ I , . Xn be integrable i.i.d. random variables. Left hand side: n = 1000, Right hand side n = 10 000. Hint: Use suitable stopping times K and apply the martingale convergence theorem (Theorem 11.4) to the stopped process XK . Intuitively, this is the symmetric simple random walk whose vertical transitions are all blocked away from the vertical axis. The next theorem shows that if the Xn are integrable, then the process of partial sums can go to infinity only with a linear speed. 21.1 Computer simulation of a Brownian motion. Hence, it remains to show measurability of If . statement is trivially true if the roles of E and T are interchanged. Let W := (W1 , . The discrete stochastic integral of H with respect to X is the stochastic process H ·X defined by (H ·X) estimate "(t + s)n# − sn! ≤ tn + 2 ≤ 3tn) from (21.36) (since t ≤ N)) *) Kn Kn 4 * E (T¯tK+sn ,n − T¯sKn ,n)4 ≤ n−2 σ −4 E (T"(t +s)n# − T sn!)) Kn = n−2 σ −4 E (T"(t +s)n#− sn! absolutely. ♣ Chapter 7 Lp -Spaces and the Radon–Nikodym Theorem In this chapter, we study the spaces of functions whose pth power is integrable. A family F ⊂ Lp (μ) is called bounded limit point? Show that the reverse inclusion to Theorem 4.19 holds, Lp (μ) ⊂ Lp (μ) if 1 ≤ p ≤ p ≤ ∞. You can search these sites by name, keywords or location and, sometimes, you can also part of the company's offerings. The additional claim follows by the Portemanteau theorem (Theorem 13.16) since N0,1 has a density; hence N0,1 (∂[a, b]) = 0. Thus X0 ≥ Sk+1 − convergence on compact sets. .) with P[Zi = i] = P[Zi = −i] = 1 11 . ...-..-.-.. Let E ∈ E with μ(E) < ∞. However, p is sufficiently contractive only if the multiplicity of the eigenvalue 1 is Theorem 24.5 Let PX be the distribution of a random measure X. If3E[X] = 0, then X is called centered. Then, by construction, P[Rn (i) = j] = r(i, j) − r(i, j − 1) = p(i, j). are i.i.d. E-valued technical statements. Let h = 0 e−t κt g dt. n=1 Reflection In the above theorem, why did we need that μ is finite? Let X be a Markov process with transition kernels (κt)t ≥0 and with (c) (−f) dμ = − f dμ. Let n ∈ N with C ⊂ Kn . We have thus reduced the problem to the one-dimensional situation and will henceforth assume d = 1. By Lemma 20.7, we have E[X0 |I] ◦ convergence in this proof, we introduce the function ⎧ 1 it x ⎪ e − 1 − itx , if x = 0, ⎪ ⎨ x2 (15.8) ft (x) := 2 ⎪ t ⎪ ⎩ − , if x = 0, 2 as well as the measures νn ∈ Mf (R), n ∈ N, νn (dx) := s is a stopping time. By concatenation of stochastic kernels we construct multi-step random experiments. The corresponding map is called a stochastic kernel. In particular, for X ∈ L1 (without exterior field (h = 0). Hence we expect the convergence in the central limit theorem to be slower. Recall the definition of the trace of a class of sets from Definition 1.25. + Xn) the market prices that we would see if all traders were risk-neutral; that is, for traders who price a future payment by its mean value. If X ∼ Nμ,C , then the following statements hold. δxi , where g has to be chosen appropriately (depending on f). , ω] = pωi . Use Doob's inequality (Exercise 21.4.1) to show that the martingale convergence theorems (a.s. convergence X. Example 17.53 Let (px)x∈N0 be numbers in (0, 1] and let X be an irreducible Markov chain on N0 with transition matrix p(x, y) = ⎧ ⎪ ⎨ ⎪ ⎩ px , 1 − px , 0, if y = x + 1, if y = 0, else. and institutional affiliations. Since the distribution of (X1 , . Hence μ∗ (E ∩ A) + μ∗ (E ∩ Ac) ≤ μ∗ (E) and thus A ∈ M(μ∗), which implies A ⊂ M(μ∗). Let g := |f |q−1 . Hence by the ind the assumption (16.30), we have α < 2. 20.3 Examples . As we saw in Theorem 17.17, by Xn+1 := Rn (Xn) and Yn+1 := Rn (Yn), two Markov chains X and Y are defined with transition We conclude that m αi μ(Ai) = m n αi μ(Ai ∩ Bj) i=1 j =1 i=1 = m n βj μ(Ai ∩ Bj) = i=1 j =1 n βj μ(Bj). For measurable A1 , . Hence the strategy is to truncate the Yi to obtain four initial value X0 = x by defining Xn = Fn ◦ Fn−1 ◦ · · · ◦ F1 (x). Finally, let A := {Sn = 0 for every n ∈ N} be the event of an "escape" from 0. On C([0, 1]) the norm f ∞ = supx∈[0,1] |f (x) for arbitrary A). Define the measure ν by ν(A) := for A ∈ A. In this case, we say that μ2 is stochastically larger than μ1 . We give just two different possibilities: μ = 12 δ1 + 12 δ3 and μ

Clearly, μ({0}) = 0. 2 2 Note that this is the rate function from Theorem 23.1. ♦ Next we describe formally the connection between the LDPs of Sanov and Cramér that was indicated in the previous example. The infinite series is a Gaussian process with the same covariance function as Brownian motion. (ii) For x ∈ I ◦ , define the function of difference quotients gx (y) := (ϕ(y) − ϕ(x)) y − x for y ∈ I \ {x}. By the Portemanteau theorem (Theorem 13.16(iv)), for any N ∈ N, μ(Acn,N) ≥ lim inf μN(Acn,N) ≥ 1 − lim inf μNk (Acn,N) ≥ 1 − δ/2. 504 20 Ergodic Theory Proof Define An := {Sn = 0 for all n ∈ N}. Let A' k = {τ' k = τ' k} ∈ σ(Y') ⊗ σ(Y). Then the map H : Ω → [0, ∞), ω ↦ dF (ω), g(ω)) is B(Ω)-measurable. Then σ(X) = F (IA) is trivial if and only if P is ergodic. For the latter property, it is sufficient that the chain is irreducible. Upper semicontinuity of μ implies μ(F) ≥ lim sup μ(Fn) (20.1) Remark 20.2 If I = N0 , I = N or I = Z, then (20.1) is equivalent to L {(Xn+1)n∈I } = L {(Xn)n∈I }. Hence, for μ-almost all x ∈ R, we get (14.1) then Σn = ... The n-th point of P(Ω, E) is then (15.1) Remark 20.2 Let X ∈ L1 (P) and X ⊂ (Ω, A) → (E, E). In the first section, we give the basic definitions and derive simple properties. Then [f] ≥ h] = [f] = P(G) ∈ R, by assumption, [f] = ... Choose Ω = {ω1, ω2} and A = 2Ω . Finally, let μ = Pω ... Example 12.16 (Strong law of large numbers) If Z1 , Z2 , . . . If P(A) = 0. (ii) Let (Ω, A, μ) be a σ-finite measure space and let f, g : Ω → [0, ∞] be measurable. The theorem of Kolmogorov (Theorem 9.16) states how to construct such a process. Proof Uniqueness of the n-th point is based on the following theorem. Theorem 14.32 (Fubini for transition kernels) Let (Ωi , Ai) be measurable spaces, i = 1, 2. Proof Exercise. Theorem 21.17 (Wiener–Zygmund [126]) For every γ > 1/2, almost surely the paths of Brownian motion are not Hölder-continuous of order γ at any point. Since that p is the transition matrix of an irreducible, aperiodic chain. Hence the claim follows. Applying Prohorov's theorem (i.e., Corollary 13.30) to the measures (μk)k∈N , for each n ∈ N, there exists a subsequence. Hence there exists a continuous modification. (ii) Let ν be a signed measure. Dual Spaces By the Riesz–Fréchet theorem (Theorem 7.26), every continuous linear functional F : L2 (μ) → R has a representation F(g) = ⟨f, g⟩ for some f ∈ L2 (μ). Now, by Exercise 8.3.6 ♦ Exercise 8.3.6 Let X ∈ L1(P) and G ⊂ A a σ-algebra. Show that E[X|G] ... Clearly, A is a kind of partial integration formula (Theorem 4.26). Assume that for t > 0 ... If X is a martingale, then for any stopping time τ with τ ≤ T we have E[Xτ] = E[X0]. The statement is proved in Theorem 7.26, and by Theorem 7.4, X ∈ L1 (P) and ... (iii) ⇒ (i) σ-additivity of P. Remark 22.8 Theorem 22.9 The Gaussian integral The Laplace transform ... Define ψ : [0, ∞) → R ... Then ψ is strictly convex and ψ(0) = 0, there is a unique t ∈ (0, 1) such that ψ(t) = r. (ii) Let X be a random variable with values in R. Let L and K be the sets of measurable functions. As this is an equivalence relation, the classes are either disjoint or identical. Let F be a separating class of functions. Theorem 2.4 is a strictly convex ... Since Y is a supermartingale, (Yn∧τ)n∈N0 is a supermartingale. Hence we get ... Theorem 2.10 Let F be a separating class and let (Xn) be tight. Then ... Since ... For each t > 0, define ... by (7.14). Let E be a Polish space and let (Xn)n∈N be a sequence of E-valued random variables. Lemma 2.16 If X is a simple random variable, then ... Consider the ... the symmetric simple random walk on Z. The harmonic functions are the affine functions. (4) Define D = {x ∈ R : Ex ...}. Exercise 2.2.3 ... This is a tight family of probability measures. Assume that there exist ... Let X be a random variable. Proof This is a direct consequence of Theorem 14.32. (iv) ⇒ (i) ... Theorem 5.29 (Paley–Zygmund inequality) ... Exercise 17.6.2 (5) Let (Xn) be i.i.d. with P[X1 = 1] = P[X1 = −1] = 1/2. Show that ... (4) The sequence of stopping times ... Hence Xn ↑ X. Since the events are independent, the Borel–Cantelli lemma yields ... (ii) ⇒ (iii) This is trivial. Proof First assume that X is bounded. Then ... 17.6 Recurrence and Transience of Random Walks ... Theorem 3.2 Let X be a random variable with values in N0 . Then the probability generating function ψX is ... Markov Chains ... Theorem 14.22 ... Definition 21.24 ... Lemma 6.1 ... Corollary 2.7 ... Theorem 17.50 ... By the strong Markov property, ... In particular, a sequence (Xn) is tight if and only if ... Hence (13.4) holds. Define the process ... Exercise 12.3.2 ... Martingales ... Let τ be a stopping time. Then the stopped process X τ is a martingale. ... Theorem 1.44 ... Proof See, e.g., [...]. ...

The metric d from (1.9) induces the product topology on Ω; hence, as remarked in Example 1.40, (Ω, d) is a compact metric space. ˜ Now x → G(x, y) is harmonic on E \ A. In particular (as the Beta distribution is continuous and, in particular, does not have atoms at 0 or 1), almost surely we draw infinitely many balls of each color. n→∞ fdd Proof As in (21.44) for 0 ≤ t1 ≤ t2 , λ1 , λ2 ≥ 0 and x ≥ 0, we get , + (' ' (˜n − λ Z˜ n +λ Z˜ n lim Ex e 1 t1 2 t2 = lim Ex Ex e−λ2 Zt2 Z˜ tn1 e−λ1 Zt1 n→∞ n→∞ + = lim Ex exp − , λ2 −λ1 Z˜ tn1 n ˜ Z e n→∞ n 1 λ2 (t2 − t1) + 1 t1 ⎛) λ2 + λ x 1 λ2 (t2 −t1)+1 ⎞ = exp ⎝− λ2 + λ t + 1 1 1 λ2 (t2 −t1)+1 ⎠ * = Ex exp(−(λ1 Yt1 + λ2 Yt2)) . Then, by Theorem 1.96, there exists a sequence of simple functions (fn)n∈N with fn ↑ f . Since μ is ∅-continuous, ν is also ∅-continuous and is thus a signed measure on Ω. Give an example of a measure-preserving dynamical system that is ergodic but not weakly mixing. Similarly, (14.7) holds. In contrast to the original model, we do not simply add one ball of the same color as the ball that we return. , xk) = k fi (xi) for any k ∈ N. Hence the aim is to study the 3 φ asymptotics of Zε := eφ(x)/ε με (dx) as ε → 0. If in Theorem 20.20 the random variables Xn are not integer-valued, then there is no hope that Sn = 0 for any n ∈ N with positive probability. Assume that Xˆ 0 = X0 and X˜ 0 = 0. 10]. However, this was the claim. For S, T > 0, by Lemma 21.5, two such modifications XS and XT are indistinguishable on [0, S ∧ T]; hence ΩS,T := there is a t ∈ [0, S ∧ T] with XtT = XtS is a null set and thus also Ω∞ := ΩS,T is a null set. 20.5 Mixing Ergodicity provides a weak notion of "independence" or "mixing". , E[Xn]). Furthermore, we compute the conditional distribution of the individual random variables. Now define κY,F (ω, A) = κY ,F (ω, φ(A)) for A ∈ E. Via the connection with subordinators, in the third section, we construct two distributions that play prominent roles in population genetics: the Poisson–Dirichlet distribution and the GEM distribution. Ik,n n→∞ Show that fn (x) −→ f (x) for λ-almost all x ∈ [0, 1]. Proof For k ≤ n, we have Mn (τ (ω)) ≥ Sk (τ (ω)). Proof (i) (ii) (iii) " ⇒ " This is trivial. On the other hand, σ (Y m , m ≥ n) = F2−n+1 ; hence F0+ = t >0 Ft = F2−n+1 = T n∈N is P-trivial. Under certain moment conditions, a stochastic process X allows for a modification X˜ that is continuous and equals X with probability 1 at any given time. In this case, t → N t is monotone but it is not right continuous, although (i) and (ii) hold. 1. Let (An)n∈N be a sequence in A with α = lim ϕ(An). Let F ∈ (Lp (μ)) . By Weierstraß's approximation theorem (Example 5.15), there isνa sequence (pn)n∈N of polynomials that approach the map [0, 1] → [0, 1], t → t uniformly. However, in the special case where the content was defined on an algebra in the first place, the measure on sets of the σ -algebra can be approximated arbitrarily well by sets from the algebra. 229 229 233 239 11 Martingale Convergence Theorems and Their Applications . Define ϕ + (A) := ϕ(A ∩ Ω +) and ϕ − (A) := −ϕ(A ∩ Ω −). , n Ξ∞ = (12.5) l=1 Drawing without replacements thus asymptotically turns into drawing with replacements. |n|→∞ Hence τ is mixing. and ∞ 0 Fn = n=1 1 sup Skε > 0 k k∈N 1 0 = 1 1 sup Sk > ε ∩ F = F, k∈N k * n→∞) *) hence Fn ↑ F . However, for d ≥ 2, the condition F1 ≥ F2 is weaker than μ1 ≤st μ2 . 1+(δ/2) n→∞ Var[Sn] lim (15.7) l=1 Lemma 15.42 The Lyapunov condition implies the Lindeberg condition. Furthermore, for every t ∈ I , By (i), we have P[N] Nt ∩ R ⊂ (Nr ∩ R) ⊂ N. (ii) I ⊂ R is a (possibly unbounded) interval and X and Y are almost surely right continuous. For any choice of sets Ai ∈ Ei , i ∈ N, the family (Ai)i∈N is independent; hence (Ei)i∈N is independent. Remark 5.2 (i) The definition in (ii) is sensible since, by virtue of Theorem 4.19, X ∈ Ln (P) implies that Mk < ∞ for all k = 1, . Now that we have established (iii), by Exercise 6.1.4, we see that ∞ n=1 (Yn − E[Yn]) converges almost surely. , xn) with x0 = x and xn = y. (i) In both cases, √ determine the conditional distribution of Y given X = x. On the other hand, any predictable H has the form Hn = Fn (D1 , . Note that in the last step, we used the fact that p − p/q = 1. The rate for the decay can be computed via the Legendre transform of the logarithmic moment generating function. , d − 1). It is enough to show that sup fn dμ ≥ g dμ. 472 19 Markov Chains and Electrical Networks Definition 19.21 Let I be a flow on E \ A. Example 9.13Let I = N0 and let Y1 , Y2 , . Let p be a probability vector on E 1 × E 2 with marginals p1 and p2 . Now, if Kn ↑ ∞, n → ∞, then for every N > 0, we have + P sup U˜ tKn ,n > ε t ∈[0,N] , ≤ * n→∞) N Var Z1Kn −→ 0, ε2 σ 2 hence U˜ Kn ,n ⇒ 0 in C([0, ∞)). By induction, we get rs = 0 hence p s ≥ 0. In particular, independent square integrable random variables are uncorrelated. j ∈J j ∈J / R (ii) Ai = σ (Z) = σ Z E ,R . Hence E[Xm 1B] = C∈Zm : P[C]>0 Q(C) P[C ∩ B] = P[C] Q(C) = Q(B). ♣ Exercise 3.3.2 Assume that we have a branching process Z = (Zn)n∈N0 with Z0 = 1 whose offspring distribution is given by pk = 13 · (2/3)k , k ∈ N0 . , N}, then in the new generation it will be random and binomially distributed with parameters N and k/N. with PX = μ and PXn = μn for n→∞ every n ∈ N such that Xn −→ X almost surely. From an abstract point of view, the integral is monotone and linear and fulfills the triangle inequality, which allows to use it to define normed vector spaces of functions. Then + , + , Bk = P Aj = P[Aj] = P[Aj] = P[Bk]. © 1996-2014, Amazon.com, Inc. :(y, x) F Therefore, ⎡ μx ({y}) = Ex⎣ τx1 −1 n=0 ⎤ :(x, y) ⎦ ⎡ F 1{Xn =y} ⎦ = Ex⎣ 1{Xn =y} ; τx1 =∞ ; τx1 > τy1 ⎦ = < ∞. (17.6) Remark 17.13 If I is countable, then the strong Markov property holds if and only if, for every almost surely finite stopping time τ , we have *)) * Lx (Xτ +t)t ∈I Fτ = LXτ (Xt)t ∈I := κ(Xτ , ·) . , ∈ A with An ↑ Ω and μ(An) < ∞ for all n ∈ N. n n−1 P[A] = k=0 Thus P[A] ∈ {0, 1}; hence I is trivial and therefore τ is ergodic. It takes a little more work to show that there exists a countable set G ⊂ X that is uniquely determined by X g , g ∈ G, and is an RCLL C0 (E) and a process X version of X. Hence, in (14.19) we can replace the normal distribution by any parameterized family of distributions (νt , t ≥ 0) with the property νt +s = νt ∗ νs . Takeaways Two events A and B are independent if P[A ∩ B] = P[A] · P[B]. , μn on arbitrary (even different) measurable spaces. n=1 * (G n). ♣ Exercise 8.1.2 Consider a theatre with n seats that is fully booked for this evening. At the second stage, given W , we construct a Markov chain X on Z with transition matrix ⎧ − ⎪ ⎨ Wi , if j = i − 1, pW (i, j) = Wi+ , if j = i + 1, ⎪ ⎩ 0, else. X is called lattice distributed if there are a, d ∈ R such that P[X ∈ a + dZ] = 1. be i.i.d. real random variables with characteristic function ϕ. The event {L > −∞} is invariant and hence has probability 0 or 1. Indeed, the value of the left-hand side does not change if we change the order of the sets A1 , A2 , . We formulate this principle as a theorem. (13.3) Arguing as above, this will show that μ1 (K) = μ2 (K) and will hence conclude the proof. (ii) Let Y ∼ N0,1 . ≤ mn ≤ 0, where mk := yakk −y −ak−1 is the slope on the kth interval. (17.18) −α < ∞ if and only if α > 1. (i) (ii) (iii) (iv) μn ({k}) μn (A) ψn (z) ψn (z) n→∞ −→ −→ n→∞ −→ −→ n→∞ −→ −→ μ({k}) μ(A) ψ(z) ψ(z) for all k ∈ N0 . Hence μ(A1) = lim μ(A1 \ An) = μ(A1) − lim μ(An). Let X be a random walk on E with weights C = (C(x, y), x, y ∈ E) and hence with transition probabilities p(x, y) = C(x, y)/C(x) (compare Definition 19.11). Indeed, it is easy to check that P[A] = P[B] = 1 1 2 and P[A ∩ B] = 4 . Let B1 , B2 , . Show that dP (μ, ν) = dP (μ, ν) = dP (ν, μ) for all μ, ν ∈ M1 (E). < tn , PXt −Xs = N0,t −s for all t > s. In the special case I = N0 , X is called a Markov chain. For any two points x, y ∈ Z, we thus have P(x,y)[τ˜ < ∞] = Px−y [Zn = 0 for some n ∈ N0] = 1. Then M is adapted to F and E[Mn Fn−1] = Mn−1 − E[Xn−1 (In) Fn−1] + E[Xn−1 (In + Nn) Fn−1] = Mn−1 − P[In = 1] Xn−1 (1) + P[In + Nn = 1] Xn−1 (i) i∈Λ i∈Λ = Mn−1 since P[In = i] = P[In + Nn = i] = L−d for all i ∈ Λ. For each x ∈ R, define fX (x) = R f(x, y) λ(dy). Show that A is an Abelian algebraic ring with multiplication " ∩ " and addition " ". Since 1A X ∈ L1 (P), we have X ∈ L1 (P · |A]). Define the linearly interpolated processes 1 Z˜ tn := t − n−1 tn! Z nt n!+1 − Z nt n! + Z nt n! . Building finite sums, (14.6) and (14.7) also hold if f is a simple function. , jn }, we have Pμ ◦ XJ−1 = μ ⊗ n−1 k=0 κjk ,jk+1 . On locally compact spaces, also Lipschitz continuous functions with compact support form a separating class. Let B ∈ A with μ(B) < ∞. By stationarity, P[σn < ∞] = 1 for every n ∈ N0 ; hence P[B] = 1. ∞ Hence, let A1 , A2 , . This is defined by dP (μ, ν) := max{dP (μ, ν), dP (ν, μ)}, (13.4) where dP (μ, ν) := inf{ε > 0 : μ(B) ≤ ν(B ε) + ε for any B ∈ B(E)}, (13.5) and where B ε = {x : d(x, B) < ε}; see, e.g., [14, Appendix III, Theorem 5]. Proof By Remark 21.10, X is characterized by (ii). Let P0 be a probability measure on (Ω0 , A0). In this case, g d(f μ) = (gf) dμ. ♦ Definition 9.12 (Predictable) Let I = N0 or I = N. ♠ 40 1 Basic Measure Theory Theorem 1.88 (Measurability of continuous maps) Let (Ω, τ) and (Ω , τ) be topological spaces and let f : Ω → Ω be a continuous map. , n, the sequence (Pnk)n∈N of kth marginal distributions is tight. 18.3 State space decomposition of a Markov chain with period d = 3. Since E is complete, the limit f (ω) := limn→∞ fn (ω) exists. Hence we get (by Theorem 6.25) fn − f p = gn 1 −→ 0. An algebra on a finite set Ω is a σ -algebra. Theorem 13.11 Let (E, d) be a metric space. 21.9 Pathwise Convergence of Branching Processes 553 Remark: The distribution of M can be expressed by the Kolmogorov–Smirnov formula ([101] and [157]; see, e.g., [133]) P[M > x] = 2 ∞ 2 2 (−1)n−1 e−2n x . 8.3 Regular Conditional Distribution . k=0 Hence τAn0 ↑ ∞ almost surely, and thus Fn ↓ {τx1 = ∞} (up to a null set). Instead, we choose representatively one case. Now let A ∈ G. Let U be a random variable that is uniformly distributed on (0, 1). 18.3 for an illustration of the state space decomposition of a periodic Markov chain. Let ϕ : I → R be continuous and in the interior I ◦ twice continuously differentiable with second derivative ϕ . Now, given Θ, is Φ uniformly distributed on [−π, π)? In order to compute Doob's decomposition of Y , define the first and second discrete derivatives of f : f (x) := f (x + 1) − f (x − 1) 2 and f (x) := f (x − 1) + f (x + 1) − 2f (x). Proof. Frequency 0.0978 0.0251 0.0079 0.0002 0.07 0.0727 0.0615 0.0435 0.0067 0.0189 0.0003 0.0004 0.0113 5.4 Speed of Convergence in the Strong LLN 135 Here '.' denotes a short signal while '-' denotes a long signal. Now find an example that shows that the conclusion of Kolmogorov's 0–1 law need not hold under this assumption. ♣ Exercise 9.2.3 Show that the claim of Theorem 9.35 continues to hold if X is only a submartingale but if ϕ is in addition assumed to be monotone increasing. , Xk = xk] (1 − F (y, x)) . The set of such cylinder sets is denoted by ZJ . Takeaways The ergodic theorem yields a law of large numbers for the occupation times of a positive recurrent Markov chain. Hence W is bounded m−1 in L2 and Theorem 11.10 yields Wn → W∞ in L2 and thus in L1 . However, the computations used in Sect. Proof Fix ε > 0, and choose δ > 0 such that |fn (t) − fn (s)| < ε for all n ∈ N and all s, t ∈ E with d(s, t) < δ. Then: (i) (ii) (iii) (iv) (v) |ϕX (t)| ≤ 1 for all t ∈ Rd and ϕX (0) = 1. The chain should be designed so that at each step, only a small number of transitions are possible 446 18 Convergence of Markov Chains in order to ensure that the procedure described in Example 17.19 works efficiently. Thus A \ B = A \ (A ∩ B) ∈ D. 17.7 Stochastic Ordering and Coupling . Theorem 15.15 (i) Let μ1 , μ2 , . (5.1) i,j In particular, Var[αX] = α 2 Var[X] and the Bienaymé formula holds, Var m i=1 . Then ϕ is convex if and only if ϕ (x) ≥ 0 for all x ∈ I ◦ . (note that there are 4(2n + 1) edges that connect (0 ↔ ∞) = Reff ∞ n=0 1 = ∞. Inductively, we get Var[Wn] = σ 2 n+1 k=2 m−k ≤ σ2 m < ∞. First consider the case K = R. Basic number theory then yields that, for every n ≥ n := r · x i=1 (ki /dx), there are numbers m1 , . .+ Xn)/n −→ m in probability. Indeed, in this case, D := A2 ∈ A2 : ω1 → κ(ω1 , A2) is A1 -measurable is a λ-system (exercise!). The gene frequencies k/N in this model can be described by a Markov chain X on E = {0, 1/N, . Lemma 14.23 Let κ be a finite transition kernel from (Ω1 , A1) to (Ω2 , A2) and let f : Ω1 × Ω2 → [0, ∞] be measurable with respect to A1 ⊗ A2 − B([0, ∞]). Here we denoted I (x) := y∈E I (x, y) and I (A) := x∈A I (x). 15.1 Separating Classes of Functions 333 Proof By the dominated convergence theorem, [−π,π)d e−i⟨t,x*⟩ ϕμ (t) dt = ⎛ [−π,π)d e−i⟨t,x*⟩ ⎝ lim n→∞ = lim n→∞ [−π,π)d = e−i⟨t,x*⟩ μ({y}) ⎞ ei⟨t,y*⟩ μ({y})⎠ dt |y|≤n ei⟨t,y*⟩ μ({y}) dt |y|≤n y∈Zd [−π,π)d ei⟨t,y−x*⟩ dt. + Xn is (nσ 2 , nb, nν). Proof (i) We always have E[ϕ(Xt)−] < ∞ (Theorem 7.9); hence, by assumption, E[|ϕ(Xt)|] < ∞ for all t ∈ I . Using the approximation theorems for measures, it can be shown that there is always a countably generated σ -algebra G ⊂ F such that for any A ∈ F , there is a B ∈ G with P[A B] = 0. .} ↑ Z and , n ∈ N is a consistent family. Then there is a k ∈ N and states x1 , . Then, for g ∈ Ef , F (g) = gf dμ. , gxn). By the definition of En , the map F : E N → R is measurable, n-symmetric and bounded. Then τ is measure-preserving. Then we have A \ An ↓ ∅ and n→∞ μ(A) − μ(An) = μ(A \ An) −→ 0. Then, by Ohm's rule, the current flow along the ith wire is Ii = u(1)−u(0) = R1i . As μ is additive, we obtain μ(A ∪ B) = μ(A) + μ(B \ A) μ(B) = μ(A ∩ B) + μ(B \ A). ♦ Takeaways The set of points visited by a random walk within the first n steps grows with n at a speed that is the probability of no return. ♦ Reflection In the previous example one might be tempted to assume that the Yi are uncorrelated instead of independent. In order to get an aperiodic chain, for ε > 0, define the transition matrix pε := (1 − ε)p + εI, where I is the unit matrix on E. n→∞ (i) μ is called lower semicontinuous if μ(An) −→ μ(A) for any A ∈ A and any sequence (An)n∈N in A with An ↑ A. ♣ Exercise 21.5.6 Let t ∈ (0, 1) and f0 (x) := t as well as fn (x) := Show that ∞ 2 sin(nπ t) cos(nπ x) nπ n=0 fn (x) for n ∈ N, x ∈ [0, 1]. Indeed, An := {n, n + 1, . By)x, y* =)y, x* ∈ K, denote an (undirected) edge that connects x with y. , N − 2, (18.12) and λxN−1 = rxN−2 . We infer that * n→∞) n StnN) −→ L[(Bt1 , . In particular, we know that, for small ε > 0, the mass of με is concentrated around the zeros of I . ♦ Takeaways In order to compute characteristic functions, in many cases it is enough to have a table of characteristic functions for some repertoire of standard distributions and to know how characteristic functions transform under linear maps and independent sums. As F is continuous, the kernel W := F −1 ({0}) is a closed (proper) linear subspace of V . , Mn,T ,n are independent and Poiλ/n -distributed. show that f d(Cn ϕn) n∈N (iv) Use this construction to contradict the assumption of metrizability. of compact sets in E such that μ(Knc) < n1 for all μ ∈ F and all n ∈ N. This is the so-called Doob decomposition. Macroscopically, the individual spins cannot be observed but the average magnetization can; that is, the modulus of the average of all spins, mΛ (β) = x∈E 1 π(x) x(i) . Let kn = n and Xn,l = √Yln . This justifies the term reversible. ♦ Remark 9.3 Following a certain tradition, we will often denote a stochastic process by X = (Xt)t ∈I if we want to emphasize the "time evolution" aspect rather than the formal notion of a family of random variables. By Exercise 19.1.1 (with A = {3, 5}, x = 2 and y = 3), the probability of visiting 3 before 5 is P = G(2, 3) = 13 29 . , N, ε > 0, be nonnegative numbers. Hence Ω = {1, . ♣ Exercise 4.1.2 (Sequence spaces) Now we do not assume μ(Ω) < ∞. For the first part of (iii), see, eg. 13.4. In order to formulate (and prove) this statement (de Finetti's theorem) rigorously in Sect. n=0 Hence there exists a nontrivial invariant measure μ (that is, μ({0}) can be chosen ∞ strictly positive) if and only if n=0 pn = 0. 2 ♦ i=0 Example 19.27 Asymmetric simple random walk on E = Z with p(x, x + 1) = p ∈ (12 , 1), p(x, x − 1) = 1 − p is transient. n→∞ x∈R Proof Fix x ∈ R and let Yn (x) = 1(−∞,x] (Xn) and Zn (x) = 1(−∞,x) (Xn) for n ∈ N. , Yn) are determined. Then the following implications hold: (i) ⇐⇒ (ii) ⇐⇒ (iii) ⇒ (iv) ⇐⇒ (v). We proceed in several steps. Any t ∈ Dm has a unique dyadic expansion m∈N t= m bi (t) 2−i for some bi (t) ∈ {0, 1}, i = 0, . 434 17 Markov Chains We define for k, l = 0, . Hence Zn /mn is an almost surely and L1 -convergent martingale. The smallest value μ∗ (E) that can be obtained by such an approximation is called the outer measure of E. Deletion of loops. , n} → {1, . We come back to this with the Poisson point process in Chap. We prepare for the proof of this theorem by assembling some analytic tools. However, we restrict ourselves to the case I ⊂ R. (8.2) i∈I Proof Due to the σ -additivity of P, we have P[A] = P i∈I . n→∞ k=1 Proof See, for example, [128]. ♦ Example 13.28 (i) If E is compact, then M1 (E) and M≤1 (E) are tight. Let d be a complete metric on E. Consider an arbitrary w ∈ W \{0}. Lemma 20.7 (i) A measurable map f : (Ω, A) → R, B(R) is I-measurable if and only if f ◦ τ = f. (vi) Similarly, the vague topology τv on M(E) is the 3coarsest topology such that for all f ∈ Cc (E), the map M(E) → R, μ → f dμ is continuous. Theorem 12.17 Let X = (Xn)n∈N be an exchangeable family of random variables with values in E. Corollary 12.19 (0-1 law of Hewitt–Savage) Let X1 , X2 , . (iv) Since |Zt | ≤ |Xt | + |Yt |, we have E[|Zt |] < ∞ for all t ∈ I . Then, for any ε ∈ (0, 1), n→∞ n→∞ n→∞ d˜N (f, fn) ≤ μ AN ∩ {d(f, fn) > ε} + ε μ(AN) −→ ε μ(AN). (iii) If dx = 1 for all x ∈ E, then X is called aperiodic. Remark 8.3 The specification in (8.1) for the case P[B] = 0 is arbitrary and is of no importance. 2.3 Kolmogorov's 0–1 Law . (ii) If E is Polish, then also the converse holds: F is tight ⇒ F is weakly relatively sequentially compact. Hence by the approximation theorem for measures (Theorem 1.65), thereare mutually disjoint sets An = Vn ∩ Cn ∈ A, n ∈ N, such that B ⊂ A := ∞ n=1 An and μ(A) ≤ μ(B) + ε/2. (iii) Give examples of sequences (px)x∈N0 such that the chain is (a) transient, (b) null recurrent, (c) positive recurrent, and (d) positive recurrent but ∞ n−1 exp − (1 − pk) n=0 k=0 = ∞. Thus we consider sequence spaces p = Lp (N, 2N , μ). In order to obtain criteria for the convergence of Markov chains, we thus have to understand periodicity first. , Dn−1). 31 (i) Compute the expectation and variance of 0 Bs ds. ♦ Takeaways Families of random variables are independent if the events they describe are independent (Definition 2.14). (17.28) Compare this representation of the Wasserstein metric with that of the total variation norm, 0 1 P −QT V = sup f d(P −Q) : f ∈ L∞ (E) with f ∞ ≤ 1 . We call p = (pω)ω∈Ω the weight function of μ. By Theorem 21.38, it is thus enough to show tightness of (Lx [Z˜ n], n ∈ N) in M1 (C([0, ∞))). l=0 (19.9) (iii) (Parallel connection (see Fig.19.2)) Let E = {0, 1}. For f˜, g˜ ∈ L2 (μ), define)f˜, g* ˜ :=)f, g*, where f ∈ f˜ and g ∈ g. Since σn2 −→ ∞, for every ε > 0 and for sufficiently large n ∈ N, we have 2K < εσn ; thus n→∞ |Xn,l | < ε for all l = 1, . Then i=1 An,N ↑ E for N → ∞ for all n ∈ N. (iii) Let Ω be countably infinite and let A := {A ⊂ Ω : #A < ∞ or #Ac < ∞}. lim sup Theorem 20.20 Let X = (Xn)n∈N be an integer-valued, integrable, stationary process with the property E[X1 I] = 0 a.s. Let Sn = X1 + . By Lemma 14.23, the map gA : (ω0 , ω1) → κ2 ((ω0 , ω1), dω2) 1A (ω1 , ω2) is well-defined and A0 ⊗ A1 -measurable. 1.3 The Measure Extension Theorem . We denote by τ˜ the corresponding topology induced on R and by τ the usual topology on R. Clearly, the events A1 , . Hence there is an n0 = n0 (ω) with Xn = Yn for all n ≥ n0 , whence for n ≥ n0 Lemma 5.19 2x Tn0 − Sn0 Tn − Sn = n n n→∞ −→ 0. In this sense, the binary model is a complete market. R \[−K,K]d ∞ ≤ g∞ = g By assumption of the theorem, inequality, we conclude 3 g dμ1 = 3 g dμ2 . Thus, by Slutzky's theorem, we n→∞ also have (Sn − Tn) ⇒ N0,1 . Only the onedimensional marginal distributions are determined. Theorem 17.38 An irreducible discrete Markov chain is either recurrent or transient. Proof We check the conditions of Theorem 21.40. "(v) ⇒ (iv)" This is evident. , K, where Z := K j =1 exp(−βWj) is the normalising constant. We introduce the following notation. By the composition theorem for maps (Theorem 1.80), fy,n is measurable. , XN = xN = N . A necessary condition for M < ∞ is of course that the series ∞ n=0 (1 − pn) diverge; that is, that X is recurrent. H is then a gambling strategy. ♦ Example 1.72 Let μ = δω be the Dirac measure for the point ω ∈ Ω on some measurable space (Ω, A). 17.3 Discrete Markov Processes in Continuous Time 409 For illustration, first consider the extreme situation where wn grows very quickly; for example, wn = 2n for every n ∈ N. Thus π(VT) = v0 = E[VT]. Define A∞ = n=1 P[An |Fn−1] = ∞ and A = lim supn→∞ An . Hence (since p 0 (x, z) = 0) μx p({z}) = ∞ p n+1 (x, z) = n=0 ∞ n=1 p n (x, z) = ∞ p n (x, z) = μx ({z}). We define the set of open edges as p E p := {e ∈ E : Xe = 1}. , Xn that cannot be extended to an infinite exchangeable family X1 , X2 , . For any x ∈ V , there is a unique representation x = y + z where y ∈ W and z ∈ W ⊥ . For all numbers x, y ∈ R, we have (f (x) − f (y))(g(x) − g(y)) ≥ 0. , 12. ♦ Example 7.42 If μ+ , μ− are finite measures, then ϕ := μ+ − μ− ∈ M± . Hence κ1 ⊗ κ2 (ω0 , ·) is σ -finite and is thus a transition kernel. n→∞ (ii) (μn)n∈N is tight, and there is a separating family C ⊂ Cb (E) such that f dμ = lim n→∞ f dμn for all f ∈ C. For I ∈ I, let NI be the number of clicks after time a but no later than b. Remark 15.7 Let X and Y be independent nonnegative random variables with Laplace transforms LX := LPX and LY := LPY , respectively. Lemma 6.23 There is a map h ∈ L1 (μ) with h > 0 almost everywhere. Let Y be a random variable with values in a Borel space (E, E) (hence, for example, E Polish, E = Rd , E = R∞ , E = C([0, 1]), etc.). However, it fails for semirings. 288 13 Convergence of Measures Exercise 13.2.1 Recall dP from (13.5). Proof Apply Theorem 6.19 with the convex map H (x) = x p . 19.10). + Xn ∼ bn,p and thus E[f (Sn /n)] = n f (k/n) P[Sn = k] = fn (p). , ωn] := {ω ∈ Ω : ωi = ωi for any i = 1, . In particular, if C(x, y) = 1{)x,y*∈K} , then X is called a simple random walk on (E, K). , kn is said to satisfy the Lindeberg condition if, for all ε > 0, Ln (ε) := kn ' (1 2 n→∞ E Xn,l 1 2 −→ 0. < tn . ♠ Example 9.31 Consider the situation of the preceding example; however, now with E[Yt] = 1 and Xt = ts=1 Ys for t ∈ N0 . (iii) (σ -∪-closedness) Let A, B ∈ D. Repeated application of Fubini's theorem and the translation invariance of λn yields P[X + Y ≤ x] = P[(X, Y) ∈ A] ⊗2 = 1A (u, v) fX (u) fY (v) λn (d(u, v)) = Rn ×Rn 1A (u, v) fX (u) λ (du) fY (v) λn (dv) n = = Rn Rn Rn Rn (−∞,x−v] fX (u − v) λn (du) fY (v) λn (dv) (−∞,x] = fX (u) λn (du) fY (v) λn (dv) (−∞,x] = Rn fX (u − v) fY (v) λn (dv) λn (du) (fX ∗ fY) dλn . Finally, let q = ∞. Example 17.24 (Poisson process) The Poisson process with rate α > 0 (compare Sect. Then σ τ n∈N0 h(P, τ) = h(P, τ ; P). ♦ The following theorem is due to Lindeberg (1922, see [108]) for the implication (i) ⇒ (ii) and is attributed to Feller (1935 and 1937, see [51, 52]) for the converse implication (ii) ⇒ (i). 13 in greater detail. , c(ek) (ek) = c1 (el), . Conclude that if F is a filtration and if B is a Brownian motion that is an F-martingale, then B is also an F+,∗ -martingale. Assume X1 , X2 , . This will be helpful in many places. In this case, we write x ←→p y. We define S= −∞, 1, if − 1 = D1 = D2 = . In this case, we write Xn n→∞ n→∞ ⇒ X or PXn −→ PX . However, the right hand sides are well-defined and by (16.1) are the exponentials of the left hand sides. (For finite σ -algebras F , this was shown in Example 7.39.) Indeed, let P = @ μ/μ(Ω) and Q = ν/μ(Ω). A more instructive approach is based on first constructing, independently of F , a sort of standard probability space on which we define a random variable with uniform distribution on (0, 1). Then XJI is measurable with respect to AI – AJ . Furthermore, by Lemma 15.49 we 3 n→∞ n→∞ have νn ⇒ δ0 . Polish spaces are standard spaces of measure theory and probability theory. Since E[(Xn − a)+] ≤ |a| + E[Xn+], by Lemma 11.3, E[Una,b] ≤ |a| + E[Xn+] . (ix) Let f : R → [0, ∞) be continuous. , Yn are independent as soon as we know X and record a success with probability X. i∈J By the induction hypothesis (2.8), we have μ(Ej) = ν(Ej) for every Ej ∈ Ej ∪ {∅, Ω}. A distribution on [0, 1] is uniquely characterized by its moments (see Theorem 15.4). 62 2 Independence Remark 2.15 (i) Clearly, the family (Xi)i∈I is independent if and only if, for any finite set J ⊂ I and any choice of Aj ∈ Aj , j ∈ J , we have P ' ({Xj ∈ Aj } = P[Xj ∈ Aj]. Now assume that (pt)t ≥0 are the transition probabilities of another Markov process X with the same generator q; that is, with lim s↓0 1 p s (x, y) − I (x, y) = q(x, y). , DT with values in {−1, +1} and functions fn : Rn−1 × {−1, +1} → R for n = 1, . Therefore, X0 ≥ max{S1 , . On the other hand, n→∞ μ([n, ∞)) −→ 0 and ν([n, ∞)) = ∞ for any n ∈ N. Xn,l > ε 2 Var[Sn] Var[Sn] l=1 (15.6) 358 15 Characteristic Functions and the Central Limit Theorem The array fulfills the Lyapunov condition if there exists a δ > 0 such that kn) * 1 E |Xn,l |2+δ = 0. By Corollary 2.22, the distribution P(W1 ,...,Wk+l+1) has the density x → α k+l+1 e−αSk+l+1 (x) , where Sn (x) := x1 + . (ii) It can be shown that Y is the (unique strong) solution of the stochastic (Itô-) differential equation (see Examples 26.11 and 26.31) dYt = 2 2Yt dWt , (21.48) where W is a Brownian motion. If in (4) we take open rectangles instead of open balls Br (x), we get B(Rn) = σ (E5). The moments can be read off from the derivatives at 0. Let νn,k (x) := (n − k)! n! n i1 ,...,ik =1 #{i1 ,...,ik }=k δ(xi1 ,...,xik) 13.4 Application: A Fresh Look at de Finetti's Theorem 301 be the distribution on E k that describes k-fold independent sampling without replacement (respecting the order) from (x1 , . Here we used A ∈ M(μ∗) in the last but one equality and B ∈ M(μ∗) in the last equality. √ (i) For the effective conductance between a and z, show that Ceff (a ←→ z) = 3. Show that lim supn→∞ Sn = ∞ almost surely. On the technical side, the conditional expectation is constructed via the Radon-Nikodym theorem. If the player bets on "red", she gets the stake back doubled if the ball lands in a red pocket. Case 1. ∈ A, ϕ ∞ n=1 An = ∞ ϕ(An). The central limit theorem suggests that pn (0, 0) ≈ CD n−D/2 as n → ∞ for some constant CD that depends on the dimension D. The simplest choice would be (Ω, A) = R, B(R) , X : R → R the identity map and P the Lebesgue–Stieltjes measure with distribution function F (see Example 1.56). Let us change the perspective and ask: For fixed X, which are the martingales Y (with Y0 = 0) that can be obtained as discrete stochastic integrals of X with a suitable gambling strategy H = H (Y)? ♣ 13.4 Application: A Fresh Look at de Finetti's Theorem (After an idea of Götz Kersting.) Let E be a Polish space and let X1 , X2 , . Denote by H γ ,t the set of maps [0, 1] → R that are Hölder-γ -continuous at t and define Hγ := t ∈[0,1] Hγ ,t . "(i) ⇒ (iii)" The equality follows by the individual ergodic theorem. For N ∈ N0 , define the branching process X with x ancestors and offspring distribution q by Xn−1 X0 = x and Xn := i=1 Yn−1,i . i=1 Thus Z is a Markov chain with transition probabilities p(i, j) = p∗i (j), where p∗i is the ith convolution power of p. We conclude 1 Ceff (x1 ↔ ∞) = lim Px1 [Fn] = Px1 [τx1 = ∞] = pF (x1). For N, δ > 0 and ω ∈ C([0, ∞)), let V N (ω, δ) := sup |ω(t) − ω(s)| : |t − s| ≤ δ, s, t ≤ N . Let α := ν({0}) ˜ u(1) and define ν ∈ M((0, ∞)) by ν(dx) = u(1)(1 − e−x)−1 1(0,∞) (x) ν˜ (dx). = e−t x 1 − u(t) := 0 That is, u is the Laplace transform of ν˜ which determines ν˜ and ν uniquely. Proof We may assume x = 0. By Lemma 17.46, (f (Xn))n∈N0 and (f (Yn))n∈N0 are martingales; hence we have n→∞ |f (x) − f (y)| = E(x,y)[f (Xn) − f (Yn)] ≤ 2f ∞ P(x,y)[Xn = Yn] −→ 0. The canonical map for f ∈ Lq (μ), g ∈ Lp (μ) is an isometry; that is, κ(f)p = f q . (P4) For any I ∈ I, we have E[NI] < ∞. That is, there exists a continuous map fJ : RJ → [0, ∞) such that FJ (x) = xj1 −∞ dt1 · · · xjn −∞ dtn fJ (t1 , . For e ∈ E, let CL (e) = c ∈ {0, 1}L : ck = ck (e) for k ≤ l(e) the set of all dyadic sequences of length L that start like c(e). For i ∈ N, define Ei = {ω ∈ Ω : ωi ∈ A} : A ⊂ E . Hence there is a γ > 0 such that k |ϕ(t)| > 1 −γ t 2 e 2 for all t ∈ R. As ε (Xn)n∈N is ergodic, 1An n∈N is also ergodic. κ is stochastic if Ki = 1 for all i ∈ Ω1 . (ii) Show that the space (Cb ([0, ∞)), · ∞) of bounded continuous functions, equipped with the supremum norm, is not separable. , k} =0 lim sup n! n→∞ for all l ∈ N. Then (Ym)m∈N are independent geometric random variables with parameter p (see Example 1.105(iii)). Then Sn := X1 + . There exists a unique measure μ on σ (A) such that μ(A) = μ(A) for all A ∈ A. Dominated convergence yields (1 ' n→∞ P A ∩ τ −k (B) = E [Yn 1A] −→ E [1A P[B]] = P[A] P[B]. Hence ϕ is strictly convex and ϕ (0) = E[X1] < 0. If X is measurable, we write X : (Ω, A) → (Ω , A). Then Xi := Fi−1 (U) := inf x ∈ R : Fi (x) ≥ U is a real random variable with distribution μi (see proof of Theorem 1.104). By assumption, we have τn < ∞ almost surely for all n. Then σ (X) and σ (Y) are independent given F1 as well as given F3 but not given F2 . , Xk−1) A 268 12 Backwards Martingales and Exchangeability and * n→∞) An (fk) −→ E fk (X1) A . The first condition of Theorem 21.40 is exactly (i). Thus F induces a continuous linear map F0 : V0 → R by F0 (x + N) = F (x). α α α μ({|f | > α}) ≤ This implies μ({|f | > α}) = 0 if α > F 1 ; hence f ∞ ≤ F 1 < ∞. Then ∞ ≥ E[ϕ(X)|F] ≥ ϕ(E[X |F]). I (x1) I (x1) I (x0) Correspondingly, the effective conductance is Ceff (x0 ↔ x1) = Reff (x0 ↔ x1)−1 . Since μ ⊥ ν, we get μ Ω \ M) = μ(n∈N Ωn+ > 0. Hence, for each N ∈ N, we have ∞ ∞ 1 ≤ N μ(AN) + μ AN ∩ d f, fnkl > 2−l < ∞. ♣ Chapter 20 Ergodic Theory Laws of large numbers, e.g., for i.i.d. random variables X1 , X2 , . For n ∈ N, we conclude inductively by Theorem 3.8 that ψZn+1 = ψ ◦ ψZn = ψ ◦ ψn = ψn+1 . Further, let (Ω, F , P) be a probability space and let I ⊂ R be arbitrary. (iii) Consider the Poisson distributions Poiα and Poiβ for α, β ≥ 0. Ki := j∈Ω2 j ∈Ω2 Then we can define a finite transition kernel from Ω1 to Ω2 by κ(i, A) = j ∈A Kij . Let > 0 be a Lebesgue number of the covering U ; that is, > 0 is such that, for every t ∈ I , there exists a U ∈ U such that U (t) ⊂ U . 7.3 Hilbert Spaces .. , n2 , and assume fn is linearly interpolated between these points. Further, for any n ∈ N, e∈E pe δe let Xn : Ω → E, (ωm)m∈N → ωn , be the projection on the nth coordinate. (i) Give a characterization of A as in Exercise 1.1.4 (page 11). (ii) In the case α = 2, we have: If PX is not concentrated at one point, then (16.32) implies that PX is in the domain of attraction of some distribution. Then, d formally, q = dt pt . By Theorem 1.81, we have σ E = σ ({E ∩ A : E ∈ E}) A = σ ({X−1 (E) : E ∈ E}) = σ (X−1 (E)) = X−1 (σ (E)) = {A ∩ B : B ∈ σ (E)} = σ (E) . Definition 6.20 Let p ∈ [1, ∞]. ♣ 8.3 Regular Conditional Distribution 203 Exercise 8.2.9 Let X and Y be real random variables with joint density f and let h : R → R be measurable with E[|h(X)|] < ∞. Finally, consider (−f) to n→∞ obtain the reverse inequality lim inf f dμn ≥ f dμ. By construction, Q, q − < x < q + with F (q +) ≤ F (x) + ε. .). be square-free (that is, there is no number r = 2, 3, 4, . Clearly, for any q ∈ {p, p }, the family (Xe)e∈E of random variables is independent q p p (see Remark 2.15(iii)) and Xe ∼ Berq . ♣ Exercise 18.4.6 Let N ∈ N and let E = {0, 1}N denote the N-dimensional hypercube. For every ε > 0, there exist events Aε Ω = E I , P = PX0 ε and B that depend on only finitely many coordinates and such that P[A Aε] < ε and P[B B ε] < ε. We define Tn := n Wk k=1 and interpret Wn as the waiting time between the (n − 1)th click and the nth click. Since (Ω, d) is separable (Theorem 21.30), every open set is a countable union of ε-balls. The proof presented here goes back to an idea of Dvoretzky, Erdös and Kakutani (see [40]). Then + φ(Em) = φ Ω (En \ En+1) n≥m = φ(Ω +) + ∞ m→∞ φ(En \ En+1) −→ φ(Ω +). We construct a network C with greater conductances by adding ring-shaped superconductors along ∂B. , ck ≥ 0 and c0 +. In particular, N0,1 is called the standard normal distribution. A set A ⊂ E is called compact if each open cover U ⊂ τ of A (that is, A ⊂ U ∈U U) has a finite subcover; that is, a finite U ⊂ U with A ⊂ U ∈U U . Then d is said to be the period of that state. n i=1 For every k ∈ An (ν), we have (compare (23.14)) P[ξn (X) = ν] = #An (ν) P[X1 = k1 , . Indeed, here only the parameters of the Beta distribution change. Since we have not yet shown the existence of an independent family with this distribution, we content ourselves with Ye that assume only three values {p, p , 1}. Since j =1 c c c U is closed and since U ∩ K = ∅, we get δ := d(U , K) > 0. Since we do not have a general notion of an integral at hand at this point, for the time being we restrict ourselves to presenting the convolution formula for integer-valued random variables only. ♣ Exercise 4.2.5 Let λ be the Lebesgue measure on R, p ∈ [1, ∞) and let f ∈ Lp (λ). m) * ∞ E Yn2 Lemma 5.20 ≤ 4 E[|X1 |]. For example, it can be enough to check independence for intervals, rectangles or, in the discrete case, for single points. Define Yn := n1 k=0 is uniformly integrable. A⊂ N We are done if we can show that (1.13) An . What exactly do we mean by “best prediction”? Indeed, φ is differentiable at 0 with derivative x→∞ x→∞ i m for some m ∈ R if and only if x P[|X| > x] −→ 0 and E[X 1{|X|≤x}] −→ m. First let f = 1A for A = A1 × A2 with A1 ∈ A1 and A2 ∈ A2 . Hence, it will be important to find out which paths are P-almost surely negligible. After that, the chains run together. This shows (i). Hence Ω is separable and thus Polish. In order for x = (x0 , . In the remainder of this section, assume that (E, C) is a finite electrical network. Then the following are equivalent. By the Portemanteau theorem (Theorem 13.16), μ = w-lim μnk ; hence F is recognized as weakly relatively sequentially compact. By F = σ (X), we denote the filtration generated by X. As f is continuous on the compact interval [0, 1], f is uniformly continuous. Now we extend the notion of independence from families of events to families of classes of events. Define x+ := inf C ∩ [0, ∞) as well as x− := sup C ∩ (−∞, 0] . ♠ Recall the notation · p from Definition 4.16. Then (R˜ n)n∈N0 is independent and we have) * P R˜ n ((x1 , y1)) = (x2 , y2) = pˆ (x1 , y1), (x2 , y2) . ♦ Proof Let π and ν be invariant distributions. (12.8) As for the arithmetic mean (Example 12.16), we can argue that limn→∞ An (φ) is T -measurable. We distinguish the cases where E[X] is in the interior I ◦ or at the boundary ∂I . Hence U is relatively compact; thus ρK,ε has compact support and is thus in Cc (E) for all ε ∈ (0, δ). For further reading, we recommend, e.g., [86, 118, 145, 152]. However, first we have to exclude the case where n is odd since here clearly pn (0, 0) = 0. (i) The family (Ai)i∈I is independent. Use the star-triangle transformation to remove the lower right node (left 19 513 54 in Fig. We present a general two-step construction principle that will be used in a similar form later in Chap. Hence, we define α := α1 > 0 and get an = n1/α for all n ∈ N (note that (16.21) implies a1 =

[illegible]

characteristic function φ, then φ is positive semidefinite. If (Xi−1 (Ei))i∈I is independent, then (Xi)i∈I is independent. Clearly, for any B, C ∈ Pn , B. In other words, the probability that the game has not finished up to the nth round is n at most C cos(π/N) . i=0 Then Y = (Yn)n∈Z is a stationary process. ♦ 7 Lp -Spaces and the Radon–Nikodym Theorem 182 Theorem 7.43 (Hahn’s decomposition theorem) Let φ be a signed measure. ♣ 17.2 Discrete Markov Chains: Examples Let E be countable and I = N0 . Then n=1 μ(An ∩ E) = ∞ n=1 μ(An ∩ E) = ∞ ν(An ∩ E) = ν(A ∩ E). k for x = 0. (ii) (Series connection (see Fig.19.1)) Let n ∈ N, n ≥ 2 and E = {0, . Let A1 , A2 , . Hence x → e2πin x + e−2πin x = 2 cos(2πn x) is Imeasurable but not a.s. constant. (ii) (Closedness under complements) Let A ∈ D. , Xn) ≥ φ(E[X1], . We can use the jump times of X for an alternative derivation. but X0 = X1 . This measure is called the restriction of μ to Ω . Lemma 1.42 (Uniqueness by an ∩-closed generator) Let (Ω, A, μ) be a σ -finite measure and let E ⊂ A be a π-system that generates A. For any f ∈ Cb (E, ; R) and ; any ε > 0, by the Stone–Weierstraß theorem, there exists a g ∈ C with ;f − g ;∞ < ε. Hence (|fn | − |fnε |)+ 1 < ε for all n ≥ nε . Define Xn = Xn,Θn and Yn = Yn,Θn . Then Bn ↓ ∅. Theorem 14.14 (Finite product measures) There exists a unique σ -finite mea/ sure μ on A := ni=1 Ai such that μ(A1 × · · · × An) = n μi (Ai) for Ai ∈ Ai , i = 1, . By the σ -subadditivity of β, μ∗ ∞ Gn ≤ β n=1 ∞ ≤ An n=1 Letting ε ↓ 0 yields μ∗ measure. e∈E We first define a specific code and then show that it is almost optimal. Hence we have u(t) = lim un (t) = lim un (1) ft d ν˜ n n→∞ n→∞ = u(1) ft d ν˜ = αt + (1 − e−t x) ν(dx). Λ∗ (z) = sup tz − Λ(t) = sup tz − = 2 2 t ∈R t ∈R Hence the rate function coincides with that of (23.4). Then define X = Z0 and Yk = 1, 0, if Zk < X, if Zk ≥ X. © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020 A. Example 1.77 (i) The identity map id : Ω → Ω is A – A-measurable. Let r ∈ (0, 1) and τr (x) = x +r (mod 1). 3 (g) = F (g) for : g → gf dμ is in (Lp (μ)) , and F Concluding, the map F is continuous and Ef ⊂ Lp (μ) is dense, we get F = F. 0 0 By Lemma 5.19, for m → ∞, fm (x) = m n−2 1{x x} ↑ f (x) ≤ 4 P[|X1 | > x]. ♦ In particular, we have shown the following theorem. , Xn) = n−1 (X1 + . Define λ := −n1 log(1 − p1) = −n2 log(1 − p2). We conclude A ∈ AI . In the opposite case, v0 < v0 , the strategy H − H 9.4 Discrete Martingale Representation Theorem and the CRR Model 227 ensures a risk-free profit. 19.7 that starts at x and at each step jumps to one of its neighbors at random with equal probability. There is an increasing sequence (Kn)n∈N of compact sets with Kn ↑ Rd . If x is recurrent, then we conclude that G(y, y) ≥ pl+n+k (y, y) ≥ pl (y, x)pk (x, y)G(x, x) = ∞ n=0 and hence also that y is recurrent. ˜ ♦ Example 18.7 Let E = Z and p(x, y) = 1/3 if |x−y| ≤ 1 and 0 otherwise. Here also the definition ends up with a product formula. ., Xm , Y1 , . For the moment, note that the martingale S = (1 − Sn)n∈N0 , just like the one in Example 9.31, has the structure of a product of independent random variables with expectation 1. 94 3 Generating Functions Case 1: limz↑1 ψ (z) ≤ 1. (17.30) See [60] for a comparison of different metrics on M1 (E).♦ As an example of a more involved coupling, we quote the following theorem that is due to Skorohod. To this end, we check the conditions of Theorem 1.36. ♣ Exercise 13.2.6 (Lévy metric) For two probability distribution functions F and G on R, define the Lévy distance by d(F, G) = inf ε ≥ 0 : G(x − ε) − ε ≤ F (x) ≤ G(x + ε) + ε for all x ∈ R . We interpret Di as the result of a bet that gives a gain or loss of one 216 9 Martingales euro for every euro we put at stake. Hence, for n individuals, qk∗n is the probability to have exactly k offspring. Since (19.5) implies equality, we infer f (x0) = f (y) for all y ∈ SA (x0). Find an example that shows that this condition cannot be dropped. q With the aid of Kolmogorov’s 0–1 law, we can infer the following theorem. (ii) By Theorem 13.34, it is enough to show that the sequence (Pn)n∈N is tight. , 1An (Y1)). It is of a certain theoretical interest and will be needed later that triangle functions and sums of triangle functions are characteristic functions. Then Define A := lim sup An = n→∞ n=1 k=n μ(A) = lim μ n→∞ k=n Ak ≤ lim n→∞ ∞ k=n μ(Ak) ≤ lim n→∞ ∞ k=n 2−k = 0. Proof μ is σ -finite since μ ∈ M(E). 23.2 Large Deviations Principle . A state is recurrent if and only if the expected number of visits (Green function) is infinite. i=1 Therefore, lim sup n→∞ f dμn ≤ Replacing f by (1 − f), we get (13.8). Then the following inclusion and exclusion formulas hold: μ(A1 ∪ . ♦ f (x, λ) dλn 162 6 Convergence Theorems Takeaways Consider a function of two variables that is continuous or differentiable with respect to one variable. From Exercise 15.1.3, we know that D Sn+1 = max{Z1 , . One possibility is to construct n + 1 independent random variables Z0 , . i∼j The Boltzmann distribution π on E := {−1, 1}Λ for the inverse temperature β ≥ 0 is defined by π(x) = Zβ−1 exp(−βH (x)), where the partition sum Zβ = exp(−βH (x)) is the normalising constant such x∈E that π is a probability measure. Now we compute (f ◦ τr)(x) = ∞ cn e2πin r e2πin x a.e. n=−∞ By Lemma 20.7, f is I-measurable if and only if f = f ◦ τr ; that is, if and only if cn = cn e2πin r for all n ∈ Z. p Proof For p ∈ [1, ∞), use Jensen’s inequality with φ(x) = |x|p . 19.6.) paths follow different directions for the first time, they will not have any common edge again, though some of the nodes can be visited by both paths. Let τ := inf{t ≥ 0 : Bt ≥ a} ∧ 1. Separability in metrizable spaces is equivalent to the existence of acountable base of the topology; that is, a countable set U ⊂ τ with A = U ∈U : U ⊂A U for all A ∈ τ. In fact, Kirchhoff’s rule says that I is divergence-free on E \ A. Further, let φ : G → R be convex. A map) · , · * : V × V → R is called an inner product if: (i) (Linearity))x, α y + z* = α)x, y* +)x, z* for all x, y, z ∈ V and α ∈ R. ∈ A ∞ ∞ such that A ⊂ An and μ An \ A < ε. , Xn)− can be derived in a similar way to the onedimensional case. (ii) If E1 ⊂ E2 , then σ (E1) ⊂ σ (E2). , k − 1}, we get w(i+l+1)/n − w(i+l)/n ≤ w(i+l+1)/n − wt + w(i+l)/n − wt ≤ 2c (k + 1)γ n−γ . For x ∈ R, define the left-sided limit 14.3 Kolmogorov’s Extension Theorem 317 F (x−) = supy s, is consistent. l By Lemma 5.18, this implies the claim of Theorem 5.17. Then A, B are independent ⇐⇒ P[A|B] = P[A] ⇐⇒ P[B |A] = P[B]. ♣ Exercise 15.1.3 Let n ∈ N and let X1 , . Construct a countable algebra C ⊂ Cb (R) that separates points. Then μ(E) ≤ lim inf μn (E). t0 , . Some of the most successful and established books in the series have evolved through several editions, always following the evolution of teaching curricula, into very polished texts. For n ≤ N, what is the conditional distribution of (X1 , . As in the proof of the martingale convergence theorem (Theorem 11.4), we infer the following. ♦ Remark 9.28 If we do not explicitly mention the filtration F, we tacitly assume that F is generated by X; that is, Ft = σ (Xs , s ≤ t). ♦ Example 17.28 (A variant of Pólya’s urn model) Consider a variant of Pólya’s urn model with black and red balls (compare Example 12.29). Further, let Sk = X1 + . (i) F is continuous and linear. The statements (iv) and (v) are immediate consequences of Lemma 16.25. (x) Let Ω = {1, 2, 3, 4} and A = ∅, {1, 2}, {1, 4}, {2, 3}, {3, 4}, {1, 2, 3, 4} . We want to combine the two parameters λ and ν into one parameter λν. ♣ 510 20 Ergodic Theory 20.6 Entropy The entropy H (P) of a probability distribution P (see Definition 5.25) measures the amount of randomness in this distribution. ♣ n n∈N Exercise 20.6.3 Let pi be the transition matrix of a Markov chain on the countable set Ei with entropy hi , i = 1, 2. The weak law of large numbers states that the distribution of Sn /n is concentrated in the vicinity of 3.5 (Fig. 15.1 Separating Classes of Functions 331 With the substitution y = log(x) − n, we get (note that sin(2π(y + n)) = sin(2πy)) ∞ 2 2 eyn+n (2π)−1/2 e−(y+n) /2 sin(2π(y + n)) dy m(n) = −∞ = (2π)−1/2 en 2 /2 ∞ −∞ e−y 2 /2 sin(2πy) dy = 0, where the last equality holds since the integrand is an odd function. be measurable maps Ω → E. (ii) (Linearity) Let c ∈ R. 8.2 Conditional Expectations 199 Proof First assume that E[E[X |F]2] < ∞. 538 21 Brownian Motion Then (bn,k) is an orthonormal system:)bm,k , bn,l * = 1{(m,k)=(n,l)} . On the other hand, ψ(t) = 12 + 12 φ(2t). Let A ∈ B(Rd) and ε > 0. + Yn for n ∈ N0 ; hence G(0, 0) = n=0 P[Sn = 0]. Show that for any A ∈ E and any n ∈ N, there exist pairwise disjoint sets A1 , . , n}. To this end it is enough to show that for any n ∈ N, the map ω → Yn (ω) := dn (ω0 , ω) (see (21.28)) is A-measurable. The Metropolis algorithm for this chain accepts the proposal of the reference chain with probability 1 if π(x i) ≥ π(x). The Haar functions bn,k are one such choice: Let b0,1 ≡ 1 and for n ∈ N and k = 1, . Definition 13.21 Let F, F1 , F2 , . P |X| ≥ ε ≤ f (ε)) *) * In the special case f (x) = x 2 , we get P |X| ≥ ε ≤ ε−2 E X2 . “(ii) ⇒ (iv)” Convergence of the total masses follows by using the test function 1 ∈ Lip(E; [0, 1]). For any A ∈ A, by (7.8) and (7.7) with h = 1A\E , 1−g f dμ = A Hence f = A∩E c g d(μ + ν) = ν(A \ E) = νa (A). If μ ∈ M(E), then we define f dμ := Re(f) dμ + i Im(f) dμ if both integrals exist and are finite. Assume that K > 0 is large enough such that f (x) = 0 for x ∈ (−K/2, K/2)d and such d d that μi R \ (−K, K) < ε, i = 1, 2. Hence there are sequences (fn)n∈N and (gn)n∈N in E+ such that fn ↑ f and gn ↑ g. ♦ L1 Remark 6.10 If fn −→ f and fn −→ g, then f = g almost everywhere. Assume the conditions of Theorem 12.24 are in force. Corollary 13.30 Let E be a compact metric space. Then: (i) φ is continuous on I ◦ and hence measurable with respect to B(I). We want to estimate the probability that there exists a point x ∈ C p (0) with distance n from the origin. A family (Xi)i∈I of random variables with values in E is called exchangeable if *) *) L X(i) i∈I = L (Xi)i∈I for any finite permutation : I → I . In other words, H must be predictable. Hence the intersection is nonempty. 2 k! Hint: Expand the bracket expression, sort the terms by the different mixed moments and compute by combinatorial means the number of each type of summand. The value of the portfolio, which is the new stochastic process, changes as the stock price changes. ♦ n=1 Example 19.4 For x ∈ E, let τx := inf{n > 0 : Xn = x}. (ii) If d ≤ 2, then Px0 [τx1 < τx0] = 12 . Corollary 21.41 Let (Xi , i ∈ I) and (Yi , i ∈ I) be families of random variables in C([0, ∞)). Define X(ω) := inf{n ∈ N : ωn = 1} − 1, where inf ∅ = ∞. In order for the problem to be interesting, assume also that the distribution π cannot be constructed directly too easily. , 0) be the first unit vector in Zd . Let I = [a, b] ⊂ R be an interval and let λ be the Lebesgue measure on I . an 1 + (x/a)2 Then Caua is called the Cauchy distribution with parameter a. Proof Step 1. For J ⊂ I , let ΩJ := × Ωj and AJ = Aj . For such z, define hz ∈ C by hz (y) = f (x) + f (z) − f (x) Hz (y) Hz (z) for all y ∈ E. Indeed, let (Xn)n∈N be an independent family of random variables n→∞ with Xn ∼ Ber1/n . As in the preceding example, we use a compactness argument. 1.4 Measurable Maps 37 Theorem 1.80 (Composition of maps) Let (Ω, A), (Ω , A) and (Ω , A) be measurable spaces and let X : Ω → Ω and X : Ω → Ω be measurable maps. Hence ΞN = ki=1 (Ni /N)δei . Partial sums of independent centred random variables are an important example. K→∞ Then E[YiK] = E[ZiK] = 0 and Var[ZiK] −→ 0 as well as Var[YiK] ≤ σ 2 , i ∈ N. l=1 In this case, (Xn,l) also satisfies the Lyapunov condition. Dropping the assumption of recurrence is easier, as the following theorem shows. an distribution with Corollary 16.30) If P*X is in the domain of attraction)of a stable * index α, then E |X|β < ∞ for all β ∈ (0, α) and E |X|β = ∞ if β > α and α < 2. 2.1 Independence of Events . (ii) The distribution PX of X is uniquely determined by ψX . , km k1 ! · · · km ! km is the multinomial coefficient and pk = p1k1 · · · pm . (iii) Cauchy distribution: Caur ∗ Caus = Caur+s for all r, s > 0. Proof (i) Let x ∈ Rn and A := {(u, v) ∈ Rn × Rn : u + v ≤ x}. be measurable maps with the property that fn −→ n→∞ f , but not fn −→ f almost everywhere. Then there is a unique σ -finite measure μ : σ (A) → [0, ∞] such that μ(A) = μ(A) for all A ∈ A. Hence we may assume f ≥ 0 and g ≥ 0 and (to avoid trivialities) f + gp > 0. 26.4 for a computer simulation of Feller’s branching diffusion. , 0, 1) are left eigenvectors for the eigenvalue 1. Then m + kd ∈ N(x, y) for every k ≥ nx . ♦ Example 7.14 Let G, Xpand Y be as in Example 7.13. Show that Wt := e−t Xt , t ≥ 0, is a martingale. + Xn−1)), *) 1 1 E Yn−1 F−n = X(1) + . In the special case of the Fourier basis b0 (x) = 1 and bn (x) = 2 cos(nπx), n ∈ N, this construction goes back to Paley and Wiener [125,3 Theorem XLV, page 154]. Choose J ⊃ J˜ := J ∪ {j } . are measurable, then meas fn −→ f ⇐⇒ ˜ fn) n→∞ d(f, −→ 0. , Utn } of the covering U := {Ut , t ∈ I } of I . Hence Theorem 1.81 yields that Y is measurable. Things would be easy if the individual coordinates of the chain were independent one-dimensional random walks. k *) fi (Xi) A = E fi (X1) A . Remark 1.14 (i) lim inf and lim sup can be rewritten as lim inf An = ω ∈ Ω : #{n ∈ N : ω ∈ An } < ∞ , n→∞ lim sup An = ω ∈ Ω : #{n ∈ N : ω ∈ An } = ∞ . (ii) If (A˜i)i∈I is an independent family of σ -algebras and if each Xi is A˜i - Ai measurable, then (Xi)i∈I is independent. Now | log(x) − (x − 1)| ≤ |x − 1|2 for all x ∈ C with n→∞ |x − 1| ≤ 12 . In this sense, aperiodicity has the flavour of an irreducibility condition which is needed in order that two independent chains started in arbitrary states could meet each other. For r > 0 let Mr (Xn) = E[|Xn |r] be the rth absolute moment. ♦ = A×B Example 8.32 Let μ1 , μ2 ∈ R, σ1 , σ2 > 0 and let Z1 , Z2 be independent and Nμi ,σ 2 -distributed (i = 1, 2). 18.4), the distribution will never be exactly the invariant distribution. 14.2 Finite Products and Transition Kernels 307 (ii) Evidently, Z E ,R ⊂ Z R ⊂ A; hence also σ (Z E ,R) ⊂ σ (Z R) ⊂ A. Now assume p0 +p1 < 1. Hence we have X ∼ PPPμ . 8.3 Regular Conditional Distribution 207 Clearly, fX (x) > 0 for PX -a.a. x ∈ R and fX−1 is the density of the absolutely continuous part of the Lebesgue measure λ with respect to PX . There is some freedom in the choice of the class of test functions so it can be adapted to the individual problem. P ◦ (X In the following, assume that (Ω, A, P) is a probability space and τ : Ω → Ω is a measurable map. Let K := ν(Ω)/δ and δ < ε/(2K). Let Ln = x1 + . Definition 1.9 A class of sets A ⊂ 2Ω is called a semiring if (i) ∅ ∈ A, (ii) for any two sets A, B ∈ A the difference set B \ A is a finite union of mutually disjoint sets in A, (iii) A is ∩-closed. Analogously, a CFP φ is called infinitely divisible if, for every n ∈ N, there is a CFP φn such that φ = φnn . Then τ n is a n→∞ stopping time and τ n ↓ τ ; hence Bτ n −→ Bτ almost surely. Similarly, we get (d/dt) qp t (x, y). First, however, we have to consider a more general situation. (Here Ep [#TL] denotes the expected value of #TL , which we define formally in Chap. Therefore, (Ω, A, P, τr) is ergodic. | 3 9 24 39 5 15 78 39 21 30 39 27 | | 58 58 29 29 29 29 29 29 58 29 29 58 | | | 3 9 24 68 5 15 78 97 21 30 68 27 | | 58 58 29 29 29 29 29 29 58 29 29 58 | | | 5 15 11 7 18 25 14 7 93 21 7 45 | | | 58 58 29 29 29 29 29 29 58 29 29 58 | | | 3 9 19 20 10 30 40 20 21 60 20 27 | | 29 29 29 29 29 29 29 29 29 29 29 29 | | | 3 9 24 107 5 15 78 68 21 30 165 27 | | 58 58 29 58 29 29 29 29 58 29 58 58 | / \ 11 33 15 27 14 55 27 27 135 81 27 215 116 116 29 58 29 58 29 58 116 58 58 116 19.5 Network Reduction 487 Takeaways There are essentially two possibilities to systematically determine the effective resistance between two points in an electrical network. Finally, consider the measure μ = q∈Q δq . We will show (Ω, A, P, τr) is ergodic ⇐⇒ r is irrational. Then E[X2] = E[X] = P[X = 1] = p and thus Var[X] = p(1 − p). Now (17.8) is implied by P[Xn > a] + *) 1 P[Xn = a] = E f τ, (Xτ +m)m≥0 2) * 1 = E0 φ(τ, Xτ) 1{τ ≤n} ≥ P0 [τ ≤ n] . Define fn = f (· , xn). For further reading, consult, e.g., [31, 32] or [74]. Then ∅ = Ω \Ω ∈ A, and hence A is a ring. ♦ Theorem 2.31 If X and Y are independent Z-valued random variables, then PX+Y = PX ∗ PY . ♣ 18.3 Markov Chain Monte Carlo Method 445 Exercise 18.2.2 Consider the bivariate process (X, Y) that was constructed from X˜ and Y˜ in Example 18.6. Show that (X, Y) is a coupling with transition matrix p. , Yn are independent and Yi ∼ Berp for any i = 1, . (That is, in Theorem 3.2(iii), the assumption ψ(z) < ∞ for some z > 1 cannot be dropped.) ♣ 3.2 Poisson Approximation Lemma 3.6 Let μ and (μn)n∈N be probability measures on (N0 , 2N0) with generating functions ψ and ψn , n ∈ N. The matrix A can be chosen to be lower triangular. This is the smallest σ -algebra with respect to which all Xi are measurable. n→∞ Definition 7.2 Let p ∈ [1, ∞] and let f, f1 , f2 , . Example 16.2 ∗n (i) δx is infinitely divisible with δx/n = δx for every n ∈ N. ♦ Takeaways Assume we are given a pointwise convergent sequence of nonnegative functions. , kr }) = dx . (I ndependence) If σ (X) and F are independent, then E[X |F] = E[X]. The conditions of the theorem are fulfilled as Fn ↑ Ω. 11.3 Example: Branching Process . Proof “(ii) ⇒ (iii) ⇐⇒ (iv)” This is obvious. For #J = 0, the statement (2.8) holds by assumption of this theorem. , N − 1, and the corresponding eigenvectors are the x k from above. Hence also, t (x, y) = pt (x, y) − p 0 t q(ps − p s) (x, y) ds. Exercise 21.1.1 Show the claim of Remark 21.7. ♣ Exercise 21.1.2 Let X = (Xt)t ≥0 be a real-valued process with continuous paths. Recall that a finite permutation is a bijection : I → I that leaves all but finitely many points unchanged. Theorem 16.14 (Lévy–Khinchin formula on [0, ∞)) Let μ ∈ M1 ([0, ∞)) and 3 let u : [0, ∞) → [0, ∞), t → − log e−t x μ(dx) be the log-Laplace transform μ. ♣ Exercise 21.2.3 Let B be a Brownian motion and σ > 0. Then * n→∞) n Ssn −→ L[Bt − Bs] L St − for any t > s ≥ 0. Consider now the situation where X does not necessarily jump to one of its nearest neighbors but where we still have E0 [|X1 |] < ∞ and E0 [X1] = 0. If A ⊂ B, then U(A) ⊃ U(B); hence μ∗ (A) ≤ μ∗ (B). n=1 Hence, by the monotone limit theorem, we can interchange the summation and the integral and obtain) * ∞ ∞ ∞ E Yn2 −2 ≤ n 1{x x} dx n2 0 n=1 n=1 ∞ = 0 ∞ −2 n 1{x x} dx n=1 ≤ 4 0 ∞ P[|X1 | > x] dx = 4 E[|X1 |]. Further, let Xi : Rn → R, (x1 , . Here N0 ⊂ R is a ν-null set with μ(R \ N0) = 0. Note that qn = ψn (0) = ψ(qn−1) for all n ∈ N and qn ↑ q. Then, for ε > 0 and k, N ∈ N, choose numbers Kε and δN,k,ε such that ε sup Pi {ω : |ω(0)| > Kε } ≤ 2 i∈I and 0 1 1 N sup Pi ω : V (ω, δN,k,ε) > ≤ 2−N−k−1 ε. 18.4 for a computer simulation of the curve β → m(β). Assume that (fn)n∈N does not converge to f in L1 . Hence, for p 23 P[#CN < ∞] = ∞ *) P there is a closed circle γ ∈ Γn n=2N ≤ ∞ n n · 3(1 − p) N→∞ −→ 0. For this purpose, it suffices to show that, for every k = 1, . Step 3. ♣ Exercise 14.4.4 Show that a continuous real convolution semigroup (νt : t ≥ 0) with νt ((−∞, 0)) = 0 for some t > 0 is nonnegative. 12.2 Backwards Martingales .. The function I is lower semicontinuous, strictly convex (in the interval where it is finite) and has its unique minimum at I (0) = 0. Show that Jn ≤ 4n+1 . Hence these are the natural classes of sets to be considered as events in probability theory. (21.9) As in the proof of Lemma 21.3(iii), we infer (with K := C0 2(1−γ)n0) |Xt (ω) − Xs (ω)| ≤ K |t − s|γ for all s, t ∈ D. , Xn−1 and Dn but not on the full information inherent in the values D1 , . Or consider bond percolation on Z2 . As you see, the argument follows a pattern similar to the proof of Carathéodory’s theorem. If, in addition, E is locally compact, then Cc (E)∩Lip1 (E; [0, 1]) is separating for M(E). Then φX φY = φX φZ ; hence X + Y = X + Z. As examples, we have shown the reflection principle and Lévy’s arcsine law. As a composition of measurable maps, ω → d(f (ω), g(ω)) is measurable.) Let f, f1 , f2 , . (i) X is a martingale if and only if, for any locally bounded predictable process H , the stochastic integral H ·X is a martingale. We define the transition matrix pˆ on Zd by p(x, d for y − x ∈ {−1, 0, 1} . 0 In fact, it is obvious that X is centered and Gaussian (since it is a limit of the Gaussian processes of partial sums) and has the given covariance function. (i) For n ∈ N, define φn = δ1/n − δ2/n . Convolution semigroups are a special application and yield real valued processes with independent and stationary increments. ♦ Example 9.14 Let I = N0 and let D1 , D2 , . Theorem 5.4 (Independent L(P)-random variables are uncorrelated) Let X, Y ∈ L1 (P) be independent. Assume there exists a measurable f with fn −→ f μ-almost everywhere. (iii) If E[|XY |] < ∞ and Y is measurable with respect to F , then E[XY |F] = Y E[X |F] (iv) (v) (vi) (vii) (viii) and E[Y |F] = E[Y |Y] = Y. Hence, without loss of generality, we will consider only sequences in M≤1 (E). , An . Show the following: (i) The functions C0 , Sn , Cn , n ∈ N form an orthogonal system in L2 ([0, 1], λ). Note, however, that those statements that we make explicitly about martingales usually cannot be adapted easily to sub- or supermartingales (such as (ii) in the preceding theorem). Since μ1 and μ2 are finite measures, for the function ρC,ε from Lemma 13.10, we have 0 ≤ ρC,ε ≤ 1 ∈ L1 (μi), i = 1, 2, for all ε > 0. The technique of coupling from the past allows for drawing exactly according to the desired distribution. Furthermore, X−∞ = E[X0 F−∞], n→∞ where F−∞ = ∞ F−n . , Yi−1] = ni=1 E[Yi2] (as in Example 10.2). Let I be an arbitrary index set. In particular, V is no larger than the multiplicity of the eigenvalue with second largest modulus. 17.4 Discrete Markov Chains: Recurrence and Transience . Exercise 6.2.1 Let H ∈ L1 (μ) with H > 0 μ-a.e. (see Lemma 6.23) and let (E, d) be a separable metric space. If A2L occurs, then there are two points x 1 , x 2 on the boundary of BL such that for any i = 1, 2, there is an infinite self-intersection free open path πx i starting at x i that avoids x 3−i . We now come to another 0–1 law for independent events and for independent σ algebras. ♣ Exercise 13.2.2 Show that the topology of weak convergence on Mf (E) is coarser than the topology induced on Mf (E) by the total variation norm (see n→∞ n→∞ Corollary 7.45). Here) · , ·* denotes the usual inner product on Rn . n) Theorem 6.24 A family F ⊂ L1 (μ) is uniformly integrable if and only if the following two conditions are fulfilled. Hint: First show that) * P |Xn | > n for infinitely many n = 0 ⇐⇒ X1 ∈ L1 (P). Define Fn := τA0 < τx1 . x↑1 (n) In this case, ψX is uniquely determined by the derivatives ψX (1), n ∈ N. We quote the following strengthening of Cramér’s Theorem(see [31, Theorem 2.2.3]). ♣ Exercise 15.4.3 Show that, for α > 2, the function φα (t) = e−|t | is not a characteristic function. We will use this Exercise later in Example 17.27. Thus we can make the following definition that is the final definition for the integral of measurable functions. k=1 Since c ≥ 0, we have (Sk + c)2 1Ak ≥ (t + c)2 1Ak .

[illegible]

distinguished. 5.2)♣ Sn − Proof Without loss of generality, assume E[Xi] = 0 for all i ∈ N and thus X1 + ··· + Xn . 2 2 2 2 Hence the average free energy of a particle is 1 + m 1 + m 1 − m 1 − m (1 1 p (m) = − m2 − mm + p − 1 log + log . Using the moment criterion, here we have shown that also the process of partial sums converges and that the limit is Brownian motion. Be careful, the cases r = 12 and r = 12 are different. This implies ∞ n=1 Bn = ∅, contradicting the assumption. m=1 Clearly, for n ≥ m,) * E (Xtm − Xtn)2 = E - n B ξk 1[0,t] , bk k=m+1 = n B 1[0,t] , bk k=m+1 Since ∞ 2 k=1)1[0,t] , bk * n C B ξl 1[0,t] , bl l=m+1 C2 ≤ ∞ B C2 1[0,t] , bk . 21.6 The Space C([0, ∞)) 545 "⊃" We have to show that open subsets of (Ω, d) are in A := (B(R))⊗[0,∞) . Example 1.37 (Compare Example 1.30(iii).) Let Ω be a countable set, and define A = {A ⊂ Ω : #A < ∞ or #Ac < ∞}, μ(A) = 0, ∞, if A is finite, if A is infinite. Definition 13.3 A σ -finite measure μ on (E, E) is called (i) locally finite or a Borel measure if, for any point x ∈ E, there exists an open neighborhood U x such that μ(U) < ∞, (ii) inner regular if μ(A) = sup μ(K) : K ⊂ A is compact for all A ∈ E, (13.1) (iii) outer regular if μ(A) = inf μ(U) : U ⊃ A is open for all A ∈ E, (iv) regular if μ is inner and outer regular, and (v) a Radon measure if μ is an inner regular Borel measure. Proof Let A ∈ Fσ and t ∈ I . In particular, X is uniformly continuous on D; hence it can be extended to [0, 1]. (16.33) Theorem 16.29 Let PX be in the domain of attraction of an α-stable distribution (that is, assume that condition (ii) or (iii) of Theorem 16.28 holds), and assume that (an)n∈N is such that C := lim n→∞ n U (an) ∈ (0, ∞) an2 exists. , An ∈ A with A ⊂ ni=1 Ai . Assume fn ↑ f a.e. for n → ∞. Example 19.3 Let X be transient and let a ∈ E be a transient state (that is, a is not absorbing). (ii) In particular, if FA (x, y) > 0 for all x, y ∈ E \ A, and if there is an x0 ∈ E \ A such that f (x0) = sup f (E), then f (x0) = f (y) for any y ∈ E \ A. By Remark 8.26, measurability holds for all B ∈ B(R) and hence κ is identified as a stochastic kernel. In this chapter, we investigate continuity properties of paths of stochastic processes and show how they ensure measurability of some path functionals. We have used the convergence theorems from the last section to show that the integral is continuous or differentiable, respectively, if a regularity assumption is fulfilled. This finishes the proof. If f1 = f2 on A, then f1 = f2 . In order to show (2.3), we compute P[A1 ∩ A2 ∩ A3] = 3 3 #(A˜ 1 × A˜ 2 × A˜ 3) #A˜ i = = P[Ai]. (Recall that inf f (A) = inf{f (x) : x ∈ A}.) If K ⊂ E is compact and nonempty, then f assumes its infimum on K. Exercise 13.4.2 Show that a family (Xn)n∈N of random variables is exchangeable if and only if, for every choice of natural numbers 1 ≤ n1 < n2 < n3 . Similarly, choose C2 ∈ C with B2 ⊂ C2 ⊂ A2 . For s, t ∈ [0, 1], we have - n n . (i) d := dx = dy for all x, y ∈ E. But modern random number generators produce sequences that for many purposes are close enough to really random sequences.) Define r(i, 0) = 0, r(i, j) = p(i, 1) + . Define Sn := n ∈ N0 . 4.1 Construction and Simple Properties 97 3 Remark 4.5 By Lemma 4.3(iii), we have I(f) = f dμ for any f ∈ E+ . Definition 15.54 Let C be a (strictly) positive definite symmetric real d × d matrix and let μ ∈ Rd . We first show that A0 is a σ -algebra by checking (i)–(iii) of Definition 1.2: (i) Clearly, Ω ∈ A0 . In the special case where for every n, the individual φn,l are equal and where n→∞ kn −→ ∞, equation (16.4) holds automatically if the product converges to a continuous function. Evidently, #Fn < ∞ for all n ∈ N. For F ∈ F , we C 3 3C thus have E[fF 1F] = F fF dP = Q(F) = F f dP = E[f 1F]; hence fF = E[f |F]. In particular, the probability generating function X is characterized by its derivatives ψX(n) (1), n ∈ N, and thus by the moments of X. In this case, A can be chosen to equal the exchangeable σ -algebra E or the tail-σ -algebra T . ∈ A ∞ ∞ Ai , we have μ(A) ≤ μ(Ai) . Hence we define the effective resistance between x0 and x1 as Reff (x0 ↔ x1) = 1 1 u(x1) − u(x0) = =− . Then we say that φ is an isomorphism of measurable spaces. ♦ 20.6 Entropy 513 Takeaways The entropy is an important characteristic of a dynamical system. For different temperatures (that is, for different values 23.4 Varadhan's Lemma and Free Energy 607 of β) these can be very different states. Definition 15.24 A function f : Rd → C is called positive semidefinite if, for all n ∈ N, all t1 , . Theorem 10.4 Let X be as in Definition 10.3. Then, for n ∈ N0 ,)X*n = n *) E (Xi − Xi−1)2 Fi−1 (10.3) i=1 and E[)X*n] = Var[Xn − X0]. (v) This follows from (i) and (ii) with X = X+ − X− . In particular, φ need not be locally Hölder-γ -continuous. , ωl−1)) Ajl+1 κjl−1 ,jl+1 (ωl−1 , dωl+1) fl+1 (ωl+1) = PL (A). be an independent family of exponentially distributed random variables with parameter α > 0; hence P[Wn > x] = e−αx . k1 ! · · · km ! 1 In order to show (5.20), note that the event in (5.20) implies L = n and that L and (Mn,1 , . Define f (ω) = f (ω) else. Now define nk := nkk . Corollary 7.27 Let (V , · , · *) be a linear vector space with complete semi-inner product. The following theorem confirms the conjecture mentioned above and also gives conditions under which we cannot expect that infinitely many of the events occur. ∪ Aik). ♣ Exercise 23.2.3 Let E = R. " ⇒ " Let X be exchangeable. We now consider σ -algebras that are generated by more than one map. In particular, Rad1/2 is called the Rademacher distribution. By definition, μ∗ is σ -subadditive; hence we conclude by Theorem 1.36 that μ∗ is also σ -additive. , n, be measurable spaces. Use Exercise 15.3.2 to show that n→∞ μn −→ δ0 . If X is an L2 -martingale, then X2 is a submartingale and the corresponding increasing process is called the variance process or square variation process. Since μ is additive (and thus n=1 monotone), we have by (ii) m μ(An) = μ n=1 It follows that m An ≤ μ(A) for any m ∈ N. Summing up, we have α = lim φ(Am) ≤ lim φ(Em) = φ(Ω +). , xN−1 , then we define x0 := r xN := λ−1 xN−1 and get that in fact xr = λ−1 xN−1 . Evidently, S˜ n = T˜ K,n + U˜ K,n . Let Aj0 , . (See [82, Lemma 4.1].) Q Let M(E) be the space of all measures on E endowed with the σ -algebra = σ IA : A ∈ Bb (E) . 536 21 Brownian Motion Let H = L2 ([0, 1]) be the Hilbert space of square integrable (with respect to Lebesgue measure λ) functions [0, 1] → R with inner product)f, g* = [0,1] f (x)g(x) λ(dx) √ and with norm f =)f, f * (compare Sect. Thus (fn)n∈N is also a Cauchy sequence in measure; hence it converges in measure by Corollary 6.15. Assume in addition that the yk are chosen such that φ is convex on [0, ∞). 15, we will see how the variance determines the size of the typical deviations of the arithmetic mean from the expectation. In the following, let X be a Markov chain on the countable space E with transition matrix p. It is easy to see that G is a Dynkin system. Let q1 , q2 , . 0 Using dominated convergence, we conclude that lim sup Pn ([−K, K]) ≤ α c −1 n→∞ = α −1 lim sup n→∞ 1 1 1 − Re(φn (t/K)) dt 0 lim 1 − Re(φn (t/K)) dt n→∞ 0 1 1 − Re(f (t/K)) dt. In fact, in many applications it is useful to have subtle dependence structures in order to couple Markov chains with different initial chains. Let T1 , T2 , . Determine the set of all invariant distributions. Let us construct a code C = (c(e), e ∈ E) that is efficient in the sense that it minimizes the expected length of the code (of a random symbol) Lp (C) := pe l(e). , Vxn of E and define g := max(gx1 , . (21.29) (ii) For all η, ε > 0 and N ∈ N, there is a δ > 0 such that Pi ω : V N (ω, δ) > η ≤ ε for all i ∈ I. By symmetry, we have P[B1−τ > a |τ < 1] = 12 ; hence P[B1 > a] = P[B1 > a τ < 1] P[τ < 1] = Pa [B1−τ > a] P[τ < 1] = 1 P[τ < 1] . n=0 Hence we get μ({n}) = μ({0}) n−1 pk k=0 and (note that the sum is a telescope sum) μ({0}) ∞ n=0 (1 − pn) n−1 k=0 pk = μ({0}) 1 − ∞ pn . In particular, Var[X] = Cov[X, X]. ... By the composition theorem (Theorem 1.80), for any measurable map f : (Ω, A) → R, B(R) the maps f ! and "f # are also A – 2Z -measurable. Thus, * using) the tower property, we infer E[XY] = E[E[X |F]Y] and E XE[X |F] = E E[XE[X |F] F] =) * E E[X |F]2 . ♦ In most cases the class of open sets that generates the Borel σ -algebra is too big to work with efficiently. n=1 Proof " ⇐ " Assume that (i), (ii) and (iii) hold. Choose a maximal sequence U (sorted by decreasing lengths) of disjoint intervals and show that each U ∈ U is in (x − 3a, x + 3a) for some (x − a, x + a) ∈ U . ∈ Z R be pairwise disjoint and let A ∈ Z R k with A ⊂ ∞ k=1 A . , An ∈ A such that μ A Ak < ε. If E[X] ∈ I ◦ , then let t + := D + φ(E[X]) be the maximal slope of a tangent of φ at E[X]. However, this is not very plausible. e∈E e∈E pe l(e) = − pe log2 (qe) ≥ H2 (p). (ii) For A ∈ A , we denote {X ∈ A } := X−1 (A) and P[X ∈ A] := P[X−1 (A)]. The converse implication is false. Hence we assume now that p ∈ (0, 1) and θ (p) > 0. If g : Ω → [0, ∞] is measurable, then (by Theorem 4.15) (7.4) 2 √1 e −x /2 2π g dν = gf dμ. Then (with p(x, y) = C(x, y)/C(x)), u(1) = 1 · p(1, 2) + 0 · p(1, 0) = = R(1, 0) C(1, 2) = C(1, 2) + C(1, 0) R(1, 0) + R(1, 2) Reff (1 ↔ 0) . We assume that F is countably generated; that is, there exist countably many sets A1 , A2 , . Then the Metropolis matrix p of q and π is irreducible with unique invariant distribution π. Hence we only have to show that κ is surjective. Reflection In the situation of Theorem 19.35, come up with an example such that n→∞ E[Wi+ − Wi−] > 0 but still Xn −→ −∞ holds. Using a contour argument, as for percolation (see [127]), one can show that (for d ≥ 2) there exists a critical value βc = βc (d) ∈ (0, ∞) such that 0 m(β) > 0, = 0, if β > βc , if β < βc . i=1 Define probability measures Px on (Rd)N0 , (B(Rd))⊗N0 by Px = P◦(S x)−1 . However, φ cannot be a (random) signed measure. ♠ Definition 23.6 (Rate function) A lower semicontinuous function I : E → [0, ∞] is called a rate function. Hence A is predictable and monotone decreasing, A0 = 0, and M is a martingale. If |E| ≥ 2, then there is no absorbing state. Repeat a random experiment with possible outcomes e ∈ E and probabilities pe for e ∈ E infinitely often (see Example 1.40 and Theorem 1.64). , XN) ∈ A˜ ε , (Xn , Xn+1 , . Clearly, P[A] = 12 and P[B] = 12 . Hence φ is a topological isomorphism. , jn ∈ I are pairwise distinct. , 1[1/2,1)(τrn−1 (x)) Clearly, we have #φn ([0, 1)) = #Pn . , xn ∈ K such that K ⊂ U := n 1 U x j . Further, with the choice a , we can change X into a martingale. However, the characterization in terms of Laplace transforms is a bit simpler in the case of locally compact Polish spaces considered here. e 2π a Proof By the scaling property of Brownian motion (Corollary 21.12), without loss of generality, we may assume T = 1. Any subsequent person takes his or her reserved seat if it is free and otherwise picks a free seat at random. 3 p→∞ (i) If |f |p dμ < ∞ for some p ∈ (0, ∞), then f p −→ f ∞ . Let pe be the probability of the symbol e ∈ E. Then z = limn→∞ wn − x = c by continuity of the norm (Lemma 7.23). This is possible, as F is right continuous. , μN be an orthonormal basis of left eigenvectors for the eigenvalues λ1 , . (i) (ii) (iii) (iv) B0 = 0, B has independent, stationary increments (compare Definition 9.7), Bt ∼ N0,t for all t > 0, and t → Bt is P-almost surely continuous. In order to find a necessary and sufficient condition on the growth of (wn), we need more subtle methods that appeal to the above example of the explosion of a Markov process. , Xk)-measurable and Sn − Sk is σ (Xk+1 , . However, this is the statement of Pólya's theorem. For A ⊂ Ω − , A ∈ A, we would have φ(A) ≤ 0 since α ≥ φ(Ω + ∪ A) = φ(Ω +) + φ(A) = α + φ(A). A signed content ν is an additive set function that is the difference ν = ν + − ν − of two finite contents. 3.1 Definition and Examples . A little differently from the usual convention, assume that Θ takes values in [0, π) and Φ in [−π, π). t ∈[0,T] Show Doob's inequalities: *) *) (i) For any p ≥ 1 and λ > 0, we have λp P |X|∗T ≥ λ ≤ E |XT |p . ♦ 14.1 Product Spaces 305 Definition 14.6 Let I = ∅ be an arbitrary index set, let (E, E) be a measurable space, let (Ω, A) = (E I , E ⊗I) and let Xt : Ω → E be the coordinate map for every t ∈ I . We will see that this is indeed possible. (13.2) 276 13 Convergence of Measures Definition 13.4 We introduce the following spaces of measures on E: M(E) := Radon measures on (E, E) , Mf (E) := finite measures on (E, E) , M1 (E) := μ ∈ Mf (E) : μ(E) = 1 , M≤1 (E) := μ ∈ Mf (E) : μ(E) ≤ 1 . Define Xnλ =)λ, Xn *, Snλ =)λ, Sn* * and S∞ ∼ N0,C . (i) Show by a direct computation using only the definition of stability that |φ(t) − 1| ≤ C|t|α for t close to 0 (for some C < ∞). Let A ∈ A with μ(A) = 0.3Then Z = {A, Ac } ∈ Z and fZ = 1Ac ν(Ac)/μ(Ac). As E is locally compact, there is a compact set L with K ⊂ L◦ ⊂ L ⊂ G. Proof Let φ be such a coupling. The distribution of a sum of independent random variables can be computed using a convolution formula (Theorem 2.31). However, F (x) =)x, f * for all x ∈ V by the definition of F0 and) · , · *0 . is a martingale. i∈I (iii) Let μ be a σ -finite measure on A, and assume every Ei is also a π-system. 19.2 Parallel connection of six resistors. , jn ∈ {1, . Note that φ(t) = ψ(t) for |t| ≤ π/2 and φ(t) = 0 for |t| > π/2; hence φ 2 = φ · ψ. Now apply (16.2) with zn = |φn (t)|2 . (iii) Determine E[ei t1 X1 +i t2 X2] and E[eit1 Y1 +it2 Y2] for t1 = 12 and t2 = 2. For every ω ∈ Ω, we say that the map I → E, t → Xt (ω) is a path of X. Hence also (Xn (k), k = 1, . 0 = ν(An) − ν(An) = An By Theorem 4.8(i), f2 1An = f1 1An μ-a.e. As f1 > f2 on An , we infer μ(An) = 0 and μ({f1 > f2 }) = μ An = 0. Further, we write f (x) = f (x1 , . 367 16.2 Stable Distributions .. Let I be the 19.3 Finite Electrical Networks 469 corresponding current flow. (vii) (Two-sided exponential distribution) If X and Y are independent expθ distributed random variables, then it is easy to check that X − Y ∼ exp2θ . A map φ : E → E is called Hölder-continuous of order γ (briefly, Hölder-γ -continuous) at the point r ∈ E if there exist ε > 0 and C < ∞ such that, for any s ∈ E with d(s, r) < ε, we have d (φ(r), φ(s)) ≤ C d(r, s)γ . 286 13 Convergence of Measures D Note that (13.7) implies F (∞) ≤ lim infn→∞ Fn (∞). Then *) Sn−1 + M . i=1 (ii) There is a modification of W such that t → Wt is almost surely continuous (see Remark 21.7). By Theorem 7.33, we get that ν = νa has a density with respect to μ. We say that (με)ε>0 satisfies a large deviations principle (LDP) with rate function I if (LDP 1) lim inf ε log(με (U)) ≥ − inf I (U) for every open U ⊂ E, (LDP 2) lim sup ε log(με (C)) ≤ − inf I (C) for every closed C ⊂ E. (Here P could be an arbitrary measure on (Ω, A).) ∗ (ii) If (An)n∈N is independent and ∞ n=1 P[An] = ∞, then P[A] = 1. Each of the x individuals has an exponentially distributed lifetime with parameter 1. , 2n }, n ≥ n0 . Theorem 8.36 Let E be a Polish space with Borel σ -algebra E. The numbers r1 , r2 , . On the other hand, ∞) *) * E 1A Yn E[X |F] = E 1A 1{Yn =k 2−n } k 2−n E[X |F] k=1 = ∞) * E 1A 1{Yn =k 2−n } k 2−n X k=1) * n→∞ = E 1A Yn X −→ E[1A Y X]. Then there exists an F ∈ V and a subsequence (Fnk)k∈N with k→∞ Fnk (x) −→ F (x) at all points of continuity of F. 1.1 Classes of Sets In the following, let Ω = ∅ be a nonempty set and let A ⊂ 2Ω (set of all subsets of Ω) be a class of subsets of Ω. Then Xσ ≥ E[Xτ Fσ], and, in particular, E[Xσ] ≥ E[Xτ]. First consider l = n. By construction, for any f ∈ Cc (E), we infer f dμ = lim f dμknn . 2.3) is transient. Takeaways Consider an i.i.d. sequence of random variables with values in a finite set. For any x ∈ R, we have {X∗ ≤ x} ∈ T ((Xn)n∈N); n→∞ hence P[X∗ ≤ x] ∈ {0, 1}. n=2 ∞ An as a countable n=1 An = A1 (A2 \A1) ((A3 \A1)\A2) (((A4 \A1)\A2)\A3) . For any p ∈ [1, ∞], N is a subvector space of Lp (μ). , StnN − StnN−1) −→ L[(Bt1 − Bt0 , . (17.33) In fact, let f : {0, . Consequently, hn (P, τr ; P) ≤ log(2n). As with the voter model (see Example 11.16) that is closely related to Wright's model, we can argue that the limit X∞ can take only the stable values 0 and 1. Hence Evidently, Z ◦ τ = Z; hence F ∈ I. We first investigate convergence with respect to · p . The powers of M are easy to compute: Mψ = 0 1 , −1 2 Mψ2 = −1 2 , −2 3 Mψ3 = −2 3 , −3 4 and inductively Mψn = −(n − 1) −n n n+1 . < e−cn/2 k=−n Therefore, − RW = ∞ 1 k−1 n− 0 −1 ≤ n=1 k=−n 1 k−1 + ∞ e−cn/2 < ∞ a.s. n=n− 0 n=1 k=−n + Similarly, there is an n+ 0 = n0 (ω) with n k > ecn/2 for all n ≥ n+ 0. In the first section, we study which probability measures on R are infinitely divisible and give an exhaustive description of this class of distributions by means of the Lévy–Khinchin formula. This is often called the mean field assumption. are i.i.d. real random variables, then (PSn /n)n∈N satisfies an LDP with rate function Λ∗ . ♦ 19.4 Recurrence and Transience 0 479 1 R(0, 1) = 1/3 2 R(1, 2) = 2/9 3 R(2, 3) = 4/27 Reff (0 ↔ 2) = 5/9 Reff (0 ↔ 3) = 19/27 Fig. Reflection In the proof of Theorem 5.10, where did we use the independence of T ? Hence the measure PX is characterized by its values on I. Then, by symmetry, also f ≥ 0 and hence f ≡ 0. If E has a finite diameter diam(E), then dW (P , Q) ≤ (diam(E) + 1)dP (P , Q) for all P , Q ∈ M1 (E). n! Thus, for |h| < 1/(3α),) n * *1/n) · |h| · e/n lim sup E |X|n · |h|n /n! = lim sup (2π n)−1/2 E |X|n n→∞ n→∞ ≤ lim sup (2π n)−1/2 (e/3)n = 0. Hint: Let π = 0 be an invariant measure for X and abbreviate Pπ = π({x})Px x∈E (note that, in general, this need not be a finite measure). and thus h(P, τ) = 0. Since f nk −→ f for k → ∞, we can choose a subsequence (fnkl)l∈N such that μ Al ∩ d f, fnkl > 1/ l < 2−l for any l ∈ N. In the following, we assume that there exists an α ∈ (0, 2] such that U (x) x α−2 is slowly varying at ∞. By Lemma 4.6(iii), we infer (f +g)+ dμ+ f − dμ+ Hence (f + g) dμ = g − dμ = (f +g)− dμ+ + (f + g) dμ − = + f dμ − = f + dμ+ g + dμ. Theorem 1.15 (Intersection of classes of sets) Let I be an arbitrary index set, and assume that Ai is a σ -algebra for every i ∈ I . The probabilistic meaning of this fact is that as a continuous function log(φ(t)) is uniquely defined and thus there exists only one continuous function φ 1/n = exp(log(φ)/n). 442 18 Convergence of Markov Chains ˆ Yˆ) such that Yˆ is independent of By Step 1, there exists a successful coupling (X, Xˇ and Z1 , Z2 , . The family (Ei)i∈I is called independent if, for any finite subset J ⊂ I and any choice of Ej ∈ Ej , j ∈ J , we have P + j ∈J , Ej = P[Ej]. (i) F is the distribution function of a 12 -stable distribution. Here φ(t) = D1 D i=1 cos(ti). For any c > 0, we have |f − g|p = |f − g|p 1{|f −g|≤c} + |f − g|p 1{|f −g|>c} ≤ cp + cp −p |f − g|p . Definition 12.20 (Conditional independence) Let (Ω, F , P) be a probability space, let A ⊂ F be a sub-σ -algebra and let (Ai)i∈I be an arbitrary family of sub-σ -algebras of F . Takeaways For a Markov process, the conditional distribution at time t given the full history until some time s < t is a function of the state at time s only and can be described by a stochastic kernel κs,t . Further, let D1 , . What is the probability that the first position changes infinitely often? It is natural to use the entropy in order to quantify also the randomness of a dynamical system. ♣ 13.2 Weak and Vague Convergence 289 Exercise 13.2.8 Show that (13.4) defines a metric on M1 (E) and that this metric induces the topology of weak convergence. It is easy to see that α := inf{h(x) : |x| ≥ 1} = 1 − sin(1) > 0. Clearly, it is a crucial requirement for any strategy that the decision for the next stake depend only on the information available at that time and not depend on the future results of the gamble. k,N=1 Hence the claim follows. , p − 1}, n→∞ log(d + 1) − log(d) 1 # i ≤ n : ai = d −→ . ♣ Exercise 8.3.2 Let p, q ∈ (1, ∞) with p1 + q1 = 1 and let X ∈ Lp (P) and Y ∈ Lq (μ). Then lim n→∞ fn dμ = f dμ, where both sides can equal +∞. For λ ≥ 0, applying (21.45) twice yields κt (x, dy)κs (y, dz) e−λz = λy κt (x, dy) exp − λs + 1 λ = exp − = exp − = λs+1 λ λs+1 t + 1 x λx λ(t + s) + 1 κt +s (x, dz) e−λz . Clearly, for 1 log Pn ((x, ∞)) n 1 ≥ sup lim inf log Pn ([y, ∞)) = − inf I (y) n→∞ y>x n y>x −I (x) ≥ lim inf n→∞ 1 log Pn ((−∞, x)) ≥ − inf I (y) for x < 0. What is the voltage u(x) at x ∈ E \ A? Show that the statement of the source coding theorem holds for b-adic prefix codes with H2 (p) replaced by Hb (p). n=0 Takeaways A state of a Markov chain is called recurrent if the chain returns to it almost surely. ♣ Exercise 5.3.6 We want to check the efficiency of the Morse alphabet. ◦ F1 (x). Then u(x) = Ex [u(XτA)]. Reflection Can we relax the condition in (viii) that the Xn be dominated to uniform integrability of (Xn)n∈N ? ♦ We collect the most important rules of expectations in a theorem. Choose h and δ(ε) > 0 as in (ii) and C as C in (i). Assume there exists an a > 0 such that for any A ∈ A either μ(A) = 0 or μ(A) ≥ a. 2 Lemma 15.47 If (i) of Theorem 15.44 holds, then kn) * lim log φn (t) − E eit Xn,l − 1 = 0. ♣ 7.4 Lebesgue's Decomposition Theorem In this section, we employ the properties of Hilbert spaces that we derived in the last section in order to decompose a measure into a singular part and a part that is absolutely continuous, both with respect to a second given measure. ♣ Exercise 17.3.3 Show that, almost surely, infinitely many balls of each color are ∞ 1 drawn if = ∞. (ii) For almost all ω ∈ Ω, the map x → f (ω, x) is continuous at the point x0 . For simplicity, assume that Λ = {0, . i=1 i=1 Hence Theorem 2.21 (and Theorem 2.13(i)) yields the independence of (Xi)i=1,...,n . Theorem 11.20 (Kesten–Stigum [95]) Let m > 1. Formally, the set of edges is a subset of the set of subsets of Zd with two elements: E = {x, y} : x, y ∈ Zd with x − y2 = 1 . Then P[Nn,t +1 ≥ l] = E[hn,l (Nn,t)] and hn,l (k) is monotone increasing both in k and in n. l=1 Hence Ym is σ (Xm,l , l ∈ N)-measurable and thus (Ym)m∈N is independent. Compute ψ (1) and the extinction probability. For any ω ∈ Ω, let κ(ω, ·) be the probability measure on (Ω, A) with distribution function F˜ (· , ω). (Hint: Use Exercise 21.2.3 and the optional sampling theorem.) (ii) τb has a 12 -stable distribution with Lévy measure √ ν(dx) = b/(2π) x −3/2 1{x>0} dx. Let F0 = {∅, Ω} and Fn = σ (Y1 , . Define I := := and N 0 Q ∩ I, (Q ∩ I) ∪ max I, r∈ I Nr . , XT). Lemma 15.50 If (i) of Theorem 15.44 holds, then n→∞ ft dνn −→ − t2 . Thus, the theorem is in fact an improvement of Corollary 16.8. The benefit of this theorem will become clear through the following observation. Hence μ(B) = μ(A) + ni=1 μ(Ci) ≥ μ(A) and thus μ is monotone. We say that (Xn,l) = Xn,l , l = 1, . Definition 18.1 (i) For x, y ∈ E, define N(x, y) := n ∈ N0 : pn (x, y) > 0 . Define ST := T Xi . Then I (x1) is the total flow into the network and −I (x0) is the total flow out of the network. The fact that in the complex plane φ has a singularity at t = −iθ implies that the power series of φ about 0 has radius of convergence θ . 5.5 The Poisson Process . In this case, we denote Z ◦ X−1 := φ (even if the inverse map X−1 itself does not exist). Proof Let k1 , . Let Hn be the number of euros to bet in the nth gamble. Let Ξ∞ be a random variable (with values in M1 (E)) such that PΞ∞ = w-lim PΞnl (X) for a suitable subsequence (nl)l∈N . ♣ 5.2 Weak Law of Large Numbers 121 Exercise 5.1.4 Let X1 , X2 , . For all A ∈ A with μ(A) < δ/(2K), we obtain h dμ ≤ Kμ(A) + h dμ < δ; {h≥K} A hence 3 A |f | dμ ≤ ε for all f ∈ F . Then f is Lebesgue integrable on I with integral b f dλ = f (x) dx. Let δ := d(K, Lc) > 0 and let ρK,δ be as in Lemma 13.10. Then f is the characteristic function of a probability measure. Finally, for k ∈ {1, . However, the map ω → Zt (ω) := |Xt (ω) − Xt (ω0)| ∧ 1 is A-measurable. This is trivial. Theorem 12.26 (de Finetti representation theorem) The family X = (Xn)n∈N is exchangeable if and only if there is a σ -algebra A ⊂ F and an A-measurable random variable Ξ∞ : Ω → M1 (E) with the property that given Ξ∞ , (Xn)n∈N is i.i.d. with L[X1 |Ξ∞] = Ξ∞ . , p−1}. Then ψ(x, y) = x 1/p + y 1/p is concave. be independent and identically distributed with PXˆ i = μ. The random variable (Sk +c) 1Ak is σ (X1 , . ε n 123 0.15 0.10 0.05 3.3 3.33 3.37 3.4 3.43 3.47 3.5 3.53 3.57 3.6 3.63 3.67 3.7 0.00 0.00 0.05 0.10 0.15 0.20 0.20 0.25 0.25 5.2 Weak Law of Large Numbers 3.3 3.33 3.37 3.4 3.43 3.47 3.5 3.53 3.57 3.6 3.63 3.67 3.7 3.2 3.3 3.4 3.5 3.6 Fig. Example 20.31 Assume that P is a product measure with marginals π on E. Then (since E[Xn+1 − Xn Fn] = 0) E[(H ·X)n+1 Fn] = E[(H ·X)n + Hn+1 (Xn+1 − Xn) Fn] = (H ·X)n + Hn+1 E[Xn+1 − Xn Fn] = (H ·X)n . Combining the Greek words gives rise to the name ergodic theory, which studies laws of large numbers for possibly dependent, but stationary, random variables. t almost surely and hence X is a modification of X. Why is it not enough to assume in the theorem that (Xn)n∈N be uncorrelated (instead of pairwise independent)? ♣ Exercise 18.4.2 Show (18.17). 18.3 Markov Chain Monte Carlo Method .. Hint: Show that the difference of two independent recurrent random walks is a recurrent random walk. By the one-dimensional central limit theorem, n→∞ we have PSnλ −→ N0,)λ,Cλ* = P)λ, S∞ * . Definition 1.103 (Distributions) Let X be a random variable. , Dn . Here we have the connection to Shannon's theorem. In particular, un (1) > 0 for sufficiently large n. We can compute the effective resistances in parallel and serial connections. n→∞ n→∞ (iv) μ = v-lim μn and {μn , n ∈ N} is tight. , XT −1 , ±1) and VT± = gT (X1 , . are identically distributed. p) *) * *) p (ii) For any p > 1, we have E |XT |p ≤ E (|X|∗T)p ≤ p−1 E |XT |p . We have thus shown the following theorem of Pólya [134]. In this case, it is enough to compute at each step F (0) and F (1) since F is constant if the values coincide. Definition 2.3 (Independence of events) Let I be an arbitrary index set and let (Ai)i∈I be an arbitrary family of events. 1 n n k=0 Xk diverges 16.2 Stable Distributions 389 Hint: Compute the density of F , and show that the Laplace transform is given by √ λ → e− 2λ . If in addition μ is σ -subadditive, then μ∗ (A) = μ(A) for all A ∈ A. ♣ lim n→∞ n→∞ integrable function on Exercise 4.2.3 Let f ∈ L1 ([0, ∞), λ) be a Lebesgue [0, ∞). For any weak 366 15 Characteristic Functions and the Central Limit Theorem limit point Q for (PXn)n∈N and for any λ ∈ Rd , we have λ* Q(dx) ei)λ,x* = E eiX . The general case, can be inferred by the usual approximation arguments (see Theorem 1.96((i))). As a shorthand, we write sup L(φ) for the map x → sup{f (x) : f ∈ L(φ)}. As F is tight, for every ε > 0, there is a K < ∞ with Fn (x) − Fn (−∞) < ε for all n ∈ N and x < −K. 3 3 3 (a) 3 (f + g) dμ 3= f dμ + g dμ. X is called the multi-period binomial model or the Cox–Ross–Rubinstein model (without interest returns). Now σ ({Fn , n ∈ N}) = 2Ω , and 22 1 Basic Measure Theory hence E = {Fn , n ∈ N} is a π-system that generates 2Ω and such that μ(Fn) < ∞ for all n ∈ N. " ⇐ " This is trivial. Show that the map Φ : L1 (F∞) → M, X∞ → (E[X∞ |Fn])n∈N is an isomorphism of vector spaces. For ε > 0, let μ {|fm − fn | > ε} ≤ ε−1 fm − fn 1 −→ 0 for m, n → ∞. Assume that the following conditions are satisfied. The Lebesgue integral does not do that. Compare Example 12.3(iii). Taking the formal derivative ∞ √ d X˙ t := Xt = ξ0 + 2 ξn cos(nπ t) dt n=1 we get independent identically distributed Fourier coefficients for all frequencies. In particular, the assumption that the factors are of Lebesgue– Stieltjes type can be dropped. + Xn)2k−1 ≤ d2k−1 nk−1 and) * (2k)!) *k E (X1 + . Proof The first equation holds by definition. The probabilities of the possible random observations will be described in terms of the distribution of the corresponding random variable, which is the image measure of P under X. Now lim inf n→∞ 1 log Pn (x − ε, ∞) ≥ −I (x) > −I (x + ε) n 1 log Pn [x + ε, ∞) . (ii) If σ, τ ≥ 0, then σ + τ is also a stopping time. Evidently, μ∗ (∅) = 0 and μ∗ is monotone. Then 1 1 n→∞ Xk = f ◦ τ k −→ E[X0 I] n n n−1 n−1 k=0 k=0 In particular, if τ is ergodic, then 1 n n−1 P-a.s. n→∞ Xk −→ E[X0] P-a.s. k=0 Proof If τ is ergodic, then E[X0 |I] = E[X0] and the supplement is a consequence of the first statement. Assume for every T > 0, there are numbers α, β, C > 0 such that *) E |Xt − Xs |α ≤ C|t − s|1+β for all s, t ∈ [0, T]. Then X is called *) *) (v) stationary if L (Xs+t)t ∈I = L (Xt)t ∈I for all s ∈ I , and (vi) a process with stationary increments if X is real-valued and) *) * L Xs+t +r − Xt +r = L Xs+r − Xr (If 0 ∈ I , then it is enough to consider r = 0.) for all r, s, t ∈ I. ♦ Theorem 15.6 (Laplace transform) A finite measure μ on [0, ∞) is characterized by its Laplace transform Lμ (λ) := e−λx μ(dx) for λ ≥ 0. , ck). 12.3 De Finetti's Theorem . Hint: (i) Compute the Laplace transforms LY1 , . Then E[X1] = −i φ (0) < ∞. (14.18) (14.19) To this end, define stochastic kernels κt (x, dy) := δx ∗ N0,t (dy) for t ∈ [0, ∞) where N0,0 = δ0 . ∞ (λeit)n it = eλ(e −1) . Exercise 3.3.1 Assume that we have a branching process Z = (Zn)n∈N0 with Z0 = 1 whose offspring distribution is given by p0 = 1/3 and p2 = 2/3. Reflection Check that each of the families E1 , . ♠♠ Owing to the last theorem, it is natural to define the convolution of two probability measures on Rn (or more generally on an Abelian group) as the distribution of the sum of two independent random variables with the corresponding distributions. Proof The random variables (Xn+)n∈N and (Xn−)n∈N again form pairwise independent families of square integrable random variables (compare Remark 2.15(ii)). < tn , we have Px ◦ (Xt0 , . At the other end of the scale, the strongest notion is "i.i.d.". A locally compact, separable metric space is manifestly σ -compact and there even exists a countable basis U of the topology consisting of relatively compact open sets. Finally, let Px := L[Xx] be the distribution of Xx . 3.1 Definition and Examples Definition 3.1 (Probability generating function) Let X be an N0 -valued random variable. Show that Y has distribution Q. , ωn ∈ {0, 1}, n ∈ N). ∈ A with AN ↑ Ω and μ(AN) < ∞ for all N ∈ N. 3 n→∞ If we define un (t) := (1−e−t x) νn (dx), then (as in (16.1)) un (t) −→ u(t) for all t ≥ 0. P (iii) Let n = Tε . (23.21) 606 23 Large Deviations Proof This is left as an exercise. Proof By Theorem 14.12(iii), the finite-dimensional distributions determine P uniquely. , ωn] is compact. (iii) Let (Yn)n∈Z be i.i.d. real random variables and let c1 , . Remark 7.48 As F is continuous, for any δ > 0, there exists an ε > 0 such that |F (f)| < δ for all f ∈ V with f < ε. Then: (i) F := {r ∈ [0, 1] : ψ(r) = r} = {q, 1}. = e∈A˜ i pe . An alternative strategy of proof is: Sort the values of the Xi by size M = X(1) > D X) (2) > . , Sn } − Mn ◦ τ) 1{Mn >0}) * = E (Mn − Mn ◦ τ) 1{Mn >0} *) ≥ E Mn − Mn ◦ τ = E[Mn] − E[Mn] = 0. " ⇒ " Assume F is uniformly integrable. (3.10) Proof The map f : x → (1 + x)α is holomorphic up to possibly a singularity at x = −1. We call (Zd , E p) a percolation model (more precisely, a model for bond percolation, in contrast to site percolation, where vertices can be open or closed). 18.1 Periodicity of Markov Chains We study the conditions under which a positive recurrent Markov chain X on the countable space E (and with transition matrix p), started in an arbitrary μ ∈ M1 (E), converges in distribution to an invariant distribution π; that is, n→∞ μpn −→ π. Now assume inf f (K) < ∞. ♦ However, this is the CFP of br,p rν r,p Not every infinitely divisible distribution is of the type CPoiν , however we have the following theorem. , Mn,m = km · P[L = n] n n! km −α α p1k1 · · · pm e k1 ! · · · km ! n! m k i λ e−λi i . By what we have shown already, there exists a compact set C ⊂ A ∩ Kn with λ((A ∩ K n) \ C) < 1; hence λ(C) > L. Exercise 11.2.1 For p = 1, the statement of Theorem 11.10 may fail. By Theorem 2.26, the two random variables are independent, and *) *) *) E (Sk + c) 1Ak (Sn − Sk) = E (Sk + c) 1Ak E Sn − Sk = 0. n→∞ n→∞ By assumption, |fn | ≤ h and fn −→ f (· , x0) almost everywhere. Let E := g −1 ({1}). That is, e−t Xt converges in distribution to a random variable W with W ∼ exp1 . The Markov process X on N0 with this semigroup is called a Poisson process with (jump) rate θ . (i) X is called a random variable with values in (Ω , A). Proof A = 2Ω is a σ -algebra with E ⊂ A. Let N := max n ∈ N0 : X1 + . Hence there exists an n0 = 1 we conclude that n1 n−1 A i=0 i pn pnε − ε n0 (ω) such that n−1 i=0 1Ai ≥ 2 for all n ≥ n0 . By Corollary 9.34, Y is a submartingale, and (by Theorem 9.39) so are H ·Y and (1 − H)·Y . By Remark 1.17, this implies σ (E1) ⊂ σ (E4). 6.3 Exchanging Integral and Differentiation 161 By assumption (ii), we have n→∞ gn −→ f (· , x0) μ-almost everywhere. We define F := σ (Bi , i ∈ I). Assuming rs = 0 for all s ≤ (n − 1)t, we conclude nt sup rs ∞ ≤ 2λ ru ∞ du ≤ 2λt sup rs ∞ = 0, s≤nt s≤nt (n−1)t s = ps for all hence rs = 0 for all s ≤ nt. , k, the limit Ξ∞ (Al) = limN→∞ ΞN (Al) exists in a suitable sense. n=1 We have to show A ∈ M(μ∗); that is, μ∗ (A ∩ E) + μ∗ (Ac ∩ E) ≤ μ∗ (E) Let Bn = n for any E ∈ 2Ω . In Definition 4.16, for measurable f : Ω → R, we defined f p := 1/p |f | dμ p for p ∈ [1, ∞) and f ∞ := inf K ≥ 0 : μ(|f | > K) = 0 . Step 2. Apply a voltage of 1 at the origin and 0 at the endpoints of the paths at the nth stage. (Hint: Assume the contrary and show that the corresponding random variable would have variance zero.) ♣ α Exercise 15.4.4 Let X1 , X2 , . (16.32) Theorem 16.28 (i) If PX is in the domain of attraction of some distribution, then there exists an α ∈ (0, 2] such that (16.32) holds. are i.i.d. and ∼ Rad1/2 ; that is, P[Ri = 1] = 1 − P[Ri = −1] = 1 . We show P[A] = 0. Is this true also for P[· |X = x]? By Lemma 1.51, μ is a measure and μ is σ -finite since μ is σ -finite. Let An = En ∩ {f1 > f2 } for n ∈ N. Without loss of generality, we may assume that X is the canonical process on the probability space (Ω, A, P) = E N0 , B(E)⊗N0 , P . Remark 4.14 We still have to show that ν is a measure. ♦ Remark 6.11 Both L1 -convergence and almost everywhere convergence imply convergence in measure. ♦ Remark 6.4 Almost everywhere convergence implies convergence in measure: For ε > 0, define Dn (ε) = d(f, fm) > ε for some m ≥ n . (18.17) Apart from the double zero at 1, we get the zeros σ cos πk/N), k = 1, . −n be Exercise 7.4.1 For every x ∈ (0, 1], let x = (0, x1 x2 x3 . Definition 1.29 Let A be a semiring. However, since X is a supermartingale, for every s > t, we have Xt ≥ E[Xs |Ft] Q+ s↓t, s>t −→ t |Ft] = X t E[X in L1 . k=1 For c = Var[Sn]/t ≥ 0, we obtain P[A] ≤ Var[Sn] c(t + c) tc Var[Sn] + c2 = 2 . Show that λ = v-lim μn but that (μn)n∈N does not converge weakly. + Tn−1 ≤t and s ≤t . ♣ k=1,...,n k=1,...,n 5.5 The Poisson Process 139 5.5 The Poisson Process We develop a model for the number of clicks of a Geiger counter in the (time) interval I = (a, b]. , d, the map yi → f (x1 , . Definition 15.21 Let (E, d) be a metric space. < γN of F such that F (y0) < ε, F (yN) > F (∞) − ε and yi − yi−1 < ε for all i. If there is no infinite open component, then the water may wet only a thin layer at the surface. Hence we have constructed a (random) subgraph (Zd , E p) of (Zd , E). 440 18 Convergence of Markov Chains Definition 18.5 A bivariate process ((Xn , Yn))n∈N0 with values in E × E is called a coupling if (Xn)n∈N0 and (Yn)n∈N0 are Markov chains, each with transition matrix p. Then F (x) =)x, f * for all x ∈ V . Definition 7.24 (Orthogonal complement) Let V be a real vector space with inner product) · , · *. (ii) For t ∈ I and ε > 0, let Uε (t) := {s ∈ I : |s − t| < ε}. What goes wrong if ∩-stability is missing? Indeed, let fn −→ f and meas fn −→ g. (23.15) Proof We consider the set of possible values for the n-tuple (X1 , . ♠♠ Example 5.21 (Monte Carlo integration) Let f : [0, 1] → R be a function and 31 assume we want to determine the value of its integral I := 0 f (x) dx numerically. Theorem measure) There exists a uniquely determined measure 1.55 (Lebesgue λn on Rn , B(Rn) with the property that λn ((a, b]) = n (bi − ai) for all a, b ∈ Rn with a < b. Other important convergence theorems for integrals follow in Chaps. n If m, n ∈ N, then write mn if m is a divisor of n; that is, if m ∈ N. , αm ∈ (0, ∞), and for mutually disjoint sets A1 , . Proof This is obvious. In order to show that DE is a λ-system, we check the properties of Definition 1.10: (i) Clearly, Ω ∈ DE . By Theorem 13.16, it is enough to show that n→∞ f dμn −→ f dμ. Then the bivariate Markov chain Z := ((X˜ n , Y˜ n))n∈N0 has the transition matrix p defined by p (x1 , y1), (x2 , y2) = p(x1 , x2) · p(y1 , y2) . Finally, there are more exercises and some new illustrations. A prominent role will be played by the complex exponential function exp : C → C, which can be defined either by Euler's formula exp(z) = exp(u) cos(v) + i sin(v) n or by the power series exp(z) = ∞ n=0 z /n!. ˜ B˜ ⊂ {1, . ♣ 19.5 Network Reduction Example 19.32 Consider a random walk on the graph in Fig. The statement can be derived more simply than by direct computation if ∗n (see Example 3.4(ii)). For continuous time, however, in general, some work has to be done to establish the strong Markov property. However, an infinitely divisible real random variable X is not simply the difference of two infinitely divisible nonnegative random variables, as the normal distribution shows. ♣ Chapter 3 Generating Functions It is a fundamental principle of mathematics to map a class of objects that are of interest into a class of objects where computations are easier. It should • be random and independent for disjoint intervals, • be homogeneous in time in the sense that the number of clicks in I = (a, b] has the same distribution as the number of clicks in c + I = (a + c, b + c], • have finite expectation, and • have no double points: At any point of time, the counter makes at most one click. Let f : Z → R be an arbitrary map. Then z = 0 and F (z) = F (v) − F (y) = F (v) = 0. , n,/let (Ωi , A/ i , Pi) be a probability space. We can interpret X + 1 as the waiting time for the first success in a series of independent random experiments, any of which yields a success with probability p. 5.) Takeaways Independent coin tosses decide if an edge of Zd is retained (probability p) or removed. , 0) and x N = (0, . Proof (i), (ii) and (iii) These are evident. However, we show that, for open A this can be done at least asymptotically. For example, in the preceding theorem, claim (i) holds with the words "submartingale" and "supermartingale" interchanged, claim (iv) holds for submartingales if the minimum is replaced by a maximum, and so on. Exercise 20.5.1 Show that "strongly mixing" implies "weakly mixing", which in turn implies

= (Xt)t ∈[0,1] that is defined by the covariance function Γ (s, t) = s ∧ t − st. Remark 5.31 Condition (5.15) is sharp in the sense that for any increasing sequence −2 2 (an)n∈N with ∞ n=1 an (log n) = ∞, there exists a sequence of pairwise independent, square integrable, centered random variables X1 , X2 , . However, if F1n turns out to be a constant map (e.g., F1n ≡ x ∗ for some random x ∗), then we will also have F1m ≡ x ∗ for all m ≥ n. n=1 Clearly, C is a countable set of compact sets in E, and C is stable under formation of unions. Inductively, we get (finite) additivity. As an application, we (Zε) e get the statistical physics principle of minimising the free energy. (iii) X has the time-homogeneous Markov property (MP): For every A ∈ B(E), every x ∈ E and all s, t ∈ I , we have *) Px Xt +s ∈ A Fs = κt (Xs , A) Px -a.s. Here, for every t ∈ I , the transition kernel κt : E × B(E) → [0, 1] is the stochastic kernel defined for x ∈ E and A ∈ B(E) by κt (x, A) := κ x, {y ∈ E I : y(t) ∈ A} = Px [Xt ∈ A] . Proof Let f : E → R be bounded and Lipschitz continuous with constant K. Lemma 15.37 Let X1 , X2 , . Inductively, we get n κk (0, ϕn (A1 × · · · × An)) n1 n−1 = n κk ϕn−1 (A1 ×···×An−1) n−1 0, d(y1, . Let Γ(x) := inf Iμ (Ex) (where Iμ (ν) = H (ν |μ) is the 602 23 Large Deviations relative entropy). Hence we have (Ω, A, P, τ) is ergodic ⇐⇒ gcd(r, n) = 1. (2.12) Consequently, the edges in E \ E p are called closed. Similarly, we can use the first inequality in Lemma 23.12 to get *) P ξn (X) ∈ A ≥ (n + 1)−#Σ exp − n inf Iμ (A ∩ En) and thus lim inf n→∞) * 1 log P ξn (X) ∈ A ≥ − lim sup inf Iμ (A ∩ En) . With this interpretation, it is particularly evident that it is natural to construct new stochastic processes by considering investment strategies for the stock. Proof Let ϕ be the CFP of μ. n (iii) κ = 0 Ak , n ∈ N converges almost surely. ∈ A with An ↑ A. 154 6 Convergence Theorems By construction, 3 {|f |>gε } (|f | − gε/2)+ dμ ≤ ε/2 and + gε/2 1{|f |>gε } ≤ |f | − gε/2 1{|f |>gε } ; hence also {|f |>gε } gε/2 dμ ≤ {|f |>gε } (|f | − gε/2)+ dμ ≤ ε . Define νn = n→∞ nμn ∈ Mf ([0, ∞)). Let I ⊂ R be an interval. n 20.2 Ergodic Theorems 499 Let ε > 0 and F := {Z > ε}. k=1 Hence μ is subadditive. In particular, letting c = f − gp we obtain p 1/p f − gp ≤ cp μ(Ω) + cp −p f − gp = (1 + μ(Ω))1/p f − gp . However, we have to make sure also that the simultaneous occurrence of A1 and A2 does not change the probability that A3 occurs. Then PX =: Poiλ is called the Poisson distribution with parameter λ. For all t ≥ 0, the function (compare (15.8)) ft : [0, ∞) → [0, ∞), x → 1−e−tx , 1−e−x t, if x > 0, if x = 0, 374 16 Infinitely Divisible Distributions is continuous and bounded (by t ∨ 1). (2.17) This part of the proof is the most difficult one. Accordingly, assume F is uniformly integrable (but not necessarily μ(Ω) < ∞). 16.2, we will show that this is true for all α ∈ (0, 2]. For measurable g : Ω → R, let Ug be the set of points of discontinuity of g. By the Hahn-Banach theorem n→∞ of functional analysis (see, e.g., [87] or [174]), F can be extended to a continuous linear functional on ∞ . By construction, X is stationary with respect to P lim P and ←− n→∞ n)n∈N0 −1 = P ◦ (Xn)n∈N0 −1 . On the other hand,/ each Xi is continuous and thus measurable with respect to B − Bi . That is, we obtain the following lemma. n n n−1 n−1 k=0 k=0 n By Birkhoff’s ergodic theorem, we have Yn −→ P[B] almost surely. Similarly, the claim follows for (IA1 , . As a by-product we get a discrete version of the celebrated Black-Scholes formula. Proof The space L2 (μ) fulfills the conditions of Corollary 7.27. Then use the approximation arguments as in Exercise 4.2.5. ♣ 4.3 Lebesgue Integral Versus Riemann Integral We show that for Riemann integrable functions the Lebesgue integral and the Riemann integral coincide. By the Borel– n=1 Cantelli lemma (Theorem 2.7), we infer lim sup n→∞ 1 sup Bt − Bn , t ∈ [n, n + 1] = 0 n Hence X is also continuous at 0. 19.6). , BtN). (14.5) i=1 n μi := μ1 ⊗ · · · ⊗ μn := μ is called the product measure of the μi . Proof Assume that there does exist a topology that induces almost everywhere meas convergence. ♣ wn n=0 17.4 Discrete Markov Chains: Recurrence and Transience In the following, let X = (Xn)n∈N0 be a Markov chain on the countable space E with transition matrix p. As a simple application of the strong Markov property, we show the reflection principle for random walks. (x) Somewhat more generally, there is no nontrivial infinitely divisible distribution that is concentrated on a bounded interval. TV ; ; Thus, for every x ∈ E, we have ;δx pn − π ;T V > 0. ∈ M≤1 (E). If x ∈ E, then denote by x i,σ the state in which at site i the spin is changed to σ ∈ {−1, +1}; that is, x i,σ (j) = σ, if j = i, x(j), if j = i. At the first stage, a probability distribution Ξ on E is drawn at random. only be computed numerically (Fig. 13.1 A Topology Primer 3 Here come the details. x≥0, λ∈I dλ x≥0, λ∈I Thus F fulfills the assumptions of Theorem 6.28. L1 Remark 6.9 If fn −→ f , then in particular L1 3 n→∞ fn dμ −→ 3 f dμ. Clearly, |YiK | ≤ K for every i. Then (X, Y) is a successful coupling and Xn = Yn for n ≥ τ . The set of such rectangular cylinders for which in addition Aj ∈ Ej for all j ∈ J holds will be denoted by ZJE ,R . 17.5 Application: Recurrence and Transience of Random Walks 419 Hence we only have to compute the asymptotics of P0 [Yt1 = 0] for large t. Therefore, (iii) does n→∞ not hold. Under certain assumptions on the continuity of the paths, however, the two notions coincide. We assume that there exists a map ψ : [0, ∞) → [0, 1] n→∞ that is continuous in 0 and such that ψn −→ ψ pointwise. (i) Let p ∈ [0, 1] and P[X = 1] = p, P[X = 0] = 1 − p. Show that the following hold. The final step is to show convergence in path space. Furthermore, we have pf = f on E \ A if and only if pf ˜ = f on E \ A. 416 17 Markov Chains J Clearly, Y1 = (Y11 , . Proof “ ⇒ ” This is obvious. n→∞ Proof Let (fN)N∈N be a sequence in Cc (E; [0, 1]) with fN ↑ 1. This is Donsker’s theorem that is known in the physics literature as the invariance principle. ε→0 Note that ρC,ε −→ 1C . Furthermore, P[A ∩ B] = ˜ #(A˜ × B) #A˜ #B˜ = · = P[A] · P[B]. Our intuitive understanding of an edge is a connection between two points x and y and not an (unordered) pair {x, y}. Note that R(x) > 0, as A is open. As we want to use only zeros and ones (and no gap-like symbols), we have to arrange the code in such a way that no code is the beginning of the code of a different symbol. If, on the other hand, (Nt , t ≥ 0) is a Poisson process, then (Nt − Ns , (s, t] ∈ I) has properties (P1)–(P5). (ii) Consider the Bernoulli distributions Berp and Berq for p, q ∈ [0, 1]. Define u(0) = 0 and u(2) = 1. 11.2 Martingale Convergence Theorems .. One can use second moments and Chebyshev’s inequality to establish a weak law of large numbers. For extensive literature on stochastic aspects of mathematical finance, we refer to the textbooks [9, 42, 48, 57, 86, 102, 121] or [160]. = 2 and Var[X] < ∞, let E[X] = 0. Unlike a, of s depends on every coordinate if it is finite. (ii) If F , G ⊂ L1 (μ) are uniformly integrable, then (f + g : f ∈ F, g ∈ G), (f − g : f ∈ F, g ∈ G) and {|f | : f ∈ F } are also uniformly integrable. Definition 9.10 A stochastic process X = (Xt , t ∈ I) is called adapted to the filtration F if Xt is Ft -measurable for all t ∈ I . For probability measures on R, weak convergence is tantamount to convergence of distribution functions at all points of continuity of the limiting function. ♣ 488 19 Markov Chains and Electrical Networks a z Fig. , ωn] = B2−n (ω) = {ω ∈ Ω : d(ω, ω) < 2−n }. = ω : |ω(0)| ≤ Kε , V (ω, δN,k,ε) ≤ k 548 21 Brownian Motion By the Arzelà–Ascoli theorem, Cε := CN,ε is relatively compact in C([0, ∞)) N∈N and we have Pi (Cεc) ≤ ∞ ε + Pi ω : V N (ω, δN,k,ε) > 1/k ≤ ε 2 for all i ∈ I. In the special cases f (x) = x 2 and f (x) = |x|, the second derivative is f (x) = 2 and f (x) = 2·1{0} (x), respectively. With the notation of Theorem 1.53, this completion is Ω, M(μ∗), μ∗ M(μ∗) . − ∗ b− = b− Exercise 3.1.1 Show that br,p r+s,p for r, s ∈ (0, ∞) and p ∈ (0, 1]. 2.1 Independence of Events 57 Proof This is left as an exercise. ♦ The following two definitions make sense also for more general index sets I that are partially ordered. Existence Let X+ = X ∨ 0 and X− = X+ − X. (iii) Since |αf + βg| ≤ |α| · |f | + |β| · |g|, Lemma 4.6(i) and (iii) yield that αf +βg ∈ L1 (μ). .}) + μ({N + 1, N + 2, . be distribution functions of probability mean−∞ sures on R. For B ∈ A, P[B |F] := E[1B |F] is called a conditional probability of B given the σ -algebra F . By k=1 n→∞ n→∞ (ii) Clearly, ϕY (t) = ∞ *) P[N = n] E ei)t,X1 +...+Xn * n=0 = ∞ P[N = n] ϕX (t)n = fN (ϕX (t)). Rather, we just used the fact that the differences (ΔX)n := Xn − Xn−1 take only the values −1 and +1. Hence t =0 X is the required Markov process. , αN ∈ (0, ∞) and mutually disjoint sets A1 , . Therefore, by Theorem 19.25, is finite if and only if Rw w X is transient ⇐⇒ − + < ∞ or Rw < ∞. Now assume that a prefix code is given. Indeed, ϕr,p (t) = erψp (t), where (br/n,p ψp (t) = log(p) − log 1 − (1 − p)eit . Since {A} ∈ U(A), we have μ (A) ≤ μ(A). r k! k − = CPoi If we had br,p rν for some ν ∈ Mf (N), then we would have ν({k}) = k (1 − p) /k. We conclude that Nn1 ,t +1 ≤st Nn2 ,t +1 which completes the induction and the proof of the theorem. Use the martingale convergence theorem to n=1 n2 show the strong law of large numbers for (Xn)n∈N . Definition 1.46 (Outer measure) A set function μ∗ : 2Ω → [0, ∞] is called an outer measure if (i) μ∗ (∅) = 0, and (ii) μ∗ is monotone, (iii) μ∗ is σ -subadditive. We denote the effective conductance from x1 to ∞ by Ceff (x1 ↔ ∞) := C(x1) inf pF (x1 , A0) : A0 ⊂ E with |E \ A0 | < ∞, A0 x1 . That is, Z ∈ Z is a finite subset of A such that the sets C ∈ Z are pairwise disjoint and C∈Z C = Ω for all Z. Thus also Ac ∈ Ai for any i ∈ I . (5–12) Exhaustion arguments similar to that in (4) also work for rectangles. ♠ Example 19.28 Symmetric simple random walk on E = Z2 is recurrent. Let L > 0 and let (Xn)n∈N be a martingale with the property Xn+1 − Xn ≤ L a.s. (11.2) Define the events C := (Xn)n∈N converges as n → ∞ , A+ := lim sup Xn < ∞ , n→∞ A− := lim inf Xn > −∞ , n→∞ F := sup)X*n < ∞ . Of course, you do not want to invent a new code for every message but rather use one code that allows for the shortest average coding of the messages for the particular information source. Since we will not need the statement here, we leave its proof as an exercise (compare Exercise 17.6.6; see also [39, Section 6.5]). For example, such a process can be used to describe the motion of a particle immersed in water or the change of prices in the stock market. 260 12 Backwards Martingales and Exchangeability (iii) For x ∈ E N , define the nth empirical distribution by ξn (x) = n1 ni=1 δxi (recall that δxi is the Dirac measure at the point xi). In this case, κn is called the family of n-step transition probabilities. The maximal entropy of a probability measure on N points is achieved by the uniform distribution and is log(N). 15.2 illustrate this. ∗1/n Proof Apply Theorem 16.6, where ϕn is the CFP of μn . In many cases, the lower bound is a lot easier to show than the upper bound. 482 19 Markov Chains and Electrical Networks Step-by-Step Reduction of the Network Having seen how to compute u(x) from the effective resistances, we now turn to the systematic computation of these effective resistances. At first reading, the reader might wish to skip this rather analytically flavored chapter. Here the statement is trivial (choose a = 0, b = 1 and c = −E[Y]). Due to the analogy of (19.3) to Green’s formula in continuous ˜ is called the Green function for the equation space potential theory, the function G (p − I)f = 0 on E \ A. Define p(x, y) = 1 q(x, y) + I (x, y) λ for x, y ∈ E, if λ > 0 and p = I otherwise. For n ∈ N0 , define YnN := Yn . In particular, a probability measure μ on R is uniquely determined by its distribution function F : R → [0, 1], x → μ((−∞, x]). That is, (2.8) holds with J replaced by J ∪ {j }. For n ∈ N choose μn ∈ M1 (R) such that μ∗n n = μ and let ϕn the CFP of μn . ∈ A with (a, b] ⊂ ∞ (a(k), b(k)]. Denote by μ0n ∈ M1 (M1 (Σ)) the corresponding a priori distribution of x; that is, of the n-particle system. Since 1 ∧ x ≤ 2(1 − e−x) for all x ≥ 0, clearly (1 ∧ x) ν(dx) ≤ 2 (1 − e−x) ν(dx) ≤ 2u(1) < ∞. Remark 24.11 The definition of the PPP (and its construction in the following theorem) still works if (E, E, μ) is only assumed to be a σ -finite measure space. By the triangle inequality, f dμ1 − f dμ2 ≤ f dμ1 − g dμ1 + g dμ1 − g dμ2 + g dμ2 − f dμ2 ≤ ε (μ1 (E) + μ2 (E)). It remains to show that PNt = Poiαt . n + 1 − ns Proof Compute ψ(s) = ∞ 2−k−1 s k = k=0 1 . Hence, the random walk X = (Xn)n∈N := (XnN)n∈N fulfills the condition from Step 2. Then lim sup sup Fn (x) − F (x) = 0 almost surely. One can even construct infinite products if all factors are probability spaces (Theorem 14.39). That is, two points x, y ∈ E are connected by an edge if they differ in exactly one coordinate. By Fatou’s lemma, lim sup gk dμ = k→∞ g dμ − lim inf k→∞ (g − gk) dμ ≤ g dμ − lim (g − gk) k→∞ dμ = 0. , Xtn+1 ∈ Bn+1 Ftn = EXs [f (X)] . Since ϕ r (t) = 0 for all t ∈ R, we have ϕ r → 1 for r → 0 and thus the semigroup is continuous. F is called right continuous if F (x) = limn→∞ F (xn) for all x ∈ Rd and every sequence (xn)n∈N in Rd with x1 ≥ x2 ≥ x3 ≥ . be identically distributed. This shows (ii). Then Y is manifestly a σ (X)-martingale. k→∞ (13.11) Assume that we can show that there is a measure μ on the Borel σ -algebra E of E such that μ(A) = sup α(C) : C ∈ C with C ⊂ A for all A ⊂ E open. meas With our definition of stochastic convergence we have fn −→ f . Since μ is lower semicontinuous, we have Fμ (∞) = μ(R); hence Fμ is indeed a (possibly defective) distribution function if μ is a (sub-) probability measure. Clearly, σ (E) = 2Ω but E is not a π-system. j ∈Λ: i∼j Here i ∼ j indicates that i and j are neighbors in Λ (that is, coordinate-wise mod N, we also speak of periodic boundary conditions). This implies lim sup P[A ∩ τ −n (B)] − P[A] P[B] |n|→∞ ≤ lim sup P[Aε ∩ τ −n (B ε)] − P[Aε] P[B ε] + 4ε = 4ε. In particular, F and σ (X) are independent. Define Mn = max{0, S1 , . Proof The proof makes use of the Radon-Nikodym theorem (Corollary 7.34). If c− = c+ , then μ is a Cauchy distribution. We now take the waiting times as the starting point and, based on them, construct the Poisson process. Hence, for n ∈ N0 , pl+n+k (y, y) ≥ pl (y, x) pn (x, x) pk (x, y). ♣ Exercise 9.2.2 Let (Xn)n∈N0 be a predictable F-martingale. As (Zn)n∈N is decreasing, by (ii) also (E[Zn F])n∈N decreases to some limit, say, Z. (ii) Let E(ω) be a property that a point ω ∈ Ω can have or not have. The procedure of defining two families of random variables that are related in a specific way (here “≤”) on one probability space is called a coupling. / Then (Ω, τ) is Polish and B = Bi . Define the evaluation map Xt : Ω → R, ω → ω(t), (21.27) that is, the restriction of the canonical projection R[0,∞) ; → R to ; For f, g ∈ C [0, ∞) and n ∈ N, let dn (f, g) := ;(f − g) Ω. Further, define Z∗R = ∞ N=1 N An : A1 , . Monotonicity is implied by convexity. Theorem 18.8 Let X be an arbitrary aperiodic and irreducible random walk on Zd with transition matrix p. (21.46) Hence Y is a martingale, and the first centered moments are Ex [(Yt − x)2] = 2t, Ex [(Yt − x)3] = 6t 2 , Ex [(Yt − x)4] = 24t 3 + 12t 2 t 2 , (21.47) Ex [(Yt − x)5] = 120t t 4 + 120t 2 t 3 , Ex [(Yt − x)6] = 720t t 5 + 1080t 2 t 4 + 120t 3 t 3 . Show that v-lim μn = 0 but n→∞ that (μn)n∈N does not converge weakly. That is, Xn = n Rk , where R1 , R2 , . Define Yn := k ci Xn−i . Give an example of a nonnegative martingale X with E[Xn] = 1 for all n ∈ N but such that n→∞ Xn −→ 0 almost surely. In this chapter, we establish a similar stability property for martingales that are stopped at a random time. What is the probability P that this Markov chain visits 1 before it visits 0? Consider first the simplest situation of symmetric simple random walk X on ZD . We define an Nm 0 -valued random variable Y = (Y1 , . In addition to the properties we required here, Brownian motion has continuous paths; that is, the maps t → Xt are almost surely continuous. (v) (Gamma distribution) Again it suffices to consider the case θ = 1. For n ∈ N0 and ε > 0, let Aεn := Sm > Sn + ε for all m ≥ n + 1 ∩ B. Show that the entropy H (p) is minimal (in fact, zero) if p = δe for some e ∈ E. 6.2 Uniform Integrability 155 Proof “ ⇐ ” Assume there is an H with the advertised properties. x=−∞ Now what about symmetric simple random walk in dimension D = 2 or in higher dimensions? , ωn]. ♣ A ∈ B(R) is a μ-null set, then dx Exercise 13.1.7 (Fundamental theorem of calculus) (Compare [37].) Let f ∈ L1 (Rd), μ = f λd and let C ⊂ Rd be open, convex and bounded with 0 ∈ C. , tn ∈ Rd and all y1 , . (i) The Lindeberg condition holds. , E[Xn] = ϕ E[X1] , . Then the set function E[X] : B(E) → [0, ∞], A → E[X(A)] is a measure. Exercise 17.2.1 (Discrete martingale problem) Let E ⊂ R be countable and let X be a Markov chain on E with transition matrix p and with the property that, for any x, there are at most three choices for the next step; that is, there exists a set Ax ⊂ E of cardinality 3 with p(x, y) = 0 for all y ∈ E \ Ax . Definition 14.16 Let (Ω, A, μ) be a measure space and (Ω , A) a measurable space. i=1 For j ∈ {σk , . That is, we consider a two-stage experiment: At the first stage we choose a realization of i.i.d. random variables W = (Wi−)i∈Z on (0, 1) and let Wi+ := 1 − Wi− . m=1 n=m m→∞ n=m However, for every m ∈ N (since log(1 − x) ≤ −x for x ∈ [0, 1]), by upper continuity of P P + ∞ + , N ∞ 1 − P[An] Acn = lim P Acn = N→∞ n=m = exp ∞ n=m n=m n=m ∞ log 1 − P[An] ≤ exp − P[An] = 0. Theorem 5.3 (Rules for expectations) Let X, Y, Xn , Zn , n ∈ N, be real integrable random variables on (Ω, A, P). 20.4 Application: Recurrence of Random Walks . 2 Reff (0 ↔ 1) (19.15) In particular, in the case R (0, 1) = ∞ (or equivalently Reff (0 ↔ 1) = Reff (0 ↔ x) + Reff (x ↔ 1)), we have Reff (0 ↔ x) = R (0, x) and Reff (1 ↔ x) = R (1, x), hence u(x) = Reff (0 ↔ x) . For a1 < b1 < n (ai , bi], define a2 < b2 < . Theorem 11.4 (Martingale convergence theorem) Let (Xn)n∈N0 be a submartingale with sup{E[Xn+] : n ≥ 0} < ∞. Under some mild conditions on the continuity of φ, the main contribution to the integral comes from those points x that are not too unlikely (for με) and for which at the same time φ(x) is large. n=1 Hence I : L2 ([0, 1]) → L2 (P), f → I (f) is an isometry. Corollary 19.16 Let X be a Markov chain on E with edge weights C. In some of the examples, the elements of the generating class are simpler sets such as rectangles or compact sets. For n ∈ N0 and t ≥ 0, define the probability fn (t) := P1 [Xt > n]. ♣ n→∞ Exercise 13.2.14 Let μ, μ1 , μ2 , . To this end, we first study more general set functions that assign nonnegative numbers to subsets. ; ; 3 Note that ;f˜;p and f˜ dμ do not depend on the choice of the representative f ∈ f˜. n→∞ Hence the characteristic function can be expanded about any point t ∈ R in a power series with radius of convergence at least 1/(3α). nπ This representation of the Brownian motions goes back to Paley and Wiener who also show that along a suitable subsequence the series converges uniformly almost surely and hence the limit X is indeed continuous, see [125, Theorem XLIII, page 148]. E[ϕ(Xt)+] = E ϕ E[Xt ∗ Ft] (iii) This is evident since x → |x|p is convex. The remainder of this section is devoted to the proof of this theorem. Using the binomial theorem, we get (note that the mixed terms) * with odd moments vanish since E Y1Kn = 0) *) *) *2) E (TnKn)4 = n E (Y1Kn)4 + 3n(n − 1) E (Y1Kn)2 ≤ nKn2 σ 2 + 3n(n − 1)σ 4 . Further, for any n ∈ N and x ≥ 2an , H (x)/x ≥ nk=1 (1 − ak /x)+ ≥ n/2; hence we have H (x)/x ↑ ∞. Example 6.29 (Laplace transform) Let X be a nonnegative random variable on (Ω, A, P). In this chapter, we define the integral by an approximation scheme with simple functions. (2.15) We need a 0–1 law similar to that of Kolmogorov. By Definition 4.13, a measurable function f : Ω → [0, ∞) is called a density of ν with respect to μ if ν(A) := f 1A dμ for all A ∈ A. , BtN − BtN−1)]. Indeed, ϕα,γ = ϕα,γ . We now compute ∞ GY := ∞ ∞ P0 [Yt = 0] dt = 0 0 = ∞ n=0 *) P0 X2n = 0, Tt = 2n dt n=0 p2n (0, 0) ∞ e−t 0 t 2n dt = G(0, 0). We formalize the description given above. , ωk] ∩ Bn = ∅ for all k, n ∈ N. If we place n indistinguishable particles independently according to λ on the random positions z1 , . Hence, let ν 0 μ. Hence, for efficiency, those symbols that appear more often get a shorter code than the more rare symbols. Assume that k ∈ N and let ϕ : E k → R be measurable with E[|ϕ(X 1 , . Then λ can be extended uniquely to a measure λ∗ on B ∗ (Rn) = σ B(Rn) ∪ N , where N is the class of subsets of Lebesgue–Borel null sets. (16.31) A function H : (0, ∞) → (0, ∞) is called slowly varying at ∞ if lim x→∞ H (γ x) =1 H (x) for all γ > 0. (21.5) on the dyadic rational numbers and then to The idea is first to construct X extend it continuously to [0, 1]. , en−1)}) = P {e0 } × . n∈Z be the canonical process on Ω = E Z . Remark 6.3 Let A1 , A2 , . 1{x(j)=x(i)} − 2j : j ∼i Hence π(x i)/π(x) = exp − 2β j ∼i 1{x(j)=x(i)} − 12 , and this expression is easy to compute as it depends only on the 2d neighboring spins and, in particular, does not require knowledge of the value of Zβ . Let π be the invariant ˜ Then, clearly, the product measure π ⊗π ∈ M1 (E ×E) is an (and distribution of X. (b) 3 αf dμ = α f3 dμ for α ≥ 0. (P3) If J ⊂ I with I ∩ J = ∅ for all I, J ∈ J with I = J , then (NJ , J ∈ J) is an independent family. Using the generalized binomial theorem (see Lemma 3.5), we get (since we have (1 − 4p(1 − p))1/2 = |2p − 1|) G(0, 0) = 1 |2p−1| , if p = 12 , ∞, if p = 12 . be Lp -integrable martingales. We write 3 L1 (μ) := L1 (Ω, A, μ) := f : Ω → R : f is measurable and |f | dμ < ∞ . Hence it can be developed in a power series about 0 with radius of convergence at least 1: f (x) = ∞ f (k) (0) k=0 k! xk for |x| < 1. + bk = n and bi ≤ Bi for all i) is given by the generalized hypergeometric distribution Bk B1 ··· b1 bk . Theorem 17.15 (Reflection principle) Let Y1 , Y2 , . 2 f (Xn) − f (Xn−1) = 10.2 Optional Sampling and Optional Stopping 233 Summing up, we get the discrete Itô formula: f (Xn) = f (x0) + n f (Xi−1)(Xi − Xi−1) + i=1 = f (x0) + (F ·X)n + n 1 i=1 n i=1 2 f (Xi−1) (10.6) 1 F . Now let μ be infinitely divisible with CFP ϕ. Theorem 21.28 (Brownian motion, L2 -approximation) There is a continuous ˜ X is a Brownian motion and we have version X of X. This implies E[Z1 E] = E[Z1]. Let θ := − Cov[X,Y Var[Y] . Then n! 12.2 Backwards Martingales 265 E[ϕ(X) E] = E[ϕ(X) T] = lim An (ϕ) n→∞ a.s. and in L1 . This connection • in some cases allows us to distinguish between recurrence and transience by means of easily computable quantities, and • in other cases provides a comparison criterion that says that if a random walk on a graph is recurrent, then a random walk on any connected subgraph is recurrent. For two square summable sequences (an)n∈N and (bn)n∈N0 , let ha,b := b0 + ∞ n=1 (an Sn + bn Cn). ♣ 598 23 Large Deviations Exercise 23.2.2 Let E = R. (−∞,x] (ii) In (i), replace μ by PX and ν by PY . If λ-almost everywhere f > 0, then μ(A) = A f dλ > 0 if λ(A) > 0; hence μ ≈ λ. .} in such a way that Y is a random walk and (X, Y) is a successful coupling. Definition 2.14 (Independent random variables) The family (Xi)i∈I of random variables is called independent if the family (σ (Xi))i∈I of σ -algebras is independent. Since I is discrete, g is measurable. In particular, G(x, y) = Px [XτA = y] for x ∈ E \ A and y ∈ A. “(ii) ⇒ (i)” This follows from Lemma 1.31(iv). We start by recalling the Arzelà–Ascoli characterization of relatively compact sets in C([0, ∞)) (see, e.g., [37, Theorem 2.4.7] or [174, Theorem III.3]). = pe e∈E If q = p, then there is an e ∈ E with pe > 0 and qe = pe . The supplement is trivial. Let X˜ be a Markov chain with transition matrix p(x, ˜ y) = 12 p(x, y) + 12 1{x} (y). Applying (7.12) with g ≡ 1 ∈ L∞ (μ) yields κ(f)∞ ≥ f 1 . Now let U := {Ux : x ∈ E} ⊂ U.} If E is a locally compact metric space and if U ⊂ E is open and K ⊂ U is compact, then there exists a compact set L with K ⊂ L◦ ⊂ L ⊂ U . By Jensen’s inequality (for x → |x|p), we have 1 |Xk |p . Somewhat * more formally, we could write: If x ∈ Ei for some i, then Px Xn ∈ Ei+n (mod d) = 1. 147 147 153 160 7 Lp -Spaces and the Radon–Nikodym Theorem . (ii) Show that dH is complete if (E, d) is complete. ♦ Example 1.71 Let λ be the Lebesgue measure (more accurately, the Lebesgue– Borel measure) on Rn , B(Rn) . Can you come up with a counterexample for the case μ(Ω) = ∞? Any of the questions raised above can be answered using this comparison technique. (ii) If X, Y ∈ L2 (P), then XY ∈ L1 (P) since |XY | ≤ X2 + Y 2 . Since fnk −→ f in measure, by Corollary 6.13, l→∞ there exists a further subsequence (fnkl)l∈N of (fnk)k∈N with fnkl −→ f almost 152 6 Convergence Theorems everywhere. ♣ Exercise 15.2.2 Show that there are real random variables X, X and Y, Y with the D D D properties (i) X = X and Y = Y , (ii) X and Y are independent, (iii) X + Y = X + Y , and (iv) X and Y are not independent. (17.13) y=x Then we define q(x, x) = − q(x, y). This is the kernel on [0, ∞) whose Laplace transform is given by ∞ κt (x, dy) e−λy = ψt (λ)x . It is enough to consider the case f ≥ 0. If κ2 is a kernel from (Ω1 , A1) to (Ω2 , A2), then we define the product κ1 ⊗ κ2 similarly by formally understanding κ2 as a kernel from (Ω0 × Ω1 , A0 ⊗ A1) to (Ω2 , A2) that does not depend on the Ω0 -coordinate. Then ST = n=1 Sn 1{T =n} . In particular, ϕ = (ϕ 1/n)n is infinitely divisible. By φ(t) = x∈ZD ei)t,x* p(0, x) denote the characteristic function of a single transition. In particular, we write XJ := XJI . In particular, the paths are almost surely nowhere differentiable. Theorem 19.22 (Thomson’s (or Dirichlet’s) principle of minimization of energy dissipation) Let I and J be unit flows from A1 to A0 (that is, I (A1) = J (A1) = 1). Definition 7.35 Let μ and ν be two measures on (Ω, A). Use Exercise 13.1.5 to show that lim r↓0 μ(x + rC) = 0 rd for λd -almost all x ∈ A. , k}, we have *) P Xjr ∈ Air for all r = 1, . In this case, we have ' ((' p n (x, y)p(y, x) = Px Xn = y; τx1 > n; Xn+1 = x = Px τx1 = n + 1 . In fact, in many cases, it is enough to consider moments, Laplace transforms or characteristic functions. , Dk,dk ∈ A i=1 Ck,i = Bk ⊂ A k such that Ak \ Bk = di=1 Dk,i . (ii) A map F : E N → E is called n-symmetric if F = F for all ∈ S(n). 468 19 Markov Chains and Electrical Networks Definition 19.14 A flow I : E × E → R on E \ A is called a current flow if there exists a function u : E → R with respect to which Ohm’s rule is fulfilled: I (x, y) = u(x) − u(y) R(x, y) for all x, y ∈ E, x = y. , k + m} = m+ n−m 1Am ◦ τ k . , k + 1) ⊂ σ (Xm,l , l ∈ N). Let E = N and let X = (Xt)t ≥0 be a Markov process on E with Q-matrix ⎧ ⎪ ⎨ x, q(x, y) = −x, ⎪ ⎩ 0, if y = x + 1, if y = x, else. For any finite I = {i1 , . If we had μ = 0, then there would exist (since μ(E) = 0) points x1 , x2 ∈ E with μ({x1 }) > 0 and μ({x2}) < 0. Clearly, for every y ∈ E, (x1 , y) + this would imply μ({x1}) p μ({x2}) p (x2 , y) < μ({x1 }) p (x1 , y) + μ({x2 }) p (x2 , y) ; hence ; ; ; μ p; TV = μ({x}) p(x, y) y∈E x∈E 0 Ex [τx1] for all x ∈ E. (ii) For anyA, B ∈ DE with A ⊂ B, we have (B\A)∩E = (B∩E)\(A∩E) ∈ δ(E). Let Ei = f −1 ([yi−1 , yi)) for i = 1, . I should like to thank all those who read the manuscript and the German original version of this book and gave numerous hints for improvements: Roland Alkemper, René Billing, Dirk Brüggemann, Anne Eisenbürger, Patrick Jahn, Arnulf Jentzen, Ortwin Lorenz, L. In both cases, let ν ∗0 := δ0 . s>0 Then P[A] ∈ {0, 1}. Definition 17.12 Let I ⊂ [0, ∞) be closed under addition. Now use the fact that πpn = π and μx pn = μx for all n ∈ N to conclude that even π({y}) = π({x})μx ({y}) holds. ∞ n=1 Gn ∞ β(An) ≤ ε + n=1 ≤ ∞ n=1 μ ∞ μ∗ (Gn). Then B1 ⊂ C1 ⊂ A1 and C1 ∈ C. Thus we have gn q ≤ F p . Compare Theorem 3.2(iii). B B) n n* C C Cov Xs , Xt = E ξk 1[0,s] , bk ξl 1[0,t] , bl k=1 = n l=1 CB C B E[ξk ξl] 1[0,s] , bk 1[0,t] , bl k,l=1 = n B CB C 1[0,s] , bk 1[0,t] , bk k=1 C B −→ 1[0,s] , 1[0,t] = min(s, t). Then Xn (ω) = X0 (τ n (ω)). A A Proof Let X : A → Ω, ω → ω be the canonical inclusion; hence X−1 (B) = A∩B for all B ⊂ Ω. Firstly, one can solve the system of linear equations belonging to the Markov chain killed in the two points. Assume that f is convex on [0, ∞). Sn = E[X1] > 0 almost surely. Clearly, we have pn0 (x0 , y) = (pA)n0 (x0 , y) for all y ∈ A. ♣ Exercise 11.2.6 Let the notation be as in Exercise 11.2.5. However, instead of (11.2) we make the weaker assumption ' (E sup Xn+1 − Xn < ∞. , Sn }, n ∈ N. Example 17.18 (Random walk on Z) Let E = Z, and assume p(x, y) = p(0, y − x) for all x, y ∈ Z. (2.7) j ∈J Example 2.12 As in Example 2.4, let (Ω, A, P) be the product space of infinitely many repetitions of a random experiment whose possible outcomes e are the elements of the finite set E and have probabilities p = (pe)e∈E . k=1 We show that μ((a, b]) ≤ ∞ μ (a(k), b(k)]. Therefore, lim sup μn (E \ K) ≤ lim sup μn (E) − ρL,K dμn n→∞ n→∞ = μ(E) − ρL,K dμ ≤ μ(E \ L) < ε. 4 5 For N ∈ N, the random variables XN := 2−N 2N |X| ∧ N and YN := −N 4 N 5 2 2 |Y | ∧ N take only finitely many values and are independent as well. ♦ Definition 7.40 (Signed measure) A set function ϕ : A → R is called a signed measure on (Ω, A) if it is σ -additive; that is, if for any sequence of pairwise disjoint sets A1 , A2 , . Clearly, (Ω, A, P, τr) is a measure-preserving dynamical system. Show that convergence in probability implies almost everywhere convergence. For every α > 0, 1 ν({|f | > α}) ≤ α 1 1 1 = F (1{|f |>α}) ≤ F 1 · 1{|f |>α} 1 = F 1 · μ({|f | > α}). Then the following statements hold. Let C ⊂ Cb (E; K) be a family that separates points; that is, stable under multiplication and that contains 1. By approximating f by simple functions, we see that the right-hand side in (8.18) is F -measurable (see Lemma 14.23 for a formal argument). .) ⊂ En for n ∈ N; hence T ⊂ E. Em B∈Pn+ n \E n−1 B⊂Em m n \ E n−1) ≥ 0. Let N > 0. Exercise 15.4.1 Let X and Y be nonnegative random variables with 1 lim sup E[|X|n]1/n < ∞, n→∞ n 1 lim sup E[|Y |n]1/n < ∞, n→∞ n 354 15 Characteristic Functions and the Central Limit Theorem and E[Xm Y n] = E[Xm] E[Y n] for all m, n ∈ N0 . If the offspring number has a second moment, then Zn grows of order mn and Zn /mn is uniformly integrable. The essential assumptions are that the summands are independent, each summand contributes only a little to the sum and the sum is centered and has variance 1. (i) E[X |F] is F -measurable. The elements of M≤1 (E) are called sub-probability measures on E. ♠♠♠ Remark 8.15 Intuitively, E[X |F] is the best prediction we can make for the value of X if we only have the information of the σ -algebra F . Denote by Ω = × Ωi the set of maps ω : I → Ωi such that ω(i) ∈ Ωi i∈I i∈I for all i ∈ I . Thus we define a new probability measure PB on (B, 2B) by PB [C] = #C #B for C ⊂ B. Now restrict ϕ to that flat piece and inductively reduce its dimension until reaching a point, the case that has already been treated above. However, here we only sketch the proof since we do not want to go into the details of signed measures and signed contents. For example, VT could be a (European) call option with maturity T and strike price K ≥ 0. Finally, we present a third method of studying recurrence and transience of random walks that does not rely on the Euclidean properties of the integer lattice but rather on the Fourier inversion formula. We introduce the notion of weak convergence of probability measures on general (mostly Polish) spaces and derive the fundamental properties. , Xn) such that ξn (X) = ν: n 1 An (ν) := k = (k1 , . If the X1 , X2 , . k→∞ Furthermore, for open A and for C ∈ C with C ⊂ A, α(C) = lim μnk (C) ≤ lim inf μnk (A), k→∞ k→∞ hence μ(A) ≤ lim infk→∞ μnk (A). I would be grateful for further suggestions, errors, etc. Klenke, Probability Theory, Universitext, 257 258 12 Backwards Martingales and Exchangeability Strictly speaking, the definition of exchangeability makes sense only if we identify Y = (Xi)i∈I and Y = (X(i))i∈I as random variables on the product space E I . ♣ Exercise 19.5.4 For the graph of Fig. For irreducible chains all states are in the same class: either positive recurrent, null recurrent or transient. : Ω → E be measurable with respect to A - B(E). For convenience, also the diagonal is drawn. k=1 In the case d = 1, the following theorem goes back to Bochner (1932). Note that the limit does not depend on the choice of the subsequence and is thus unique. ˜ y) = ϕ −1 (x) − ϕ −1 (y) for x, y ∈ R defines a metric on R such that In fact, d(x, ϕ and ϕ −1 are continuous. Klenke, Probability Theory, Universitext, 515 516 21 Brownian Motion (ii) indistinguishable if there exists an N ∈ A with P[N] = 0 such that {Xt = Yt } ⊂ N for all t ∈ I. If an ↓ inf f (K) is strictly monotone decreasing, then K ∩ f −1 ([−∞, an]) = ∅ is compact for every n ∈ N and hence the infinite intersection also is nonempty: f −1 (inf f (K)) = K ∩ ∞ f −1 ([−∞, an]) = ∅. Let Sn = n−1 Xk k=0 denote the nth partial sum. 16.1 Lévy–Khinchin Formula 375 Example 16.15 For an infinitely divisible distribution μ on [0, ∞), we can compute the Lévy measure ν by the vague limit ν = v-lim nμ∗1/n n→∞ (0,∞) (16.7) . In the reduced network, the effective resistances are easy to compute: If {a, b, c} = {0, 1, x}, then Reff (a ↔ b) = 1 1 + R (a, b) R (a, c) + R (b, c) −1 . Hence 1 ∩ {τ ≤ n}. Show that f is (properly) Riemann integrable if and only if f is λ-a.e. continuous. Denote by Pn the partition that is generated by the sets n−1 l=0 τ (Ail), i1 , . We say that X explodes. Since μ1 and μ2 are locally finite, for every x ∈ K, there exists an open set Ux x with μ1 (Ux) < ∞ and μ2 (Ux) < ∞. In other words, (X1 , . As uniform limits of continuous functions are continuous, (21.25) implies that X is continuous. Loosely speaking, at the boundaries of closed sets, mass can immigrate but not emigrate. This implies that E ∈ DB for any B ∈ δ(E). 17.3 Discrete Markov Processes in Continuous Time . Proof Clearly, 0 and Y = X − a are submartingales. (p) (1 − p) i i=A If we define p = (1 + b)p∗ , then p ∈ (0, 1) and 1 − p = (1 − p∗)(1 + a). 5.2 Rolling a die n times: For a fixed realisation, the values of Sn /n converge to 3.5. We have n on the horizontal axis. , Xn (ω)? . E12 are countable. Then Z := X ∧ Y = (min(Xt , Yt))t ∈I is a supermartingale. Since μ is upper semicontinuous (Theorem 1.36), there is a δ > 0 such that μ(Bδ) ≤ μ(B)+ε. We then introduce martingales and the discrete stochastic integral. k=1) * Therefore, Ex τx1 = 1 π({x}) < ∞, and thus X is positive recurrent. ♦ Proof (ii) As μ and the outer measure μ∗ coincide on σ (A) and since μ(A) is finite, by the very definition of μ∗ (see Lemma 1.47) there exists a covering B1 , B2 , . Here also the convergence has exponential speed and the rate is determined by the second largest eigenvalue of p. + − For any 3measurable 3map f : Ω → R, we have 3 f ≤ |f | and f ≤3|f |,− which ± implies 3 f dμ ≤ |f | dμ. n→∞ n→∞ By Lévy’s continuity theorem, as a continuous limit of CFPs, ϕ is a CFP. Hence convergence in measure plus relative sequential compactness in L1 yields convergence in L1 . By the triangle inequality, d(x, z) + d(z, y) ≥ d(x, y) for all x, y ∈ E and z ∈ F . Further, we agree on the following notation for spaces of continuous functions: C(E) := f : E → R is continuous , Cb (E) := f ∈ C(E) is bounded , Cc (E) := f ∈ C(E) has compact support ⊂ Cb (E). For the first equivalence, we distinguish two cases. Hence, in this case, the speed of convergence is known precisely. The following proof shows that this formal argument can t =0 be made rigorous. Evidently, gn (x) = gm (x) for every x ≤ m ∧ n; hence there exists a g ∈ Ω with g(x) = gn (x) for every x ≤ n for every n ∈ N. The number σ := Var[X] is called the standard deviation of X. If we simulate such a chain and let it run long enough this should give a sample that is distributed approximately like π. By Theorem 6.25, we thus have lim E[Yn] = 0. Without loss of generality, assume f ∞ ∈ (0, ∞). Now let (fn)n∈N be a sequence of simple functions with fn ↑ f . , Cn ∈ A such that B \ A = ni=1 Ci . In particular, if the network is irreducible, an electrical potential is uniquely determined by the values on A. Due to the monotonicity, we can make the following definition. A complete normed vector space is called a Banach space. Let X be as above and let Z be a σ (X)-measurable real random variable. Then, for any i = 1, . Evidently, the distribution function of (X1 , X2) is given by F . , L}d and denote by EL = {e = x, y∗ ∈ E : x, y ∈ BL } the set of those edges with both vertices lying in BL . Then u is also 2n-times differentiable at 0 and u(2k−1)(0) = 0 for k = 1, . Let F = σ (X) be the filtration generated by X and define F := + (Ft)t ≥0 by Ft = s>t Fs . Rather, we have to peel off the negative portions layer by layer. 22.2 Skorohod’s Embedding Theorem . Then the following statements hold: (i) A is ∩-closed. 5.5) has the Q-matrix q(x, y) = α(1{y=x+1} − 1{y=x}). Define the left continuous inverse of F : F −1 (t) := inf{x ∈ R : F (x) ≥ t} for t ∈ (0, 1). = Poiμ1 (A) ∗ Poiμ2 (A) ∗ . ♣ 15.2 Characteristic Functions: Examples Recall that Re(z) is the real part of z ∈ C. 458 18 Convergence of Markov Chains An alternative approach to the eigenvalues can be made via the roots of the characteristic polynomial χN (x) = det(p − xI), x ∈ R. The map ν → m(ν) is continuous; hence EA is open (respectively closed) if A is open (respectively closed). ♦ Remark 9.29 Let F and F be filtrations with Ft ⊂ Ft for all t, and let X be an F -(sub-, super-) martingale that is adapted to F. In the following theorem, a whole bunch of such statements will be hung on a coat hanger (French: portemanteau). Here we give a construction for X that could actually be used to implement a computer simulation of X. 15.5 The Central Limit Theorem . ♦ 8.3 Regular Conditional Distribution 205 Definition 8.28 Let Y be a random variable with values in a measurable space (E, E) and let F ⊂ A be a sub-σ -algebra. Proof (i) Let x ∈ I ◦ . 25.4 Dirichlet Problem and Brownian Motion .. ˆ Then) ˆ * 1 ϕ(t) ˆ := E et X1 = R et x eτ x μ(dx) = 1 ϕ(t + τ). In addition, we assume that there is an exterior magnetic field of strength h. A We will show f ≤ 0. Ceff (A0 ↔ A1) ≥ Ceff (A0 ↔ A1). However, if we choose c = 0, then Sk2 1A˜ k ≥ t 2 1A˜ k . 19.13). α Remark 15.27 In fact, ϕα,r is a characteristic function for every α ∈ (0, 2] (α = 2 corresponds to the normal distribution), see Sect. ◦ Fn (x) = Fn ◦ . By the Kolmogorov–Chentsov theorem (Theorem 21.6(ii)), for ε > 0 and γ ∈ (0, β/α), there exists a K such that, for every i ∈ I , we have) * P |Xti − Xsi | ≤ K |t − s|γ for all s, t ∈ [0, N] ≥ 1 − ε. The transition probabilities are given by stochastic matrices. On An , we have Sτn +1 < L − ε. Takeaways Almost everywhere (almost sure) convergence implies stochastic convergence. ♦ C(y) (19.7) Definition 19.11 Let (E, K), C and X be as in Example 19.10. , Xn) is independent and hence, by Theorem 2.13(ii), (Xn)n∈N is independent as well. For t ∈ [0, 1] and n ∈ N0 , define t Bn (t) = bn and Electrical Networks . In statistical physics, one is often interested in integrating with respect to με (where 1/ε is interpreted as "size of the system") functions that attain their maximal values away from the zeros of I . Proof Let ϕ : R → [0, 1], t → (t ∨ 0) ∧ 1. Hence, using the triangle f dμ1 − f dμ2 ≤ |f − g| dμ1 + |f − g| dμ2 ≤ ε 2f ∞ + 2ε + μ1 (Rd) + μ2 (Rd) . 5.1 Moments 119 (iii) Let μ ∈ R and σ 2 > 0, and let X be normally distributed, X ∼ Nμ,σ 2 . k=1 Again, by the ergodic theorem, 1 (20.6) Rn ≤ P[Am I] a.s. n→∞ n m→∞ Since Am ↓ A and P[Am I] −→ P[A|I] almost surely by Theorem 8.14(8.14)), the claim follows from (20.5) and (20.6). (v) Now let τ be an arbitrary stopping time. be i.i.d. real random variables that satisfy the condition of Cramér's theorem (Theorem 23.3); i.e., Λ(t) = log(E[et X1]) < ∞ 596 23 Large Deviations for every t ∈ R. ♦ (i) Let μ be a measure on R, B(R) with density f with respect to the 3 Lebesgue measure λ. Rd A map f : → R is called partially continuous at x = (x1 , . Denote)v + N , w + N *0 :=)v, w* to obtain a Hilbert space (V0 ,) · , · *0). 592 23 Large Deviations Let X̂ 1 , X̂ 2 , , 2n) are independent and identically distributed. The meaning of (P5) is explained by the following calculation. To show continuity at t = 0, consider lim sup Xt = lim sup t → ∞ t ↓ 0 = lim sup n→∞ 1 Bt t 1 1 Bn + lim sup sup Bt − Bn , t ∈ [n, n + 1] . The bounded harmonic functions are constant, but what are the unbounded harmonic functions? A Markov process (Xt)t ∈I with distributions (Px , x ∈ E) has the strong Markov property if, for every a.s. finite stopping time τ , every bounded B(E)⊗I − B(R) measurable function f : E I → R and every x ∈ E, we have Ex) * f ((Xτ +t)t ∈I) Fτ = EXτ [f (X)] := EI κ(Xτ , dy) f (y). The procedure imitates the proof that Ω is compact. If σ 2 ≥ 0 and b ∈ R, then (σ 2 , b, ν) is called a canonical triple. If all level sets I −1 ([−∞, a]), a ∈ [0, ∞), are compact, then I is called a good rate function. (iii) If A1 , A2 , . On the other hand, again by Theorem 12.17, *) n→∞ An (ϕk−1) −→ E ϕk−1 (X1 , . Now we show that r must equal 0, which contradicts the assumption Pp [N ≥ 3] > 0. By computing the cases Xn = Xn−1 − 1 and Xn = Xn−1 + 1 separately, we see that for all n ∈ N f (Xn−1 + 1) − f (Xn−1 − 1) (Xn − Xn−1) 2 1 1 + f (Xn−1 + 1) + f (Xn−1 − 1) − f (Xn−1) 2 2 1 = f (Xn−1)(Xn − Xn−1) + f (Xn−1) 2 1 = Fn · (Xn − Xn−1) + Fn . For ν ∈ Mf (R), we can define ν *n = ν(R)n (ν/ν(R))*n (and ν *n = 0 if ν = 0). In order to simulate a chain X that converges to π, we take a reference chain with transition matrix q and use the Metropolis algorithm: If the chain with transition matrix q proposes a transition from the present state x to state y, then we accept this proposal with probability π(y) q(y, x) ∧ 1. In Example 10.19 (Equation (10.7)) for the case r = 12 , and Example 10.16 for the case r = 12 , it was shown that, for every μ ∈ M1 (E), n→∞ μpn −→ (1 − m(μ))δ0 + m(μ)δN . Then A ∈ T = ∞ σ (Xn , Xn+1 , . Indeed, by the monotonicity principle, we have (E,K) Reff (Z2 ,L2) (0 ↔ ∞) ≥ Reff (0 ↔ ∞) = ∞. Takeaways An event that is described by a sequence X1 , X2 , . We will construct a binomially distributed random variable by throwing a Poiλ -distributed number T of balls in ni boxes and count the number of nonempty boxes. Choose an open set U ⊃ (A ∩ [−N, N]n)c such that λn (U \ (A ∩ [−N, N]n)c) < ε/2, and let K := [−N, N]n \ U ⊂ A. Try and fill the details in this argument. , n} ≥ # k ≤ n : Sl = Sk for all l > k = n 1A ◦ τ k . The supplementary statement is simple and is left as an exercise. Indeed, for the case where f (x) = ∞ for all x ∈ K, the statement is trivial. n→∞ As B is right continuous, we have F (Bτ +t)t ≥0 −→ F (Bτ +t)t ≥0 almost surely and in L1 (P) and thus ') n→∞) * (E Ex F (Bτ n +t)t ≥0 Fτ n − Ex F (Bτ +t)t ≥0 Fτ n ' (n→∞ ≤ Ex F (Bτ n +t)t ≥0 − F (Bτ +t)t ≥0 −→ 0. , θn ∈ (0, ∞). For any two points a, b ∈ I with a < b, we have #(Aε ∩ (a, b)) ≤ ε−1 (D + ϕ(b) − D + ϕ(a)); hence Aε ∩ (a, b) is a finite set. Then *) E X0 1{Mn >0} ≥ 0 for every n ∈ N. Hence the Borel–Cantelli lemma yields P[A] = 1. ♠ Remark 6.6 In general, convergence in measure does not imply almost everywhere convergence. (ii) Let m ∈ N(x, y). More precisely, let the pairwise distinct points e1 , . Then Pnk −→ Q, k → ∞. Then B1 ⊂ D1 = U ∈UD U . f cx+d cd For two linear rational functions f and g, we have Mf ◦g = Mf · Mg . Furthermore, C˜ := {g˜ : g ∈ C} ⊂ Cb (E; C) is an algebra that separates points and is closed under complex conjugation. , Nk be the corresponding absolute frequencies. It is easy to check that ε → |λε,N/2 | is monotone decreasing and that ε → |λε,1 | is monotone increasing. (ii) For almost all ω ∈ Ω, the map I → R, x → f (ω, x) is differentiable with derivative f . Let Bn := A\ ni=1 Ai . 11.2 Snapshot of a voter model on an 800 × 800 torus. Inductively, we get the statement for F (n) since n d ≤ (n/ε)n e−n < ∞ for x ≥ 0 and λ ≥ ε. Now let H be progressively measurable and bounded. It is easy to check that Y is indeed a random walk with transition matrix p. The product σ -algebra is the smallest σ -algebra such that all coordinate maps are measurable. ♦ Theorem 8.4 If P[B] > 0, then P[· |B] is a probability measure on (Ω, A). Theorem 8.12 E[X |F] exists and is unique (up to equality almost surely). In Polish spaces, a partial converse is true. We will come back to this connection in the framework of the martingale convergence theorem that will provide an alternative proof of the Radon–Nikodym theorem (Corollary 7.34). μ is called a • • • • content if μ is additive, premeasure if μ is σ -additive, measure if μ is a premeasure and A is a σ -algebra, and probability measure if μ is a measure and μ(Ω) = 1. That is, μn − μT V −→ 0 implies μn −→ μ weakly. For any binary prefix code C = (c(e), e ∈ E), we have Lp (C) ≥ H2 (p). , r}2 /(rdx2) is known; see, e.g., [45]. ♣ Exercise 1.1.2 Give a counterexample that shows that, in general, the union A ∪ A of two σ -algebras need not be a σ -algebra. 18.4 Speed of Convergence 453 For a practical implementation, there are two main problems: (1) The full map Fn has to be generated and has to be composed with F . Proof (Outer regularity) Step 1. By the approximation theorem for measures (Theorem 1.65), there exists an N ∈ N and mutually disjoint sets F1 , . ∞) * P (pi N) = P {p1 · · · pk n} i=1 n=1 = ζ (s) −1 (p1 · · · pk) −s ∞ n−s n=1 = (p1 · · · pk)−s = k P[pi N] . For n ∈ N, there exists a μn ∈ M1 (R) such that μ*n n = μ. ♣ Chapter 18 Convergence of Markov Chains We consider a Markov chain X with invariant distribution π and investigate conditions under which the distribution of Xn converges to π for n → ∞. Due to translation invariance, we have (#BL)−1 Ep [#TL] = r for any L. ♦ Definition 7.6 Let G be a convex set. In this case, we have VT = (XT − K)+ . ♣ 2.4 Example: Percolation Consider the d-dimensional integer lattice Zd , where any point is connected to any of its 2d nearest neighbors by an edge. A map f : E → C is measurable if and only if Re(f) and Im(f) are measurable (see Theorem 1.90 with C ∼ = R2). For b > 0, define Hb (p) := − pe logb (pe) e∈E with the convention 0 logb (0) := 0. Apart from the inequalities, the important results for probability theory are Lebesgue's decomposition theorem and the Radon–Nikodym theorem in Sect. + p(i, j) for i, j ∈ E, and define Yn by Rn (i) = j ⇐⇒ Un ∈ [r(i, j − 1), r(i, j)). In particular, every characteristic function is uniformly continuous. The analogue of Theorem 13.34 holds for C ⊂ Cb (E; C). 273 274 281 290 300 224 Contents xiii 14 Probability Measures on Product Spaces . P0 XτN = −N = Rw,eff (−N ↔ N) Rw,eff (0 ↔ −N) + Rw,eff (0 ↔ N) Again, since X is transient, we infer) *) * n→∞ P0 Xn −→ −∞ = P sup{Xn : n ∈ N0 } < ∞) * = lim P sup{Xn : n ∈ N0 } < N N→∞) * ≤ lim sup P XτN = −N N→∞ = Rw + −. Then μ∗ ((A ∩ B) ∩ E) + μ∗ ((A ∩ B)c ∩ E) = μ∗ (A ∩ B ∩ E) ∪ (Ac ∩ B c ∩ E) ∪ (A ∩ B c ∩ E) ≤ μ∗ (A ∩ B ∩ E) + μ∗ (Ac ∩ B c ∩ E) + μ∗ (A ∩ B c ∩ E) = μ∗ (B ∩ E) + μ∗ (B c ∩ E) = μ∗ (E). n→∞ (ii) ϕ is differentiable at 0 with ϕ (0) = i m if and only if (X1 + . (17.32) and Proof (The proof follows the exposition in [100, Section 3]) Since bni ,pi ({0}) = (1−pi)ni , conditions (17.31) and (17.32) are clearly necessary for bn1 ,p1 ≤st bn2 ,p2 . 21.3. Indeed, there we needed only right continuity of the paths and a certain continuity of the distribution as a function of the starting point, which is exactly the Feller property. Proof "(ii)" This follows from (i) since Theorem 17.11 yields uniqueness of X. Exercise 12.1.1 Let n ∈ N. Then the relative entropy of ν ∈ M1 (Σ) is H (ν |μ) = 1−m 1+m log(1 + m) + log(1 − m). How does this work in detail? ♣ Exercise 23.2.7 Let (Xt)t ≥0 be a random walk on Z in continuous time that makes a jump to the right with rate 12 and a jump to the left also with rate 12 . Takeaways For (sub-) martingales, Doob's upcrossing lemma bounds the number of upcrossings over a given interval. How stable is weak convergence if we pass to image measures under some map ϕ? Hence, let F be continuous and linear. e−λ (xi) (Poisson distribution) Clearly, ϕPoiλ (t) = n! n=0 Corollary 15.14 The following convolution formulas hold. 17.7 Stochastic Ordering and Coupling 429 Compute the invariant measure and show the following using Theorem 17.52: (i) If r ∈ 0, 12 , then X is positive recurrent. If f ∈ C, then also |f | = f ∞ lim pn f 2 /f 2∞ n→∞ is in the closure C of C in Cb (E; R). d Differentiating the power series termwise yields dt pt (x, y) = q(x, y). Define the waiting time for the first "success" in the mth row of the matrix (Xm,n)m,n by Ym := inf n ∈ N : Xm,n = 1 − 1. We compute the Laplace transform for these kernels. In particular, for α > 2, t → e−|t| is not a CFP. By (20.4), we conclude that, for every x ∈ E, Pπ [X ∈ A] = Eπ [Px [X ∈ A]] = Px [X ∈ A]. ∈ A, • ∪-closed (closed under unions) if A ∪ B ∈ A whenever A, B ∈ A, • σ -∪-closed (closed under countable unions) if ∞ n=1 An ∈ A for any choice of countably many sets A1 , A2 , . By Wald's identity (Theorem 5.5), we have E[ST] = E[T] E[X1]; hence ' ( ) * Var[ST] = E ST2 − E[ST]2 = E[T] Var[X1] + E T 2 − E[T]2 E[X1]2 , as claimed. Similarly, we get Y ≥ Y almost surely. Lemma 4.6 Let f, g, f1 , f2 , . Manifestly, all three notions of infinite divisibility are equivalent, and we will use them synonymously. Equation (2.9) says that, for any choice 64 2 Independence of real numbers (xi)i∈I , the events (Xi−1 ((−∞, xi]))i∈I are independent. That is, for any F -measurable Y with E[Y 2] < ∞,) *) * E (X − Y)2 ≥ E (X − E[X |F])2 with equality if and only if Y = E[X |F]. By Lemma 15.12(iv) and (ii), ϕSn∗ (t) = ϕ Now 1 − t2 2n n n→∞ −→ e−t 2 /2 t √ nσ 2 n . Here one needs assumptions on the regularity of the paths t → Xt (ω); for example, right continuity. Then the measure μ := pn μn ∈ Mf (Rd) has n=1 characteristic function ϕμ = ∞ (15.3) pn ϕμn . In this case, (Xi)i∈I is independent if and only if, for any finite J ⊂ I and any choice of xj ∈ E, j ∈ J ,) * P Xj = xj for all j ∈ J = P[Xj = xj]. + Xn of i.i.d. integrable random variables is n · E[X1]. , x d) ∈ Rd and y = 1 d d (y , . Proof Let ε > 0. , WK ∈ R. Example 18.6 (Independent coalescence) The most important coupling is Markov chains that run independently until they coalesce: Let X and Y be independent chains with transition matrix p until they first meet. If we have (A) for every state, then the Markov chain is called aperiodic. Definition 9.7 An E-valued stochastic process X = (Xt)t ∈I is called (i) real-valued if E = R, (ii) a process with independent increments if X is real-valued and for all n ∈ N and all t0 , . It is easy to check that A is a semiring. (ii) Let X be a Markov chain with invariant distribution π. (12.9) j ∈J Then the family (Ai)i∈I is called independent given A. Exercise 21.5.1 Use the representation of Brownian motion (Wt)t ∈[0,1] as a random linear combination of the Schauder functions (Bn,k) to show that the Brownian bridge Y = (Yt)t ∈[0,1] = (Wt − tW1)t ∈[0,1] is a continuous, Gaussian process with covariance function Cov[Yt , Ys] = (s ∧ t) − st. Theorem 15.44 (Central limit theorem of Lindeberg–Feller) Let (Xn,l) be an independent centered and normed array of real random variables. If x, y ∈ Zd are nearest neighbors (that is, x − y2 = 1), then we denote by e =)x, y* =)y, x* the edge that connects x and y. Strictly speaking, this gives the Poisson process only on the time interval (0, 1], but it is clear how to move on: We perform the same procedure independently for each of the intervals (1, 2], (2, 3] and so on and then collect the jumps (see also Exercise 5.5.1). Then E is a π-system that generates 2Ω . Proof We give the proof for σ -algebras only. |Fn | ≤ maxx∈{x0 −n,...,x0 +n} |F (x)|), and A := i=1 2 i n∈N0 Hence f (X) := (f (Xn))n∈N0 = M + A is the Doob decomposition of f (X). Hence, there is an ε > 0 such that P[X1 < −2ε] > ε. (1.11) 1.3 The Measure Extension Theorem Inductively, we get μ∗ (E ∩ Bn) = implies that 25 n i=1 μ ∗ (E ∩ Ai). On the other hand, (Xj J)−1 (Ej) ∈ AJ for all j ∈ J and Ej ∈ Ej . are independent Rd -valued random variables with distributions μi := PXi , i = 1, . We define the conditional probability given B for any A ∈ A by ⎧ ⎨ P[A ∩ B] , P[B] P[A|B] = ⎩ 0, if P[B] > 0, (8.1) otherwise. On the other hand, if r is rational, then there exists some n ∈ Z \ {0} with e2πin r = e−2πin r = 1. Therefore, P(x,y)[τ < ∞] = 1 for all initial points (x, y) ∈ E × E of Z. Then A → μ(A) := ω∈A pω defines a σ -finite measure on 2Ω . In order for the sequence (fn)n∈N of measurable maps Ω → E to converge almost everywhere, it is sufficient that one of the following conditions holds. , N; hence lim sup n→∞ f dμn ≤ 3 ε + N f (yi) F (yi) − F (yi−1) ≤ 4 ε + f dμ. N→∞ This completes the proof. We may consider Brownian motion as the canonical process on the space Ω := C([0, ∞)) of continuous paths. The sets A ∈ A are called events. This property is reflected by the assumption that the jump rate q(x, x + 1) equals x. Takeaways We have adapted the convergence theorems of the last chapter to Lp convergence. Proof See, e.g., [54, Chapter XV.7]. 226 9 Martingales Proof We show that there exist FT −1 -measurable random variables VT −1 and HT such that VT = VT −1 + HT (XT − XT −1). By the strong law of large numbers, we infer Shannon's theorem: 1 1 n→∞ − log πn = Yi −→ H (p) n n n almost surely. ♣ Exercise 15.4.6 Let X1 , X2 , . Thus Y satisfies the condition of Theorem 21.6 (Kolmogorov–Chentsov) with α = 4 and β = 1. , kn , n ∈ N) be an array of real random variables with CFPs ϕn,l . For (iii), note that τ − s peeks into the future by s time units (in fact, {τ − s ≤ t} ∈ Ft +s), while τ + s looks back s time units. Since E is locally compact, there exists a compact set K ⊂ E with K ◦ ⊃ L and a ρL,K ∈ Cc (E) with 1L ≤ ρL,K (x) ≤ 1K . As F satisfies the usual exists and is RCLL. If Y1 ∈ L2+δ (P) for some δ > 0, then n * * n→∞)) E |Xn,l |2+δ = n−(δ/2) E |Y1 |2+δ −→ 0. , ωn ∈ E that the probability of the event [ω1 , . 2 For the inequality compute 1 P[B1 > a] = √ 2π ∞ a 1 1 ≤ √ 2π a e−x ∞ a 2 /2 dx x e−x 2 /2 dx = 1 −a 2 /2 e dx = √ . Take the integral with respect to the other variable. , Ifn) : n ∈ N; f1 , . n (x1 , y1), (x2 , y2) > 0. f ∈F A A If μ(Ω) < ∞, then (ii) is equivalent to (iii): (iii) For all ε > 0, there is a δ(ε) > 0 such that |f | dμ ≤ ε for all A ∈ A with μ(A) < δ(ε). If in (ii) the measure is a probability measure for all ω1 ∈ Ω1 , then κ is called a stochastic kernel or a Markov kernel. P{−n,−n+1,...} X *) = P X0 ∈ A−n , X1 ∈ A−n+1 , . ∞ (ii) n=0 |An | = ∞ almost surely. In order to apply the results on finite electrical networks from the last section, we henceforth assume that A0 ⊂ E is such that E \ A0 is finite. Proof "(i) ⇒ (ii)" Let A, A1 , A2 , . To this end, we want to apply Kolmogorov's moment criterion. Proof (i) Let m := sup f (SA (x0)). Hence there exist g and h with gn ↑ g and hn ↓ h. The basic idea is that it is energetically favorable for particles to be oriented in the same direction. How can we construct a probability space on which all these random variables are defined? For 0 ≤ b < c ≤ ∞ and t ∈ R, let γb,c,t be the linear path in C from b − ibt to c − ict, let δb,t be the linear path from b to b − ibt and let #c,t be the linear path from c − ict to c. Takeaways Consider a stochastic process that can take only two values at time t + 1 given the full history up to time t. For A ⊂ E, let τ := τA := inf τx x∈A be the stopping time of the first entrance to A. X is transient if and only if r = 1/2, in which case we have μ1 = μ2 . ♣ Exercise 5.3.5 Let b ∈ {2, 3, 4, . Now define μ : σ (A) → [0, ∞], A → μ∗ (A). Although this flaw might seem inevitable in the MCMC method, it is in fact, at least theoretically, possible to use a very similar method that allows perfect sampling according to the invariant distribution π, even if we do not know anything about the speed of convergence. n→∞ m≥n Brownian Motion and White Noise The construction of Brownian motion via Haar functions has the advantage that continuity of the paths is straightforward. Thus (n(1 − |ϕn (t)|2))n∈N is bounded for t ∈ [−ε, ε]. 7 Lp -Spaces and the Radon–Nikodym Theorem 180 Since ν is finite and upper semicontinuous (Theorem 1.36), we have ∞ ν(A) = lim ν n→∞ Ak ≥ inf ν(An) ≥ ε > 0. As a direct application of Theorem 16.5, we give a complete description of the class of infinitely divisible probability measures on [0, ∞) in terms of their Laplace transforms. Definition 24.6 Let X be a random measure on E. Proof (i) Assume that A, B ∈ A. We conclude that Ac ∈ AI . In a similar way to (viii), we define μf (A) = n i=1 bi f (x) dx. 6. The eigenvalue 1 has the multiplicity 1. Let μ and ν be arbitrary (but nonzero) 3σ -finite measures.3Then there exist measurable functions g, h : Ω → (0, ∞) with g dμ = 1 and h dν = 1. Usually it is difficult to show in a specific situation that the extension to 2Ω is impossible. 20.1 Definitions 495 Definition 20.6 (i) τ is called measure-preserving if *) P τ −1 (A) = P[A] for all A ∈ A. Proof Clearly, ft is continuous on R \ {0}. Let X = (Xt)t ∈I be a real-valued, adapted stochastic process with E[|Xt |] < ∞ for all t ∈ I . -.-. In fact, these sets form a π-system that generates B(Rn) (see Theorem 1.23). ∈ L1 (μ) be nonnegative and such that lim fn dμ n→∞ exists. Clearly, M = M is the trace σ -algebra of M Hence M(E) ∈ M. Let p, p ∈ [0, 1] with p < p . Here we used Fτ ⊃ Fσ , the tower property and the monotonicity of the conditional expectation (see Theorem 8.14). , X(n−1) − X(n) |X(n) does not depend on X(n) and that it equals the (unconditional) distribution of the ordered values of X1 , . The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Define Z := lim sup n→∞ 1 Sn . Remark 21.7 The statement of Theorem 21.6 remains true if X assumes values in some Polish space (E,) since in the proof we did not make use of the assumption that the range was in R. If we let f = 1E in (7.8), then we get μ(E) = 0. Let F ⊂ A be a σ -algebra and let κX,F be a regular conditional distribution of X given F . Proof (i) Let AJ = σ × Ej : Ej ∈ Ej ∪ {Ωj } for every j ∈ J . Also assume that we follow a bounded trading strategy that cannot use future information. Theorem 5.5 (Wald's identity) Let T , X1 , X2 , . If only (i) and (ii) hold and)x, x* ≥ 0 for all x, then) · , · * is called a positive semidefinite symmetric bilinear form, or a semi-inner product. Furthermore, (An)n∈N is an independent family with the property ∞ ∞ 1 * P[An] = 6 = ∞. For n ∈ N, define An := {[ω1 , . The monotonicity of μ∗ now μ∗ (E) = μ∗ (E ∩ Bn) + μ∗ (E ∩ Bnc) ≥ μ∗ (E ∩ Bn) + μ∗ (E ∩ Ac) = n μ∗ (E ∩ Ai) + μ∗ (E ∩ Ac). Then X is the Markov chain with transition matrix ⎧ wi , if j = i − 1, pw (i, j) = wi+ , if j = i + 1, ⎩ 0, else. Then there exists an increasing sequence K1 ⊂ K2 ⊂ K3 ⊂ . 16 1 Basic Measure Theory Definition 1.34 Let A, A1 , A2 , . Define Yn (ω) := − log(pXn (ω)). Thus it comes in very handy that it is sufficient to check measurability on a generator of A by the following theorem. For some applications, however, a decomposition in trigonometric functions is preferable. lim sup an n→∞ k=1 See [22]. be real random variables with Xn ⇒ X. ♦ Example 20.9 (Rotation) Let Ω = [0, 1), let A = B(Ω) and let P = λ be the Lebesgue measure. ♦ If f, g : Ω → R with f (ω) ≤ g(ω) for any ω ∈ Ω, then we write f ≤ g. Proof " ⇒ " Let ν be totally continuous with respect to μ. (i) (ii) (iii) (iv) E[Xi] = μi for all i = 1, . < tn , the family (Nti − Nti−1 , i = 1, . Therefore, since pn+k (x, x) ≥ pn (x, y) pk (y, x), we have G(x, x) ≥ ∞ pn (x, y) pk (y, x) = pk (y, x) G(x, y) = ∞. k=1 See Fig. Show (using Theorem 4.15) that xf (x) λ(dx). Theorem 8.5 Let A, B ∈ A with P[A], P[B] > 0. , Xn) and Sn := X1 + . We give an introduction to the basic concepts and then study certain examples in more detail. 1 1 11 18 36 45 2 Independence . Taylor's theorem (Lemma 15.31) yields eit x − 1 − itx = − t 2x2 + R(tx) 2 15.5 The Central Limit Theorem 361 with |R(tx)| ≤ 16 |tx|3 . Further, let X1 , X2 , . 7 Lp -Spaces and the Radon–Nikodym Theorem 176 Theorem 7.29 (Uniqueness of the density) Let ν be σ -finite. (21.33) fdd The aim of this section is to strengthen this convergence statement to weak convergence of probability measures on C([0, ∞)). As X assumes values in I ⊂ [0, ∞), we have 0 ≤ E[X 1A] = E[E[X |F] 1A] = 0; hence X1A = 0. = 1[0,t] (x) for x ∈ (0, 1) \ {t}. + Xn−1 , Mnε := max{0, S1ε , . Then ; ; n→∞ ;(δx − δy)pn ; ≤ 2 P(x,y)[Xn = Yn] −→ 0. Let α > 0 and let L be a Poiα random variable. Klenke, Probability Theory, Universitext, 147 148 6 Convergence Theorems fz (ω) = d(f (ω), z) and gz (ω) = d(g(ω), z) are also measurable. Proof Define Sn = ni=1 Xi for n ∈ N. Then P[X1 = x1] = g0 (x1)/N and P[Xk+1 = xk+1 |X1 = x1 , . Hence) * 0 ≤ E f (X) − f (Y) g(X) − g(Y) = E[f (X)g(X)] − E[f (X)] E[g(Y)] + E[f (Y)g(Y)] − E[f (Y)] E[g(X)] = 2 Cov[f (X), g(X)]. lim i=1 In particular, Ap ∩ Aq = ∅ if p = q and thus (Berp)⊗N ⊥ (Berq)⊗N . , m}. Note that E[f (X)] = E[f (Y)] and E[g(X)] = E[g(Y)]. Let Y1 , . The corresponding eigenvalues are λn = σ cos n π N for n = 1, . Then (Xn − An)n∈I is a martingale. Definition 6.16 A family F ⊂ L1 (μ) is called uniformly integrable if sup inf 0≤g∈L1(μ) f ∈F |f | − g + dμ = 0. N∈N a,b∈Q+ 0≤at Xs (ω) t (ω) = 0. To every canonical triple, by (16.8) there corresponds an infinitely divisible random variable. Defining Ŝn = X̂ 1 + . ∩ Aik), {i1 ,...,ik }⊂{1,...,n} k=1 μ(A1 ∩ . We now extend the result to a larger class of ranges for Y . (21.28) n=1 Theorem 21.30 d is a complete metric on Ω := C [0, ∞) that induces the topology of uniform convergence on compact sets. AN ˜ fn) −→ 0 if and only if d˜N (f, fn) −→ 0 for all N ∈ N. Lemma 6.1 Let f, g : Ω → E be measurable with respect to A – B(E). In particular, we have Ex [Yt] = x, Ex [Yt2] = 2x t + x 2 , Ex [Yt3] = 6x t 2 + 6x 2 t + x 3 , Ex [Yt4] = 24x t 3 + 36x 2 t 2 + 12x 3 t + x 4 , Ex [Yt5] = 120x t 4 + 240x 2 t 3 + 120x 3 t 2 + 20x 4 t + x 5 , Ex [Yt6] = 720x t 5 + 1800x 2 t 4 + 1200x 3 t 3 + 300x 4 t 2 + 30x 5 t + x 6 . Proof For p = ∞, the equivalence of (i) and (ii) is a simple consequence of the triangle inequality. If μ is an invariant measure, then the equations for μp = μ read μ({n}) = pn−1 μ({n − 1}) μ({0}) = ∞ for n ∈ N, μ({n})(1 − pn). We infer ; ;(f − g)1{−K,K}d ; < ε ∞ and ; ;(f − g)1 d ; ˜ ∞ ≤ f˜ ∞ = f ∞ + ε. By the preceding corollary, (fF : F ∈ I) is uniformly integrable with respect to P and thus also with respect to μ. Hence we get mβ,h = h ⇒ h = β −1 − 1 T − Tc for T → ∞, (23.24) where the Curie temperature Tc = 1 is the critical temperature for spontaneous magnetization. If the latter were the case, a ramification into more than two values in one time step would be possible. n! Corollary 15.33 (Method of moments) Let X be a real random variable with α := lim sup n→∞ 1) n *1/n E |X| < ∞. Assume that for any finite J ⊂ I, any choice of Aj ∈ Aj and for all j ∈ J , P ' ( ) * Aj A = P Aj A j ∈J almost surely. Now choose a finite subcovering {U1 , . Furthermore, δ(E) ⊂ σ (E). However, τr is not mixing: Since r is irrational, there exists a sequence kn ↑ ∞ such that τrkn (0) ∈ 1 3 , 4 4 for n ∈ N. As C separates points, for any z ∈ E \ {x}, there exists an Hz ∈ C with Hz (z) = Hz (x) = 0. By Fatou's lemma (Theorem 4.21) with 0 as a minorant, we thus get |fnk | dμ < ∞. When a particle dies, it has two offspring. Takeaways Consider the conditional probability of some event B given a σ -algebra. max i=1,...,N ε→0 Proof The sum and maximum differ at most by a factor N: max i=1,...,N aεi ≤ ε log N aεi ≤ ε log(N) + max i=1,...,N i=1 ε log(aεi). Let Ω = E [0,∞) and A = B(E)⊗[0,∞) . ♣ 594 23 Large Deviations 23.2 Large Deviations Principle The basic idea of Cramér's theorem is to quantify the probabilities of rare events by an exponential rate and a rate function. ♦ Reflection Maybe it comes as a surprise that in the definition of stochastic convergence, the set A of finite measure pops up. The resulting measure is called Lebesgue measure (or sometimes Lebesgue–Borel measure) λ on n R , B(Rn) . 12.1 Exchangeable Families of Random Variables . The relation (23.24) is called the Curie–Weiss law. The following frequencies for letters in German texts are taken from [11, p. If g = 1A1 ×A2 for some A1 ∈ A1 and A2 ∈ A2 , then clearly Ig (ω1) = 1A1 (ω1)κ(ω1 , A2) is measurable. This is a materialdependent constant (chromium bromide (CrBr) 37 Kelvin, nickel 645 K, iron 1017 K, cobalt 1404 K). If in (v) the sum ω∈Ω pω equals one, then μ is a probability measure. Proof The p.g.f. of the Poisson distribution is ψ(z) = eλ(z−1) (see (3.4)). Theorem 23.13 (Sanov [150]) Let X1 , X2 , . Then Yn −→ X. ♣ Exercise 1.5.4 Show that F : R2 → [0, 1] is the distribution function of a (uniquely determined) probability measure μ on (R2 , B(R2)) if and only if (i) F is monotone increasing and right continuous (ii) F ((−x1 , y2)) + F ((y1 , −x2)) → 0 and F (x) → 1 for all y1 , y2 ∈ R and for x = (x1 , x2) → ∞, (iii) F ((y1 , y2)) − F ((y1 , x2)) − F ((x1 , y2)) + F ((x1 , x2)) ≥ 0 for all x1 ≤ y1 and x2 ≤ y2 . n→∞ n Remark5.13 The strong law of large numbers implies the weak law. , Xtnk) ⇒ (Xt1 , . Examples for measure preserving dynamical systems are stationary stochastic processes, i.i.d. random variables, rotations and so on. For a proof of (ii) ⇒ (i) see, e.g., [155, Theorem III.4.3]. ♠♠ Corollary 21.12 (Scaling property of Brownian motion) If B is a Brownian motion and if K = 0, then (K −1 BK 2 t)t ≥0 is also a Brownian motion. Hence C0 is dense in Cb (E; R). Now if X is not a martingale, then in some cases, we can replace X by a different process X that is a martingale and such that the distributions PX and PX are equivalent; that is, have the same null sets. On the other hand, n→∞ in this case, there is also some kind of recurrence property, namely Sn /n −→ 0 almost surely by the ergodic theorem. Hence ω = (ω1 , ω2 , . Hence Aω˜ 1 = i=1 (A) ∈ A2 and fω˜ 1 = f ◦ i1 is measurable with respect to A2 . Now let X, Y, Z be independent real random variables with characteristic functions ϕX = ϕY = ϕ and D ϕZ = ψ. We interpret [ω1 , . 1 2 Expanding the cosine function in a Taylor series, we get cos(ti) = 1 − 2 ti + O(ti4) ; 1 hence 1 − φ(t) = 2D t22 + O(t42). PX = P−X if and only if ϕ is real-valued. Since (14.6) and (14.7) hold for f − and f +, by Remark 14.17 and 14.18 this is true for f also. t ∈R √ Under the assumption that F = F˜ holds, n/2 Dn converges in distribution to M. Similarly, (14.7) holds for f = 1A . (β > 1) where so-called spontaneous magnetization occurs (that is, magnetization without an exterior field). Hence X−1 (A) ∈ σ (X−1 (E)) for any A ∈ σ (E) and thus X−1 (σ (E)) ⊂ σ (X−1 (E)). It is the price multiplied by the number of stocks in the portfolio. This follows from Corollary 8.22 and the fact that X−n = E[X0 F−n] for any n ∈ N0 . Note that Z∞ is the only choice since σ 2 > 0. Then 1{x←→p y} is a random variable. However, it is not too hard to give an estimate that shows that there a c = cD such that p2n (0, 0) ≤ c n−D/2 , which implies G(0, 0) ≤ ∞exists −D/2 c n n=1 n < ∞ (see, e.g., [53, page 361] or [59, Example 6.31]). 2.3). On the other hand, if μ is a Lebesgue–Stieltjes measure, this is μ = μF for some F , then #{n ∈ N : xn ∈ (−K, K]} = F (K) − F (−K) < ∞ for all K > 0; hence (xn)n∈N does not have a limit point. Let X1 , X2 , . Now let A ⊂ E. Let E be the exchangeable σ -algebra and let T be the tail σ -algebra. ♦ 13.3 Prohorov's Theorem 291 Recall that a family F of measures is called weakly relatively sequentially compact if every sequence in F has a weak limit point (in the closure of F). Similarly, for x ∈ E n , denote x = (x(1), x(2), . In the case where Xn+1 − Xn takes three values, the system has three equations and is thus overdetermined. Define A2L,0 in a similar way to A2L ; however, we now consider all edges e ∈ EL p p as closed, irrespective of whether Xe = 1 or Xe = 0. Then the family (Pξn (X))n∈N of distributions of empirical measures satisfies an LDP with rate n and rate function Iμ := H (·|μ). 19.3 Electrical network on Z2 . , Xn = kn] μ({x})nν({x}) = #An (ν) x∈Σ = #An (ν) exp n ν(dx) log μ({x}) = #An (ν) exp − n[H (ν) + H (ν |μ)] . 1 Furthermore, by Theorem 12.10 (with k = n and ϕ(X1 , . Let τ n = inf{t ≥ 0 : Xt = n} = Sn for n ∈ N. 46 1 Basic Measure Theory Theorem 1.104 For any distribution function F , there exists a real random variable X with FX = F . If c0 , . Then X and Y are indistinguishable. Some results may be free, while others may require a fee to unlock the information.Reverse Address LookupA reverse address lookup is another type of search you can do if you only have part of the information about the number you need to find. n→∞ n→∞ Thus q ∈ F . Let TN → ∞ with E[XTN] ≥ E[X0] for all N ∈ N. n→∞ (13.7) If F, F1 , F2 , . For s ≥ 0, define the Mellin transform of PX by mX (s) = E[Xs] (with values in [0, ∞]). , n}, define −k it x Yk (t, h, x) = k! h e e ihx − k−1 (ihx)l l=0 l! . , FN ∈ Fn and thus F := F1 ∪ . F :(x, y) > 0 (otherwise y would not be visited). Hence, ν({x}) = ⎧ ⎪ ⎪ ⎨ ⎪ ⎪ ⎩ 1 2, 8 , n 2x2 0, if x = 0, if x 2 ∈ Z is odd, else. Recall that, by Ohm's rule, the resistance of a wire is the quotient of the potential difference and the current flow. , N − 1, be the Nth roots of unity and let the corresponding (right) eigenvectors be x k := θk0 , θk1 , . ♣ 2.3 Kolmogorov's 0–1 Law With the Borel–Cantelli lemma, we have seen a first 0–1 law for independent events. 20.6 Entropy 511 Then the entropy of Pn is (using stationarity of π in the third line) H (Pn) = − p(0, x) log(p(0, x)) x0 ,...,xn−1 ∈E =− p(0, x) log(π({x0 })) + x0 ,...,xn−1 ∈E = H (π) − n−2 n−2 . By the monotone convergence theorem (applied to the measure (1−g)(μ+ν); that is, the measure with density (1−g) with respect to μ + ν), we obtain that (7.8) holds for all measurable f ≥ 0. Example 8.8 In the production of certain electronic devices, a fraction of 2% of the production is defective. (ii). , Mn,m = km = n! km pk1 · · · pm . By Taylor's formula, for every t ∈ (−ε, ε), n−1 |t|2n−1 t 2k (2k) u (0) ≤ u(t) − (2k)! (2n − 1)! k=0 sup u(2n−1) (θ t) . As 15.4 Characteristic Functions and Moments 351 E[|X|k] < ∞ by assumption, the dominated convergence theorem implies h→0 E[Yk (t, h, X)] −→ E[(iX)k eit X] = ϕ (k) (t). Together with (23.19), this implies (23.17). = exp − λ + 1/t However, the function ψtx is the Laplace transform of the compound Poisson distribution CPoi(x/t) exp1/t (see Definition 16.3). (ii) Show that P[{Θ ∈ · }|Φ = φ] = U[0,π) for almost all φ. 2 2 2 2 As W is linear, we have (wm + wn)/2 ∈ W ; hence 12 (wm + wn) − x ≥ c. 8.3 Regular Conditional Distribution 209 A separable topological space whose topology is induced by a complete metric is called a Polish space. , xk) = (x1 , x1 + x2 , . A similar argument for the right-hand side yields continuity of ϕ at x. (i) The family (fZ : Z ∈ Z) is uniformly integrable in L1 (μ) and ν(Ω) for any Z ∈ Z. ♣ Exercise 20.3.2 Let p = 2, 3, 5, 6, 7, 10, . Exercise 8.2.1 Show the assertions of Remark 8.16. F (x) = inf F is monotone increasing, F is right continuous and monotone increasing. Theorem 5.23 (Glivenko–Cantelli) Let X1 , X2 , . (i) μ = w-lim μn . n→∞ Then gn −→ 0 in measure, and (gn)n∈N is uniformly integrable since gn ≤ sup n→∞ p 2p (|fn |p + |f |p). For Z ∈ Z, define a function fZ : Ω → R by fZ (ω) = C∈Z: μ(C)>0 ν(C) 1C (ω). Lemma 11.18 W is a martingale. In fact, if for 0 ∈ J ⊂ I finite, we define PJ as the projection of PJ ∪{0} to E J , then the family (PJ : J ⊂ I finite) is projective. This is precisely what memory is and is thus in contrast with the Markov property of X. That is, p(e1 ,e2) = pe11 and e2 ∈E 2 pf 1 ,f 2) = pf2 2 for all e1 ∈ E 1 , f 2 ∈ E 2. 20.5 Mixing .. Hence ∞ X = X + n n n=1 n=1 N=1 Yn n=1 converges a.s. " ⇒ " Assume that ∞ n=1 Xn converges a.s. Clearly, this implies (i) (otherwise, by the Borel–Cantelli lemma, |Xn | > K infinitely often, contradicting the assumption). (ii) For transient X, there can be more than one invariant measure. Choose three pairwise distinct points x 1 , x 2 , x 3 ∈ BL \ BL−1 with Pp [Fx 1 ,x 2 ,x 3] > 0. Now assume that Q is absolutely continuous with respect to P . Let x, y ∈ E, and let (X, Y) be a successful coupling. with Tn ∼ expn . For every n ∈ N, let an be the leading digit of the p-adic expansion of q n . Exercise 1.2.1 Let A = {(a, b] ∩ Q : a, b ∈ R, a ≤ b}. Clearly, E[T∞] = n=0 1/wn < ∞; r < ∞] = 1. ♣ Exercise 15.2.3 Let X be a real random variable with characteristic function ϕ. n→∞ x∈E N Now let f1 , . As μ is regular (Theorem 13.6), there is a compact set L ⊂ E with μ(E \ L) < ε. μ-almost 280 13 Convergence of Measures (ii) For any ε > 0, there is a compact set Kε with μ(Ω \ Kε) < ε such that the restricted function f is continuous. (14.4) J ⊂I countable Hint: Show that the right-hand side is a σ -algebra. 2 = δ/R2 = 27 and R R 5 125 Step 8. Then the following statements are equivalent. For irreducible chains, all states have the same period. Proof First assume that (Nt , t ≥ 0) is a Poisson process with intensity α ≥ 0. Hence, it is enough to show (14.8) for sets of this type. (21.42) Proof The exact formulas for the first six moments are obtained by tenaciously computing the right-hand side of (21.40). Later, with a lot of additional effort, we will achieve a sharp bound in Theorem 22.11. We come back to Polish spaces in the context of convergence of measures in Chap. , σi2n (see Example 1.105(ix)). Hence also β(A1 ∪ A2) = sup α(C) : C ∈ C with C ⊂ A1 ∪ A2 ≤ β(A1) + β(A2). Thus also X = infn∈N Xn is A – B(R)-measurable and is hence a random variable. In particular, if f = g3 almost then f dμ = g dμ. ♣ 18.4 Speed of Convergence 459 √ Exercise 18.4.3 Let ν(dx) = π2 1 − x 2 1[−1,1] (x) dx. In fact, in this case, the reflection principle can be derived also in an elementary way via a bijection that changes the signs of those Yi with i > τ . √ In the following, let i = −1 be the imaginary unit. ♦ #{1, 2, 3} 3 Motivated by this example, we make the following definition. .)) = ωn be the projection of Ω to the nth coordinate. , Xk) = E (F dνn,k (X) . Then there exists a random variable Y : Ω → [0, 1] such that, for all finite J ⊂ N, *) P Xj = 1 for all j ∈ J Y = Y #J . (We have done this already for d = 1, see Exercise 13.3.4 for d ≥ 2.) In a second step, the statement is lifted to sequence spaces RN . We infer that X is recurrent if and only if 3 −2 t 2 τγ1 . (ii) The series defining ha,b converges in L2 ([0, 1], λ). (ii) Infer the optional sampling theorem for right continuous supermartingales by using the analogous statement for discrete time (Theorem 10.11); that is, Xσ ≥ E[Xτ |Fσ]. Define have E U the F∞ -measurable events 1 0 a,b C = lim inf Xn < a ∩ lim sup Xn > b ⊂ U a,b = ∞ n→∞ n→∞ and C = C a,b . We have to show that P[◦ (XL 1 L for some l = 1, . E fn (tX) X2n ≤ |t| θ∈(0,1) θ |t| θ∈(0,1) Now Fatou's lemma implies)) * * * E X2n = E fn (0)X2n ≤ lim inf E fn (tX)X2n t →0 ≤ lim inf gn (t) = 2n u2n) (0) < ∞. On the n space (Ω, A, P) := ×i=1 Ωi , ni=1 Ai , ni=1 Pi , the coordinate maps Xi : Ω → Ωi are independent with distribution PXi = Pi . (iii) (Positive definiteness))x, x* > 0 for all x ∈ V \ {0}. π(x−i) π(y−i) Thus the Gibbs sampler is a reversible Markov chain with invariant measure π. Corollary 7.22 (L2 (μ),) · , · *) is a real Hilbert space. Next we show that Y has a continuous version. n→∞ Proof Essentially this is a simple conclusion of the Borel–Cantelli lemma. Check that f is Höldery -continuous for any γ ∈ (0, 1). Then F ◦ X = 1A . Then X = (Xn)n∈N0 is a square integrable martingale with respect to F = σ (X) (why?) and *) 2 2 . The vector p = (pω)ω∈Ω is called a probability vector. The validity of (2.2) follows as in Example 2.1(i). In many cases of interest, these quotients are easy to compute even though π(x) and π(y) are not. Clearly, 0 ∈ G; hence G = ∅. lim sup sup Fn (x) − F (x) ≤ N N n→∞ x∈R n→∞ Letting N → ∞, the claim follows. A stable distribution is characterised by its index α and a skewness parameter (and, of course, a scale parameter); see Remark 16.23. 17.1 Definitions and Construction .. Then either P = Q or 3 P ⊥ Q. , Zn that are uniformly distributed on [0, 1] (see, e.g., Corollary 2.23 for the construction). Reflection Give an example of a current that is not an electrical current. Rather, F β is asymmetric and has a global minimum mβ,h with the same sign as h. νa has a density with respect to μ, and dνa is A-measurable and finite μ-a.e. dμ Corollary 7.34 (Radon–Nikodym theorem) Let μ and ν be σ -finite measures on (Ω, A). By the Ionescu–Tulcea theorem (Theorem 14.35), the projective limit P := {−n, −n+1,...} exists. On this tree, random walk is transient. Furthermore, by the n→∞ n y 0, we have Similarly, lim inf lim 1 1 log Pn ([−x, ∞)) n 1 1 = lim log Pn ((−∞, x)) = lim log Pn ((−∞, x])) = 0 = −I (0). In this section, we derive two simple criteria that prepare us for important applications such as the law of large numbers (Chap. ♠ Theorem 6.28 (Differentiation lemma) Let I ⊂ R be a nontrivial open interval and let f : Ω × I → R be a map with the following properties. By Vd denote the set of monotone increasing, bounded right continuous functions on Rd . Again, by Theorem 6.19, it is enough to show that E[f (Yn)] ≤ C for every n ∈ N. In a second step, this random variable will be transformed by applying the inverse map F −1 : Let Ω := (0, 1), A := B(R) Ω and let P be the Lebesgue measure on (Ω, A) (see Example 1.74). By Exercise 5.3.4, the entropy is subadditive: hm+n ≤ hm + hn for m, n ∈ N. ♣ Exercise 17.7.3 Let n ∈ N, p ∈ (0, 1) and λ > 0. , Xsn = in) *) * = E E[1{Xt =i} Fsn] 1A = E E[1{Xt =i} Xsn] 1A) *) * = E P[Xt = i Xsn = in] 1A = P Xt = i Xsn = in P[A]. Show that νt = lims→t νs for all t > 0. Proof The proof is based on a Taylor expansion of the logarithm, | log(z) − (z − 1)| ≤ |z − 1|2 for z ∈ C with |z − 1| 0 with |ϕ(t)| > 12 for all t ∈ [−ε, ε]. Then, by the monotone convergence theorem, μ1 (dω1) · · · μ(A) ˜ = ≤ μn (dωn) 1A ((ω1 , . Denote by hn the entropy of Pn . That is, there exists a square integrable martingale X that converges almost surely but without lim)X*n < ∞ almost surely. Thus gε = max{f1 |, . Come up with an example that shows that this condition cannot simply be dropped. n∈N Let g ∈ E+ with g ≤ f . 18.3 Markov Chain Monte Carlo Method 447 • Each atom i ∈ Λ has a magnetic every s ∈ [0, 1] with |s − t| < δ. We consider also as a map N → N by defining (k) = k for k > n. ♣ Exercise 1.5.2 Give an example of two normally distributed random variables X and Y such that (X, Y) is not (two-dimensional) normally distributed. Note, however, that in (17.11) we have required neither independence of the random variables (Rn (x), x ∈ E) nor that all Rn had the same distribution. By the upcrossing [a, b] between times −n and 0. Theorem 21.42 (Kolmogorov's criterion for weak relative compactness) Let (Xi , i ∈ I) be a sequence of continuous stochastic processes. ∈ G and A := A1 ∪ A2 ∪ . However, we now want to develop a systematic framework for the description and construction of multi-stage experiments. 587 588 594 598 603 24 The Poisson Point Process . ♦ Example 20.28 Let Ω = [0, 1), A = B([0, 1)) and let P = λ be the Lebesgue measure on ([0, 1), B([0, 1))). Show that A is a continuous linear operator from L2 (μ1) to L2 (μ2). We show the optional sampling and optional stopping theorems in the second section. (15.4) k=1 Hint: (i) Define the characteristic functions (see Theorem 15.13) ϕ1 (t) = ϕ2 (t) = (1 − t/2)+ . ♦ Theorem 13.23 Let μ, μ1 , μ2 , . (iii) E is complete and there is a summable sequence (εn)n∈N such that ∞ n=1 μ(A ∩ {d(fn , fn+1) > εn }) < ∞ for all A ∈ A with μ(A) < ∞. Define Bn = An \ A ∈ A for all n ∈ N. Let A, A1 , A2 , . (i) Show the validity of Helly's theorem with V replaced by Vd . Define X := AW +μ. An F-martingale Y is thus determined uniquely by the terminal values YT (and vice versa). Each path of the process X of partial sums that ends above a corresponds to a unique path that reaches a but ends below a. Hence, what is the fair price π(VT) for which a trader would offer (and buy) the contingent claim VT ? Then {|f |>aε } |f | dμ ≤ {|f |> gε/2 } |f | dμ + { gε/2 >aε } gε/2 dμ < ε. (ii) X is mixing if and only if X is aperiodic. Our first task is to give precise definitions. For general measurable f , the statement follows by the usual approximation arguments. Essentially it is necessary and sufficient that the state space of the chain cannot be decomposed into subspaces • that the chain does not leave • or that are visited by the chain periodically; e.g., only for odd n or only for even n. To this end, we check (i)–(iii) of Definition 1.2. (i) Clearly, Ω ∈ Ai for every i ∈ I , and hence Ω ∈ AI . Let X be binomially distributed, X ∼ bn,p . n→∞ Remark 13.32 The implication (ii) in Theorem 13.29 is less useful but a lot simpler to prove. i=1 A function f : I → R is called Riemann integrable if there exists a t such that the limits of the lower sums and upper sums are finite and coincide. For n ∈ N0 , define Xn := nm=1 Ym . + Tn−1 and Xt = sup{n ∈ N0 : Sn ≤ t}. Definition 1.1 A class of sets A is called • ∩-closed (closed under intersections) or a π-system if A ∩ B ∈ A whenever A, B ∈ A, ∞ • σ -∩-closed (closed under countable1 intersections) if n=1 An ∈ A for any choice of countably many sets A1 , A2 , . The function f is called the density of μ and plays a role similar to the weight function p in (v). Proof Clearly, every dn is a complete metric on (C([0, n]), · ∞). ♦ Remark 12.9 Denote by T = n∈N σ (Xn+1 , Xn+2 , . The case μ = δ0 is trivial. We show that for any f ∈ Cb (E; R), any x ∈ E and any ε > 0, there exists a gx ∈ C with gx (x) = f (x) and gx (y) ≤ f (y) + ε for all y ∈ E. Theorem 2.47 (Uniqueness of the infinite open cluster) For any p ∈ [0, 1], we have Pp [N ≤ 1] = 1. In order to apply the Poisson approximation theorem (Theorem 3.7), for fixed n ∈ N, we decompose the interval (0, t] into 2n disjoint intervals of equal length, * I n (k) := (k − 1)2−n t, k2−n t , k = 1, . Intuitively, this should be x n . As the minimum of x independent exp1 -distributed random variables is expx -distributed, the waiting time for the next branching event is expx -distributed. We say that a point x ∈ Zd is a trifurcation point if • x is in an infinite open cluster C p (x), • there are exactly three open edges with endpoint x, and • removing all of these three edges splits C p (x) into three mutually disjoint infinite open clusters. n→∞ "(iv) ⇒ (iii)" (for finite μ) Assume that μ(A) < ∞ for every A ∈ A and that μ is ∅-continuous. Furthermore, the text has been extended carefully in many places. ♣ Exercise 21.1.4 Let X = (Xt)t ≥0 be a stochastic process on (Ω, F , P) with values in the Polish space E and with right continuous paths. Then (aX + bY) is a martingale. "(ii) ⇒ (iii)" Assume (ii). Show that the Chebyshev polynomials of the second kind are orthonormal with respect to ν; that is, Um Un dν = 1{m=n} . k=0 We get |f (Sn /n) − f (p)| ≤ ε + 2f ∞ 1{|(Sn /n)−p|≥δ} and thus (by Theorem 5.14 with V = p(1 − p) ≤ 14) |fn (p) − f (p)| ≤ E[|f (Sn /n) − f (p)|] , + Sn ≤ ε + 2f ∞ P − p ≥ δ n ≤ ε+ f ∞ 2 δ2 n n→∞ for any p ∈ [0, 1]. , τk+1 }, we have (H ·Y)j = (H ·Y)σk . In particular, a random walk on Z with centered increments is recurrent (Chung– Fuchs theorem, compare Theorem 17.41). Example 20.12 Let (Xn)n∈N0 be i.i.d. and let Xn (ω) = X0 (τ n (ω)). Definition 14.20 Let n ∈ N. Denote by LI := LC I := 1 I (x, y)2 R(x, y) 2 x,y∈E the energy dissipation of I in the network (E, C). The ith individual in the nth generation has Xn,i offspring (in the (n + 1)th generation). Also one can allow for different distributions in the respective factors. ∈ AI . (21.36) Note real random variables X, Y with E[X] = E[Y] = 0) that, * for) independent * and E X4 , E Y 4 < ∞ and for a ∈ [−1, 1], we have)) *) *) * *) * E (aX + Y)4 = a 4 E X4 + 6 a 2 E X2 E Y 2 + E Y 4) *) *) *) * ≤ E X4 + 6 E X2 E Y 2 + E Y 4 = E[(X + Y)4]. Exercise 1.1.1 Let A be a semiring. However, it is not a Markov chain with respect to the so-called annealed measure P[X ∈ ·]. , m and distribution p. (i) The distribution on R with density f (x) = ⎧ ⎪ ⎪ ⎨ ⎪ ⎪ ⎩ 1 α 1+α |x| , 1) 1+β xβ , (1 − 0, if x < −1, if x > 1, else. fdd Let Q be a limit point for (Pnk)k∈N along some subsequence (nk). Define C = D ∩ K. This is the reason why we studied independence of classes of events in the last section. ♦ Example 15.19 Let ϕ(t) = (1 − 2|t|/π)+ be the characteristic function of the distribution "N.N." from Theorem 15.13 (with a = π/2) and let ψ be the characteristic function from the preceding example. Theorem 5.17 (Etemadi's strong law of large numbers (1981)) Let X1 , X2 , . Show that X is a random measure if and only if we have P[X(B) 0, there exists a sequence (H n)n∈N in E such that + T E 0 Step 1. 15.1 Separating Classes of Functions 335 Assume there is an ε0 > 0 with mX (ε0) < ∞ (respectively mX (−ε0) < ∞). The event B from above can be written as B = Zn+1 < Zn only finitely often . ♦ Example 1.63 (Infinite product measure, continuation of Example 1.40) Let E be a finite set and let Ω = E N be the space of E-valued sequences. We have extended this notion to families of events and even to families of classes of sets. Then f is harmonic on E \ A. If (Xi)i∈I is independent, then (fi ◦ Xi)i∈I is independent. Proof Clearly, E[X] is finitely additive. ♣ Exercise 21.5.4 Let d ∈ N. (1) Choose a random coordinate I according to some distribution (qi)i∈Λ . Generating functions are the ideal tool for the analysis of such processes. Assume that we had νs ⊥ μ. (ii) Show that E[(Gn (t) − Gn (s))4] ≤ C (t − s)2 + |t − s|/n for some C > 0. Define E := ∞ n=1 Kn . In general, the growth to infinity of Zn can also be slower than mn . In this section, we develop a formal framework for the quantification of probabilities of rare events in which the complete theory of large deviations can be developed. (iii) First assume X ≥ 0 and Y ≥ 0. Assume that there is an integrable dominating function 0 ≤ g ∈ L1 (μ) n→∞ with |fn | ≤ g almost everywhere for all n ∈ N. ♠ Definition 21.2 Let (E, d) and (E , d) be metric spaces and γ ∈ (0, 1]. = 1, let dn = n an E[sin(X/an)] for all n ∈ N. In Sect. F ∈Fn Then F := ∞ n=1 Fn μ∗ (A) ≤ ∈ U(A) and μ(F) ≤ F ∈F ∞ μ(F) ≤ n=1 F ∈Fn ∞ μ∗ (An) + ε. Using Lemma 23.9, we get lim sup n→∞ 1 log Pn (C) n ≤ lim sup n→∞ 1 log Pn (−∞, x−] + Pn [x+ , ∞) n 1 1 log Pn [x+ , ∞) = max lim sup log Pn (−∞, x−] , lim sup n→∞ n n→∞ n = max − I (x−), −I (x+) = − inf I (C). (ii) This is obvious since Pϕ(X) = PX ◦ ϕ −1 . Hence X is ergodic. Corollary 6.13 Let (E, d) be a separable metric space. For ω0 ∈ Ω0 and n ∈ N,let Aω0 ,n := {ω1 ∈ Ω1 : κ2 ((ω0 , ω1), Ω2) < n}. First we consider the situation of a simple shift: Let Ω = E N0 , where E is a finite set equipped with the product σ -algebra A = (2E)⊗N0 . By the monotone convergence theorem (Theorem 4.20), we infer E[|XY |] = lim E[XN YN] = lim E[XN] E[YN] N→∞ = lim E[XN] N→∞ N→∞ lim E[YN] = E[|X|] E[|Y |] < ∞. , n, are called the kth moments and kth absolute moments, respectively, of X. Theorem 19.35 (Solomon [158]) Assume that E[| log(0)|] < ∞. For any measurable set A ⊂ E, define μ(A) := sup sup μ˜ m (A ∩ Wn) = sup μ˜ n+1 (A ∩ Wn). For a formal proof along the lines of this heuristic, see Sect. Clearly, P[B |A] = E[1B |A] for all B ∈ A. Warning: One of the implications is rather difficult to show. 19.6 whose effective resistance Reff 3 Z . At each step, choose one of the N balls uniformly at random. Hence also U := V ∩ W is open. Analogously, define P[A|X = x] = E[1A X = x] for A ∈ A. and n 2 n t t t2 1− t −ϕ √ −ϕ √ ≤ n 1 − 2 2 2n 2n nσ nσ ≤n n→∞ t 2 t −→ 0. In short, every random observation (the technical term is random variable) is a measurable map. Then we call f dλ the Lebesgue integral of f . Proof Let Ω = Ω + Ω − be a Hahn decomposition. The dashed vertical line indicates the critical inverse temperature. (21.16) It is enough to consider continuous bounded functions F that depend on only finitely many coordinates t1 , . By we obtain n lim E[eλZn /2] = n→∞ 1 1 1 1+λ = + = E[e−λW], 1 + 2λ 2 2 1 + 2λ where in the last step we assumed that PW = 12 δ0 + 1 2 exp1/2. This implies (using the triangle inequality; see Theorem 5.3(v)) ∞ ∞) * * *)) *) E |ST | = E |Sn | 1{T =n} = E |Sn | E 1{T =n} n=1 ≤ ∞ n=1) * E |X1 | n P[T = n] = E[|X1 |] E[T]. Owing to the independence of increments and since the density of the standard normal distribution is bounded by 1, we get *k *k)) P[B ∈ AN,n,i] = P |B1/n | ≤ N n−γ = P |B1 | ≤ N n−γ +1/2 ≤ N k nk(−γ +1/2). , n,) * F{i} (x) = P[Xi ≤ x] = P Ri−1 × (−∞, x] × Rn−i μj (R) = μi (−∞, x] . For simplicity, assume that Σ is a finite set and λ = UΣ is the uniform distribution on Σ. Since M(μ∗) is a σ -algebra (Lemma 1.52), it is enough to show A ⊂ M(μ∗). Moreover, X ∼ Nμ,C ⇐⇒ (iii) ⇐⇒ (iv). ♦ Takeaways Consider a two-step random experiment where in the first step we choose a probability measure Ξ∞ on some space E. n→∞ (iv) Finally, show that Mn ⇒ M. Step 7. ∈ A with A ⊂ ∞ i=1 Ai . The same is true for (iii) and (iv). Definition 13.12 (Weak and vague convergence) Let E be a metric space. , n since u is even. (Of course, this is wishful thinking. (19.6) Equation (19.6) is sometimes called the equation of detailed balance. In practice, for a large space E, computing the spectral gap is often extremely difficult. For x ∈ Rd and A ⊂ Rd , define Ex := m−1 ({x}) = {ν ∈ E : m(ν) = x} and EA = m−1 (A) = {ν ∈ E : m(ν) ∈ A}.