Probability Sequence Notes

Kenneth Tay

Overall Tricks

For computing expectations:

1. Consider the use of the moment generating function (either $\mathbb{E}[e^{tX}]$ or $\mathbb{E}[t^X]$) or the cumulant generating function ($\log \mathbb{E}[e^{tX}]$).
2. For non-negative X, try the formula $\mathbb{E}X = \int_0^\infty P(X \geq t)dt = \int_0^\infty P(X > t)dt$. If $p > 0$, we also have $\mathbb{E}[Y^p] = \int_0^\infty py^{p-1}P(Y > y)dy$.

To prove (in)equalities on the limit of a sequence of random variables $\{X_n\}$:

1. Consider picking a subsequence with properties that allow you to prove what you want, then use interpolation to fill in the gaps.
2. If there is weak convergence, consider using the Portmanteau Theorem.

To prove a.s. convergence of i.i.d. series:

1. Try Kolmogorov's Three-Series Theorem.
2. If the $\{X_n\}$ are non-negative, try the generalized Borel-Cantelli lemmas (see 310A section on convergence & SLLN/WLLN).

To find the limit (in probability) of a sequence of random variables:

1. Try to use the Law of Large Numbers (or the triangular array version of LLN).

To find the limiting distribution of a sequence of random variables:

1. Check if you can use Lindeberg's or Lyapounov's CLT.
2. If the CLT condition doesn't hold, try using truncation to make it hold.
3. Consider looking at the characteristic functions and see if you can derive the limit.
4. Sometimes the characteristic function $\varphi(t/n)$ ends up being an integral with no closed form. Consider applying the Mean Value Theorem to get $\varphi(t/n) = \varphi(0) + \varphi'(\varepsilon_n)\dfrac{t}{n}$ for some $\varepsilon_n \in (0, t/n)$.

5. Approximation tricks are useful (e.g. $\log(1-x) \approx -x$ for small x, $1 - \Phi(x) \approx \dfrac{\phi(x)}{x}$ for large x).

6. Consider using the Portmanteau Theorem.

7. If we are looking at many events of small independent probability, the limit is probably a Poisson. Consider using Stein's method, or Poisson convergence (Durrett Thm 3.6.1).

To prove that function F is a distribution function:

1. Show that there exists a random variable X such that $P(X \leq x) = F(x)$.

2. Go directly through the definition (i.e. monotonically increasing in each variable, right-continuous, $F(\infty) = 1, F(-\infty) = 0$, and $\Delta_A(F) \geq 0$ for all finite rectangles A).

Martingale questions:

1. If a stopping time is involved, try to find a martingale and use the Optional Stopping Theorem.

2. For a stopping time τ which takes on countably many values $\{a_1, a_2, \dots\}$, we can rewrite X_τ as $\displaystyle\sum_{i=1}^{\infty} X_{a_i} 1\{\tau = a_i\}$.

3. When trying to find a martingale, think of the linear, quadratic and exponential martingales.

4. If the question involves X_∞, think about how to involve the Martingale Convergence Theorem.

Markov chain questions:

1. If the state space is finite with $p(x, y) > 0$ for all x, y, the Perron-Frobenius Theorem might come in handy.

2. In computing expectations, it may be worth trying to find a relationship between $\mathbb{E}X_{n+1}$ and $\mathbb{E}X_n$ (or consider $\mathbb{E}[X_{n+1} - X_n \mid X_n]$).

3. It is worth checking if the Markov chain is a martingale. This will help us to prove convergence of the Markov chain.

For questions with stopping times:

1. Find a martingale and use the Optional Stopping Theorem for the stopped process.

2. To show that a stopping time is a.s. finite, the easiest way is to show that it has finite expectation.

3. Alternatively, try to use facts of the underlying process to show that the stopping time is a.s. finite (e.g. $\limsup_{t\to\infty} W_t = +\infty$, $N_t/t \overset{a.s.}{\to} \lambda$).

4. To get $\mathbb{E}T$ from $\mathbb{E}[e^{-aT}]$, take a derivative w.r.t. a.

For questions with limits of random walks:

1. To find the limiting distribution of functionals of random walks, use Donsker's invariance principle to show that they converge to the standard Brownian motion, and hope to use the continuous mapping theorem.

   Specifically, say $X_1, X_2, \ldots$ i.i.d. with mean 0 variance 1, and $S_n = X_1 + X_2 + \cdots + X_n$. If we let

$$S_n(t) = \begin{cases} S_k/\sqrt{n} & \text{if } t = k/n, \\ \text{linear interpolation} & \text{otherwise,} \end{cases}$$

   then S_n converges in distribution to the standard Brownian motion on $[0,1]$.

2. To find the limiting distribution of a bounded, continuous function of a martingale differences sequence, use the martingale CLT. (Verify the theorem's hypotheses.) There's usually a 3-step process:

   (a) We have some martingale $S_{k,n}$. Let $S^*_{t,n}$ be the linear interpolation of $S_{k,n}$. Show that CLT conditions are met so that $S^*_{t,n} \Rightarrow W_t$.

   (b) Hope that for your functional f, there is an equivalent g such that $g(S^*_{t,n}) = f(S_{k,n})$. This is the case for many functionals, e.g. $\max_{0 \leq t \leq 1} S^*_{t,n} = \max_{0 \leq k \leq n} S_{k,n}$.

   (c) Conclude using the continuous mapping theorem.

Some results with sums of RVs:

1. (Durrett Thm 2.5.7) Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = \sigma^2 < \infty$. Let $S_n = X_1 + \cdots + X_n$. If $\varepsilon > 0$, then $\dfrac{S_n}{\sqrt{n}(\log n)^{1/2+\varepsilon}} \xrightarrow{a.s.} 0$, and $\limsup\limits_{n\to\infty} \dfrac{S_n}{\sqrt{n \log\log n}} = \sigma\sqrt{2}$ a.s.

2. (Durrett Thm 2.5.8) Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^p < \infty$ for some $1 < p < 2$. Let $S_n = X_1 + \cdots + X_n$. Then $\dfrac{S_n}{n^{1/p}} \xrightarrow{a.s.} 0$.

3. (Durrett Thm 2.5.9) Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X_i| = \infty$. Let $S_n = X_1 + \cdots + X_n$. Let a_n be a sequence of positive numbers s.t. a_n/n is increasing. Then $\limsup\limits_{n\to\infty} |S_n|/a_n = 0$ or ∞, according as $\sum\limits_n P(|X_1| \geq a_n) < \infty$ or $= \infty$.

Miscellaneous tricks:

1. For random variables $X \in [0,1]$, we have $X^2 \leq X$. This can be useful in inequalities involving variances.
2. If we are counting things, we can try to represent the random variable as a sum of indicator variables.
3. If a predictable sequence is involved, the martingale transform or Lenglart's bound might come in useful.

Misc facts

- See 310A Lec 1 for birthday problem approximation.
- For sets, $\limsup A_n = \{A_n \text{ infinitely often}\} = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{m=n}^{\infty} A_n$, and $\liminf A_n = \{A_n \text{ all but finitely often}\} = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{m=n}^{\infty} A_n$.

- (Durrett Ex 2.3.1) $P(\limsup A_n) \geq \limsup P(A_n)$ and $P(\liminf A_n) \leq \liminf P(A_n)$.
- (Billingsley Prob 14.5, 310A HW4) **Definition of Lévy distance:** For 2 distribution functions F and G, $d(F,G) = \inf\{\varepsilon : G(x-\varepsilon) - \varepsilon \leq F(x) \leq G(x+\varepsilon) + \varepsilon \text{ for all } x\}$. This is a metric, and $F_n \Rightarrow F$ iff $d(F_n, F) \to 0$.
- (Billingsley Prob 20.25, 310A HW7) For 2 random variables X and Y, define $d(X,Y) = \inf\{\varepsilon : \mathbb{P}(|X-Y| \geq \varepsilon) \leq \varepsilon\}$.
  - d is a metric. Hence, $d(X,Y) = 0$ iff $X = Y$ w.p. 1.
  - $X_n \xrightarrow{P} X$ iff $d(X_n, X) \to 0$.
  - (Billingsley Prob 21.15, 310A HW7) This metric is equivalent to $d_1(X,Y) = \mathbb{E}\left[\frac{|X-Y|}{1+|X-Y|}\right]$.
- (Durrett Thm 2.5.2) **Kolmogorov's maximal inequality**: Let $X_1, \dots, X_n$ be independent random variables, each with zero mean and finite variance. Then for each $\lambda > 0$,

$$P\left(\max_{1\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}\sum_{k=1}^n \text{Var } X_k.$$

- (310B Lec 16) **Wald's equation for sums of i.i.d. random variables**: Let $X_1, X_2, \dots$ be i.i.d. random variables with $\mathbb{E}X_1 = \mu$, $\mathbb{E}|X_1| < \infty$. Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Let τ be a stopping time w.r.t. $\{\mathcal{F}_n\}$. Let $S_n = \sum_{i=1}^n X_i$.

  If $\mathbb{E}\tau < \infty$, then $\mathbb{E}[S_\tau] = \mu\mathbb{E}\tau$.
- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n Y_n \xrightarrow{P} XY$.
- (300B HW1) **Uniform over the n-sphere:** If U_n is the random vector drawn uniformly from the unit sphere in n-dimensions, we may parametrize U_n as $U_n = \frac{1}{N}(X_1, \dots, X_n)$, where $X_i \overset{iid}{\sim} \mathcal{N}(0,1)$, $N^2 = \sum_{i=1}^n X_i^2$.