Probability — Worked Exercises
in Preparation for the Qualifying Exam

T.-Y. Li (kellty@pku.edu.cn)

References

[1] Durrett, R. *Probability: Theory and Examples*

[2] Williams, D. *Probability with Martingales*

[3] Walsh, J.B. *Knowing the Odds: An Introduction to Probability*

[4] Varadhan, S.R.S. *Probability Theory*

[5] Gut, A. *Probability: A Graduate Course*

[6] Chung, K.L. *A Course in Probability Theory*

[7] Chow, Y.S. & Teicher, H. *Probability Theory: Independence, Interchangeability, Martingales*

[8] Shiryaev, A.N. *Problems in Probability*

[9] Chaumont, L. & Yor, M. *Exercises in Probability: A Guided Tour from Measure Theory to Random Processes, via Conditioning*

[10] Kallenberg, O. *Foundations of Modern Probability*

[11] Brémaud, P. *Discrete Probability Models and Methods: Probability on Graphs and Trees, Markov Chains and Random Fields, Entropy and Coding*

0 Courses

1. https://www.math.cmu.edu/users/ttkocz/teaching/1920/prob-grad-notes.pdf
2. https://math.mit.edu/~nsun/f19-18675.html
3. https://www.stat.berkeley.edu/~aldous/205A
4. https://statweb.stanford.edu/~adembo/stat-310b
5. https://people.math.wisc.edu/~roch/grad-prob
6. https://www.math.ucla.edu/~biskup/275c.1.21s
7. http://www.statslab.cam.ac.uk/~james/Lectures/ap.pdf
8. http://www.hairer.org/notes/Markov.pdf
9. https://www.mat.univie.ac.at/~bruin/ET3.pdf
10. https://people.bath.ac.uk/maspm/bm11.html
11. http://www.math.leidenuniv.nl/~spieksma/SPspring08.html
12. http://www.stat.cmu.edu/~cshalizi/almost-none
13. https://user.eng.umd.edu/~abarg/MDP

Contents

0 Courses **-1**

1 Basic tools **1**
1.1 Best constant approximation . . . 1
1.2 Integration — layer cake representation . . . 1
1.3 Generalized second Borel–Cantelli lemma . . . 2
1.4 Equality contained in conditional expectation . . . 2
1.5 Correlation inequality and independent copies . . . 2
1.6 Taking advantage of characteristic functions . . . 3
1.7 Inversion formula for point masses . . . 4

2 Stochastic convergence **4**
2.1 Convergence in probability from the perspective of metrics . . . 4
2.2 Convergence of random series — Lévy's equivalence theorem . . . 5
2.3 Series of nonnegative random variables . . . 5
2.4 Converse of strong law of large numbers . . . 6
2.5 Asymptotic behavior of Gaussian maxima . . . 6
2.6 Law of iterated logarithm . . . 7

3 Martingales **7**
3.1 Switch at a stopping time . . . 7
3.2 Optimal stopping with finite horizon . . . 8
3.3 Martingales derived from differentiation . . . 9
3.4 Strong law of large numbers . . . 9

4 Markov chains **9**
4.1 First passage decomposition . . . 9
4.2 Number of visits . . . 9
4.3 Superharmonicity and recurrence . . . 10
4.4 Green's function — potential theory . . . 11

5 Stationary sequences **12**
5.1 Invariant sets and functions . . . 12
5.2 Criteria for ergodicity . . . 12

6 Brownian motion **12**
6.1 Chaining and continuous modification of stochastic process . . . 12
6.2 Nonsmoothness of sample path . . . 14
6.3 Reflection principle and arcsine law . . . 14

1 Basic tools

1.1 Best constant approximation

Let m be a median of X, i.e., $\mathbb{P}\{X \le m\} \ge 1/2$ and $\mathbb{P}\{X \ge m\} \ge 1/2$.

1. $m \in \arg\min_x \mathbb{E}|X - x|$.

   Proof. If $a < b$, then $|X-b| - |X-a| = \begin{cases} (b-a)(1 - 2\cdot \mathbb{1}_{\{X \ge b\}}) + 2(a-X)\mathbb{1}_{\{a<X<b\}} \\ (b-a)(2\cdot \mathbb{1}_{\{X\le a\}} - 1) + 2(b-X)\mathbb{1}_{\{a<X<b\}} \end{cases}$. This implies that $x \mapsto \mathbb{E}|X-x|$ is nonincreasing on $(-\infty, m]$ and is nondecreasing on $[m, \infty)$. □

2. $|\mathbb{E}X - m| \le \sqrt{\operatorname{Var}(X)}$.

   Proof. $|\mathbb{E}X - m| \le \mathbb{E}|X-m| \le \mathbb{E}|X-x| \le \sqrt{\mathbb{E}|X-x|^2}$, $\forall x$. □

3. $\mathbb{E}X = \arg\min_x \mathbb{E}|X-x|^2$.

   Proof. $\mathbb{E}|X-x|^2 = \mathbb{E}|X - \mathbb{E}X - (x - \mathbb{E}X)|^2 = \mathbb{E}|X-\mathbb{E}X|^2 + |x - \mathbb{E}X|^2$. □

1.2 Integration — layer cake representation

1. (Integrability) Let $X \ge 0$. Then $\mathbb{E}X < \infty$ if and only if $\sum \mathbb{P}\{X > n\} < \infty$.

   Proof. Note that $\mathbb{E}X = \int_0^\infty \mathbb{P}\{X > x\}\,\mathrm{d}x = \sum_{n=1}^\infty \int_{n-1}^n \mathbb{P}\{X > x\}\,\mathrm{d}x$, where for $n-1 \le x \le n$ one has

   $$\mathbb{P}\{X > n\} \le \mathbb{P}\{X > x\} \le \mathbb{P}\{X > n-1\}.$$

   Therefore, $\sum_{n=1}^\infty \mathbb{P}\{X > n\} \le \mathbb{E}X \le 1 + \sum_{n=1}^\infty \mathbb{P}\{X > n\}$. □

2. (Exponential decay) If $\mathbb{P}\{X > x\} = O(q^x)$ as $x \to \infty$ for some $q \in (0,1)$, then $\mathbb{E}\,\mathrm{e}^{tX} < \infty$ for some $t > 0$. The converse is also true, due to Chernoff's bound.

   Proof. Since $\mathbb{P}\{\mathrm{e}^{tX} > n\} = \mathbb{P}\{X > \log(n)/t\} \lesssim q^{\log(n)/t} = n^{\log(q)/t}$, it suffices that $t < -\log(q)$. □

3. If $X \perp\!\!\!\perp Y$, then $\mathbb{E}|X+Y| - \mathbb{E}|X-Y| = 2\int_0^\infty \big(\mathbb{P}\{X > u\} - \mathbb{P}\{X < -u\}\big)\big(\mathbb{P}\{Y > u\} - \mathbb{P}\{Y < -u\}\big)\mathrm{d}u$. (Shepp) If X and Y are i.i.d., then $\mathbb{E}|X+Y| \ge \mathbb{E}|X-Y|$, with equality holding if and only if $X \overset{d}{=} -X$.

   Proof. To begin with, denote by P and Q the distributions of X and Y, respectively. Then

   $$\begin{aligned} \mathbb{E}(X+Y)^+ = \int_0^\infty \mathbb{P}\{X+Y > t\}\,\mathrm{d}t &= \int_{\mathbb{R}^3} \mathbb{1}_{[t>0]}\mathbb{1}_{[x+y>t]}\,\mathrm{d}P(x)\mathrm{d}Q(y)\mathrm{d}t \\ &= \int_{\mathbb{R}^3} \mathbb{1}_{[y>-u]}\mathbb{1}_{[x>u]}\,\mathrm{d}P(x)\mathrm{d}Q(y)\mathrm{d}u \\ &= \int_{-\infty}^\infty \mathbb{P}\{X > u\}\mathbb{P}\{Y > -u\}\,\mathrm{d}u \\ &= \int_0^\infty \big(\mathbb{P}\{X>u\}\mathbb{P}\{Y>-u\} + \mathbb{P}\{X > -u\}\mathbb{P}\{Y > u\}\big)\mathrm{d}u. \end{aligned}$$

   Similarly, $\mathbb{E}(X+Y)^- = \int_0^\infty \mathbb{P}\{X+Y < -t\}\,\mathrm{d}t = \int_0^\infty \big(\mathbb{P}\{X<u\}\mathbb{P}\{Y<-u\} + \mathbb{P}\{X<-u\}\mathbb{P}\{Y<u\}\big)\mathrm{d}u$. By symmetry, $\begin{cases} \mathbb{E}(X-Y)^+ = \int_0^\infty \big(\mathbb{P}\{X>u\}\mathbb{P}\{Y<u\} + \mathbb{P}\{X>-u\}\mathbb{P}\{Y<-u\}\big)\mathrm{d}u, \\ \mathbb{E}(X-Y)^- = \int_0^\infty \big(\mathbb{P}\{X<u\}\mathbb{P}\{Y>u\} + \mathbb{P}\{X<-u\}\mathbb{P}\{Y>-u\}\big)\mathrm{d}u. \end{cases}$ Therefore,

   $$\begin{aligned} \mathbb{E}|X+Y| - \mathbb{E}|X-Y| &= \big(\mathbb{E}(X+Y)^+ + \mathbb{E}(X+Y)^-\big) - \big(\mathbb{E}(X-Y)^+ + \mathbb{E}(X-Y)^-\big) \\ &= \big(\mathbb{E}(X+Y)^+ - \mathbb{E}(X-Y)^+\big) - \big(\mathbb{E}(X-Y)^- - \mathbb{E}(X+Y)^-\big) \\ &= \int_0^\infty \big(\mathbb{P}\{X>u\} + \mathbb{P}\{X>-u\}\big)\big(\mathbb{P}\{Y>u\} - \mathbb{P}\{Y<-u\}\big)\mathrm{d}u \\ &\quad - \int_0^\infty \big(\mathbb{P}\{X<u\} + \mathbb{P}\{X<-u\}\big)\big(\mathbb{P}\{Y>u\} - \mathbb{P}\{Y<-u\}\big)\mathrm{d}u \\ &= 2\int_0^\infty \big(\mathbb{P}\{X>u\} - \mathbb{P}\{X<-u\}\big)\big(\mathbb{P}\{Y>u\} - \mathbb{P}\{Y<-u\}\big)\mathrm{d}u, \end{aligned}$$

   since $\mathbb{1}_{(-u,\infty)} - \mathbb{1}_{(-\infty,u)} = \mathbb{1}_{(u,\infty)} - \mathbb{1}_{(-\infty,-u)}$. □

1.3 Generalized second Borel–Cantelli lemma

1. (Paley–Zygmund) If $X \geq 0$ with $0 < \mathbb{E}X < \infty$, then $\mathbb{P}\{X > t\,\mathbb{E}X\} \geq (1-t)^2(\mathbb{E}X)^2/\,\mathbb{E}X^2$, $\forall\, t \in [0,1]$.

   Proof. $\mathbb{E}X = \mathbb{E}X\mathbb{1}_{\{X>t\,\mathbb{E}X\}} + \mathbb{E}X\mathbb{1}_{\{X\leq t\,\mathbb{E}X\}} \leq \sqrt{\mathbb{E}X^2\,\mathbb{P}\{X>t\,\mathbb{E}X\}} + t\,\mathbb{E}X$. □

2. (Chung–Erdős) If $A_1,\ldots,A_n$ are events, then $\mathbb{P}(\bigcup_{k=1}^n A_k) \geq [\sum_{k=1}^n \mathbb{P}(A_k)]^2/\sum_{i,j=1}^n \mathbb{P}(A_i\cap A_j)$.

   Proof. Apply the Paley–Zygmund inequality to $X = \sum_{k=1}^n \mathbb{1}_{A_k}$ with $t=0$. □

3. (Kochen–Stone) If $\sum_{k=1}^\infty \mathbb{P}(A_k) = \infty$, then $\mathbb{P}(A_n \text{ i.o.}) \geq \limsup_{n\to\infty}[\sum_{k=1}^n \mathbb{P}(A_k)]^2/\sum_{i,j=1}^n \mathbb{P}(A_i\cap A_j)$.

   First Proof. Let $x_n = [\sum_{k=1}^n \mathbb{P}(A_k)]^2$ and $y_n = \sum_{i,j=1}^n \mathbb{P}(A_i\cap A_j)$. By the Chung–Erdős inequality, we have $y_n \geq x_n \to \infty$ as $n\to\infty$, and then using the fact that $\sum_{i,j=m+1}^n \mathbb{P}(A_i\cap A_j) \leq y_n - y_m$,

   $$\mathbb{P}\Big(\bigcup_{k=m+1}^\infty A_k\Big) = \lim_{n\to\infty}\mathbb{P}\Big(\bigcup_{k=m+1}^n A_k\Big) \geq \limsup_{n\to\infty}\frac{(\sqrt{x_n}-\sqrt{x_m})^2}{y_n-y_m} = \limsup_{n\to\infty}\frac{x_n}{y_n}.$$

   Letting $m\to\infty$ completes the proof. □

   Second Proof. Let $X_n = \sum_{k=1}^n \mathbb{1}_{A_k}$ and $Y_n = X_n/\mathbb{E}X_n$. Then $\{A_n \text{ i.o.}\} = \{\lim X_n = \infty\} \supset \{Y_n > t \text{ i.o.}\}$ for any $t\in(0,1)$, since $\lim \mathbb{E}X_n = \infty$. Therefore,

   $$\mathbb{P}(A_n \text{ i.o.}) \geq \lim_{n\to\infty}\mathbb{P}(\textstyle\bigcup_{k=n}^\infty\{Y_k>t\}) \geq \limsup_{n\to\infty}\mathbb{P}\{Y_n>t\},$$

   where $\mathbb{P}\{Y_n>t\} \geq (1-t)^2/\mathbb{E}Y_n^2 = (1-t)^2(\mathbb{E}X_n)^2/\,\mathbb{E}X_n^2$ by the Paley–Zygmund inequality. □

1.4 Equality contained in conditional expectation

Let X and Y be integrable random variables.

1. If $X \overset{d}{=} Y = \mathbb{E}[X|\mathscr{G}]$, then $X \overset{\text{a.s.}}{=} Y$.

   Proof. First, consider the special case when X and Y are square integrable. Since $\mathbb{E}X^2 = \mathbb{E}Y^2 = \mathbb{E}\,XY$, we have $\mathbb{E}(X-Y)^2 = 0$ and thus $X \overset{\text{a.s.}}{=} Y$. For the general case, we will show that

   $$a\vee X\wedge b \overset{\text{a.s.}}{=} a\vee Y\wedge b,$$

   and conclude by letting $a\searrow -\infty$ and $b\nearrow\infty$. By Jensen's inequality, $\mathbb{E}[a\vee X\,|\,\mathscr{G}] \geq a\vee Y$, where the equality must hold for $\mathbb{E}\,a\vee X = \mathbb{E}\,a\vee Y$. Finally, $a\vee X\wedge b \overset{d}{=} a\vee Y\wedge b = \mathbb{E}[a\vee X\wedge b\,|\,\mathscr{G}]$. □

2. If $\mathbb{E}[X|Y] = Y$ and $\mathbb{E}[Y|X] = X$, then $X \overset{\text{a.s.}}{=} Y$.

   Proof. Let $h:\mathbb{R}\to\mathbb{R}$ be bounded and strictly increasing, e.g., $h = \arctan$. Since

   $$\{X\neq Y\} = \{(X-Y)\big(h(X)-h(Y)\big) > 0\},$$

   it suffices to show $\mathbb{E}\,(X-Y)\big(h(X)-h(Y)\big) = 0$. To see this, $\mathbb{E}\,Yh(X) = \mathbb{E}\{\mathbb{E}[Y|X]h(X)\} = \mathbb{E}\,Xh(X)$. □

1.5 Correlation inequality and independent copies

1. (Harris–FKG / Chebyshev–Kimball) Let $f,g:\mathbb{R}^n\to\mathbb{R}$ be nondecreasing functions, and $\boldsymbol{X} = (X_1,\ldots,X_n)$ be a random vector with independent coordinates. Then $\mathbb{E}\,f(\boldsymbol{X})g(\boldsymbol{X}) \geq \mathbb{E}f(\boldsymbol{X})\,\mathbb{E}g(\boldsymbol{X})$.

   Proof. First, consider the case $n=1$. Let X' be an independent copy of X. Taking the expectation of

   $$\big(f(X)-f(X')\big)\big(g(X)-g(X')\big) \geq 0$$

leads to the desired result. Then we perform induction on n. Define $f_1(x_1) = \mathbb{E}[f(\boldsymbol{X})|X_1 = x_1]$ and $g_1(x_1) = \mathbb{E}[g(\boldsymbol{X})|X_1 = x_1]$, which preserves monotonicity. It follows from the inductive hypothesis that

$$\mathbb{E}[f(\boldsymbol{X})g(\boldsymbol{X})|X_1] \ge f_1(X_1)g_1(X_1),$$

where $\mathbb{E}\, f_1(X_1)g_1(X_1) \ge \mathbb{E}f_1(X_1)\,\mathbb{E}g_1(X_1) = \mathbb{E}f(\boldsymbol{X})\,\mathbb{E}g(\boldsymbol{X})$. □

2. (Kac) If $\mathbb{E}\,\mathrm{e}^{\sqrt{-1}(sX+tY)} = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}sX}\,\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tY}$ for any s and t, then X and Y are independent.

Proof. Let ξ and η be independent random variables such that $\xi \overset{d}{=} X$ and $\eta \overset{d}{=} Y$. We have

$$\mathbb{E}\,\mathrm{e}^{\sqrt{-1}(sX+tY)} = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}sX}\,\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tY} = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}s\xi}\,\mathbb{E}\,\mathrm{e}^{\sqrt{-1}t\eta} = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}(s\xi+t\eta)},$$

and thus $(X,Y) \overset{d}{=} (\xi,\eta)$ by the uniqueness of characteristic functions. □

3. If ϕ is a characteristic function, then so are ϕ^2, $|\phi|^2$, and $\operatorname{Re}\phi$.

Proof. Suppose that $\phi(t) = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tX}$ for some random variable X. Let X' be an independent copy of X. Then $\phi(t)^2 = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}t(X+X')}$ and $|\phi(t)|^2 = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}t(X-X')}$. Let $Y = X\mathbb{1}_{\{U=1\}} - X'\mathbb{1}_{\{U=0\}}$ for $U \sim \mathrm{Bernoulli}(1/2)$ independent of $\{X, X'\}$. Then $\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tY} = \frac{1}{2}(\phi(t)+\phi(-t)) = \operatorname{Re}\phi(t)$. □

1.6 Taking advantage of characteristic functions

Given a random variable X, denote $F_X(x) = \mathbb{P}\{X \le x\}$ and $\phi_X(t) = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tX} = \mathbb{E}\cos(tX) + \sqrt{-1}\,\mathbb{E}\sin(tX)$.

1. (Constancy and independence) In each of the following cases, X is almost surely a constant:
(a) $|\phi_X| \equiv 1$; (b) $X \perp\!\!\!\perp X$; (c) $X \perp\!\!\!\perp Y$ and $X+Y$ is a constant.

Proof. By the uniqueness of characteristic functions, it suffices that $\phi_X \equiv 1$.
(a) For every $t \in \mathbb{R}$, note that $|\phi_X(t)|^2 = [\mathbb{E}\cos(tX)]^2 + [\mathbb{E}\sin(tX)]^2 \le \mathbb{E}\cos^2(tX) + \mathbb{E}\sin^2(tX) = 1$ with equality holding only if $\cos(tX) \overset{\text{a.s.}}{=} c_t$ and $\sin(tX) \overset{\text{a.s.}}{=} s_t$ for some constants c_t and s_t, which means that

$$tX \in (\pm\arccos(c_t) + 2\pi\mathbb{Z}) \cap (\{\arcsin(s_t),\, \pi - \arcsin(s_t)\} + 2\pi\mathbb{Z}).$$

Then let t varies. (b)&(c) can be reduced to (a). □

2. (Second moment) $\frac{11}{24}\mathbb{E}[X^2; |X| < \frac{1}{t}] \le \frac{1}{t^2}(1 - \operatorname{Re}\phi_X(t))$, $\forall t > 0$. It follows that $\mathbb{E}X^2 < \infty$ if $\phi_X''(0)$ exists.

Proof. Note that $1 - \cos u \ge \frac{u^2}{2} - \frac{u^4}{24}$, so $\int_{-\infty}^{\infty}(1-\cos(tx))\,\mathrm{d}F_X(x) \ge \int_{-1/t}^{1/t}(\frac{1}{2} - \frac{t^2x^2}{24})t^2x^2\,\mathrm{d}F_X(x)$ where $\frac{1}{2} - \frac{t^2x^2}{24} \ge \frac{11}{24}$. As $t \to 0$, we have $1 - \operatorname{Re}\phi_X(t) = -\frac{1}{2}(\phi_X(t) + \phi_X(-t) - 2\phi_X(0)) \sim -\frac{1}{2}\phi_X''(0)t^2$. □

If $\phi_X(t) = 1 - ct^2 + o(t^2)$ as $t \to 0$ for some constant $c \in \mathbb{R}$, then $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 2c$. In particular, $X \overset{\text{a.s.}}{=} 0$ if $\phi_X(t) = 1 + o(t^2)$. As a corollary, $\phi(t) = \mathrm{e}^{-|t|^\alpha}$ is *not* a characteristic function for any $\alpha > 2$.

Proof. We have $\phi_X'(t) = \sqrt{-1}\,\mathbb{E}\,X\mathrm{e}^{\sqrt{-1}tX}$ and $\phi_X''(t) = -\mathbb{E}\,X^2\mathrm{e}^{\sqrt{-1}tX}$. Then put $t = 0$. □

3. $\mathbb{E}|X|^r = K_r\int_{-\infty}^{\infty}\frac{1-\operatorname{Re}\phi_X(t)}{|t|^{r+1}}\,\mathrm{d}t$ for $r \in (0,2)$, where K_r is a constant only depending on r.
(Shepp) If X and Y are i.i.d., then $\mathbb{E}|X+Y|^r \ge \mathbb{E}|X-Y|^r$, with equality holding if and only if $X \overset{d}{=} -X$.

Proof. Let $K_r = 1\big/\int_{-\infty}^{\infty}\frac{1-\cos u}{|u|^{r+1}}\,\mathrm{d}u$, which can be shown to be $\frac{\Gamma(r+1)}{\pi}\sin\frac{r\pi}{2}$. Then

$$|x|^r = K_r\int_{-\infty}^{\infty}\frac{1-\cos(xt)}{|t|^{r+1}}\,\mathrm{d}t,$$

and thus $\mathbb{E}|X|^r = \int_{-\infty}^{\infty}|x|^r\,\mathrm{d}F_X(x)$ can be evaluated by Fubini's theorem. Based on such a formula, Shepp's inequality follows from the fact that $1 - \operatorname{Re}\phi_X^2 \ge 1 - |\phi_X|^2$, with equality holding if and only if $\phi_X^2 \ge 0$ if and only if ϕ_X is real-valued if and only if $X \overset{d}{=} -X$. □

See `https://artofproblemsolving.com/wiki/index.php/2021_IMO_Problems/Problem_2` for fun.

1.7 Inversion formula for point masses

1. Let $\hat{\mu}(t) = \int_{\mathbb{R}} \mathrm{e}^{\sqrt{-1}tx}\,\mathrm{d}\mu(x)$ for μ a probability measure on $\mathbb{R}$. Then $\mu\{a\} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{-\sqrt{-1}at}\hat{\mu}(t)\,\mathrm{d}t$.

   Proof. Fix $a \in \mathbb{R}$. By Fubini's theorem, $\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{-\sqrt{-1}at}\hat{\mu}(t)\,\mathrm{d}t = \int_{\mathbb{R}} \big(\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{\sqrt{-1}(x-a)t}\,\mathrm{d}t\big)\mathrm{d}\mu(x)$, where $\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{\sqrt{-1}(x-a)t}\,\mathrm{d}t = \frac{1}{2T}\int_{-T}^{T}\cos((x-a)t)\,\mathrm{d}t \to \mathbb{1}_{[x=a]}$ and the dominated convergence applies. ☐

2. If $X \sim P$, $Y \sim Q$, and $X \perp\!\!\!\perp Y$, then $\mathbb{P}\{X = Y\} = \sum_x P\{x\}Q\{x\}$.
   Note that $P\{x\} > 0$ for at most countably many x.

   Proof. $\mathbb{P}\{X = Y\} = \mathbb{E}\,\mathbb{1}_{\{X=Y\}} = \iint \mathbb{1}_{[x=y]}\,\mathrm{d}P(x)\,\mathrm{d}Q(y) = \sum_x (P\otimes Q)(\{x\}\times\{x\}) = \sum_x P\{x\}Q\{x\}$. ☐

3. Let $\phi_X(t) = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tX}$. Then $\mathbb{P}\{X = x\} = 0$ for all x if and only if $\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\phi_X(t)|^2\,\mathrm{d}t = 0$.

   Proof. Consider μ to be the distribution of $X - X'$, where X' is an independent copy of X. Combining the previous results, $\sum_x \mathbb{P}\{X = x\}^2 = \mathbb{P}\{X - X' = 0\} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\phi_X(t)|^2\,\mathrm{d}t$. ☐

   Therefore, the distribution of X has no point mass if $\phi_X(t) \to 0$ as $t \to \infty$, which can be derived by the Riemann–Lebesgue lemma when a probability density function exists. However, the converse is false, e.g., $2\sum_{k=1}^{\infty} X_k/3^k$ has the Cantor distribution if $X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim}$ Bernoulli(1/2), whose characteristic function is given by $t \mapsto \prod_{k=1}^{\infty}\frac{1}{2}(1 + \mathrm{e}^{2\sqrt{-1}t/3^k})$ and has the same value on $\{3^n\pi\}_{n=0}^{\infty}$.

2 Stochastic convergence

2.1 Convergence in probability from the perspective of metrics

The Ky Fan metric is defined as $\alpha(X,Y) = \inf\big\{\varepsilon > 0 : \mathbb{P}\{|X-Y| > \varepsilon\} \le \varepsilon\big\}$ for random variables X and Y. Also, introduce $\beta(X,Y) = \mathbb{E}\frac{|X-Y|}{1+|X-Y|}$ and $\gamma(X,Y) = \mathbb{E}\min\{|X-Y|,\,1\}$.

1. (Triangle inequality) $\alpha(X,Z) \le \alpha(X,Y) + \alpha(Y,Z)$.

   Proof. $\mathbb{P}\{|X-Z| > \varepsilon_1 + \varepsilon_2\} \le \mathbb{P}\{|X-Y| > \varepsilon_1\} + \mathbb{P}\{|Y-Z| > \varepsilon_2\}$. ☐

   One can check that α,β,γ are metrics indeed.

2. (Equivalence) $\alpha^2/(1+\alpha) \le \beta \le 2\alpha/(1+\alpha)$ and (trivially) $\beta \le \gamma \le 2\beta$.

   Proof. Write $\alpha = \alpha(X,Y)$, $\beta = \beta(X,Y)$, and $T = |X-Y|$. On one hand, $\beta \ge \frac{\varepsilon}{1+\varepsilon}\mathbb{P}\{T > \varepsilon\} \xrightarrow{\varepsilon\to\alpha} \frac{\alpha^2}{1+\alpha}$. On the other hand, $\beta = \int_0^1 \mathbb{P}\{\frac{T}{1+T} > u\}\,\mathrm{d}u = \int_0^\infty \mathbb{P}\{T > t\}\frac{\mathrm{d}t}{(1+t)^2} \le \int_0^\alpha \frac{\mathrm{d}t}{(1+t)^2} + \int_\alpha^\infty \alpha\frac{\mathrm{d}t}{(1+t)^2} = \frac{2\alpha}{1+\alpha}$. ☐

3. $X_n \xrightarrow{\mathbb{P}} X \iff \gamma(X_n, X) \to 0 \iff \beta(X_n, X) \to 0 \iff \alpha(X_n, X) \to 0$.

   Proof. $|X_n - X| \xrightarrow{\mathbb{P}} 0 \iff \min\{|X_n - X|,\,1\} \xrightarrow{\mathbb{P}} 0 \iff \mathbb{E}\min\{|X_n - X|,\,1\} \to 0$. ☐

4. (Uniqueness) If $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$, then $X \overset{\text{a.s.}}{=} Y$.

5. If $\{X_n\}$ is Cauchy in that $\mathbb{P}\{|X_n - X_m| > \varepsilon\} \xrightarrow{m,n\to\infty} 0$ ($\forall\varepsilon > 0$), then X_n converges in probability.

   Proof. Now that $\alpha(X_n, X_m) \xrightarrow{m,n\to\infty} 0$, we choose $\{n_j\}$ such that $\sup_{m>n_j}\mathbb{P}\{|X_m - X_{n_j}| > 2^{-j}\} \le 2^{-j}$. Then $A_j = \{|X_{n_{j+1}} - X_{n_j}| > 2^{-j}\}$ satisfy that $\sum\mathbb{P}(A_j) \le \sum 2^{-j} < \infty$, so the first Borel–Cantelli lemma implies that $A = \{A_j \text{ i.o.}\}$ occurs with probability zero. Next we restrict ourselves to $A^{\complement}$, on which $\lim_{k\to\infty}\sum_{j=k}^{\infty}|X_{n_{j+1}} - X_{n_j}| \le \lim_{k\to\infty}\sum_{j=k}^{\infty}2^{-j} = 0$ and thus $\lim_{k\to\infty}X_{n_k} = X_{n_1} + \sum_{j=1}^{\infty}(X_{n_{j+1}} - X_{n_j})$ exists and is finite. Finally, it must hold that X_n converges to $X = \limsup_{k\to\infty}X_{n_k}$ in probability, as $\{|X_n - X| > \varepsilon\} \subset \{|X_n - X_{n_k}| > \varepsilon/2\} \cup \{|X_{n_k} - X| > \varepsilon/2\}$. ☐

2.2 Convergence of random series — Lévy's equivalence theorem

Let $S_n = \sum_{i=1}^n X_i$ where X_i's are independent random variables.

1. (Ottaviani–Skorokhod) It holds for $\lambda > 0$ and $\mu > 0$ that

$$\mathbb{P}\Big\{\max_{m<j\le n}|S_j - S_m| > \lambda + \mu\Big\} \min_{m<k\le n} \mathbb{P}\{|S_n - S_k| \le \mu\} \le \mathbb{P}\{|S_n - S_m| > \lambda\}.$$

Proof. Note that $\{|S_n - S_m| > \lambda\} \supset \bigcup_{k=m+1}^n \big(\{\inf\{j > m : |S_j - S_m| > \lambda + \mu\} = k\} \cap \{|S_n - S_k| \le \mu\}\big)$. □

2. (Etemadi) $\mathbb{P}\big\{\max_{m<j\le n}|S_j - S_m| > 3\lambda\big\} \le 2\,\mathbb{P}\{|S_n - S_m| > \lambda\} + \max_{m<k\le n} \mathbb{P}\{|S_k - S_m| > \lambda\}$, $\forall \lambda > 0$.

Proof. From 1, $\mathbb{P}\big\{\max_{m<j\le n}|S_j - S_m| > 3\lambda\big\} - \mathbb{P}\{|S_n - S_m| > \lambda\} \le \max_{m<k\le n} \mathbb{P}\{|S_n - S_k| > 2\lambda\}$, but $\mathbb{P}\{|S_n - S_k| > 2\lambda\} \le \mathbb{P}\{|S_n - S_m| > \lambda\} + \mathbb{P}\{|S_k - S_m| > \lambda\}$. □

3. If S_n converges in probability, then S_n converges almost surely.

Proof. It follows from $\mathbb{P}\{|S_n - S_m| > \lambda\} \xrightarrow{m,n\to\infty} 0$ that $\mathbb{P}\big\{\max_{m<j\le n}|S_j - S_m| > 3\lambda\big\} \xrightarrow{m,n\to\infty} 0$ by Etemadi's inequality. Then

$$\mathbb{P}\big\{\sup_{j,k>m}|S_j - S_k| > 6\lambda\big\} \xrightarrow{m\to\infty} 0.$$

However, $\sup_{j,k>m}|S_j - S_k|$ decreases with m and thus admits a pointwise limit Z. The uniqueness of the limit in probability forces that $Z \overset{\text{a.s.}}{=} 0$, whence $\{S_n\}_{n=1}^\infty$ is a Cauchy sequence and converges. □

4. If S_n converges in distribution, then S_n converges in probability.

Proof. Since any Cauchy sequence in probability is convergent in probability, it suffices that $Y_j = S_{n_j} - S_{m_j}$ converges to zero in probability, or equivalently $Y_j \xrightarrow{d} 0$, for all sequences $\{n_j\}$ and $\{m_j\}$ with $n_j > m_j$. For $|t|$ small enough,

$$\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tS_{m_j}}\,\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tY_j} = \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tS_{n_j}},$$

where $\lim_{j\to\infty} \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tS_{m_j}} = \lim_{j\to\infty} \mathbb{E}\,\mathrm{e}^{\sqrt{-1}tS_{n_j}}$ is nonzero. Hence, $\mathbb{E}\,\mathrm{e}^{\sqrt{-1}tY_j} \to 1$ for t in a neighborhood of 0. We then conclude by Lévy's continuity theorem. □

2.3 Series of nonnegative random variables

Let $S_n = \sum_{i=1}^n X_i$ where $X_i \ge 0$ are independent. Then $S_n \nearrow S_\infty$.

1. Kolmogorov's zero–one law ensures that $\{\sum_{n=1}^\infty X_n < \infty\}$ is $\mathbb{P}$-trivial. The following are equivalent: (a) $\sum_{n=1}^\infty X_n < \infty$ a.s.; (b) $\sum_{n=1}^\infty \big(\mathbb{P}\{X_n > 1\} + \mathbb{E}[X_n; X_n \le 1]\big) < \infty$; (c) $\sum_{n=1}^\infty \mathbb{E}\frac{X_n}{1+X_n} < \infty$.

Proof. By Kolmogorov's three-series theorem, (a) $\iff \sum_{n=1}^\infty \big[\mathbb{P}\{X_n > 1\} + \mathbb{E}Y_n + \operatorname{Var}(Y_n)\big] < \infty$, where $Y_n = X_n \mathbb{1}_{\{X_n \le 1\}}$. Since $\operatorname{Var}(Y_n) \le \mathbb{E}Y_n^2$ and $Y_n^2 \le Y_n$, we obtain that (a) $\iff$ (b). As for (b) $\iff$ (c), note that $\frac{1}{2}\big(\mathbb{P}\{X_n > 1\} + \mathbb{E}[X_n; X_n \le 1]\big) \le \mathbb{E}\frac{X_n}{1+X_n} < \mathbb{P}\{X_n > 1\} + \mathbb{E}[X_n; X_n \le 1]$. □

2. (Chi-squares) Suppose $\sqrt{X_n} \sim \mathcal{N}(\mu_n, \sigma_n^2)$. In other words, $X_n = (\mu_n + \sigma_n Z_n)^2$ where $Z_n \sim \mathcal{N}(0,1)$.

   (a) If S_n converges in L^1, then $\sum_{n=1}^\infty (\mu_n^2 + \sigma_n^2) < \infty$.

   Proof. $\mathbb{E}S_\infty = \sum_{n=1}^\infty (\mu_n^2 + \sigma_n^2)$. □

   (b) If $\sum_{n=1}^\infty (\mu_n^2 + \sigma_n^2) < \infty$, then S_n converges in L^p for any $p \in [1, \infty)$.

   Proof. $\sum \|X_n\|_{L^p} \le \sum \big(\mu_n^2 + 2|\mu_n\sigma_n| \|Z_n\|_{L^p} + \sigma_n^2 \|Z_n^2\|_{L^p}\big)$ where $2|\mu_n\sigma_n| \le \mu_n^2 + \sigma_n^2$. □

3. A useful fact is that $S_\infty \overset{\text{a.s.}}{=} \infty \iff 0 = \mathbb{E}\,\mathrm{e}^{-S_\infty} = \prod_{n=1}^\infty \mathbb{E}\,\mathrm{e}^{-X_n}$. Also, $S_\infty < \infty$ a.s. if $\mathbb{E}S_\infty < \infty$.

(a) Suppose $\sqrt{X_n} \sim \mathcal{N}(0, \sigma_n^2)$. Then $S_\infty \overset{\text{a.s.}}{=} \infty \iff \sum_{n=1}^\infty \sigma_n^2 = \infty$.

Proof. $\prod \mathbb{E}\,\mathrm{e}^{-X_n} = \prod \mathbb{E}\,\mathrm{e}^{-\sigma_n^2 Z_n^2} = \prod(1+2\sigma_n^2)^{-1/2}$ where $\prod(1+2\sigma_n^2) \geq 1 + 2\sum \sigma_n^2$. □

(b) Suppose X_n is exponentially distributed with rate λ_n. Then $S_\infty \overset{\text{a.s.}}{=} \infty \iff \sum_{n=1}^\infty \frac{1}{\lambda_n} = \infty$.

Proof. $1/\prod \mathbb{E}\,\mathrm{e}^{-X_n} = 1/\prod \frac{\lambda_n}{\lambda_n+1} = \prod(1+\frac{1}{\lambda_n}) \geq 1 + \sum \frac{1}{\lambda_n}$. □

2.4 Converse of strong law of large numbers

Let $X, X_1, X_2, \ldots$ be i.i.d., $S_n = \sum_{i=1}^n X_i$, and $p > 0$.

1. $\mathbb{E}|X|^p < \infty \iff \lim |X_n|^p/n \overset{\text{a.s.}}{=} 0 \iff X_n/n^{1/p} \xrightarrow{\text{a.s.}} 0$.

   Proof. For any $\varepsilon > 0$, we have $\mathbb{E}|X|^p/\varepsilon < \infty \iff \sum \mathbb{P}\{|X|^p > n\varepsilon\} < \infty \iff \mathbb{P}\{|X_n|^p > n\varepsilon \text{ i.o.}\} = 0$ and $\{\limsup |X_n|^p/n > \varepsilon\} \subset \{|X_n|^p > n\varepsilon \text{ i.o.}\} \subset \{\limsup |X_n|^p/n \geq \varepsilon\}$. □

2. If $S_n/n^{1/p} \xrightarrow{\text{a.s.}} 0$ and $p \geq 1$, then $\mathbb{E}|X|^p < \infty$ and $\mathbb{E}X = 0$.

   Proof. Note that $X_n/n^{1/p} = S_n/n^{1/p} - (1-1/n)^{1/p} S_{n-1}/(n-1)^{1/p} \xrightarrow{\text{a.s.}} 0 - 0 = 0$, so $\mathbb{E}|X|^p < \infty$. Since $\mathbb{E}|X| < \infty$, Kolmogorov's SLLN gives $S_n/n \xrightarrow{\text{a.s.}} \mathbb{E}X$. Also, $S_n/n = n^{1/p-1} S_n/n^{1/p} \xrightarrow{\text{a.s.}} 0$. □

2.5 Asymptotic behavior of Gaussian maxima

Let $Z, Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, whose probability density function is $\varphi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$. Let $M_n = \max_{1\leq i\leq n} Z_i$. Since $\mathbb{P}\{M_n \leq z\} = \mathbb{P}\{Z \leq z\}^n$ and $(1-\frac{1}{n})^n \to \mathrm{e}^{-1}$, let $e_-(x) = \mathrm{e}^{-x}$ and $b_n = \inf\{b : \mathbb{P}\{Z > b\} \leq \frac{1}{n}\} \nearrow \infty$.

1. (Mills ratio) $1/z - 1/z^3 < z/(z^2+1) < \mathbb{P}\{Z > z\}/\varphi(z) < 1/z$ for $z > 0$.

   Proof. $\frac{1}{z}\mathrm{e}^{-z^2/2} - \int_z^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \int_z^\infty \frac{1}{u^2}\mathrm{e}^{-u^2/2}\,\mathrm{d}u < \frac{1}{z^2}\int_z^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u$. □

2. $\lim_{z\to\infty} \mathbb{P}\{Z > z + \theta/z\}/\mathbb{P}\{Z > z\} = e_-(\theta)$, $\forall \theta \in \mathbb{R}$.

   Proof. Since $\mathbb{P}\{Z > z\} \sim \frac{1}{z}\varphi(z)$ as $z \to \infty$, we have $\mathbb{P}\{Z > z+\frac{\theta}{z}\}/\mathbb{P}\{Z > z\} \sim \varphi(z+\frac{\theta}{z})/\varphi(z) \sim e_-(\theta)$. □

3. (Extreme value distribution) Let $a_n = 1/b_n = o(1)$. Then $\mathbb{P}\{(M_n - b_n)/a_n \leq x\} \to e_-(e_-(x))$ for $x \in \mathbb{R}$.

   Proof. $\mathbb{P}\{(M_n - b_n)/a_n \leq x\} = \big(1 - \mathbb{P}\{Z > a_n x + b_n\}\big)^n$ where $\mathbb{P}\{Z > a_n x + b_n\} \sim \frac{1}{n}e_-(x)$ using 2. □

   Recall the Fisher–Tippett–Gnedenko theorem.

4. $b_n \sim \sqrt{2\log n}$ and thus $M_n/\sqrt{2\log n} \xrightarrow{\mathbb{P}} 1$.

   Proof. For n large enough, $\mathbb{P}\{Z > \sqrt{2\log n - 2\log v_n}\} \sim \frac{v_n}{\sqrt{4\pi \log n}} \cdot \frac{1}{n}$ if $1 \leq v_n = O(\log n)$. By choosing v_n appropriately,
   $$\sqrt{2\log n - 2\log\log n} \leq b_n \leq \sqrt{2\log n - \log\log n}.$$
   Then $M_n - \sqrt{2\log n} = M_n - b_n + b_n - \sqrt{2\log n} = O_{\mathbb{P}}(a_n) + o(\sqrt{2\log n}) = o_{\mathbb{P}}(\sqrt{2\log n})$. □

5. $\mathbb{E}M_n/\sqrt{2\log n} \to 1$.

   Proof. Jensen's inequality gives $\mathrm{e}^{t\mathbb{E}M_n} \leq \mathbb{E}\mathrm{e}^{tM_n}$ for $t \in \mathbb{R}_+$. But $\mathrm{e}^{tM_n} \leq \sum_{i=1}^n \mathrm{e}^{tZ_i}$, leading to
   $$\mathbb{E}\mathrm{e}^{tM_n} \leq n\mathbb{E}\mathrm{e}^{tZ} = n\mathrm{e}^{t^2/2}.$$
   Thus, $\mathbb{E}M_n \leq \frac{1}{t}\log(n\mathrm{e}^{t^2/2}) = \frac{\log n}{t} + \frac{t}{2}$. We obtain $\mathbb{E}M_n \leq \sqrt{2\log n}$ by optimizing the upper bound over t. As for the lower bound, $0 \leq \mathbb{E}M_n^- \leq \mathbb{E}Z^- = O(1)$, and $\mathbb{E}M_n^+/\sqrt{2\log n} = \int_0^\infty \mathbb{P}\{M_n/\sqrt{2\log n} > u\}\,\mathrm{d}u$ has $\liminf \geq \int_0^\infty \liminf \mathbb{P}\{M_n/\sqrt{2\log n} > u\}\,\mathrm{d}u$ by Fatou's lemma, where $\mathbb{P}\{M_n/\sqrt{2\log n} > u\} \to \mathbb{1}_{[u<1]}$ for almost all u. This shows that $\liminf \mathbb{E}M_n/\sqrt{2\log n} = \liminf \mathbb{E}M_n^+/\sqrt{2\log n} \geq 1$. □

2.6 Law of iterated logarithm

Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a stochastic process with continuous sample paths. Denote $h_t = \sqrt{2t\log\log t}$.

1. (Upper bound derived by sub-Gaussianity) If there exist $0 < v_t = O(t)$ such that $\mathbb{P}\{X_t^* > \lambda\} \lesssim \mathrm{e}^{-\lambda^2/(2v_t)}$ for $\lambda > 0$, then $\limsup_{t\to\infty} X_t^*/h_t \le 1$ a.s., where $X_t^* = \sup_{s\le t} X_s$ is the running maximum.

   Proof. For any $t > \mathrm{e}^{\mathrm{e}}$ and $c > 1$, we have $\mathbb{P}\{X_t^* > c\,h_t\} \lesssim \mathrm{e}^{-c^2(t/v_t)\log\log t} \lesssim (\log t)^{-c^2}$. Choosing $t_n = q^n$ for some $q > 1$, it follows that
   $$\mathbb{P}\{X_{q^n}^* > c\,h_{q^n}\} \lesssim n^{-c^2}.$$
   Since $\sum n^{-c^2} < \infty$, we obtain that $\mathbb{P}\{X_{q^n}^* > c\,h_{q^n} \text{ i.o.}\} = 0$ by the Borel–Cantelli lemma. This implies that $\limsup_{n\to\infty} X_{q^n}^*/h_{q^n} \le c$ a.s.. Note that
   $$X_t^*/h_t \le X_{q^n}^*/h_{q^{n-1}} = (X_{q^n}^*/h_{q^n})(h_{q^n}/h_{q^{n-1}}),\ t \in [q^{n-1}, q^n].$$
   Thus, $\limsup_{t\to\infty} X_t^*/h_t \le c\sqrt{q}$ a.s.. Letting $c \searrow 1$ and $q \searrow 1$ completes the proof. □

2. (Lower bound) If $\limsup_{t\to\infty}(-X_t)/h_t \le 1$ a.s. and $\limsup_{t\to\infty}(X_t - X_{t/q})/h_t \ge \sqrt{(q-1)/q}$ a.s. for any $q > 1$, then $\limsup_{t\to\infty} X_t/h_t \ge 1$ a.s..

   Proof. (a.s.) $\limsup_{t\to\infty}(-X_{t/q})/h_t \le \lim_{t\to\infty} h_{t/q}/h_t = 1/\sqrt{q}$, so $\limsup_{t\to\infty} X_t/h_t \ge (\sqrt{q-1}-1)/\sqrt{q}$, where $(\sqrt{q-1}-1)/\sqrt{q} \to 1$ as $q \to \infty$. □

3 Martingales

See `https://zhuanlan.zhihu.com/p/76804737` for a fast-paced review, whose pdf version is available; see comments therein.

3.1 Switch at a stopping time

1. Let $X = (X_t)_{t\ge0}$ and $Y = (Y_t)_{t\ge0}$ be supermartingales with respect to a filtration $(\mathscr{F}_t)_{t\ge0}$. Suppose τ is a stopping time such that $X_\tau \le Y_\tau$. Define $Z_t = X_t\mathbb{1}_{\{\tau\le t\}} + Y_t\mathbb{1}_{\{\tau>t\}}$ and $W_t = X_t\mathbb{1}_{\{\tau<t\}} + Y_t\mathbb{1}_{\{\tau\ge t\}}$. Then $Z = (Z_t)_{t\ge0}$ and $W = (W_t)_{t\ge0}$ are also an $(\mathscr{F}_t)_{t\ge0}$-supermartingales.

   Proof. Write $\Delta^- X_t = X_t - X_{t-}$ and $\Delta^+ X_t = X_{t+} - X_t$. It can be seen that
   $$\begin{cases}\Delta^- Z_t = \Delta^- X_t\mathbb{1}_{\{\tau<t\}} + \Delta^- Y_t\mathbb{1}_{\{\tau\ge t\}} + (X_\tau - Y_\tau)\mathbb{1}_{\{\tau=t\}}\\ \Delta^+ W_t = \Delta^+ X_t\mathbb{1}_{\{\tau\le t\}} + \Delta^+ Y_t\mathbb{1}_{\{\tau>t\}} + (X_\tau - Y_\tau)\mathbb{1}_{\{\tau=t\}}\end{cases},$$
   so $\begin{cases}\mathbb{E}[\Delta^- Z_t|\mathscr{F}_{t-}] = \mathbb{E}[\Delta^- X_t|\mathscr{F}_{t-}]\mathbb{1}_{\{\tau<t\}} + \mathbb{E}[\Delta^- Y_t|\mathscr{F}_{t-}]\mathbb{1}_{\{\tau\ge t\}} + \mathbb{E}[(X_\tau - Y_\tau)\mathbb{1}_{\{\tau=t\}}|\mathscr{F}_{t-}] \le 0\\ \mathbb{E}[\Delta^+ W_t|\mathscr{F}_t] = \mathbb{E}[\Delta^+ X_t|\mathscr{F}_t]\mathbb{1}_{\{\tau\le t\}} + \mathbb{E}[\Delta^+ Y_t|\mathscr{F}_t]\mathbb{1}_{\{\tau>t\}} + \mathbb{E}[(X_\tau - Y_\tau)\mathbb{1}_{\{\tau=t\}}|\mathscr{F}_t] \le 0\end{cases}$. □

2. (Dubins) Let $X = (X_t)_{t\ge0}$ be a positive supermartingale with respect to a filtration $(\mathscr{F}_t)_{t\ge0}$. Denote by $U^{a,b}$ the number of upcrossings through $[a,b]$ made by $t \mapsto X_t$. Then $\mathbb{P}\{U^{a,b} \ge k\} \le (a/b)^k\mathbb{E}\min\{X_0/a, 1\}$.

   Proof. Let $\tau_0 = 0$ and $\begin{cases}\sigma_j = \inf\{t \ge \tau_{j-1} : X_t \le a\}\\ \tau_j = \inf\{t \ge \sigma_j : X_t \ge b\}\end{cases}$ for $j = 1, 2, \cdots$. Define $W^{(0)} = \min\{X/a, 1\}$ and recursively $\begin{cases}Z_t^{(j)} = W_t^{(j-1)}\mathbb{1}_{\{t<\sigma_j\}} + (b/a)^{j-1}(X_t/a)\mathbb{1}_{\{t\ge\sigma_j\}}\\ W_t^{(j)} = Z_t^{(j)}\mathbb{1}_{\{t\le\tau_j\}} + (b/a)^j\mathbb{1}_{\{t>\tau_j\}}\end{cases}$ so that by 1 they are supermartingales with respect to $\mathscr{F}_t = \sigma(X_s : s \le t)$. In order to bound $\mathbb{P}\{U^{a,b} \ge k\} = \mathbb{P}\{\tau_k < \infty\} = \lim_{t\to\infty}\mathbb{P}\{\tau_k < t\}$, note that $(b/a)^k\mathbb{P}\{\tau_k < t\} \le \mathbb{E}W_t^{(k)} \le \mathbb{E}W_0^{(k)} = \mathbb{E}W_0^{(0)}$. □

3. (Random walk) Let $S_n = \sum_{i=1}^n \varepsilon_i$ with ε_i's taking values in $\{\pm 1\}$. For any $s = (s_n)_{n\ge 0}$, define $\tau_k(s) = \inf\{n : s_n = k\}$ and $\Theta_k(s) = \big(s_n \mathbb{1}_{[n\le \tau_k(s)]} + (2k - s_n)\mathbb{1}_{[n>\tau_k(s)]}\big)_{n\ge 0}$. If S satisfies the reflection principle that $\Theta_k(S) \overset{d}{=} S$ for $k = 0, 1, 2, \cdots$, then S is a symmetric simple random walk.

   Proof. It suffices that $S_{0:n} = (S_0, S_1, \ldots, S_n)$ is uniformly distributed on $\Lambda^n = \{s_{0:n} = (s_0, s_1, \ldots, s_n) : s_0 = 0,\ s_i - s_{i-1} = \pm 1\ (\forall i)\}$. Let s be a possible path with $s_{0:n}$ as its first $(n+1)$ elements. There exist $k_1 < \cdots < k_m$ such that $\Theta_{(s)} = \Theta_{k_m} \circ \cdots \circ \Theta_{k_1}$ transforms s to have $(0, 1, \cdots, n)$ as its first $(n+1)$ elements. Then $\Theta_{(s)}(S) \overset{d}{=} S$, so $\mathbb{P}\{S_{0:n} = s_{0:n}\} = \mathbb{P}\{\Theta_{(s)}(S)_{0:n} = \Theta_{(s)}(s)_{0:n}\} = \mathbb{P}\{S_{0:n} = (0, 1, \cdots, n)\}$. □

4. (Converse of optional stopping theorem) Let $M = (M_t)_{t\ge 0}$ be an integrable stochastic process adapted to a filtration $(\mathscr{F}_t)_{t\ge 0}$. Then M is a martingale if $\mathbb{E}M_\tau = \mathbb{E}M_0$ for every bounded stopping time τ.

   Proof. Let $s < t$. If $A \in \mathscr{F}_s$, then $\tau = s\mathbb{1}_A + t\mathbb{1}_{A^\complement}$ is a stopping time. Thus,

$$0 = \mathbb{E}M_t - \mathbb{E}M_\tau = \mathbb{E}[M_t - M_s; A].$$

   Since A is arbitrary, we conclude that $\mathbb{E}[M_t|\mathscr{F}_s] = M_s$. □

3.2 Optimal stopping with finite horizon

Let $Y = (Y_n)_{n=0,1,\cdots,N}$ be an integrable stochastic process adapted to a filtration $(\mathscr{F}_n)_{n=0,1,\cdots,N}$. Then the Snell envelope $U = (U_n)_{n=0,1,\cdots,N}$ is recursively defined by $U_N = Y_N$ and $U_n = Y_n \vee \mathbb{E}[U_{n+1}|\mathscr{F}_n]$ for $n < N$. Denote by $\mathcal{S}_{t_0}^{t_1}$ the set of stopping times τ with $t_0 \le \tau \le t_1$.

1. U is a supermartingale and $U_n \le X_n$ for all n if X is a supermartingale such that $X_n \ge Y_n$ for all n.

   Proof. First, $X_N \ge Y_N = U_N$. If $X_{n+1} \ge U_{n+1}$, then $X_n \ge \mathbb{E}[X_{n+1}|\mathscr{F}_n] \ge \mathbb{E}[U_{n+1}|\mathscr{F}_n]$, so $X_n \ge U_n$. □

2. (Value function) $\sup_{\tau\in\mathcal{S}_n^N} \mathbb{E}[Y_\tau|\mathscr{F}_n] = U_n = \mathbb{E}[Y_{\tau_n}|\mathscr{F}_n]$, where $\tau_n = \inf\{t \ge n : Y_t = U_t\}$. Consequently, (Bellman equation) $\sup_{\tau\in\mathcal{S}_n^N} \mathbb{E}[Y_\tau|\mathscr{F}_n] = Y_n \vee \mathbb{E}\big[\sup_{\tau\in\mathcal{S}_{n+1}^N} \mathbb{E}[Y_\tau|\mathscr{F}_{n+1}]\,\big|\mathscr{F}_n\big]$ for $n < N$.

   Proof. The statement is trivial for $n = N$. We proceed backwards inductively. If $\tau \in \mathcal{S}_{n-1}^N$, then $\tau \vee n \in \mathcal{S}_n^N$ and thus $\mathbb{E}[Y_{\tau\vee n}|\mathscr{F}_n] \le U_n$. For $Y_\tau = Y_{n-1}\mathbb{1}_{\{\tau=n-1\}} + Y_{\tau\vee n}\mathbb{1}_{\{\tau\ge n\}}$, we have

$$\mathbb{E}[Y_\tau|\mathscr{F}_{n-1}] = Y_{n-1}\mathbb{1}_{\{\tau=n-1\}} + \mathbb{E}[Y_{\tau\vee n}|\mathscr{F}_{n-1}]\mathbb{1}_{\{\tau\ge n\}} \le U_{n-1}\mathbb{1}_{\{\tau=n-1\}} + \mathbb{E}[U_n|\mathscr{F}_{n-1}]\mathbb{1}_{\{\tau\ge n\}} \le U_{n-1},$$

   with equality holding when $\tau = \tau_{n-1}$, since $Y_{\tau_{n-1}} = U_{n-1}\mathbb{1}_{\{\tau_{n-1}=n-1\}} + Y_{\tau_n}\mathbb{1}_{\{\tau_{n-1}\ge n\}}$. □

   Particularly, $\mathbb{E}Y_\tau = \mathbb{E}\{\mathbb{E}[Y_\tau|\mathscr{F}_0]\} \le \mathbb{E}U_0 = \mathbb{E}Y_{\tau_0}$ for any $\tau \in \mathcal{S}_0^N$, and thus $\tau_0 = \arg\max_{\tau\in\mathcal{S}_0^N} \mathbb{E}Y_\tau$. Besides, the stopped supermartingale $U^{\tau_0} = (U_{n\wedge\tau_0})_{n=0,1,\cdots,N}$ is actually a martingale, since $U_{\tau_0} = Y_{\tau_0}$.

3. (Cayley–Moser) Suppose that Y_n's are i.i.d. copies of Y and $\mathscr{F}_n = \sigma(Y_0, Y_1, \ldots, Y_n)$. Then $\mathbb{E}[U_{n+1}|\mathscr{F}_n] = A_{N-n}$ is a constant that depends only on $N - n$. Moreover,

   (a) $A_n = \log\big(n + O(\log n)\big)$ if $Y \sim \text{Exponential}(1)$.

   (b) $A_n = 1 - 2/[n + \log(n) + O(1)]$ if $Y \sim \text{Uniform}(0, 1)$.

   Proof. Put $A_0 = -\infty$. By induction, $\mathbb{E}[U_n|\mathscr{F}_{n-1}] = \mathbb{E}[Y_n \vee A_{N-n}|\mathscr{F}_{n-1}] = \mathbb{E}[Y \vee A_{N-n}]$ since $Y_n \perp\!\!\!\perp \mathscr{F}_{n-1}$. This also leads to the recursion formula $A_{n+1} = \mathbb{E}[Y \vee A_n]$, starting from $A_1 = \mathbb{E}Y$.

   (a) Now $A_{n+1} = A_n + \mathrm{e}^{-A_n}$. Write $A_n = \log(n + x_n)$, then

$$\tfrac{1}{n+x_n} = \mathrm{e}^{-A_n} = A_{n+1} - A_n = \log(1 + \tfrac{1+x_{n+1}-x_n}{n+x_n}).$$

   Using $\frac{u}{1+u} \le \log(1+u) \le u$, we obtain that $0 \le x_{n+1} - x_n \le \frac{1}{n+x_n-1} \lesssim \frac{1}{n}$.

   (b) Now $A_{n+1} = (A_n^2 + 1)/2$. Write $A_n = 1 - 2/(n + x_n)$, then some calculation gives $x_{n+1} - x_n = \frac{1}{n+x_n-1}$, so $x_n \le \log(n) + O(1)$. It follows that $x_n - \log(n) - O(1) \ge \sum_{k=1}^n (\frac{1}{k+\log k} - \frac{1}{k}) \ge -\sum_{k=1}^\infty \frac{\log k}{k^2}$. □

3.3 Martingales derived from differentiation

Let $M(\theta) = (M_t(\theta))_{t\geq 0}$ be a martingale with respect to a filtration $(\mathscr{F}_t)_{t\geq 0}$, for any θ in a neighborhood of 0. If $M_t^{(n)}(\theta) = \frac{\partial^n}{\partial\theta^n} M_t(\theta)$ exists and $\mathbb{E}\sup_\theta |M_t^{(n)}(\theta)| < \infty$ for all t, then $(M_t^{(n)}(0))_{t\geq 0}$ is a martingale.

Proof. For $s < t$, we have $\mathbb{E}\big[\frac{\partial^n}{\partial\theta^n}\big|_0 M_t(\theta) \,\big|\, \mathscr{F}_s\big] = \frac{\partial^n}{\partial\theta^n}\big|_0 \mathbb{E}[M_t(\theta)\,|\,\mathscr{F}_s]$ by the dominated convergence. □

E.g., the exponential martingale of a Brownian motion is associated with Hermite polynomials.

3.4 Strong law of large numbers

Let $X, X_1, X_2, \dots$ be i.i.d. and $S_n = \sum_{i=1}^n X_i$.

1. (Convergence rate) If $\mathbb{E}X^2 < \infty$, then $(S_n - n\,\mathbb{E}X)/a_n \xrightarrow{\text{a.s.}} 0$ for $a_n = n^{1/2}(\log n)^{1/2+\epsilon}$ with $\epsilon > 0$.

   Proof. Let $M_n = \sum_{i=1}^n (X_i - \mathbb{E}X)/a_i$, which is an L^2-martingale adapted to $\mathscr{F}_n = \sigma(X_1, \dots, X_n)$. Using the fact that $\sup \mathbb{E}M_n^2 = \mathrm{Var}(X)\sup\sum_{i=1}^n 1/a_i^2 < \infty$, we obtain the a.s. convergence of M_n. The proof is completed by applying Kronecker's lemma. □

2. (Moment convergence) If $\mathbb{E}|X|^p < \infty$ for some $p \in [1, \infty)$, then $\bar{X}_n \xrightarrow{L^p} \mathbb{E}X$ where $\bar{X}_n = S_n/n$.

   Proof. Let $\mathscr{F}_{-n} = \sigma(\bar{X}_n, X_{n+1}, X_{n+2}, \dots)$, then $\bar{X}_n = \mathbb{E}[X_1|\mathscr{F}_{-n}] \xrightarrow{\text{a.s.}} \mathbb{E}X$. By Vitali's convergence theorem, it suffices that $\{|\bar{X}_n|^p\}_{n\geq 1}$ is uniformly integrable, but $|\bar{X}_n|^p \leq \mathbb{E}[|X_1|^p|\mathscr{F}_{-n}]$. □

4 Markov chains

Suppose throughout this section that $X = (X_0, X_1, X_2, \dots)$ is a homogeneous Markov chain with transition probabilities $\mathbb{P}(X_n = y\,|\,X_0 = x) = p^n(x, y)$ for states x, y. Denote $\mathbb{P}_x = \mathbb{P}(\,\cdot\,|X_0 = x)$ and $\mathbb{E}_x = \mathbb{E}[\,\cdot\,|X_0 = x]$.

4.1 First passage decomposition

Let $T_x = \inf\{n \geq 1 : X_n = x\}$ and $f^n(x, y) = \mathbb{P}_x\{T_y = n\}$.

1. $p^n(x, y) = \sum_{m=1}^n f^m(x, y)\, p^{n-m}(y, y)$ for $n \geq 1$. In other words,
   $P_{xy}(s) = \mathbb{1}_{[x=y]} + F_{xy}(s)P_{yy}(s)$, where $P_{xy}(s) = \sum_{n=0}^\infty p^n(x, y)s^n$ and $F_{xy}(s) = \sum_{n=0}^\infty f^n(x, y)s^n$.

   Proof. $\{X_n = y\} = \bigcup_{m=1}^n \{T_y = m,\, X_n = y\}$ and $\mathbb{P}_x(X_n = \cdot\,|\,T_y = m) = p^{n-m}(y, \cdot)$. □

2. $\mathbb{P}_x\{T_x < \infty\} = 1 - 1/G(x, x)$ where $G(x, x) = \sum_{n=0}^\infty p^n(x, x)$. Hence, $T_x < \infty$ $\mathbb{P}_x$-a.s. $\iff G(x, x) = \infty$.

   Proof. Let $s \nearrow 1$ in $F_{xx}(s) = 1 - 1/P_{xx}(s)$. □

3. $\sum_{n=0}^N p^n(x, x) \geq \sum_{n=k}^{N+k} p^n(x, x)$, $\forall k \geq 1$.

   Proof. Let $T = \inf\{n \geq k : X_n = x\}$, then $p^n(x, x) = \sum_{m=k}^n \mathbb{P}_x\{T = m\}p^{n-m}(x, x)$ for $n \geq k$. It follows that $\sum_{n=k}^{N+k} p^n(x, x) = \sum_{n=k}^{N+k}\sum_{m=k}^n \mathbb{P}_x\{T = m\}\,p^{n-m}(x, x) = \sum_{m=k}^{N+k}\mathbb{P}_x\{T = m\}\sum_{n=m}^{N+k} p^{n-m}(x, x)$, where $\sum_{n=m}^{N+k} p^{n-m}(x, x) \leq \sum_{n=0}^N p^n(x, x)$ and $\sum_{m=k}^{N+k}\mathbb{P}_x\{T = m\} \leq 1$. □

4.2 Number of visits

Let $V_n(x) = \sum_{m=1}^n \mathbb{1}_{\{X_m = x\}}$ and $T_x^{(k)} = \inf\{n > T_x^{(k-1)} : X_n = x\}$, where $T_x^{(1)} = T_x = \inf\{n > 0 : X_n = x\}$. Clearly $V_n(x) = \sum_{k=1}^\infty \mathbb{1}_{\{T_x^{(k)} \leq n\}}$. Assume that X is irreducible and recurrent, so $\mathbb{P}_x\{T_y < \infty\} = 1$, $\forall x, y$.

1. $\mathbb{E}_x V_{T_x}(y) = \frac{\mathbb{P}_x\{T_y < T_x\}}{\mathbb{P}_y\{T_x < T_y\}}$ for $x \neq y$.

   Proof. $\mathbb{E}_x V_{T_x}(y) = \sum_{k=1}^\infty \mathbb{P}_x\{T_y^{(k)} < T_x\} = \sum_{k=1}^\infty \mathbb{P}_x\{T_y < T_x\}\prod_{j=1}^{k-1}\mathbb{P}_x(T_y^{(j+1)} < T_x\,|\,T_y^{(j)} < T_x)$ where $\mathbb{P}_x(T_y^{(j+1)} < T_x\,|\,T_y^{(j)} < T_x) = \mathbb{P}_y\{T_y < T_x\}$ due to the strong Markov property. □

2. $\mathbb{E}_x V_{T_x}(y)\,\mathbb{E}_y V_{T_y}(z) = \mathbb{E}_x V_{T_x}(z)$.

   Proof. Since the stationary distribution is unique up to constant multiples, $\mathbb{E}_x V_{T_x}(\cdot) \propto \mathbb{E}_y V_{T_y}(\cdot)$. □

3. $\frac{V_n(y)}{V_n(z)} \xrightarrow{\mathbb{P}\text{-a.s.}} \frac{1}{\mathbb{E}_y V_{T_y}(z)} = \frac{\mathbb{E}_x V_{T_x}(y)}{\mathbb{E}_x V_{T_x}(z)}$, and thus $\frac{n}{V_n(z)} = \sum_y \frac{V_n(y)}{V_n(z)} \xrightarrow{\mathbb{P}\text{-a.s.}} \sum_y \frac{\mathbb{E}_x V_{T_x}(y)}{\mathbb{E}_x V_{T_x}(z)} = \frac{\mathbb{E}_x T_x}{\mathbb{E}_x V_{T_x}(z)} = \mathbb{E}_z T_z$.

   Proof. If $T_y^{(k)} \le n < T_y^{(k+1)}$, then $\frac{k}{k+1} \cdot \frac{k+1}{V_{T_y^{(k+1)}}(z)} \le \frac{V_n(y)}{V_n(z)} \le \frac{k}{V_{T_y^{(k)}}(z)}$. To conclude, it suffices that $\frac{V_{T_y^{(k)}}(z)}{k} = \frac{1}{k}\big[V_{T_y}(z) + \sum_{j=1}^{k-1}\big(V_{T_y^{(j+1)}}(z) - V_{T_y^{(j)}}(z)\big)\big] \xrightarrow{\text{a.s.}} \mathbb{E}_y V_{T_y}(z)$ by the strong law of large numbers. □

4. $\frac{\mathbb{E}_x V_n(y)}{\mathbb{E}_x V_n(z)} \to \frac{1}{\mathbb{E}_y V_{T_y}(z)} = \frac{\mathbb{E}_x V_{T_x}(y)}{\mathbb{E}_x V_{T_x}(z)}$.

   Proof. Let $L_y^{(n)} = \max\{m \le n : X_m = y\}\mathbb{1}_{\{T_y \le n\}}$. Then the last exit decomposition gives

$$\begin{aligned}\mathbb{E}_x V_n(z) &= \textstyle\sum_{m=1}^n \mathbb{P}_x\{X_m = z\} = \sum_{m=1}^n p^m(x,z)\\ &= \textstyle\sum_{m=1}^n \mathbb{P}_x\{X_m = z,\, T_y > m\} + \sum_{m=1}^n \sum_{\ell=1}^m \mathbb{P}_x\{X_m = z,\, L_y^{(m)} = \ell\}\\ &= \textstyle\sum_{m=1}^n p^m_{\setminus y}(x,z) + \sum_{\ell=1}^n \sum_{m=\ell}^n p^\ell(x,y) p^{m-\ell}_{\setminus y}(y,z),\end{aligned}$$

   where $p^m_{\setminus y}(x,z) = \mathbb{P}_x\{X_m = z,\, T_y > m\}$. Since $\mathbb{E}_x V_n(y) \nearrow \infty$ and $\sum_{m=1}^\infty p^m_{\setminus y}(x,z) = \mathbb{E}_x V_{T_y}(z) - \mathbb{1}_{[y=z]}$, we obtain that $\frac{\mathbb{E}_x V_n(z)}{\mathbb{E}_x V_n(y)} \to \sum_{m=\ell}^\infty p^{m-\ell}_{\setminus y}(y,z) = \mathbb{1}_{[y=z]} + \mathbb{E}_y V_{T_y}(z) - \mathbb{1}_{[y=z]} = \mathbb{E}_y V_{T_y}(z)$. □

4.3 Superharmonicity and recurrence

A function f is said to be superharmonic if $f(x) \ge \sum_y p^1(x,y) f(y)$ for all x, and to be harmonic if there are only equalities. Suppose that X is irreducible.

1. $x \mapsto \mathbb{P}_x\{T_A < \infty\}$ is superharmonic, where $T_A = \inf\{n \ge 1 : X_n \in A\}$ for A a subset of the state space.

   Proof. By the one-step forward analysis, $\mathbb{P}_x\{T_A < \infty\} = \sum_{y\in A} p^1(x,y) + \sum_{y\notin A} p^1(x,y)\,\mathbb{P}_y\{T_A < \infty\}$. □

2. X is recurrent if and only if every bounded superharmonic function is constant.

   Proof. Let f be a bounded superharmonic function so that $Y_n = f(X_n)$ is a bounded supermartingale converging a.s. to some Y_∞. If X is recurrent, then for any x we have a.s. $X_n = x$ i.o., and thus $Y_\infty \overset{\text{a.s.}}{=} f(x)$, which forces f to be constant. Conversely, if X is transient, then take $f(x) = G(x,z) = \sum_{n=0}^\infty p^n(x,z)$ for a fixed z. We have $\sum_y p^1(x,y) f(y) = f(x) - \mathbb{1}_{[x=z]}$, so f is a nonconstant superharmonic function. Note that $f \le G(z,z) < \infty$. As an alternative, one may consider $f(x) = \mathbb{P}_x\{T_{\{z\}} < \infty\}$ for a fixed z. □

3. (birth-and-death) Let the state space be $\mathbb{N}$, and $p^1(x,y) = b_x \mathbb{1}_{[y=x+1]} + d_x \mathbb{1}_{[y=x-1]}$ where $b_x + d_x = 1$ and $d_0 = 0$. Then X is recurrent if and only if $\sum_{x=0}^\infty \prod_{y=1}^x \frac{d_y}{b_y} = \infty$.

   Proof. Let $h(x) = \mathbb{P}_x\{T_{\{0\}} < \infty\}$. We have $h(0) = h(1) = b_1 h(2) + d_1$ and $h(x) = b_x h(x+1) + d_x h(x-1)$ for $x > 1$, which can be written as $h(x) - h(x+1) = \frac{d_x}{b_x}\big(h(x-1) - h(x)\big)$. Then it's easily seen that $1 - h(x) = (1 - h(1)) g(x)$, where $g(0) = 1$ and $g(x) = \sum_{z=0}^{x-1} \prod_{y=1}^z \frac{d_y}{b_y}$ for $x \ge 1$. If $g(\infty) = \infty$, then the boundedness of h entails that $h(1) = 1$, in which case X is recurrent. Conversely, if $g(\infty) < \infty$, then the superharmonic function $g(\infty) - g$ is not constant, so X is transient. □

   Second Proof. Note that $\tilde{g}(X_{n\wedge\tau})$ is a martingale, where $\tilde{g} = g\,\mathbb{1}_{\{0\}^\complement}$ and $\tau = \inf\{n : X_n \in \{0, M\}\}$ for some $M \in \mathbb{N}$. One can apply the optional stopping theorem to obtain that $\mathbb{P}_x\{T_{\{0\}} > T_{\{M\}}\} = g(x)/g(M)$ if $0 < x < M$. Letting $M \to \infty$ gives $\mathbb{P}_x\{T_{\{0\}} = \infty\} = g(x)/g(\infty)$. □

4.4 Green's function — potential theory

Suppose that X is irreducible. Let $G_A(x,y) = \mathbb{E}_x \sum_{n=0}^{\epsilon_A - 1} \mathbb{1}_{\{X_n = y\}}$, where A is a subset of the state space S, and $\epsilon_A = \inf\{n : X_n \notin A\}$. In particular, $G_S(x,y) = G(x,y) = \sum_{n=0}^{\infty} p^n(x,y)$. Write $\mathsf{P}f = \sum_y p^1(\cdot,y)f(y)$ for a function f on S which is either bounded or nonnegative. Note that $\mathsf{P}^n f = \sum_y p^n(\cdot,y)f(y) = \big(x \mapsto \mathbb{E}_x f(X_n)\big)$.

1. Assume that X is recurrent and $0 < \#A < \#S$.

   (a) $G_A(x,y) < \infty$, $\forall x, y \in S$.

   Proof. Note that $G_A(x,y) = 0$ if $x \in A^\complement$ or $(x,y) \in A \times A^\complement$. For $x, y \in A$, we have $G_A(x,y) \le \mathbb{E}_x \epsilon_A$. Since $\mathbb{P}_x\{\epsilon_A > n_x\} < 1$ for some $n_x \in \mathbb{N}$ by the recurrence, we have

   $$\mathbb{P}(\epsilon_A > n_A \,|\, X_0 \in A) \le \max_{x\in A} \mathbb{P}_x\{\epsilon_A > n_A\} \le \max_{x\in A} \mathbb{P}_x\{\epsilon_A > n_x\} = a < 1$$

   for $n_A = \max_{x\in A} n_x$, and thus $\mathbb{P}_x\{\epsilon_A > kn_A\} = \prod_{j=1}^{k} \mathbb{P}_x(\epsilon_A > jn_A \,|\, \epsilon_A > (j-1)n_A) \le a^k$ for every $k \in \mathbb{N}$, which implies that $\mathbb{E}_x \epsilon_A \lesssim n_A \sum a^k < \infty$. □

   (b) $(1 - \mathsf{P})G_A(\cdot,y) = \mathbb{1}_{\{y\}}$ on A, for any $y \in A$.

   Proof. For any $x \in A$, we have

   $$G_A(x,y) - \mathbb{1}_{\{y\}}(x) = \sum_{z\in A} \mathbb{E}_x \sum_{n=1}^{\epsilon_A - 1} \mathbb{1}_{\{X_1 = z,\, X_n = y\}} = \sum_{z\in A} p^1(x,z)G_A(z,y)$$

   by the strong Markov property, but $G_A(z,y) = 0$ for $z \notin A$. □

   (c) For any function ϱ on A, the Poisson equation $\begin{cases} (1-\mathsf{P})\psi = \varrho \text{ on } A \\ \psi = 0 \text{ on } A^\complement \end{cases}$ has a unique solution ψ given by $\sum_{y\in A} G_A(\cdot,y)\varrho(y)$, as G_A is the fundamental solution suggested by 1b.

   Proof. It remains to show the uniqueness. If ψ is a solution to the Poisson equation, then for any $x \in A$,

   $$\begin{aligned} \sum_{y\in A} G_A(x,y)\varrho(y) &= \sum_{y\in A} G_A(x,y)\big(\psi(y) - \sum_{z\in A} p^1(y,z)\psi(z)\big) \\ &= \sum_{z\in A} \psi(z) \sum_{y\in A} G_A(x,y)\big(\mathbb{1}_{[y=z]} - p^1(y,z)\big) \\ &= \sum_{z\in A} \psi(z)\mathbb{1}_{[x=z]} \;= \psi(x), \end{aligned}$$

   since $G_A(x,z) - \mathbb{1}_{[x=z]} = \sum_{y\in A} \mathbb{E}_x \sum_{n=1}^{\epsilon_A - 1} \mathbb{1}_{\{X_{n-1} = y,\, X_n = z\}} = \sum_{y\in A} G_A(x,y)\, p^1(y,z)$. □

2. Assume here that X is transient, whence $G(x,y) < \infty$, $\forall x, y \in S$.

   (a) $\mathsf{P}^n G(\cdot,y)(x) \to 0$ as $n \to \infty$, $\forall x, y \in S$.

   Proof. Proceeding the same way as in 1b, we have $(1-\mathsf{P})G(\cdot,y) = \mathbb{1}_{\{y\}}$, so

   $$\mathsf{P}^n G(\cdot,y)(x) - \mathsf{P}^{n+1} G(\cdot,y)(x) = \mathsf{P}^n \mathbb{1}_{\{y\}}(x) = \mathbb{E}_x \mathbb{1}_{\{y\}}(X_n) = p^n(x,y).$$

   Therefore, $\mathsf{P}^n G(\cdot,y)(x) = G(x,y) - \sum_{k=0}^{n-1} \big(\mathsf{P}^k G(\cdot,y)(x) - \mathsf{P}^{k+1} G(\cdot,y)(x)\big) = \sum_{k=n}^{\infty} p^k(x,y) \to 0$. □

   (b) (Riesz) Let $f : S \to \mathbb{R}_+$ be superharmonic in that $f \ge \mathsf{P}f$. Then $h = \lim_{n\to\infty} \mathsf{P}^n f$ exists pointwise and is harmonic, and $f(x) = h(x) + \sum_y G(x,y)\, q(y)$ for all x, where $q = f - \mathsf{P}f$ represents charges.

   Proof. The sequence $f \ge \mathsf{P}f \ge \mathsf{P}^2 f \ge \cdots \ge \mathsf{P}^n f \ge \cdots \ge 0$ admits a P-invariant limit. Next, notice that $f - \mathsf{P}^n f = \sum_{k=0}^{n-1} \mathsf{P}^k q = \sum_{k=0}^{n-1} \sum_y p^k(\cdot,y)\, q(y) \nearrow \sum_y G(\cdot,y)\, q(y)$. □

5 Stationary sequences

Recall that a measurable transformation T on a measure space $(S,\mathscr{S},\mu)$ is said to preserve μ if $\mu\circ T^{-1}=\mu$, and to be ergodic for μ if all T-invariant sets are μ-trivial, i.e., $\mu(I)\mu(I^{\complement})=0$ for any $I\in\mathscr{S}$ such that $T^{-1}(I)=I$. A sequence $\xi=(\xi_0,\xi_1,\xi_2,\dots)$ of random variables is said to be stationary if θ preserves $\mathbb{P}\circ\xi^{-1}$, and to be ergodic if θ is ergodic for $\mathbb{P}\circ\xi^{-1}$, where $\theta:(x_0,x_1,\dots)\mapsto(x_1,x_2,\dots)$ is the shift operator. We are primarily interested in the case $\xi_n=X\circ\varphi^{\circ n}$ for some transformation φ on $(\Omega,\mathscr{F},\mathbb{P})$ that is $\mathbb{P}$-preserving and $\mathbb{P}$-ergodic.

5.1 Invariant sets and functions

Let T be a transformation on $(S,\mathscr{S},\mu)$ which is measure-preserving. Suppose that μ is complete.

1. (σ-algebras) $\mathscr{I}_T^\mu=\{A\in\mathscr{S}:\mu(T^{-1}(A)\Delta A)=0\}$ is the completion of $\mathscr{I}_T=\{I\in\mathscr{S}:T^{-1}(I)=I\}$.

   Proof. On one hand, for $A\in\mathscr{I}_T^\mu$ we have $C=\{T^{-n}(A)\text{ i.o.}\}\in\mathscr{I}_T$ such that $\mu(A\Delta C)=0$. To see this, let $B=\bigcup_{n=0}^\infty T^{-n}(A)$. Then $\mu(A\Delta B)\le\sum_{n=1}^\infty\mu(A\Delta T^{-n}(A))\le\sum_{n=1}^\infty\sum_{k=1}^n\mu(T^{-(k-1)}(A)\Delta T^{-k}(A))=0$, and $\mu(B\Delta C)=\sum_{n=1}^\infty\mu(T^{-(n-1)}(B)\setminus T^{-n}(B))=\infty\cdot\mu(B\setminus T^{-1}(B))$, where $B\setminus T^{-1}(B)\subset A\setminus T^{-1}(A)$ has measure zero. On the other hand, if $J\in\mathscr{S}$ satisfies that $\mu(J\Delta I)=0$ for some $I\in\mathscr{I}_T$, then $\mu(T^{-1}(J)\Delta J)\le\mu(T^{-1}(J)\Delta T^{-1}(I))+\mu(I\Delta J)=0$. □

2. $f:S\to\mathbb{R}$ is $\mathscr{I}_T^\mu$-measurable if and only if $f\circ T\overset{\text{a.e.}}{=}f$, and is $\mathscr{I}_T$-measurable if and only if $f\circ T=f$.

   Proof. Denote $I_{a,b}=f^{-1}((a,b])=\{a<f\le b\}$. Clearly $T^{-1}(I_{a,b})=\{a<f\circ T\le b\}$. Note that $\{f\circ T\ne f\}=\bigcup_{r\in\mathbb{Q}}\big(\{f\circ T<r<f\}\cup\{f<r<f\circ T\}\big)$. Also, $f_n=2^{-n}\lfloor 2^n f\rfloor$ converges to f pointwise as $n\to\infty$, of which invariance can carry over to the limit. □

5.2 Criteria for ergodicity

Let $\xi=(\xi_1,\xi_2,\dots)$ be a stationary sequence.

1. Let $\eta=(\eta_1,\eta_2,\dots)$ where $\eta_k=g(\xi_k,\xi_{k+1},\dots)=g\circ\theta^{\circ(k-1)}(\xi)$ for some measurable function $g:\mathbb{R}^\infty\to\mathbb{R}$. Then η is also stationary. Moreover, if ξ is ergodic, then so is η.

   Proof. Introduce $G=(g\circ\theta^{\circ(k-1)})_{k=1,2,\cdots}$, which satisfies that $G\circ\theta=\theta\circ G$. We have $\xi\overset{d}{=}\theta(\xi)$, so $\eta=G(\xi)\overset{d}{=}G\circ\theta(\xi)=\theta(\eta)$. For any $J\in\mathscr{B}_{\mathbb{R}}^\infty$ such that $\theta^{-1}(J)=J$, called θ-invariant, $I=G^{-1}(J)$ is also θ-invariant, and thus $\mathbb{P}\{\eta\in J\}=\mathbb{P}\{\xi\in I\}$ should be either 0 or 1 as long as ξ is ergodic. □

2. ξ is ergodic if and only if $\frac{1}{n}\sum_{i=1}^n\mathbb{1}_B(\xi_i,\dots,\xi_{i+k-1})\xrightarrow[n\to\infty]{\text{a.s.}}\mathbb{P}\{(\xi_1,\dots,\xi_k)\in B\}$, $\forall B\in\mathscr{B}_{\mathbb{R}}^k$, $\forall k=1,2,\cdots$.

   Proof. Denote $\nu=\mathbb{P}\circ\xi^{-1}$. The stated property translates into $\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{B\times\mathbb{R}^\infty}\circ\theta^{\circ(i-1)}\xrightarrow{\nu\text{-a.s.}}\nu(B\times\mathbb{R}^\infty)$. In conjunction with Birkhoff's ergodic theorem, this yields $\mathbb{E}^\nu[\mathbb{1}_{B\times\mathbb{R}^\infty}|\mathscr{I}_\theta]=\mathbb{E}^\nu\mathbb{1}_{B\times\mathbb{R}^\infty}$, indicating that $B\times\mathbb{R}^\infty$ is independent of $\mathscr{I}_\theta$ under ν. The "only if" part is now trivial. As for the "if" part, notice that $\mathscr{I}_\theta\perp\!\!\!\perp\sigma\big(\bigcup_{k=1}^\infty\{B\times\mathbb{R}^\infty:B\in\mathscr{B}_{\mathbb{R}}^k\}\big)=\mathscr{B}_{\mathbb{R}}^\infty\supset\mathscr{I}_\theta$. □

6 Brownian motion

6.1 Chaining and continuous modification of stochastic process

1. (Talagrand) Let $X=(X_t)_{t\in T}$ where T is a countable set equipped with a metric ρ such that

   $$\mathbb{P}\{|X_t-X_s|>\varepsilon\}\le f(\varepsilon,\Delta),\quad\forall\varepsilon>0,\ \forall t,s\in T:\rho(t,s)\le\Delta.$$

   For any increasing sequence $\{T_n\}_{n=0}^\infty$ of subsets of T with $\bigcup T_n=T$, if $T_0=\{t_0\}$, then

   $$\mathbb{P}\Big\{\sup_{t\in T}|X_t-X_{t_0}|>\sum_{n\ge1}\varepsilon_n\Big\}\le\sum_{n\ge1}\#T_n\max_{s\in T_n}\#\{t\in T_n:\rho(t,s)\le\Delta_n\}\,f(\varepsilon_n,\Delta_n),\quad\forall\varepsilon_n>0,$$

   where $\Delta_n=2\sup_{t\in T}\rho(t,T_{n-1})$.

Proof. Define $\pi_n(t) = \arg\min_{s\in T_n} \rho(t,s)$, which $= t$ for sufficiently large n. Using the relation that $X_t - X_{t_0} = \sum_{n\geq 1}(X_{\pi_n(t)} - X_{\pi_{n-1}(t)})$, we have

$$\Big\{\sup_{t\in T}|X_t - X_{t_0}| > \sum_{n\geq 1}\varepsilon_n\Big\} \subset \bigcup_{t\in T}\bigcup_{n\geq 1}\{|X_{\pi_n(t)} - X_{\pi_{n-1}(t)}| > \varepsilon_n\} \subset \bigcup_{n\geq 1}\bigcup_{t,s\in T_n:\rho(t,s)\leq\Delta_n}\{|X_t - X_s| > \varepsilon_n\},$$

where $\Delta_n \geq \rho(t,T_n) + \rho(t,T_{n-1}) = \rho(t,\pi_n(t)) + \rho(t,\pi_{n-1}(t)) \geq \rho(\pi_n(t),\pi_{n-1}(t))$. □

2. (Kolmogorov–Chentsov) Assume that $X = (X_t)_{t\in[0,1]^d}$ admits some constants $a,b,K\in(0,\infty)$ for which

$$\mathbb{E}\,|X_t - X_s|^a \leq K\|t-s\|_\infty^{d+b}, \quad \forall t,s\in[0,1]^d.$$

It's immediate for 1 that $f(\varepsilon,\Delta) = K\Delta^{d+b}/\varepsilon^a$ applies when $\rho(t,s) = \|t-s\|_\infty$, by Markov's inequality. Let $c\in(0,b/a)$. Denote the dyadic lattice by $D = \bigcup D_n$ where $D_n = \{k/2^n : k = 0,1,\cdots,2^n-1\}^d$.

(a) The path $t\in D\mapsto X_t$ is Hölder continuous of order c, with probability one.

Proof. Note that if $q,r\in D : \|q-r\|_\infty < 2^{1-m}$, then there exist $q_0,r_0\in D_m : \|q_0-r_0\|_\infty \leq 2^{-m}$, $q\in D(q_0,m) = q_0 + 2^{-m}D$, $r\in D(r_0,m) = r_0 + 2^{-m}D$. Hence, with $L = 1 + 2/(2^c-1)$,

$$\begin{aligned}
&\bigcup_{\substack{q,r\in D:\\ \|q-r\|_\infty<2^{-M}}}\{|X_q - X_r| > L\,\|q-r\|_\infty^c\} = A_M \\
&\subset \bigcup_{m>M}\ \bigcup_{\substack{q,r\in D:\\ 2^{-m}\leq\|q-r\|_\infty<2^{1-m}}}\{|X_q - X_r| > L\cdot 2^{-cm}\} \\
&\subset \bigcup_{m>M}\ \bigcup_{\substack{q_0,r_0\in D_m:\\ \|q_0-r_0\|_\infty\leq 2^{-m}}}\ \bigcup_{q\in D(q_0,m)}\ \bigcup_{r\in D(r_0,m)}\begin{pmatrix}\{|X_q - X_{q_0}| > 2^{-cm}/(2^c-1)\}\\ \cup\{|X_{q_0} - X_{r_0}| > 2^{-cm}\}\\ \cup\{|X_{r_0} - X_r| > 2^{-cm}/(2^c-1)\}\end{pmatrix} \\
&= \bigcup_{m>M}\ \bigcup_{\substack{q_0,r_0\in D_m:\\ \|q_0-r_0\|_\infty\leq 2^{-m}}}\begin{pmatrix}\{|X_{q_0} - X_{r_0}| > 2^{-cm}\}\\ \cup\{\sup_{q\in D(q_0,m)}|X_q - X_{q_0}| > 2^{-cm}\sum_{n\geq1}2^{-cn}\}\\ \cup\{\sup_{r\in D(r_0,m)}|X_r - X_{r_0}| > 2^{-cm}\sum_{n\geq1}2^{-cn}\}\end{pmatrix},
\end{aligned}$$

whose probability is bounded by

$$\begin{aligned}
&\sum_{m>M}2^{dm}\cdot 3^d\cdot\Big(f(2^{-cm},2^{-m}) + 2\sum_{n\geq1}2^{dn}\cdot 9^d\cdot f(2^{-cm-cn}, 2\cdot 2^{-m-(n-1)})\Big) \\
&= 3^dK\sum_{m>M}2^{(ac-b)m}\Big(1 + 9^d 2^{1+2(d+b)}\sum_{n\geq1}2^{(ac-b)n}\Big) \lesssim 2^{(ac-b)M}.
\end{aligned}$$

Now $\sum\mathbb{P}(A_M) < \infty$, so $\mathbb{P}(A_M\text{ i.o.}) = 0$ by the Borel–Cantelli lemma. For $\omega\in\{A_M\text{ i.o.}\}^\complement = \bigcup A_M^\complement$, let

$$M_*(\omega) = \inf\{M : \omega\in A_M^\complement\}.$$

$\forall t,s\in D$ can be connected by $s = s_0\leftrightarrow s_1\leftrightarrow\cdots\leftrightarrow s_n = t$ with $\|s_i - s_{i-1}\|_\infty < 2^{-M_*(\omega)}\|t-s\|_\infty$ and $n\leq N(\omega) = 1 + 2^{M_*(\omega)}$. It follows that $|X_t(\omega) - X_s(\omega)| \leq N(\omega)\cdot L\cdot 2^{-cM_*(\omega)}\|t-s\|_\infty^c$. □

(b) A continuous process $\tilde{X} = (\tilde{X}_t)_{t\in[0,1]^d}$ agreeing with X on D a.s. is a modification of X. Note that the Hölder continuity of $\tilde{X}$ on D extends with the same order to the entire cube $[0,1]^d$.

Proof. For any $t\in[0,1]^d$, choose a sequence $\{t_n\}\subset D$ with $t_n\to t$. Then $\tilde{X}_{t_n} \overset{\text{a.s.}}{=} X_{t_n}$. Since $\tilde{X}_{t_n}\to\tilde{X}_t$ by continuity and $X_{t_n}\xrightarrow{L^a} X_t$, the uniqueness of limits in probability entails $\tilde{X}_t \overset{\text{a.s.}}{=} X_t$. □

(c) The Brownian path is Hölder continuous of any order $< 1/2$, with probability one.

Proof. If $Z\sim\mathcal{N}(0,t)$, then $\mathbb{E}\,Z^{2k} = (2k-1)!!\cdot t^k$ for every positive integer k. □

6.2 Nonsmoothness of sample path

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion.

1. The path $t \mapsto B_t$ is *nowhere* Hölder continuous of any order $> 1/2$, with probability one.

   Proof. (Dvoretsky–Erdős–Kakutani) Fixing $\alpha > 1/2$ and $C > 0$, it suffices to show that the event $A_n = \bigcup_{s\in[0,1]} \bigcap_{t:|t-s|\leq m/n} \{|B_t - B_s| \leq C|t-s|^\alpha\}$ has probability zero for $n \gg 1$, where $m > 1/(\alpha - 1/2)$. On A_n we have $\max_{k<j\leq k+m} |B_{j/n} - B_{(j-1)/n}| \leq 2Cm^\alpha/n^\alpha$ for some $0 \leq k \leq n-m$. To conclude, note that $A_n \uparrow$ and that $(n-m+1)\,\mathbb{P}\{|\mathcal{N}(0,1/n)| \leq M/n^\alpha\}^m \leq n\cdot(\frac{2M}{\sqrt{2\pi}} n^{1/2-\alpha})^m \lesssim n^{1-(\alpha-1/2)m} \to 0$. □

2. (Paley–Wiener–Zygmund) The path $t \mapsto B_t$ is *nowhere* differentiable, with probability one.

6.3 Reflection principle and arcsine law

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion, starting from x under $\mathbb{P}_x$. Denote $M_t = \max_{s\leq t} B_s$.

1. $B \overset{d}{=} 2B^T - B$ for any stopping time T, where $B^T = (B_{t\wedge T})_{t\geq 0}$.

   Proof. By conditioning, assume $T < \infty$. Then by the strong Markov property, $B^{(T)} = (B_{T+t} - B_T)_{t\geq 0}$ is a standard Brownian motion starting from 0 and $B^{(T)} \perp\!\!\!\perp (B^T, T)$, so $(B^T, T, B^{(T)}) \overset{d}{=} (B^T, T, -B^{(T)})$. Therefore, $B = B^T + B^{(T)}_{(\cdot - T)^+} \overset{d}{=} B^T - B^{(T)}_{(\cdot-T)^+} = 2B^T - B$. □

2. $M_t \overset{d}{=} |B_t|$ for any $t \geq 0$, under $\mathbb{P}_0$.

   Proof. Using the reflection principle with $T_a = \inf\{t : B_t = a\}$, we have $\{M_t \geq a\} = \{T_a \leq t\}$ and hence

   $$\mathbb{P}_0\{M_t \geq a,\, B_t \leq b\} = \mathbb{P}_0\{2a - B_t \leq b\} = \mathbb{P}_0\{\mathcal{N}(0,t) \geq 2a - b\}, \text{ where } a \geq b \vee 0.$$

   Thus, $\mathbb{P}_0\{M_t \in \mathrm{d}a,\, B_t \leq b\} = \frac{2}{\sqrt{2\pi t}} \mathrm{e}^{-(2a-b)^2/(2t)} \mathbb{1}_{[a\geq b\vee 0]}$. Letting $b \leftarrow a$ completes the proof. □

3. Recall that $\arcsin(\frac{\xi}{\sqrt{\xi^2+\eta^2}}) \sim \mathrm{Uniform}(-\frac{\pi}{2}, \frac{\pi}{2})$ for $\xi, \eta \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

   (a) $L = \sup\{t \leq 1 : B_t = 0\}$ satisfies that $\mathbb{P}_0\{L < t\} = \frac{2}{\pi} \arcsin \sqrt{t}$ for $t \in [0,1]$.

   Proof.
   $$\begin{aligned}\mathbb{P}_0\{L < t\} &= \mathbb{P}_0\{\max\nolimits_{s\in[t,1]} B_s < 0\} + \mathbb{P}_0\{\min\nolimits_{s\in[t,1]} B_s > 0\} \\ &= \mathbb{P}_0\{\max\nolimits_{s\in[t,1]}(B_s - B_t) < -B_t\} + \mathbb{P}_0\{-\min\nolimits_{s\in[t,1]}(B_s - B_t) < B_t\} \\ &= \mathbb{P}\{\sqrt{1-t}\,|\xi| < -\sqrt{t}\,\eta\} + \mathbb{P}\{\sqrt{1-t}\,|\xi| < \sqrt{t}\,\eta\} \\ &= \mathbb{P}\{\sqrt{1-t}\,|\xi| < \sqrt{t}\,|\eta|\} \\ &= \mathbb{P}\{\tfrac{\xi^2}{\xi^2+\eta^2} < t\} \quad = \tfrac{2}{\pi}\arcsin\sqrt{t}.\end{aligned}$$
   □

   (b) $\tau = \inf\{t : B_t = M_1\}$ satisfies that $\mathbb{P}_0\{\tau \leq t\} = \frac{2}{\pi} \arcsin \sqrt{t}$ for $t \in [0,1]$.

   Proof. We have $(B_{t-s} - B_t)_{s\in[0,t]} \overset{d}{=} (B_s - B_0)_{s\in[0,t]}$, and thus $M_t - B_t \overset{d}{=} M_t - B_0$. With this in mind, $\mathbb{P}_0\{\tau \leq t\} = \mathbb{P}_0\{M_t - B_t \geq \max_{s\in[t,1]} B_s - B_t\} = \mathbb{P}\{\sqrt{t}\,|\eta| \geq \sqrt{1-t}\,|\xi|\} = \mathbb{P}\{\frac{\xi^2}{\xi^2+\eta^2} \leq t\}$. □