Theorems in Probability

Zi Yin

Department of Electrical Engineering, Stanford University

September 24, 2015

Contents

1 General Theorems on Measure Theory

1.1 Integration and Expectation

1. *Independence* (c.f. [1] p.55): If $\mathcal{A}_1, \mathcal{A}_2, .., \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system then $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent.

2. *Fatou's Lemma*: For nonnegative f_n, $\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$.

3. *Change of variable*: If f is continuous, g is one-one, g' exists and is continuous, then

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(y)dy$$

   For higher dimensions $g : \mathbb{R}^k \to \mathbb{R}^k$, substitute $g'(x)$ with $|J_g(x)|$ the determinant of the Jacobian of g.

4. *Exchangibility of derivative and integration* (c.f. [1] p.212): Suppose $f(\omega, x)$ has derivative $f'(\omega, x)$ with respect to x, and $|f'(\omega, x)| \leq g(\omega)$ for all ω and x, where g is integrable. Then $\frac{d(\int f(\omega,x)\mu d\omega)}{dx} = \int f'(\omega, x)\mu d\omega$.

5. $\mathbb{E}[|X|^p] = \int_0^\infty px^{p-1}\mathbb{P}(|X| > x)dx = \int_0^\infty px^{p-1}\mathbb{P}(|X| \geq x)dx$.

6. *Radon-Nikodym* (c.f. [1] p.423, [3] p.165): If μ and ν are two σ-finite measures on $(\Omega, \mathcal{F})$ such that $\nu \ll \mu$, then there exists f measurable on $\mathcal{F}$, such that $\int_A h d\nu = \int_A f h d\mu, \forall A \in \mathcal{F}, h$ measurable. Moreover, f is unique up to a null set with respect to μ. $f = \frac{d\nu}{d\mu}$ is called the Radon-Nikodym derivative.

1.2 Uniform Integrability

7. (c.f. [3] p.45): X_n U.I. $\Rightarrow \sup \mathbb{E}[|X_n|] < \infty$.

8. *U.I. of collection of conditional expectation* (c.f. [3] p.165): For any $X \in L^1(\Omega, \mathcal{F}, P)$, the collection $\{\mathbb{E}[X|\mathcal{H}] : \mathcal{H} \subset \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$ is U.I.

9. (c.f. [3] p.48, p.200): If $p > 1$, $\sup \mathbb{E}[|X_n|^p] < \infty \Rightarrow X_n$ U.I.

1.3 Moments and Characteristic Function

10. *Uniqueness of moment generating function* (c.f. [1] p.285): Suppose that μ and ν are two probability measures on $[0, +\infty)$ (one sided). If

$$\int_0^\infty e^{-sx}\mu dx = \int_0^\infty e^{-sx}\nu dx, s \geq s_0$$

then $\mu = \nu$.

11. *Uniqueness and inversion of characteristic function* (c.f. [1] p.346): Suppose $\phi(t) = \mathbb{E}[e^{itX}] = \int e^{itx}\mu dx$ is the characteristic function of X with distribution μ. Then $\phi_1 = \phi_2 \Rightarrow \mu_1 = \mu_2$.
Moreover, if $\mu(a) = \mu(b) = 0$,

$$\mu(a, b] = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt$$

12. *Method of Moments* (c.f. [1] p.388): Let μ be a probability measure on the real line having finite moments $m_n = \int_{-\infty}^{\infty} \mu dx$ of all orders. If the power series $\sum_k m_k r^k/k!$ has a positive radius of convergence then μ is the unique probability measure with moments $M_0, M_1, ...$

    Remark: A counter example is the log-normal where $X = e^N, N \sim N(0, 1)$, where all its moments exist but no positive ROC for the moment generating function.

1.4 Zero One Laws

13. *Borel-Cantelli*:
    (a) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, $\mathbb{P}(A_n \text{ i.o.}) = 0$.
    (b) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and A_n independent, $\mathbb{P}(A_n \text{ i.o.}) = 1$.

14. *Kolmogorov*: If $\{X_n\}$ mutually independent and $\mathcal{T} = \cap_{n=0}^{\infty}\sigma(X_i, i \geq n)$ is the tail σ-algebra, then $\mathbb{P}(A), A \in \mathcal{T}$ is either 0 or 1.

15. *Hewlett-Savage* (c.f. [3] p.224): Define an *exchangable σ-algebra* to be $\mathcal{E} = \cap_m \mathcal{E}_m$, where

    $\mathcal{E}_m = \{A : \omega = (\omega_1, \omega_2, ...) \in A \Rightarrow (\omega_{\pi(1)}, \omega_{\pi(2)}, .., \omega_{\pi(m)}, \omega_{m+1}...) \in A\}$

    Suppose $\mathcal{E}$ is the exchangable σ-algebra of iid random variables ξ_i, $\omega_i(\omega) = \xi_i(\omega)$. Then $\mathbb{P}(A), A \in \mathcal{E}$ is either 0 or 1.

16. Suppose φ is a measure preserving homomoerphism:

$$\varphi : \Omega \to \Omega, \mathbb{P}(\varphi^{-1}(A)) = \mathbb{P}(A), \forall A \in \mathcal{F}$$

If φ is ergodic, *i.e.*

$$\forall X \in L^1(\Omega, \mathcal{F}, P), \frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m(\omega)) \stackrel{a.c.}{\to} \mathbb{E}[X],$$

then for the invariant σ-algebra $\mathcal{I} = \{A : \mathbb{P}(\varphi^{-1}(A)\Delta A) = 0\}$, $\forall A \in \mathcal{I}$, $\mathbb{P}(A)$ is either 0 or 1.

2 Convergence Theorems

2.1 Basic Theorems

1. *Relationships between convergence*:

   (a) Converge a.c. $\Rightarrow$ converge in probability $\Rightarrow$ weak convergence.

   (b) Converge in L^p $\Rightarrow$ converge in L^q $\Rightarrow$ converge in probability $\Rightarrow$ converge weakly, $p \geq q \geq 1$.

   (c) Convergence in KL divergence $\Rightarrow$ Convergence in total variation $\Rightarrow$ strong convergence of measure $\Rightarrow$ weak convergence, where

      i. $\mu_n \overset{TV}{\to} \mu$ means $\lim ||\mu_n - \mu||_{TV} = 0$, where

      $$||\mu - \nu||_{TV} = \sup_{||f||_\infty \leq 1} \{\int f d\mu - \int f d\nu\}$$

      which also equals

      $$||\mu - \nu||_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

      ii. $\mu_n \to \mu$ strongly if $\lim \mu_n(A) = \mu(A), \forall A \in \mathcal{F}$.

2. *Subsequence of a.c. convergence*: If $X_n \overset{p}{\to} X$, then there exists an subsequence n_k, $X_{n_k} \overset{a.c.}{\to} X$.

3. *Equivalence of convergence in probability and a.c. convergence* (c.f. [1] p.290): Let $S_n = \sum_{i=1}^n X_i$. If $\{X_n\}$ is independent, then S_n converges a.c. iff S_n converges in probability.

4. *When a.c. convergence implies L^1 convergence*: Monotone convergence (MCT), Dominated convergence (DCT), Uniform integrability (U.I.).

5. *Vitali* (c.f. [3] p.46): If $X_n \overset{p}{\to} X$, then X_n is U.I. iff $X_n \overset{L^1}{\to} X$, which is again equivalent to X, X_n integrable and $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$.

6. *Scheffé* (c.f. [1] p.215): Suppose $\mu_n(A) = \int_A \delta_n d\mu$ and $\mu(A) = \int_A \delta d\mu$ for densities δ_n and δ. If $\mu_n(\Omega) = \mu(\Omega) < \infty$, and $\delta_n \to \delta$ a.c., then

   $$\sup_{A \in \mathcal{F}} |\mu(A) - \mu_n(A)| \leq \int_\Omega |\delta - \delta_n| d\mu \to 0$$

7. *Slutsky*: If $X_n \Rightarrow X$ and $X_n - Y_n \Rightarrow 0$, then $Y_n \Rightarrow X$.

   Remark: $Y_n \Rightarrow c$ is equivalent to $Y_n \stackrel{p}{\to} c$ if c is a constant, in the sense that $\lim_n \mathbb{P}(|Y_n - c| > \epsilon) = 0$.

8. *Skorohod* (c.f. [1] p.333): Suppose $\mu_n \Rightarrow \mu$ where μ_n and μ are probability measures on the real line. Then there exist some Y_n and Y on a common probability space $(\Omega, \mathcal{F}, P)$ such that $Y_n(\omega) \to Y(\omega), \forall \omega \in \Omega$, and Y_n, Y have distributions μ_n, μ.

2.2 Weak Convergence

9. *Portmanteau* (c.f. [2] p.16): The following five conditions are equivalent concerning weak convergence of probability measures:

   (a) $\mathbb{P}_n \Rightarrow \mathbb{P}$;

   (b) $\int f d\mathbb{P}_n \to \int f d\mathbb{P}$ for any bounded continuous function f;

   (c) $\limsup_n \mathbb{P}_n(F) \leq \mathbb{P}(F)$ for all closed set F;

   (d) $\liminf_n \mathbb{P}_n(G) \geq \mathbb{P}(G)$ for all open sets G;

   (e) $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for all P-continuous set A.

10. *Helly selection*: For $\{F_n\}$ a sequence of distribution functions, there exists a subsequence $\{F_{n_k}\}$, such that there exists a right-continuous non-decreasing function F, $\lim F_{n_k}(x) = F(x)$ at all continuity points of F. Moreover, F is a distribution function if and only if $\{F_n\}$ is tight.

11. *Continuous mapping preserves weak convergence* (c.f. [1] p.380): Suppose h is measurable and the discontinuity set has measure 0. If $\mu_n \Rightarrow \mu$, then $\mu_n h^{-1} \Rightarrow \mu h^{-1}$, where $\mu h^{-1}(A) \stackrel{def}{=} \mu(h^{-1}(A))$.

12. *Characteristic functions and convergence in distribution* (c.f. [1] p.383): $\mu_n \Rightarrow \mu$ iff $\varphi_n(t) \to \varphi(t)$.

13. *Necessary and sufficient conditions for multivariate weak convergence* (c.f. [1] p.383): Suppose $X_n \in \mathbb{R}^k$, $X_n = (X_{n1}, ..., X_{nk}), X = (X_1, ..., X_k)$. $X_n \Rightarrow X$ iff $\sum_{i=1}^k t_i X_{ni} \Rightarrow \sum_{i=1}^k t_i X_i$ for every $\mathbf{t} = (t_1, ..., t_k) \in \mathbb{R}^k$.

2.3 Convergence of Random Series

14. (*c.f. [1] p.289*): Suppose $\{X_n\}$ independent with $\mathbb{E}[X_i] = 0, \forall i$. Further if $\sum Var(X_n) < \infty$, then $\sum X_n$ converges a.c.

15. *Kolmogorov three-series theorem* (c.f [1] p.290): Suppose $\{X_n\}$ is independent. Consider the three series $\sum \mathbb{P}(|X_n| > c), \sum \mathbb{E}[|X_n^{(c)}|]$, and $\sum Var(X_n^{(c)})$, where $X_n^{(c)} = X_n 1_{\{|X_n| \leq c\}}$. Then $\sum X_n$ converges a.c. implies above series converge for all c. On the other hand, if the above three series converge for some positive c, then $\sum X_n$ converges a.c.

3 Inequalities

3.1 Basic Inequalities

1. *Markov*: $\mathbb{P}(|X| > \alpha) \leq \frac{1}{\alpha^k}\mathbb{E}[|X|^k]$.
2. *Chebyshev*: $\mathbb{P}(|X - \mu| > \alpha) \leq \frac{1}{\alpha^2}Var(X)$.
3. *Jensen*: If f is convex, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.
4. *Hölder*: $\mathbb{E}[|XY|] \leq ||X||_p||Y||_q$, $\frac{1}{p} + \frac{1}{q} = 1, p \geq 1$.
5. *Minkowski*: $||X + Y||_p \leq ||X||_p + ||Y||_p$, $p \geq 1$.
6. *Lyapounov*: $||X||_p \leq ||X||_q$, $0 < p \leq q$.

3.2 Maximal Inequalities

7. *Kolmogorov* (c.f. [1] p.287): Suppose $\{X_n\}$ independent with zero mean and finite second moments. Then for $\alpha > 0$,

$$\mathbb{P}(\max_{1 \leq k \leq n} |S_k| \geq \alpha) \leq \frac{1}{\alpha^2}Var(S_n)$$

8. *Etemadi* (c.f. [1] p.288): Suppose $\{X_n\}$ independent, for $\alpha > 0$,

$$\mathbb{P}(\max_{1 \leq k \leq n} |S_k| \geq \alpha) \leq 3 \max_{1 \leq k \leq n} \mathbb{P}(|S_k| \geq \frac{\alpha}{3})$$

4 Asymptotics

4.1 LLN, CLT, LIL and Extreme Values

1. *Strong LLN*: Suppose $X_1, X_2, ...$ are iid random variables with finite first moment. Then with probability 1, $S_n/n \to \mathbb{E}[X_1]$.

2. *Law of iterated logarithm* (c.f. [1] p.154): Suppose $X_1, .., X_n$ are iid simple random variables with mean 0 and variance 1. Then

$$\mathbb{P}(\limsup_n \frac{S_n}{\sqrt{2n \log\log n}} = 1) = 1$$

3. *Glivenko-Cantelli*: Suppose X_n is a stationary ergodic process, then

$$||F_n - F||_\infty \stackrel{a.c.}{=} 0$$

where $F_n(x) = \frac{1}{n}\sum 1_{(-\infty,x]}(X_i)$ is the empirical distribution function.

4. *Lindeberg CLT* (c.f. [1] p.359): Suppose $\{X_{nk}\}$ is a triangular array. Let $S_n = \sum_{i=1}^{r_n} X_{ni}$. If for all $X_{nk}, 1 \le k \le r_n$,

$$\mathbb{E}[X_{nk}] = 0, \sigma_{nk}^2 = \mathbb{E}[X_{nk}^2], s_n^2 = \sum_{i=1}^{r_n} \sigma_{ni}^2$$

and the Lindeberg condition holds for all $\epsilon > 0$:

$$\lim_{n\to\infty} \sum_{i=1}^{r_n} \frac{1}{s_n^2} \int_{|x_{ni}|\ge \epsilon s_n} x_{ni}^2 d\mathbb{P}_{X_{ni}} = 0$$

Then $S_n/s_n \Rightarrow N(0,1)$.

5. *Fisher-Tippett-Gnedenko*: Suppose $X_1, X_2, ..$ are iid random variables, and $M_n = \max\{X_1, .., X_n\}$. If there exists a sequence of pairs of reals (a_n, b_n), $a_n > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}(\frac{M_n - b_n}{a_n} \le x) = F(x)$$

where F is non-degenerate, then F can only be one of the following three distributions:

   (a) Gumbel: $F(x) = e^{-e^{-x}}$;

(b) Fréchet:

$$F(x) = \begin{cases} 0, x \leq 0 \\ e^{-x^{-\alpha}}, x > 0 \end{cases} \tag{1}$$

(c) reversed Weibull:

$$F(x) = \begin{cases} e^{-(-x)^{\alpha}}, x \leq 0 \\ 1, x > 0 \end{cases} \tag{2}$$

6. Suppose $X_1, X_2, ..$ are iid random variables with mean 0 and variance 1, and $S_n = \sum_{i=1}^{n} X_i$. For each $\epsilon > 0$, let $N(\epsilon) = \inf\{n : S_k/k < \epsilon, \forall k > n\}$. Then $\epsilon^2 N(\epsilon)$ converges in distribution to a 1-DoF chi-square distribution.
Remark: It is related to the Brownian hitting time $\sup\{t \geq 0 : B_t = t\}$.

4.2 Stein-Chen Method

7. *Wasserstein metric*: the distance $d_{\mathcal{H}}(X, Y)$ between two random variables with respect to a set of test functions $\mathcal{H}$ is defined by

$$d_{\mathcal{H}}(X, Y) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|$$

When $\mathcal{H} = \{h : |h(x) - h(y)| \leq |x - y|, \forall x, y\}$, this distance is defined to be the Wasserstein distance.

4.2.1 Gaussian Approximation

8. *Stein's Lemma*: Define a differential operator $\mathcal{D}$ by

$$\mathcal{D}(f)(x) = f'(x) - xf(x)$$

If $\mathbb{E}[\mathcal{D}(f)(Z)] = 0$ for all absolutely continuous function f with $||f'||_\infty < \infty$, then Z is a standard Gaussian random variable.
Conversely, if Z is a standard Gaussian random variable, then $\mathbb{E}[\mathcal{D}(f)(Z)] = 0$ for all absolutely continuous function f with $\mathbb{E}[|f'(Z)|] < \infty$.

9. If W is a random variable and Z is a standard Gaussian random variable, define the set of functions $\mathcal{F} = \{f : ||f||_\infty \leq 2, ||f''||_\infty \leq 2, ||f'||_\infty \leq \sqrt{2/\pi}\}$. Then

$$d_W(W, Z) \leq \sup_{f \in \mathcal{F}} |\mathbb{E}[f'(W) - Wf(W)]|$$

10. *Approximation of dependency neighborhoods*: Suppose $X_1, X_2, ..$ are random variables such that $\mathbb{E}[X_i] = 0, \sigma_n^2 = Var(\sum_{i=1}^n X_i), (\mathbb{E}[|X_i|^4] < \infty$. Let $D = \max_{1 \le i \le n} |N_i|$, $S_n = \sum X_i/\sigma_n$. Then

$$d_W(S_n, Z) \le \frac{D^2}{\sigma_n^3} \sum_{i=1}^n \mathbb{E}[|X_i|^3] + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi}\sigma_n^2} \sqrt{\sum_{i=1}^n \mathbb{E}[|X_i|^4]}$$

4.2.2 Poisson Approximation

11. *Poisson characteristic operator*: For $\lambda > 0$, define operator $\mathcal{D}$ by

$$\mathcal{D}(f)(k) = \lambda f(k+1) - kf(k)$$

If for some nonnegative integer valued random variable W, $\mathbb{E}[\mathcal{D}(f)(W)] = 0$ for all bounded functions f, then $W \sim Po(\lambda)$.
Conversely, if $W \sim Po(\lambda)$, then $\mathbb{E}[\mathcal{D}(f)(W)] = 0$ for all bounded f.

12. Let $\mathcal{F} = \{f : ||f||_\infty \le \min\{1, \lambda^{-1/2}\}, \ \textit{and} \ ||\Delta f||_\infty \le \frac{1-e^{-\lambda}}{\lambda} \le \min\{1, \lambda^{-1}\}\}$, and W is an integer valued nonnegative random variable with mean λ. If $Z \sim Po(\lambda)$, then

$$d_{TV}(W, Z) \le \sup_{f \in \mathcal{F}} |\mathbb{E}[\lambda f(W+1) - W f(W)]|$$

13. *Approximation of dependency neighborhoods*: Suppose $X_1, X_2, ..$ are binary random variables with $\mathbb{P}(X_i = 1) = p_i$. Let $S_n = \sum_{i=1}^n X_i$ and $\lambda_n = \sum p_i$. Define $p_{ij} = \mathbb{E}[X_i X_j]$, and $Z \sim Po(\lambda)$. Then

$$d_{TV}(S_n, Z) \le \min\{1, \lambda^{-1}\}(\sum_{i=1}^n \sum_{j \in N_i} p_i p_j + \sum_{i=1}^n \sum_{j \in N_i - \{i\}} p_{ij})$$

4.3 Method of Types

Suppose $X_1, ..., X_n$ are iid random variables taking values from a discrete set $\mathcal{X}$. The *type* $P_{\mathbf{x}^n}$ of sequence $\mathbf{x}^n$ is the empirical distribution of $\mathbf{x}^n$. The *type class* $T(P_{\mathbf{x}^n})$ of a type $P_{\mathbf{x}^n}$ is defined to be $\{\mathbf{y}^n : \mathbf{y}^n$ has empirical distribution $P_{\mathbf{x}^n}\}$. $\mathcal{P}_n$ is the set of all types with respect to n and alphabet $\mathcal{X}$.

14. If $X_1, ..., X_n$ are drawn iid according to a distribution $Q(x)$, then the probability of $\mathbf{x}^n$ depends only on its type and equals:

$$Q^n(\mathbf{x}^n) = 2^{-n(H(P_{\mathbf{x}^n})+D(P_{\mathbf{x}^n}||Q))}$$

15. *Size of a type class T(P)*: For any $P \in \mathcal{P}_n$,

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(P)} \leq |T(P)| \leq 2^{nH(P)}$$

    Remark: Here no underlying distribution is assumed.

16. *Probability of a type class*: for any $P \in \mathcal{P}_n$ and any distribution $\mathbb{Q}$, the probability of the type class $T(P)$ under Q^n is $2^{-nD(P||Q)}$ to first order in the exponent. More precisely,

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-nD(P||Q)} \leq Q^n(T(P)) \leq 2^{-nD(P||Q)}$$

17. *LLN for empirical distribution* (c.f. [5] p.356): Suppose $X_1, ..., X_n$ are iid according to $\mathbb{P}(x), x \in \mathcal{X}$. Then,

$$D(P_{\mathbf{x}^n}||P) \overset{a.c.}{\to} 0$$

4.4 Large Deviation

18. *Berry-Esseen*: Suppose $X_1, X_2, ..$ are independent random variables with $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = \sigma_i^2 > 0$ and $\mathbb{E}[|X_i|^3] = \rho_i < \infty$. Let $s_n^2 = \sum_{i=1}^n \sigma_i^2$, $S_n = \sum_{i=1}^n X_i/s_n$. Then for Z a standard Gaussian random variable,

$$d_K(S_n, Z) \leq C_0 \psi_n$$

    Where

$$\psi_n = (\sum_{i=1}^n \sigma_i^2)^{-3/2} \cdot \sum_{i=1}^n \rho_i$$

    and $d_K(X, Y) = \sup_x \{|F_X(x) - F_Y(x)|\}$ is the Kolmogorov distance.

19. *Sanov*: Suppose $X_1, ..., X_n$ are iid according to $Q(x), x \in \mathcal{X}$. Let E be a set of probability of distributions. Then

$$Q^n(E) = Q^n(E \cap \mathcal{P}_n) \leq (n+1)^{|\mathcal{X}|} 2^{-nD(P^*||Q)}$$

Where $P^* = \underset{P \in E}{argmin} D(P||Q)$, and

$$Q^n(E \cap \mathcal{P}_n) = \sum_{\mathbf{x}^n : P_{\mathbf{x}^n} \in E} Q^n(\mathbf{x}^n)$$

20. *Hoeffding*: Suppose $X_1, ..., X_n$ are independent variables, each is a.c. bounded. Suppose for each $X_i, \mathbb{P}(X_i \in [a_i, b_i]) = 1$. Let $S_n = \sum X_i, \mu = \mathbb{E}[S_n]/n$. Then

$$\mathbb{P}(|\frac{S_n}{n} - \mu| > \epsilon) \leq 2e^{-\frac{2n^2\epsilon^2}{\sum(b_i - a_i)^2}}$$

21. *Chernoff*: Suppose $X_1, X_2, ..$ are iid random variables with $\mathbb{E}[X_1] < 0$, $\mathbb{P}(X_1 > 0) > 0$. Let $M(t) = \mathbb{E}[e^{tX_1}]$, and $\rho = \inf_t M(t)$. Then

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\sum_{i=1}^{n} X_i \geq 0) = \log \rho$$

22. *Covering Lemma* (c.f. [9] p.62): Let $(U, X, \hat{X}) \sim p(u, x, \hat{x})$ and $\epsilon' < \epsilon$. Let $(U^n, X^n) \sim p(u^n, x^n)$ be a pair of random sequences with $\lim_{n\to\infty} \mathbb{P}((U^n, X^n) \in \mathcal{T}_{\epsilon'}(U, X)) = 1$. Suppose there are $x_n \geq 2^{nR}$ many random sequences $\hat{X}^n(1), .., \hat{X}^n(x_n)$, each distributed accoding to $\prod_{i=1}^n p_{\hat{X}|U}(\hat{x}_i|u_i)$ which are conditionally independent of each other and X^n given U^n. Then $\exists\ \delta(\epsilon)$ tends to 0 as $\epsilon \to 0$, such that

$$\lim_{n\to\infty} \mathbb{P}((U^n, X^n, \hat{X}^n(m)) \notin \mathcal{T}_\epsilon(U, X, \hat{X}), \forall m = 1, .., x_n) = 0$$

if $R > I(X, \hat{X}|U) + \delta(\epsilon)$.

23. *Packing Lemma* (c.f. [9] p.46): Let $(U, X, Y) \sim p(u, x, y)$. Let $(\tilde{U}^n, \tilde{Y}^n) \sim p(\tilde{u}^n, \tilde{y}^n)$ be a pair of arbitrarily distributed random sequence. Suppose there are $x_n \leq 2^{nR}$ random sequences $X^n(1), .., X^n(x_n)$, each distributed accoding to $\prod_{i=1}^n p_{X|U}(x_i|u_i)$, which are independent of $\tilde{Y}^n$ given U^n. Then $\exists\ \delta(\epsilon)$ tends to 0 as $\epsilon \to 0$, such that

$$\lim_{n\to\infty} \mathbb{P}(\exists\ m \in \{1, .., x_n\}, (\tilde{U}^n, X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon(U, X, Y)) = 0$$

if $R < I(X, Y|U) - \delta(\epsilon)$.

4.5 KL Divergence

24. *Pythagorean* (c.f. [5] p.367): For a closed convex set of probability distributions E and distribution $\mathbb{Q} \notin E$, let $\mathbb{P}^* = \underset{\mathbb{P} \in E}{argmin} D(\mathbb{P}||\mathbb{Q})$. Then

$$D(\mathbb{P}||\mathbb{Q}) \geq D(\mathbb{P}||\mathbb{P}^*) + D(\mathbb{P}^*||\mathbb{Q})$$

25. *Pinsker*: Suppose $\mathbb{P}$ and $\mathbb{Q}$ are two probability distributions in the same space, $||\mathbb{P} - \mathbb{Q}||_{TV} = 2 \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|$. Then

$$||\mathbb{P} - \mathbb{Q}||_{TV} \leq 2\sqrt{2\ln(2) D(\mathbb{P}||\mathbb{Q})}$$

5 Conditional Expectation

1. *Independence* (c.f. [3] p.159): If $X \in L^1(\Omega, \mathcal{F}, P)$, and $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then

$$\mathbb{E}[X|\sigma(\mathcal{H}, \mathcal{G})] = \mathbb{E}[X|\mathcal{G}]$$

2. *Tower Property* (c.f. [3] p.160): If $X \in L^1(\Omega, \mathcal{F}, P)$, and $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$.

3. *Taking out what's known* (c.f. [3] p.160): Suppose $Y \in m\mathcal{G}$ and $X \in L^1(\Omega, \mathcal{F}, P)$ are such that $XY \in L^1(\Omega, \mathcal{F}, P)$. Then $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$.

4. *Law of total variation*: For any σ-algebra $\mathcal{G}$ and random variable X,

$$Var(X) = \mathbb{E}[Var(X|\mathcal{G})] + Var(\mathbb{E}[X|\mathcal{G}])$$

   Where $Var(X|\mathcal{G}) = \mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])^2|\mathcal{G}]$.

5. *Conditional Jensen* (c.f. [3] p.162): Suppose $g(\cdot)$ is a convex function on an open interval G of $\mathbb{R}$. If X is an integrable R.V. with $\mathbb{P}(X \in G) = 1$ and $g(X)$ is also integrable, then almost surely $\mathbb{E}[g(X)|\mathcal{H}] \geq g(\mathbb{E}[X|\mathcal{H}])$ for any σ-algebra $\mathcal{H}$.

6. *Conditioning decreases p-norm* (c.f. [3] p.163):

$$||X||_p \geq ||\mathbb{E}[X|\mathcal{G}]||_p, \forall p > 1$$

7. *MCT, DCT, Fatou's Lemma, conditional version* (c.f. [3] p.165).

8. *U.I. of collection of conditional expectation* (c.f. [3] p.165): For any $X \in L^1(\Omega, \mathcal{F}, P)$, the collection $\{\mathbb{E}[X|\mathcal{H}] : \mathcal{H} \subset \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$ is U.I.

9. *C.E. minimizes L^2 norm* (c.f. [3] p.170): Suppose $X \in L^2(\Omega, \mathcal{F}, P)$, $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra. If $Y = \mathbb{E}[X|\mathcal{G}]$, among all $Z \in m\mathcal{G}$, $\mathbb{E}[(X - Y)^2] \leq \mathbb{E}[(X - Z)^2]$.

10. *Definition of R.C.P.D* (c.f. [3] p.172): Let $Y : \Omega \to \mathbb{S}$ be an $(\mathbb{S}, \mathcal{S})$-valued R.V. in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. The collection $\hat{\mathbb{P}}_{Y|\mathcal{G}}(\cdot, \cdot) : \mathcal{S} \times \Omega \to [0, 1]$ is called the regular conditional probability distribution (R.C.P.D.) of Y given $\mathcal{G}$ if:

(a) $\mathbb{P}(A, \cdot)$ is a version of the C.E. $\mathbb{E}[1_{Y \in A}|\mathcal{G}]$ for each fixed $A \in \mathcal{S}$.
(b) For any fixed $\omega \in \Omega$, the set function $\hat{\mathbb{P}}_{Y|G}(\cdot, \omega)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$.

In case $\mathbb{S} = \Omega, \mathcal{S} = \mathcal{F}$ and $Y(\omega) = \omega$, we call this collection the regular conditional probability on $\mathcal{F}$ given $\mathcal{G}$, denoted by $\hat{\mathbb{P}}(A|\mathcal{G})(\omega)$.

11. *C.E. and R.C.P.D.* (c.f [3] p.174): $\mathbb{E}[X|\mathcal{G}](\omega) = \int_{\mathbb{R}} x d\hat{\mathbb{P}}_{X|\mathcal{G}}(x, \omega)$.

6 Martingale

1. (c.f. [3] p.182): Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}[\phi(X_n)] < \infty, \forall n$. If $\{X_n\}$ is a martingale then $\{\phi(X_n)\}$ is a sub-martingale. Moreover, if ϕ is non-decreasing, then $\{X_n\}$ a sub-martingale $\Rightarrow \{\phi(X_n)\}$ a sub-martingale.

2. *Martingale transform* (c.f. [3] p.183): Suppose $\{Y_n\}$ is the martingale transform of $\mathcal{F}_n$-predictable $\{V_n\}$ with respect to a sub or super martingale $(X_n, \mathcal{F}_n)$, i.e.

$$Y_n = \sum_{k=1}^{n} V_k(X_k - X_{k-1})$$

   Then

   (a) If Y_n is integrable and $(X_n, \mathcal{F}_n)$ is a martingale, then $(Y_n, \mathcal{F}_n)$ is also a martingale.

   (b) If Y_n is integrable, $V_n \geq 0$ and $(X_n, \mathcal{F}_n)$ is a sub-(sup)martingale, then $(Y_n, \mathcal{F}_n)$ is also a sub-(sup)martingale.

   (c) For the integrability of Y_n it suffices in both cases to have $|V_n| \leq c_n$ for some non-random finite constants c_n, or alternatively to have $V_n \in L^q$, and $X_n \in L^p$ for all n and some $p, q > 1$ such that $1/p + 1/q = 1$.

3. *Stopping time decomposition* (c.f. [3] p.185): Suppose $\{X_n\}$ is a sub-(sup)martingale, and $\theta \leq \tau$ are two stopping times, then

$$X_{n\wedge\tau} - X_{n\wedge\theta} = \sum_{k=1}^{n} 1_{\{\theta < k \leq \tau\}}(X_k - X_{k-1})$$

   is a sub-(sup)martingale.

4. *Doob's Decomposition* (c.f. [3] p.186): Given an integrable stochastic process $\{X_n\}$ adapted to a filtration $\{\mathcal{F}_n\}, n \geq 0$, there exists $X_n = Y_n + A_n$ such that:

   (a) $(Y_n, \mathcal{F}_n)$ is a martingale and

   (b) $\{A_n\}$ is an $\mathcal{F}_n$-predictable sequence. This decomposition is unique up to $Y_0 \in m\mathcal{F}_0$.

6.1 Inequalities

5. *Doob's Inequality* (c.f. [3] p.188): Suppose $\{X_n\}$ is a sub-martingale and $x > 0$. Define $\tau_x = \min\{k : X_k \geq x\}$. Then for any $n \geq 0$,

$$\mathbb{P}(\max_{0 \leq k \leq n} X_k \geq x) \leq x^{-1}\mathbb{E}[X_n 1_{\{\tau_x \leq n\}}] \leq x^{-1}\mathbb{E}[(X_n)_+] \leq x^{-1}\mathbb{E}[|X_n|]$$

6. L^p *maximal* (c.f. [3] p.191): If $\{X_n\}$ is a sub-martingale then for any n and $p > 1$, Then

$$\mathbb{E}[(\max_{k \leq n} X_k)_+^p] \leq q^p \mathbb{E}[(X_n)_+^p]$$

where $q = p/(p-1)$. If $\{Y_n\}$ is a martingale then for any n and $p > 1$,

$$\mathbb{E}[(\max_{k \leq n} |Y_k|)^p] \leq q^p \mathbb{E}[|Y_n|^p]$$

7. (c.f. [3] p.189): Suppose Z_n is a non-negative sub-martingale with $Z_0 = 0$. Let A_n be the predictable sequence in Doob's Decomposition, and $V_n = \max_{1 \leq k \leq n} Z_k$. Then for any stopping time τ and any $x, y > 0$,

$$\mathbb{P}(V_\tau \geq x, A_\tau \leq y) \leq \frac{1}{x}\mathbb{E}[A_\tau \wedge y]$$

Further $\mathbb{E}[V_\tau^p] \leq c_p \mathbb{E}[A_\tau^p]$, $c_p = 1 + 1/(1-p), \forall p \in (0,1)$.

8. *Azuma*: Suppose $\{X_n\}$ a sub-martingale with bounded increament, *i.e.* $|X_k - X_{k-1}| < c_k$ a.c. Then for any positive integer n and positive t,

$$\mathbb{P}(X_n - X_0 \geq t) \leq e^{(\frac{-t^2}{2\sum_{k=1}^n c_k^2})}, \mathbb{P}(X_n - X_0 \leq -t) \leq e^{(\frac{-t^2}{2\sum_{k=1}^n c_k^2})}$$

6.2 Convergence

9. *Doob's Up Crossing* (c.f. [3] p.192): Suppose $\{X_n\}$ is a sup-martingale. Then for any $a < b$,

$$(b-a)\mathbb{E}[U_n[a,b]] \leq \mathbb{E}[(X_n - a)_-] - \mathbb{E}[(X_0 - a)_-]$$

10. *Doob's Convergence* (c.f. [3] p.194): Suppose $(X_n, \mathcal{F}_n)$ is a sup-(sub)martingale with $\sup_n\{\mathbb{E}[(X_n)_-]\} < \infty$ (or $\sup_n\{\mathbb{E}[(X_n)_+]\} < \infty$). Then $X_n \overset{a.c.}{\to} X_\infty$ and $\mathbb{E}[|X_\infty|] \leq \liminf \mathbb{E}[|X_n|]$ which is finite.

11. *Bounded difference* (c.f. [3] p.195): Suppose $\{X_n\}$ is a martingale of uniformly bounded difference. Consider the two events:

$$A = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega) \in (-\infty, \infty)\}$$

$$B = \{\omega : \liminf_{n\to\infty} X_n(\omega) = -\infty, \limsup_{n\to\infty} X_n(\omega) = \infty\}$$

Then $\mathbb{P}(A \cup B) = 1$.

12. *Martingale CLT*: Suppose $(X_n, \mathcal{F}_n)$ is a martingale with bounded difference, $|X_1| < k$ and $|X_i - X_{i-1}| < k$ for all i and some constant k. Define $\sigma_k^2 = \mathbb{E}[(X_{k+1} - X_k)^2|\mathcal{F}_k]$, and let $\tau_\nu = \min\{k : \sum_{i=1}^k \sigma_i^2 \geq \nu\}$. Then $\frac{X_{\tau_\nu}}{\sqrt{\nu}}$ converges in distribution to a standard Gaussian distribution.

6.3 Uniform Integrable Martingale

13. If X_n is a sub-martingale then $\{X_n\}$ is U.I. if and only if $X_n \overset{L^1}{\to} X_\infty$. In this case, we also have $X_n \overset{a.c.}{\to} X_\infty$ and $X_n \leq \mathbb{E}[X_\infty|\mathcal{F}_n]$.

    Remark: $(X_n, \mathcal{F}_n)$ is a U.I. martingale if and only if $X_n = \mathbb{E}[X|\mathcal{F}_n]$ for some X, and $X_n \overset{a.c.}{\to} X$ in this case.

14. *Lévy's Upward Theorem* (c.f. [3] p.198): Suppose $\sup|X_n|$ is integrable, $X_n \overset{a.c.}{\to} X_\infty$ and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Then $\mathbb{E}[X_n|\mathcal{F}_n] \to \mathbb{E}[X_\infty|\mathcal{F}_\infty]$ both a.c. and in L^1.

15. *Lévy's 0-1 Law* (c.f. [3] p.199):If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, and $A \in \mathcal{F}_\infty$, then $\mathbb{E}[1_A|\mathcal{F}_n] \to 1_A$.

16. *L^p martingale convergence* (c.f. [3] p.201): Suppose X_n is a martingale and $\sup \mathbb{E}[|X_n|^p] < \infty$ for some $p > 1$, then $X_n \overset{a.c.}{\to} X_\infty$ and also $X_n \overset{L^p}{\to} X_\infty$ for some random variable X_∞.

6.4 Square Integrable Martingale

17. *Predictable compensator* (c.f. [3] p.202): Let $(X_n, \mathcal{F}_n)$ be a square integrable martingale. Suppose $X_n^2 = A_n + M_n$ in Doob's decomposition, where $A_n = X_0^2 + \sum_{k=1}^n \mathbb{E}[(X_k - X_{k-1})^2|\mathcal{F}_{k-1}]$ is the predictable compensator, denoted by $A_n = \langle X\rangle_n$, and M_n is a martingale.

18. There exist finite constants $c_q, q \in (0,1]$, such that if $(X_n, \mathcal{F}_n)$ is an L^2 martingale with $X_0 = 0$, then

$$\mathbb{E}[\sup |X_k|^{2q}] \le c_q \mathbb{E}[\langle X \rangle_\infty^q]$$

where $\langle X \rangle_\infty$ is the pointwise limit of $\langle X \rangle_n$.

19. Suppose $(X_n, \mathcal{F}_n)$ is a L^2 martingale with $X_0 = 0$. Then

    (a) X_n converges to a finite limit a.c. for ω where $\langle X \rangle_\infty(\omega)$ is finite.

    (b) $X_n(\omega)/\langle X \rangle_n(\omega) \to 0$ a.c. for $\{\omega : \langle X \rangle_\infty(\omega) < \infty\}$.

    (c) If $|X_n - X_{n-1}|$ is uniformly bounded then the converse of (a) holds, *i.e.* $\langle X \rangle_\infty < \infty$ a.c. for $\{\omega : X_n(\omega)$ converging to a finite limit$\}$.

20. *Borel Cantelli III* (c.f. [3] p.204): Consider events $A_n \in \mathcal{F}_n$ for some filtration $\{\mathcal{F}_n\}$. Let $S_n = \sum_{k=1}^n 1_{A_k}$ count the number of events occurring among the first n, with $S_\infty = \sum_{k=1}^\infty 1_{A_k}$ the corresponding total number of occurrences. Similarly, let $Z_n = \sum_{k=1}^n \xi_k$ denote the sum of the first n conditional probabilities $\xi_k = \mathbb{P}(A_k|\mathcal{F}_{k-1})$, and $Z_\infty = \sum_{k=1}^\infty \xi_k$. Then a.c.

    (a) If $Z_\infty(\omega)$ is finite, so is $S_\infty(\omega)$.

    (b) If $Z_\infty(\omega)$ is infinite, then $S_\infty(\omega)/Z_\infty(\omega) \to 1$.

6.5 Optional Stopping

21. *U.I. of stopped process* (c.f. [3] p.208): Suppose $\{Y_n\}$ is integrable and τ is a stopping time. Then $\{Y_{n\wedge\tau}\}$ is U.I. if any of the following conditions hold:

    (a) $\mathbb{E}[\tau] < \infty$ and $\mathbb{E}[|Y_n - Y_{n-1}|\mathcal{F}_{n-1}] < c$ a.c. for some constant c;

    (b) $\{Y_n 1_{\{\tau > n\}}\}$ is U.I. and $Y_\tau 1_{\{\tau<\infty\}}$ is integrable;

    (c) $\{Y_n\}$ is a U.I. sub(sup)-martingale.

22. *Optional stopping I*: Suppose $\theta < \tau$ are stopping times and X_n non-positive sub-martingales for the filtration $\mathcal{F}_n$. Then X_θ and X_τ are integrable and $\mathbb{E}[X_0] \le \mathbb{E}[X_\theta] \le \mathbb{E}[X_\tau]$.

23. *Optional stopping II*: Suppose $\theta < \tau$ are stopping times and X_n sub-martingales for the filtration $\mathcal{F}_n$ such that $X_{n\wedge\tau}$ is U.I. Then X_θ and X_τ are integrable and $\mathbb{E}[X_0] \le \mathbb{E}[X_\theta] \le \mathbb{E}[X_\tau]$.

24. *Optional stopping III* (c.f. [3] p.207): Suppose θ, τ are two stopping times such that $\tau \geq \theta$ a.c., X_θ is integrable and $\mathbb{E}[X_\tau] \geq \mathbb{E}[X_\theta]$. Then $\mathbb{E}[X_\tau|\mathcal{F}_\theta] \geq X_\theta$ a.c.

25. Suppose $\{X_n\}$ is a sub-martingale and $\{\tau_k\}$ a sequence of non-decreasing stopping times. Then $(X_{\tau_k}, \mathcal{F}_{\tau_k})$ is a sub-martingale if either $\sup \tau_k < \infty$ or $X_n \leq \mathbb{E}[X|\mathcal{F}_n]$ for some integrable X and all n.

6.6 Branching Process

26. Suppose Z_n is a branching process, *i.e.* $Z_0 = 1$ and $Z_n = \sum_{i=1}^{Z_{n-1}} N_i^{(n)}$ for some random variables $N_i^{(n)}, \mathbb{E}[N_i^{(n)}] < \infty$. If $N_i^{(n)} \stackrel{d}{=} N, \mathbb{P}(N = 0) > 0$, then almost certainly either $Z_n \neq 0$ for finitely many n, or $Z_n \to \infty$.

27. *Generating function*: $L(s) = \mathbb{E}[s^N]$ is called the *generating function* of a branching process Z_n.

28. *Associated martingales* (c.f. [3] p.214): Suppose Z_n a branching process with $0 < \mathbb{P}(N = 0) < 1$. Then $(m_N^{-n} Z_n, \mathcal{F}_n)$ is a martingale where $m_N = \mathbb{E}[N] < \infty$.
    If Z_n is super-critical, *i.e.* $m_N > 1$, $(\rho^{Z_n}, \mathcal{F}_n)$ is a martingale where $0 < \rho < 1$ is the unique solution for $L(x) = x$. In the sub-critical case, $(\rho^{Z_n}, \mathcal{F}_n)$ is a martingale where $\rho > 1$ is a solution for $L(x) = x$ if exists.

29. *Extinction probability*: Suppose $0 < \mathbb{P}(N = 0) < 1$. if $m_N \leq 1$ then $p_{ex} = 1$. If $m_N > 1$, $p_{ex} = \rho$ is the solution of $L(x) = x$. In this case, $m_N^{-n} Z_n \stackrel{a.c.}{\to} X_\infty$ and $Z_n \stackrel{a.c.}{\to} Z_\infty \in \{0, \infty\}$.

30. *Moment generating function* (c.f. [3] p.216): Consider the moment generating function for Z_n: $M_n(s) = \mathbb{E}[s^{Z_n}]$ for $s \in [0, 1]$. Then recursively $M_0(s) = s$ and $M_n(s) = L(M_{n-1}(s))$.
    The moment generating function $\hat{M}_\infty(s)$ for $(m_N^{-n} Z_n)_\infty$ is a solution of $\hat{M}_\infty(s) = L(\hat{M}_\infty(s^{1/m_N}))$.

6.7 Reversed Martingale

31. *Kakutani*: (c.f. [3] p.218): Suppose $M_n = \prod_{k=1}^{n} Y_k$, with $M_0 = 1$ and independent $Y_k > 0$ such that $\mathbb{E}[Y_k] = 1$. Further let $a_k = \mathbb{E}[\sqrt{Y_k}]$. The following statements are equivalent:

(a) $\{M_n\}$ is U.I.;

(b) $M_n \xrightarrow{L^1} M_\infty$;

(c) $\mathbb{E}[M_\infty] = 1$;

(d) $\prod_{k=1}^\infty a_k > 0$;

(e) $\sum_{k=1}^\infty (1 - a_k) < \infty$.

If any of these conditions fail, $M_\infty = 0$ a.c.

32. Let $\mathbb{P}, \mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F}_\infty)$. Let $\mathbb{P}_n, \mathbb{Q}_n$ denoting $\mathbb{P}, \mathbb{Q}$ restricted on a filtration $\{\mathcal{F}_n\} \uparrow \mathcal{F}_\infty$ .Suppose $\mathbb{Q}_n$ is absolutely continuous with respect to $\mathbb{P}_n$, and $M_n = d\mathbb{Q}_n/d\mathbb{P}_n$. Then $(M_n, \mathcal{F}_n)$ is a martingale on $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ and $M_n \xrightarrow{a.c.} M_\infty$ where M_∞ is finite a.c. If $\{M_n\}$ is U.I. then $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, and $M_\infty = d\mathbb{Q}/d\mathbb{P}$.
Moreover, generally the Lebesgue decomposition of $\mathbb{Q}$ with respect to $\mathbb{P}$ is
$$\mathbb{Q} = \mathbb{Q}_{ac} + \mathbb{Q}_s = M_\infty \mathbb{P} + 1_{\{M_\infty = \infty\}} \mathbb{Q}$$
i.e. $\mathbb{Q}_{ac}(A) = \int_A M_\infty(\omega) d\mathbb{P}$, $\mathbb{Q}_s(A) = \int_A 1_{\{M_\infty=\infty\}} d\mathbb{Q}$.

33. *Likelihood ratios* (c.f. [3] p.220): Suppose $\mathbb{P}, \mathbb{Q}$ are two measures on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$, and under both the $\mathbb{P}$ and $\mathbb{Q}$, the coodinate maps $X_n(\omega) = \omega_n$ are independent. Further suppose $\mathbb{Q} \cdot X_k^{-1}$ is absolutely continuous with respect to $\mathbb{P} \cdot X_k^{-1}$.
Let $Y_k(\omega) = \frac{d(\mathbb{Q} \cdot X_k^{-1})}{d(\mathbb{P} \cdot X_k^{-1})}(X_k(\omega))$. Then $M_\infty = \prod_k Y_k$ exists under both $\mathbb{P}$ and $\mathbb{Q}$. Moreover if $\alpha = \prod_{k=1}^\infty \mathbb{P}(\sqrt{Y_k}) > 0$ then $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$ and $d\mathbb{Q}/d\mathbb{P} = M_\infty$. If $\alpha = 0$ then $\mathbb{Q}$ is singular with respect to $\mathbb{P}$ and $M_\infty \overset{\mathbb{Q}-a.c.}{=} \infty$ and $M_\infty \overset{\mathbb{P}-a.c.}{=} 0$.

34. *Reversed martingale convergence*: Suppose X_0 is integrable, $(X_n, \mathcal{F}_n)$, $n \leq 0$ is a reversed margtingale if and only if $X_n = \mathbb{E}[X_0|\mathcal{F}_n]$ for all $n \leq 0$.
Further
$$X_n \underset{L^1}{\overset{a.c.}{\rightarrow}} \mathbb{E}[X_0|\mathcal{F}_{-\infty}] \ as \ n \to -\infty$$

35. *Lévy's downward theorem*: Suppose $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ and $X_n \xrightarrow{a.c.} X_{-\infty}$. If $\sup_n |X_n|$ is integrable, then $\mathbb{E}[X_n|\mathcal{F}_n] \xrightarrow{a.c.} \mathbb{E}[X_{-\infty}|\mathcal{F}_{-\infty}]$.

36. L^p *convergence of reversed martingale*: Suppose $(X_n, \mathcal{F}_n), n \leq 0$ is a reversed martingale. If for some positive p, $\mathbb{E}[|X_0|^p] < \infty$, then $X_n \xrightarrow{L^p} X_{-\infty}$.

37. *Hewitt-Savage 0-1 law*: (c.f. [3] p.224): The exchangable σ-algebra $\mathcal{E} = \cap_{n>0}\mathcal{E}_n$, where

$$\mathcal{E}_n = \sigma(\{A : \ \forall \omega = (\omega_1, \omega_2, ..) \in A, (\omega_{\pi(1)}, .., \omega_{\pi(n)}, \omega_{n+1}..) \in A\})$$

of a sequence of iid random variables $\xi_k(\omega) = \omega_k$ is $\mathbb{P}$-trivial.

38. *De-Finetti*: If $\xi_k(\omega) = \omega_k$ is an exchangable sequence, then conditioned on $\mathcal{E}$, the random variables ξ_k are iid.

7 Markov Chains

7.1 Canonical Construction

1. *Transition kernel* (c.f. [3] p.228): Suppose $\{X_n\}$ is an $\mathcal{F}_n$ Markov chain, and p_n is its n-th state transition kernel. For any bounded measurable function h,

$$\mathbb{E}[h(X_{n+1})|\mathcal{F}_n] = \int_{\mathcal{S}} h(y)p(X_n, dy)$$

2. *Chain rule*: Suppose $\{X_n\}$ is a Markov chain on $(\mathbb{S}, \mathcal{S})$, n-th state transition kernel $p_n(\cdot, \cdot)$ and initial distribution $\nu(A) = \mathbb{P}(X_0 \in A)$. Then for all bounded measurable functions h_l on $\mathcal{S}$ and all $k \in \mathbb{N}$,

$$\mathbb{E}[\prod_{l=0}^{k} h_l(X_l)] = \int h_0(x_0) \int h_1(x_1).. \int h_k(x_k)p_{k-1}(x_{k-1}, dx_k)..p_0(x_0, dx_1)\nu(dx_0)$$

3. *Canonical construction* (c.f. [3] p.230): If $(\mathbb{S}, \mathcal{S})$ is Borel-isomorphic, $\{p_n\}$ a set of transition kernels, and ν a σ-finite measure on $\mathcal{S}$. Then there corresponds a Markov chain X_n with initial distribution ν and transition kernel p_n, such that

$$\mathbb{P}_\nu((X_0, .., X_k) \in A) = \nu \otimes p_0.. \otimes p_{k-1}(A), \forall A \in \mathcal{S}^{k+1}$$

The space $(\mathbb{S}^\infty, \mathcal{S}^\infty, P_\nu)$ is the *canonical measurable space* of the Markov chain X_n where $\forall \omega = (\omega_0, \omega_1, ..) \in \mathbb{S}^\infty, \omega_n = X_n(\omega_0)$.

7.2 Strong Markov Property

4. *strong Markov property*: Suppose $(\mathbb{S}^\infty, \mathcal{S}^\infty, P_\nu)$ is the canonical measurable space, and X_n its corresponding Markov chain. If X_n is homogeneous, for any class of bounded measurable functions $\{h_n\}$ on $\mathcal{S}^\infty$ with $\sup_{n,\omega} |h_n(\omega)| < \infty$,

$$E_\nu[h_\tau(\theta^\tau \omega)|\mathcal{F}_\tau^X]1_{\tau<\infty} = E_{X_\tau}[h_\tau]1_{\tau<\infty}$$

where θ is the left-shift operator and τ is a $\mathcal{F}_n^X$ stopping time.

5. *shift invariance* (c.f. [3] p.234): Suppose ν is a σ-finite measure on $(\mathbb{S}, \mathcal{S})$, and $p_n(\cdot, \cdot)$ are transition kernels. If $\nu \otimes p_0(\mathbb{S} \times A) = \nu(A)$ for all $A \in \mathcal{S}$, then for all $A \in \mathcal{S}^{k+1}$,

$$\nu \otimes p_0 \otimes .. \otimes p_k(\mathbb{S} \times A) = \nu \otimes p_1 \otimes .. \otimes p_k(A)$$

6. A positive σ-finite measure μ on a Borel-isomorphic space $(\mathbb{S}, \mathcal{S})$ is invariant for homogeneous kernels $p(\cdot, \cdot)$ if and only if $\mu \otimes p(\mathbb{S} \times A) = \mu(A)$ for all $A \in \mathcal{S}$.

7.3 Countable State Space Markov Chain

7. *Definitions*:

   (a) x is *accessible* from $y \in \mathbb{S}$ if $\rho_{yx} = P_y(T_x < \infty) > 0$.

   (b) If $x \neq y$ and x, y are accessible from each other, x, y are *intercommunicate*.

   (c) A non empty set $C \subset \mathbb{S}$ is *closed* if $\forall y \in \mathbb{S} - C$, y is not accessible from any $x \in C$.

   (d) A non empty set $C \subset \mathbb{S}$ is *irreducible* if $\forall x, y \in C$, x, y are intercommunicate.

   (e) A state $y \in \mathbb{S}$ is *recurrent* if $\rho_{yy} = 1$, otherwise y is *transient*.

   (f) The *k-th return* T_y^k to state $y \in \mathbb{S}$ is recursively defined as $T_y^k = \inf\{n > T_y^{k-1} : X_n = y\}$ for $k > 0$ and $T_y^0 = 0$.

8. *Harmonic functions on Markov chains*: $f : \mathbb{S} \to \mathbb{R}$ is *(super,sub) harmonic* for a transition probability $p(\cdot, \cdot)$ if $f(x) = \sum p(x, y) f(y)$. When $f(X_0)$ is integrable, $\{f(X_n)\}$ is a (sub, sup) martingale if f is (sub, super) harmonic when f is bounded above or below.

9. *Chapman-Kolmogorov* (c.f. [3] p.235): Suppose X_n is a homogeneous Markov chain with countable state space $\mathbb{S}$, then for any $x, y \in \mathbb{S}$,

$$\mathbb{P}_x(X_n = y) = \sum_{s \in \mathbb{S}} \mathbb{P}_x(X_k = s) \mathbb{P}_s(X_{n-k} = y)$$

10. *Expected visit time*: For any $x, y \in \mathbb{S}$ and $k > 0$,

$$\mathbb{P}_x(T_y^k < \infty) = \rho_{xy} \rho_{yy}^{k-1}$$

Define $N_\infty(y)$ be the expected number of visits to y at finite time, then

$$\mathbb{E}[N_\infty(y)] = \frac{\rho_{xy}}{1-\rho_{yy}}$$

11. *Decomposition of Markov chain* (c.f. [3] p.239): A countable state space $\mathbb{S}$ of a homogeneous Markov chain can be partitioned uniquely as $\mathbb{S} = \mathbb{T} \cup \mathbf{R}_1 \cup \mathbf{R}_2 \cup ..$, where $\mathbb{T}$ is the set of all transient states and $\mathbf{R}_i$ are disjoint, irreducible closed sets of recurrent states with $\rho_{xy} = 1,\ \forall x, y \in \mathbf{R}_i$.

12. If F is a finite set of transient states then $\mathbb{P}_\nu(X_n \in F\ i.o.) = 0$ for any initial distribution ν. Hence if a finite closed set C contains at least one recurrent state, and if C is also irreducible then C is recurrent.

7.4 Ways to Show Recurrence

13. Suppose $\mathbb{S}$ is irreducible for a Markov chain $\{X_n\}$ and there exists $h : \mathbb{S} \to \mathbb{R}^+$ such that $\exists r > 0, G_r = \{x : h(x) < r\}$ is finite and non-empty, and h is super-harmonic on $\mathbb{S} - G_r$. Then X_n is recurrent.

14. Suppose $\mathbb{S}$ is irreducible for a homogeneous Markov chain X_n. Then X_n is recurrent if and only if the only non-negative super-harmonic functions on $\mathbb{S}$ are constant functions.

7.5 Invariant Measure

15. Suppose X_n is a homogeneous Markov chain, and $T_z = \inf\{k \geq 1 : X_k = z\}$. Then

$$\mu_z(y) = \mathbb{E}_z[\sum_{k=0}^{T_z-1} 1_{X_k=y}]$$

is an excessive measure, *i.e.* $\mu(y) \geq \sum \mu(x)p(x,y), \forall y \in \mathbb{S}$. Moreover if z is recurrent then $\mu_z(\cdot)$ is an invariant measure.

16. The invariant measure on a recurrent and irreducible Markov chain is unique up to a multiplicative constant.

17. If $\mu(\cdot) = c > 0$ is an invariant measure for a homogeneous Markov chain with transition probability $p(\cdot,\cdot)$, then it is *doubly stochastic*, *i.e.* $\sum_x p(x,y) = 1, \forall y$.

18. If $\mu(\cdot)$ is an invariant measure for transition probability $p(\cdot,\cdot)$, then for $\mu(x) \neq 0$, $q(x,y) \stackrel{\Delta}{=} \mu(y)p(y,x)/\mu(x)$ is a transition probability and corresponds to the reversed chain of the original Markov chain.

19. *Kolmogorov cycle condition* (c.f. [3] p.247): An irreducible Markov chain with transition probability $p(\cdot,\cdot)$ is reversible if and only if $p(x,y) > 0$ whenever $p(y,x) > 0$ and

$$\prod_{i=1}^{k} p(x_{i-1}, x_i) = \prod_{i=1}^{k} p(x_i, x_{i-1})$$

for all $k > 2$ and $x_0 = x_k$, in which case $\exists \mu$ a positive measure on $\mathbb{S}$, such that $\mu(x)p(x,y) = \mu(y)p(y,x), \forall x, y \in \mathbb{S}$.

20. *Invariant probability*: If $\pi(\cdot)$ is an invariant probability measure then $z \in \mathbb{S}$ is positive recurrent for all $z, \pi(z) > 0$. Conversely, if π is supported on an irreducible and positive recurrent set $\mathbf{R} \subset \mathbb{S}$, uniquely $\pi(z) = 1/\mathbb{E}_z[T_z], \forall z \in \mathbf{R}$.

21. *Second law of thermodynamics* (c.f. [5] p.81):

    (a) Suppose μ, ν are two initial distributions of a homogeneous Markov chain X_n with transition probability $\mathbb{P}$. Let μ_n, ν_n be the measure of n-th coodinate of $\mu \otimes p^n, \nu \otimes p^n$. Then $D(\mu_n||\nu_n) \geq D(\mu_{n+1}||\nu_{n+1})$.

    (b) Suppose X_n admits an invariant measure π. Then for any starting distribution μ, $D(\mu_n||\pi) \geq D(\mu_{n+1}||\pi)$.

7.6 Aperiodic Markov Chains

22. *Asymptotic occupation time*: For any initial distribution ν and all $y \in \mathbb{S}$,

$$\lim_{n\to\infty} \frac{1}{n} N_n(y) \stackrel{P_\nu - a.c.}{=} \frac{1}{E_y[T_y]} 1_{\{T_y < \infty\}}$$

where $N_n(y) = \sum_{k=1}^{n} 1_{\{X_k = y\}}$. Moreover, for all $x, y \in \mathbb{S}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}_x(X_k = y) = \frac{\rho_{xy}}{E_y[T_y]}$$

23. $\mathcal{I}_x = \{n \geq 1 : \mathbb{P}_x(X_n = x) > 0\}$ contains all large enough integer multiples of $d_x = gcd(\mathcal{I}_x)$ and if x, y intercommunicates, $d_x = d_y$.

24. *Coupling of independent chains* (c.f. [3] p.253): If X_n, Y_n are two copies of an aperiodic irreducible Markov chain, and further suppose $Z_n = (X_n, Y_n)$ is recurrent. Then $\tau = \min\{l \geq 0 : X_l = Y_l\}$ is finite a.c. regardless of the initial distributions (μ, ν), and

    $$||\mu_n - \nu_n||_{TV} \leq 2\mathbb{P}(\tau > n)$$

    Remark: This conclusion is stronger than the 2nd law of thermodynamics. If one can show the convergence of KL divergence then the coupling theorem is concluded via Pinsker's inequality.

25. If X_n is irreducible, positive recurrent and aperiodic, then for any $x \in \mathbb{S}$,

    $$\lim_{n\to\infty} ||\mathbb{P}_x(X_n \in \cdot) - \pi(\cdot)||_{TV} = 0$$

8 Stochastic Processes

1. *Cylindrical sets and Borel sets*: Let $\mathbb{R}^{[0,\infty)}$ be the set of all functions from $[0,\infty) \to \mathbb{R}$. $\mathcal{B}(\mathbb{R}^{[0,\infty)})$ is the Borel sets generated by the basic open sets

$$\{f \in \mathbb{R}^{[0,\infty)} : f(x_1) \in U_1, .., f(x_n) \in U_n, \forall x_1, .., x_n, \forall U_1, .., U_n \textit{ open}\}$$

2. *Kolmogorov consistency theorem*: A family of measures $\{Q_{\mathbf{t}}\}$ is consistent if:

   (a) $\mathbf{t} = (t_1, .., t_n)$, and $\mathbf{s} = \pi(\mathbf{t})$ a permutation of $\mathbf{t}$, Then $\forall A_1, .., A_n \in \mathcal{B}(\mathbb{R})$,
   $$Q_{\mathbf{t}}(A_1 \times .. \times A_n) = Q_{\mathbf{s}}(A_{\pi(1)} \times .. \times A_{\pi(n)})$$

   (b) $\mathbf{t} = (t_1, .., t_n), \mathbf{s} = (t_1, .., t_{n-1})$, then
   $$Q_{\mathbf{s}}(A) = Q_{\mathbf{t}}(A \times \mathbb{R}),\ \forall A \in \mathcal{B}(\mathbb{R}^{n-1})$$

   Then there exists a measure on $\mathbb{R}^{[0,\infty)}$, such that

   $$\mathbb{P}(\{\omega \in \mathbb{R}^{[0,\infty)} : (\omega(t_1), .., \omega(t_n)) \in A\}) = Q_{\mathbf{t}}(A), \forall n, \mathbf{t}, A \in \mathcal{B}(\mathbb{R}^n)$$

3. *Kolmogorov-Chentsov*: Suppose that $X_t : \Omega \to \mathbb{R}^{[0,\infty)}$ is a stochastic process. If there exists positive α, β, C, such that

   $$\mathbb{E}[|X_t - X_s|^\alpha] < C|t-s|^{1+\beta}, \forall 0 \le s < t \le T < \infty$$

   Then there exists a continuous stochastic process $\tilde{X}_t : \Omega \to C[0,T]$, such that $\tilde{X}_t$ is measurable, and for any $t \in [0,T], \mathbb{P}(\tilde{X}_t = X_t) = 1$.

4. *Lévy process*: For any infinitely divisible distribution μ, there exists a random process Y_t, which is almost certainly Càdlàg , *i.e.* has left limit and is right continuous, with independent and stationary increment $Y_t - Y_s$ distributed according to μ.

5. *Wiener-Khinchin*: Suppose X_t is a wide sense stationary process. Then its power spectrum density

   $$S(\omega) = \lim_{T\to\infty} \mathbb{E}[|\hat{x}_T(\omega)|^2],\ \hat{x}_T(\omega) = \frac{1}{\sqrt{T}} \int_0^T X_t e^{-i\omega t} dt$$

   equals $S(\omega) = \int_{-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau} d\tau$, where $\gamma(\tau) = \langle X_t, X_{t+\tau} \rangle = \mathbb{E}[X_t X_{t+\tau}^*]$.

9 Brownian Motion

B_t, without further clarification, denotes a standard 1-dimensional Brownian motion starting at 0.

9.1 General Properties

1. *f.d.d. of BM*: Suppose B_t is a standard Brownian motion. For any $0 \le t_1 \le t_2 \le ... \le t_n$, $(B_{t_1}, ..., B_{t_n})$ is multivariate Gaussian, with $\mathbb{E}[B_{t_i}] = 0, Cov(B_{t_i}, B_{t_j}) = \min(t_i, t_j)$.

2. *Brownian filtration*: Let $\mathcal{F}_t^0 = \sigma(\{B_s : s \le t\})$. Define $\mathcal{F}_t^+ = \cap_{s>t} \mathcal{F}_s^0$. Then $\mathcal{F}_t^+$ is a right continuous filtration, *i.e.* $\mathcal{F}_t^+ = \cap_{s>t} \mathcal{F}_s^+$, and B_t is measurable on $\mathcal{F}_t^+$.

3. Scaling and time inversion of B_t:

   (a) $W_t'(\omega) = c^{-1} B_{c^2 t}(\omega)$

   (b)
   $$W_t''(\omega) = \begin{cases} tB_{\frac{1}{t}}(\omega), t > 0 \\ 0, t < 0 \end{cases} \tag{3}$$

   Both W_t' and W_t'' are Brownian motions.

4. *Blumenthal's 0-1 law*: If $A \in \mathcal{F}_0^+$, then $\mathbb{P}(A)$ is either 0 or 1.

5. *Donsker's Invariance Principle* (c.f. [4] p.134): Let $X_1, X_2, ...$ be iid random variables with $\mathbb{E}[X_1] = 0$, Var(X_1)=1. Let $S_n = \sum_{i=1}^n X_i$. Define a random function $W^{(n)} : [0, 1] \to \mathbb{R}$ by
   $$W^{(n)}(\frac{k}{n}) = \frac{S_k}{\sqrt{n}}, k = 0, 1, .., n$$
   and linear interpolation in between. Then $W^{(n)} \Rightarrow B$ where B is a standard Brownian motion on $[0, 1]$, *i.e.* $\mu_n(A) = \mathbb{P}(W^{(n)} \in A)$, $\mu_n \Rightarrow \mu$ where μ is the Wiener measure on $C[0, 1]$.

9.2 Path Regularity

6. Almost certaily B_t has no interval of increase of decrease.

7. *Nowhere differentiability* (c.f. [6] p.306): B_t is nowhere differentiable a.c.

8. *Local maxima* (c.f. [4] p.46): The set of local maxima is almost certainly dense and countable.

9. *Zero set* (c.f. [4] p.52): Let $Z = \{t \geq 0 : B_t = 0\}$. Then almost certainly, Z is a perfect set, *i.e.* Z is closed with no isolated points.

10. *Hölder continuity* (c.f. [4] p.30): For any $\alpha < 1/2$, B_t is almost certainly locally α-*Hölder* continuous, which means for any $t \geq 0$, there exists $\epsilon > 0, c > 0$ such that

$$|B_s - B_t| < c|s-t|^\alpha, y \in \mathbb{R}^+ \cap (x-\epsilon, x+\epsilon)$$

11. *Lower bound of growth* (c.f. [4] p.32): Almost certainly,

$$\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} = +\infty, \liminf_{n\to\infty} \frac{B_n}{\sqrt{n}} = -\infty$$

12. *Quadratic Variation* (c.f. [4] p.35): Suppose there is a nested sequence of partition $0 = t_0^{(n)} \leq ... \leq t_{k_n}^{(n)} = t$ with mesh size $\sup\{t_j^{(n)} - t_{j-1}^{(n)}\}$ going to 0. Then almost certainly,

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} (B_{t_j} - B_{t_{j-1}})^2 = t$$

Remark: This implies almost certainly B_t has unbounded variation.

9.2.1 Dimension

Definitions:

(a) *Hausdorff content*: Given a metric space E and a covering $E_1, E_2, .., E_k$. The α-Hausdorff content of E is defined as:

$$\mathcal{H}_\infty^\alpha(E) = \inf\{\sum |Ei|^\alpha : E_1, E_2, .. \text{ a covering of } E\}$$

Remark: If $\alpha \leq \beta$, then $\mathcal{H}_\infty^\alpha = 0 \Rightarrow \mathcal{H}_\infty^\beta = 0$.

(b) *α-Hausdorff measure*: for any fixed $\delta > 0$,

$$\mathcal{H}^{\alpha}_{\delta}(E) = \inf\{\sum |E_i|^{\alpha} : E_1, E_2, .. \ cover\ E\ and\ |E_i| \leq \delta, \forall i\}$$

The α-Hausdorff measure of E is

$$\mathcal{H}^{\alpha}(E) = \lim_{\delta \to 0} \mathcal{H}^{\alpha}_{\delta}(E) = \sup_{\delta > 0} \mathcal{H}^{\alpha}_{\delta}(E)$$

Remark: $\mathcal{H}^{\alpha}(E)$ is either 0 or ∞.

(c) *Hausdorff Dimension*: The Hausdorff dimension of E is

$$dim(E) = \inf\{\alpha \geq 0 : \mathcal{H}^{\alpha}_{\infty}(E) = 0\} = \sup\{\alpha \geq 0 : \mathcal{H}^{\alpha}_{\infty} > 0\}$$

Remark: $dim(E) = \inf\{\alpha \geq 0 : \mathcal{H}^{\alpha}(E) = 0\} = \sup\{\alpha \geq 0 : \mathcal{H}^{\alpha}(E) = \infty\}$.

13. *Dimension of BM*: Almost certainly $\mathcal{H}^2(B[0,\infty)) = 0$. In particular, $dim(B[0,\infty)) = 2$ for d-dimensional Brownian motion, $d \geq 2$.

9.3 Maximum Process

14. *1-D distribution*: Suppose $M_t = \max_{0 \leq s \leq t} B_s$, for any t, $M_t \overset{d}{=} |N_t|$, where N_t follows Gaussian $N(0,t)$.

15. *Joint distribution of BM and Maximum process* (c.f. [8] p.10): Suppose B_t is a Brownian motion and M_t is its maximum process. Then

$$\mathbb{P}(X_t \leq x, M_t \leq y) = \Phi(\frac{x}{\sqrt{t}}) - \Phi(\frac{x - 2y}{\sqrt{t}})$$

where $\Phi(\cdot)$ is the CDF of a standard Gaussian.

16. (c.f. [7] p.73): Supose B_t is a Brownian with drift $\mu < 0$, and M_t corresponds to its maximum process. Then $M = \lim_{t \to \infty} M_t$ is finite a.c., and has exponential distribution with parameter 2μ.

9.4 Martingale Property

17. Both B_t and $B_t^2 - t$ are continuous martingales with respect to $\mathcal{F}_t^+$.

18. *Optional stopping*: Suppose X_t is a right-continuous martingale, and T a stopping time such that there exists c, $T \leq c$ almost certainly. Moreover, if $\mathbb{E}[\sup_{0\leq t\leq c+1}|X_t|] < \infty$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

19. *Wald's Lemma*: Let B_t be a standard Brownian motion, and T is a stopping time. If $\mathbb{E}[T] < \infty$, then $\mathbb{E}[B_T] = 0, \mathbb{E}[T] = \mathbb{E}[B_T^2]$.

9.4.1 Exponential Martingale and Girsanov Theorem

20. *Exponential martingale*: Suppose B_t is a Brownian motion with drift μ and variance σ. Then $V_t^\theta = e^{\theta B_t - (\mu\theta + \frac{1}{2}\sigma^2\theta^2)t}$ is a martingale for any $\theta \in \mathbb{R}$.

21. *Change of measure*: Suppose B_t is a Brownian motion with drift μ and variance σ. B_t is a Brownian motion with drift $\mu + \theta$ and variance σ under new measure $P^* : d\mathbb{P}^* = V_t^{(\theta/\sigma^2)} d\mathbb{P}$.

9.5 Stopping Times

22. T is a *stopping time* if $\{T < t\} \in \mathcal{F}_t^+, \forall t \in [0, +\infty)$. Equivalently, by right continuity, T is a stopping time if $\{T \leq t\} \in \mathcal{F}_t^+, \forall t \in [0, +\infty)$

23. *Strong Markov Property*: Suppose T is a stopping time. Let $W_t = B_{T+t} - B_T, \forall t \geq 0$. Then W_t is a Brownian motion and independent of $\mathcal{F}_T^+$ the stopped σ-algebra.

24. *Skorohod embedding*: If X is a random variable, with $\mathbb{E}[X] = 0, \mathbb{E}[X^2] < \infty$. Then there exists two random variables (U, V), $U < 0, V > 0$, which are independent of B_t such that if $T = \inf\{t : B_t \notin (U, V)\}$, then $X \overset{d}{=} B_T, \mathbb{E}[T] = \mathbb{E}[X^2]$.

25. *KMT embedding*: If $X_1, X_2, ..$ are iid random variables with mean 0 and variance 1. Moreover, $\mathbb{E}[e^{\theta|X_1|}] < \infty$ for some positive θ. Let $S_n = \sum_{i=1}^n X_i$, then there exists constants C, k, λ depending only on the distributions of X_1 such that the following is true:
    For any n, there is a Brownian motion constructed on the same space (expand if necessary) such that for any $x > 0, \mathbb{P}(\max|S_k - B_k| > C\log n + x) < ke^{-\lambda x}$.

9.6 Distributions

26. *Hitting 0 in an interval*: The probability that a standard Brownian motion hits zero in the interval $[s,t]$ is $\frac{2}{\pi}\arccos(\sqrt{\frac{s}{t}})$.

27. *Arcsin law of last zero*: Let L_t be the time of the last zero of a standard Brownian motion. Then L_t is arcsin distributed, *i.e.* $\mathbb{P}(L_t < s) = 1 - \frac{2}{\pi}\arccos(\sqrt{\frac{s}{t}}) = \frac{2}{\pi}\arcsin(\sqrt{\frac{s}{t}})$.

28. *Arcsin law*: Suppose $X(\omega) = \lambda(\{t \in [0,1] : B_t > 0\}) = \int_0^1 1_{\{B_s>0\}}(\omega)ds$ is the Lebesgue measure of time a Brownian motion $B_t(\omega)$ spends above 0. Then X is arcsin distributed, *i.e.* $\mathbb{P}(X \leq x) = \frac{2}{\pi}\arcsin(x)$.

9.6.1 Hitting Times

29. *Hitting time I*: Suppose B_t a standard Brownian motion, and $T_a = \inf\{t \geq 0 : B_t = a\}$. Then T_a is finite a.c. and follows inverse Gaussian distribution with density

$$f(x) = \frac{|a|e^{-a^2/(2t)}}{\sqrt{2\pi t^3}}$$

30. *Hitting time II*: Let $T = \sup\{t \geq 0 : B_t = t\}$, where B_t is a standard Brownian motion. Then T is chi-square distributed with one degree of freedom, *i.e.* it has density $f(x) = \frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-x/2}$.

31. *With drift*: Suppose B_t is a Brownian motion with drift μ and variance σ, and T_y be the first hitting time of y. Then

$$\mathbb{P}(T_y > t) = \Phi(\frac{y - \mu t}{\sigma t^{\frac{1}{2}}}) - e^{-\frac{2\mu y}{\sigma^2}}\Phi(\frac{-y - \mu t}{\sigma t^{\frac{1}{2}}})$$

32. *Planar BM* (c.f. [7] p.108): Suppose $B_t = (B_t^{(1)}, B_t^{(2)})$ is a 2-D Brownian motion starting at the origin. For any $a > 0$, let $\tau = \inf\{t \geq 0 : B_t^{(1)} = a\}$. Then $B_\tau^{(2)}$ follows Cauchy distribution with density $f(x) = \frac{a}{\pi(a^2+x^2)}$

9.7 Characterizations

33. *Lévy's characterization*: Suppose X_t is a continuous stochastic process such that X_t and $X_t^2 - t$ are both martingales adapted to $\mathcal{F}_t^+$. Then X_t is a Brownian motion with no drift.

34. *Quadratic variation* (c.f. [8] p.7): Suppose X_t adapted to $\mathcal{F}_t^+$ is continuous such that X_t is a martingale and X_t has quadratic variation t on $[0, t]$. Then X_t is a standard Brownian motion.

35. *Exponential martingale* (c.f. [8] p.7): Suppose X_t adapted to $\mathcal{F}_t^+$ is continuous. If
$$V_\beta(t) = e^{\beta X_t - (\beta\mu t + \frac{1}{2}\beta^2\sigma^2 t)}$$
is a martingale for any $\beta \in \mathbb{R}$, then X_t is a Brownian motion with drift μ and variance $\sigma^2 t$.

36. *Characterization function*: If X_t is a process adapted to $\mathcal{F}_s^+$, then X_t is a Brownian motion if and only if for any $0 < s < t$, the conditional expectation
$$\mathbb{E}[e^{iu(W_t - W_s)}|\mathcal{F}_s^+] = e^{-\frac{u^2(t-s)}{2}}$$

37. *martingale representation theorem*: Suppose X_t is a continuous L^2 martingale adapted to $\mathcal{F}_t^+$. Then there exists an adapted process f_t such that for any $t, X_t = \int_0^t f_s dB_s$.

38. *Ito representation*: For any $t > 0$, if X is measurable on $\mathcal{F}_t^+$, and $\mathbb{E}[X^2] < \infty$. Then there exists an adapted process $f_s, 0 \leq s \leq t$, such that $X = \mathbb{E}[X] + \int_0^t f_s dB_s$.

9.8 PDE

39. *heat equation* (c.f. [4] p.207): Suppose $u = u(x, t)$ such that $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$ and with initial condition $u(x, 0) = f(x)$. Then $u(x, t) = \mathbb{E}_x[f(B_t)]$ solves this PDE.

40. *Feynman-Kac* (c.f. [4] p.207): If $V : \mathbb{R} \to \mathbb{R}$ is a bounded measurable function, $f : \mathbb{R} \to \mathbb{R}$ is bounded and coutinuous. Define
$$u(x, t) = \mathbb{E}_x[f(B_t)e^{\int_0^t V(B_s)ds}]$$

Then we have

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2} + V(x)u(x,t)$$

and

$$\lim_{t\to 0, x\to x_0} u(x,t) = f(x_0)$$

Conversely, if u satisfies the above two equations and u is twice differentiable on $\mathbb{R}\times(0,+\infty)$, such that its first order partial derivatives are bounded on $\mathbb{R}\times(0,+\infty), \forall t>0$. Then u must have the form

$$u(x,t) = \mathbb{E}_x[f(B_t)e^{\int_0^t V(B_s)ds}]$$

41. *Ornstein-Uhlenbeck process*: If X_t is the Ornstein-Uhlenbeck process starting at x, *i.e.* $X_t = e^{-t}x + e^{-t}B_{e^{2t}-1}$, and $f:\mathbb{R}\to\mathbb{R}$ is in C^∞ with bounded derivatives. Define $u(x,t) = \mathbb{E}_x[f(X_t)]$, then we have

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - x\frac{\partial u}{\partial t}$$

and

$$u(x,0) = f(x)$$

9.9 Harmonic Functions

42. *Harmonic function*: A *Domain* in $\mathbb{R}^d$ is an open connected set. $f: U\to\mathbb{R}$ is *harmonic* if f is twice differentiable and $\Delta f = \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2} = 0$ on U.

43. *Mean value property*: If U is a domain and u is a measurable and locally bounded function on U. Then the following statements are equivalent:

    (a) u is harmonic;

    (b) For any balls contained in U, $u(x) = \frac{1}{Vol(B_r(x))}\int_{B_r(x)} u(y)dy$;

    (c) For any balls contained in U, $u(x) = \frac{1}{\sigma(\partial B_r(x))}\int_{\partial B_r(x)} u(y)d\sigma(y)$.

44. *Maximum principle*: U is a domain and u on U is harmonic. If u attains maximum in U then u must be a constant. Moreover, if u extends continuously to $\bar{U}$ and U is bounded, then u attains maximum on ∂U.

9.9.1 Dirichlet Problem

45. *Poincare Cone condition*: A domain U satisfies the Poincare Cone condition if for any $z \in \partial U$, there exists a cone C_z at z of nonzero volumn, such that for some $r > 0$, $B_r(z) \cap C_z \subset U^c$.

46. *Dirichlet problem* (c.f. [4] p.73): Let U be a domain, $\varphi : \partial U \to \mathbb{R}$ is measurable. Let

$$u(x) = \mathbb{E}_x[\varphi(B_\tau)1_{\tau<\infty}],\ where\ \tau = \inf\{t : B_t \in \partial U\}$$

Then u is harmonic. Moreover if φ is continuous and U is bounded satisfying the Poincare Cone condition, then $u \to \varphi$ on the boundary.

9.9.2 Recurrence of Brownian Motions

47. Fix $0 < r < R$, define $U = \{x \in \mathbb{R}^d : r < |x| < R\}$ be an annulus. Consider

$$u(x) = \begin{cases} |x|, d = 1 \\ \log|x|, d = 2 \\ |x|^{2-d}, d \geq 3 \end{cases} \tag{4}$$

Then u is harmonic in U.

48. *First hitting time*: (c.f. [4] p.76): Suppose B_t is a d-dimensional Brownian motion started at $x \in U = \{x \in \mathbb{R}^d : r < |x| < R\}$, and T_r, T_R the first hitting times of the inner and outer boundary. Then

$$\mathbb{P}(T_r < T_R) = \begin{cases} \frac{R-|x|}{R-r}, d = 1 \\ \frac{\log R - \log|x|}{\log R - \log r}, d = 2 \\ \frac{R^{2-d}-|x|^{2-d}}{R^{2-d}-r^{2-d}}, d \geq 3 \end{cases} \tag{5}$$

49. The d-dimensional Brownian motion B_t is
    (a) *point recurrent* if $d = 1$;
    (b) *neighborhood recurrent* if $d = 2$;
    (c) *transient* if $d \geq 3$.

9.10 Local Time

50. *Dimension of zero set*: Almost certainly the zero set $Z = \{s \in [0,t) : B_s = 0\}$ is of Hausdorff dimension 1/2. And the 1/2-Hausdorff measure $\mathcal{H}^{\frac{1}{2}}(Z) = 0$.

51. The Brownian local time at 0 is $L_t^0 = \lim_{\epsilon \to 0} \int_0^t 1_{\{-\epsilon \le B_s \le \epsilon\}} ds$. This limit exists and has the same law as $M_t = \max_{0 \le s \le t} B_s$. Moreover, $(|B_t|, L_t^0) \stackrel{d}{=} (M_t - B_t, M_t)$.

52. *Tanaka* (c.f. [7] p.222): If $W_t = \int_0^t sgn(B_s) dB_s$, then W_t is a standard Brownian motion. Moreover, $|B_t| = W_t + L_t^0$ and $L_t^0 = \tilde{M}_t = \max_{0 \le s \le t}(-W_s)$.
Remark:

    (a) The first conclusion is by Lévy's construction of BM,

    (b) The first equation is by Ito's formula on $f_\epsilon(B_t)$ where $f'_\epsilon(\cdot)$ is a continuous estimation of the Heavyside step function, and

    (c) The second equation is by the first equation and increasing properties of L_t^0 and $\tilde{M}_t$.

53. *Ray-Knight I* (c.f. [4] p.164): Suppose B_t is a standard Brownian motion and $T_a = \inf\{t \ge 0 : B_t = a\}$. Then the process $L_T^{a-t} \stackrel{d}{=} |W_t|^2, t \in [0,a]$ where W_t is a 2-D standard Brownian motion.

54. *Ray-Knight II* (c.f. [7] p.456): Let $T_a = \inf\{t \ge 0 : L_t^0 > a\}$. Then the processes $L_{T_a}^t + W_t^2 \stackrel{d}{=} (W_t + \sqrt{a})^2, \forall t \ge 0$, where W_t is a 1-D standard Brownian motion.

10 Stochastic Integration

1. *Existence of Solution*: For a SDE of the form

$$dY_t = f(t, Y_t)dt + g(t, Y_t)dB_t$$

i.e. find Y_t such that $Y_t = Y_0 + \int_0^t f(s, Y_s)ds + \int_0^t g(s, Y_s)dB_s$. A unique continuous solution in $C[0, T]$ exists if there exists $L > 0$, such that

$$\forall t, x, y, |f(t, x) - f(t, y)| + |g(t, x) - g(t, y)| \leq L|x - y|$$

And the solution is given by the L^2 limit of

$$Y_t^{(n)} = x + \int_0^t f(s, Y_s^{(n-1)})ds + \int_0^t g(s, Y_s^{(n-1)})dB_s$$

starting at $Y_t^{(0)} = x$.

2. *Geometric Brownian motion*: The solution to the SDE

$$dY_t = \mu Y_t dt + \sigma Y_t dB_t$$

is $Y_t = Y_0 e^{(\mu - \frac{1}{2}\sigma^2)t} e^{\sigma B_t}$.

3. *Bessel Process*: The solution to the SDE

$$dY_t = dB_t + \frac{n-1}{2Y_t}dt$$

is $Y_t = ||W_t||$ where W_t is a n-dimensional Brownian motion.

10.1 Formulae

4. (c.f. [4] p.189): Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuous, $t > 0$ and s_n is a mesh going to zero. Then

$$\sum_{i=0}^{n-1} f(B_{s_i})(B_{s_{i+1}} - B_{s_i})^2 \to \sum_{i=0}^{n-1} f(B_{s_i})(s_{i+1} - s_i) \to \int_0^t f(B_s)ds$$

5. *Ito's lemma*: Suppose X_t is a drift-diffusion process, *i.e.*

$$dX_t = \mu_t dt + \sigma_t dB_t$$

Then for any twice differentiable function $f(t, x)$,

$$df(t, X_t) = (\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2})dt + \sigma_t \frac{\partial f}{\partial x} dB_t$$

6. *Ito's formula* (c.f. [4] p.189): Suppose $f \in C^\infty$ and all derivatives bounded. Then

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$$

Also if $f = f(t, x)$,

$$f(t, B_t) = f(0, B_0) + \int_0^t \frac{\partial f}{\partial s}(s, B_s)ds + \int_0^t \frac{\partial f}{\partial x}(s, B_s)dB_s + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_s)ds$$

7. *General Ito lemma*:

$$df(t, B_t) = \frac{\partial f}{\partial t}(t, B_t)dt + \frac{\partial f}{\partial x}(t, B_t)dB_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, B_t)dt$$

And for square integrable martingale X_t,

$$df(t, X_t) = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)d\langle X_t\rangle$$

where $\langle X_t\rangle$ is the quadratic variation of X_t.

8. *Isometry*: Suppose f_s is a L^2 stochastic process, then

$$\mathbb{E}[(\int_0^t f_s dB_s)^2] = \int_0^t \mathbb{E}[f_s^2]ds$$

9. *Generalized Ito's Formula*: If $V_t = f(U_t)$, then

$$dV_t = f'(U_t)dU_t + \frac{1}{2}f''(U_t)(dU_t)^2$$

11 List of References

[1] Probability and Measure, *Patrick Billingsley*
[2] Convergence of Probability Measures, *Patrick Billingsley*
[3] Lecture notes, *Amir Dembo*
[4] Brownian Motion, *Peter Mörters, Yuval Peres*
[5] Elements of Information Theory, *Thomas M. Cover*
[6] Probability-Theory and Examples, *Rick Durrett*
[7] Continuous Martingales and Brownian Motion, *Daniel Revuz, Marc Yor*
[8] Brownian Models of Performance and Control, *Michael J. Harrison*
[9] Network Information Theory, *Abbas El Gamal, Young Han Kim*