STATS 310A Notes

Kenneth Tay

Basic measure theory (Lec 1-6)

- (Billingsley Prob 2.4, HW1) The union of σ-fields need not be a σ-field. The countable union of fields is a field.
- (Billingsley Prob 2.5) Let $\mathcal{A}$ be a collection of subsets of Ω. The field generated by $\mathcal{A}$ is equal to the collection of sets of the form $\displaystyle\bigcup_{i=1}^{m}\bigcap_{j=1}^{n_i} A_{ij}$, where either $A_{ij} \in \mathcal{A}$ or $A_{ij}^c \in \mathcal{A}$, and where the m sets $\bigcap_{j=1}^{n_i} A_{ij}$ are disjoint.
- **Definition of outer measure**: Let P be a probability on field $\mathcal{F}_0$. For every $A \subseteq \Omega$, define $P^*(A) := \inf \displaystyle\sum_{i=1}^{\infty} P(B_i)$, where $A \subseteq \bigcup B_i$, $B_i \in \mathcal{F}_0$. (Covers can be countably infinite in size.) Note that P^* is countably sub-additive.
- A probability measure on a field has a unique extension to the generated σ-field.
- $\pi - \lambda$ **Theorem**: If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.
- **Law of the iterated logarithm**: Let $\{Y_n\}$ be iid random variables with zero mean and variance 1. Then almost surely, $\limsup \dfrac{Y_1 + \cdots + Y_n}{\sqrt{2n \log\log n}} = 1$.
- For the simple symmetric (± 1) random walk, let $M_n = \max(S_1, \ldots, S_n)$. Then for every integer $c \geq 1$, $P(M_n \geq c) \leq P(S_n \geq c) + P(S_n > c) \leq 2P(S_n \geq c)$.
- For the simple symmetric (± 1) random walk, $P\left(\frac{S_n}{n} \geq \varepsilon\right) \leq 2\exp\left(-\frac{n\varepsilon^2}{2}\right)$. (Proof in HW2 Q5.)
- **Definition of a general measure on a field**: $\mu : \mathcal{F} \mapsto \mathbb{R}$ is a measure if $\mu(\emptyset) = 0$, μ is non-negative and countably additive.
- If μ_1 and μ_2 are measures on $\sigma(\mathcal{P})$ such that they agree on π-system $\mathcal{P}$ and are σ-finite on $\mathcal{P}$, then they agree on $\sigma(\mathcal{P})$.
- **Definition of outer measure**: A set function μ^* defined on all subsets of Ω is an outer measure if it is non-negative, monotone, countably sub-additive and $\mu^*(\emptyset) = 0$.
- If μ^* is an outer measure on subsets of Ω, then μ^* is a measure on $\mathcal{M}^*(\mu^*) = \{A \subseteq \Omega : \forall E \subseteq \Omega, \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)\}$.
- **Definition of semi-ring**: A collection of subsets $\mathcal{R}$ of Ω is a semi-ring if it contains $\emptyset$, is closed under finite intersections, and if $A, B \in \mathcal{R}$ with $A \subseteq B$, then $B \setminus A = \displaystyle\bigcup_{i=1}^{n} C_i$ with $C_i \in \mathbb{R}$, C_i's disjoint.

Examples of semi-rings: finite subsets of $[0,1]$, ∞ rectangles in $\mathbb{R}^d$ (must extend to ∞ in at least 1 direction), and finite rectangles in $\mathbb{R}^d$.

- **Extension theorem**: If $\mu : \mathcal{R} \mapsto [0,\infty]$ such that μ is countably sub-additive, finitely additive on $\mathcal{R}$ and $\mu(\emptyset) = 0$, then μ has an extension to a measure on $\sigma(\mathcal{R})$. If μ is σ-finite, this measure is unique.

Distribution functions, random variables, integration (Lec 7-10, 12)

- Let $F : \mathbb{R}^k \mapsto \mathbb{R}$ be monotonically increasing, right-continuous, such that $F(\infty) = 1$ and $F(-\infty) = 0$. If $\Delta_A(F) \geq 0$ for all finite rectangles A, then there exists a unique probability measure μ on Borel sets of $\mathbb{R}^k$ such that $\mu(A) = \Delta_A(F)$ for all A.
- (Billingsley Prob 14.8, HW4) If a distribution function F is everywhere continuous, then it is uniformly continuous.
- If X has CDF F, then $F(X)$ is a Unif$(0,1)$ variable.
- Useful proposition to manipulate inverses: Let T be a measurable map from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$. For a collection of sets $\{B_i'\}_{i\in I}$ in $\mathcal{F}'$,
  1. $[T^{-1}(B')]^c = T^{-1}(B'^c)$.
  2. $T^{-1}\left(\bigcup_{i\in I} B_i'\right) = \bigcup_{i\in I} T^{-1}(B_i')$.
  3. $T^{-1}\left(\bigcap_{i\in I} B_i'\right) = \bigcap_{i\in I} T^{-1}(B_i')$.
- **Definition of push-forward:** Suppose we have $(\Omega, \mathcal{F}, \mu)$ a measure space, $(\Omega', \mathcal{F}')$ a measurable space, $T : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ measurable. The push-forward $\mu^{T^{-1}}$, a measure on $(\Omega', \mathcal{F}')$, is defined by
$$\mu^{T^{-1}}(B') := \mu(T^{-1}(B')) \qquad \text{for } B' \in \mathcal{F}'.$$
- **Definition of simple function:** A measurable function which takes on finitely many values, i.e. exist $x_i \in \mathbb{R} \cup \pm\infty$ and a partition of Ω, $A_i \in \mathcal{F}$, such that $f(\omega) = \sum_{i=1}^{n} x_i \delta_{A_i}(\omega)$.
- For every non-negative measurable function f, there exists a non-decreasing sequence of simple functions f_n such that $f_n(\omega) \to f(\omega)$ for all ω.
- **Definition of integral:** For $f \geq 0$, define $\int f d\mu := \sup \sum_{i=1}^{n} \nu_i \mu(A_i)$, where the sup is taken over all **finite** measurable partitions $\{A_i\}$ of Ω, and $\nu_i = \inf_{\omega\in A_i} f(\omega)$.
- **Monotone Convergence Theorem:** For non-negative f, if $f_n(\omega) \nearrow f(\omega)$ for all ω, then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu = \int \lim_{n\to\infty} f_n d\mu$.
- **Fatou's Lemma:** Let $\{f_n\}$ be any sequence of non-negative measurable functions. Then $\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$.

- **Dominated Convergence Theorem:** Let $\{f_n\}$ be a sequence of measurable functions such that $f_n \to f$ a.s. and there exists an integrable function $g \geq 0$ such that $|f_n| \leq g$ a.s. Then f is measurable and integrable, and $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.

- For $X \geq 0$, $\mathbb{E}X = \int_0^\infty P(X \geq t)dt = \int_0^\infty P(X > t)dt$.

- Let X have MGF M. If $M(s)$ is finite on some non-empty interval $(-s_0, s_0)$, then M is infinitely differentiable and $M^{(k)}(0) = \mathbb{E}[X^k]$.

Product spaces (Lec 11)

Let $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ be two measurable spaces.

- **Projection mappings** are $\pi_x : (x, y) \mapsto x$ and $\pi_y : (x, y) \mapsto y$.

- **Product σ-algebra**: $\mathcal{X} \times \mathcal{Y} := \sigma(\pi_x, \pi_y)$ (i.e. the σ-algebra generated by the projection mappings).

- **Cylinder sets**: $\mathcal{C} = \{\pi_x^{-1}(A) \cup \pi_y^{-1}(B) : A \in \mathcal{X}, B \in \mathcal{Y}\}$.

- **Measurable rectangles**: $\mathcal{P} = \{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\}$.

- $\mathcal{X} \times \mathcal{Y} = \sigma(\mathcal{C}) = \sigma(\mathcal{P}) = \sigma(\{\text{all finite disjoint unions of measurable rectangles}\})$.

- **Sections:** For $A \subseteq Z$, sections are defined to be $A_x = \{y : (x, y) \in A\}$, $A_y = \{x : (x, y) \in A\}$.

- **Sectioning lemma:** If $f : (Z, \mathcal{Z}) \to (W, \mathcal{W})$ is measurable, then $f_x : Y \to W$ is measurable.
  If $A \in \mathcal{Z}$, then A_x, A_y are measurable.

- **Definition of Markov kernel:** A **Markov kernel** is a function $K(x, B) : (X, \mathcal{Y}) \to [0, 1]$ such that
  1. For every fixed x, $K(x, \cdot)$ is a probability on $(Y, \mathcal{Y})$, and
  2. For every fixed B, the map $x \mapsto K(x, B)$ is measurable.

- Let μ be a probability on $(X, \mathcal{X})$. We can define a probability $\mu \times K$ on $(Z, \mathcal{Z})$ by $\mu \times K(A) := \int K(x, A_x)\mu(dx)$.

- **Fubini's theorem:** If $f : Z \to \mathbb{R}$ is Borel-measurable and non-negative, then
  1. $x \mapsto \int f(x, y)K(x, dy)$ is $\mathcal{X}$-measurable, and
  2. $$\int f d(\mu \times K) = \int_X \left[\int_Y f(x, y)K(x, dy)\right] \mu(dx).$$

- (Durrett Thm 1.7.2 p37) **Durrett's version of Fubini:** If $f \geq 0$ or $\int |f| d\mu < \infty$ (here $\mu = \mu_1 \times \mu_2$, then
$$\int_X \int_Y f(x, y)\mu_2(dy)\mu_1(dx) = \int_{X\times Y} f d\mu = \int_Y \int_X f(x, y)\mu_1(dx)\mu_2(dy).$$

Convergence, strong & weak law of large numbers (Lec 13)

- (Billingsley Prob 20.24, HW7) In a discrete probability space, convergence in probability is equivalent to a.s. convergence.
- (Durrett Thm 2.4.1 p73) **Strong law of large numbers:** Let $X_1, X_2, \ldots$ be pairwise independent identically distributed random variables with $\mathbb{E}|X_i| < \infty$. Let $\mathbb{E}X_i = \mu$ and $S_n = X_1 + \cdots + X_n$. Then as $n \to \infty$, $S_n/n \to \mu$ a.s.
- (Durrett Them 2.4.5 p75) Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}X_i^+ = \infty$ and $\mathbb{E}X^- < \infty$. Then $S_n/n \to \infty$ a.s.
- **SLLN without i.i.d. assumption:** Let $X_1, X_2, \ldots$ be mutually independent such that the X_k's each have finite variance and $\displaystyle\sum_{k=1}^{\infty} \frac{\text{Var } X_k}{k^2} < \infty$. Then $\bar{X}_n - \mathbb{E}[\bar{X}_n] \xrightarrow{a.s.} 0$.
- **Siegmund's theorem about deviations**: Let X_i be i.i.d. with mean 0 and variance 1. For each $\varepsilon > 0$, let $m_\varepsilon = \sup\{n \geq 0 : |S_n/n| \geq \varepsilon\}$. Then for $0 \leq x < \infty$, $\mathbb{P}(\varepsilon^2 m_\varepsilon \leq x) \to 2\Phi(x) - 1$ as $\varepsilon \to 0$.
- (Durrett Thm 2.2.6) **Weak law for triangular arrays:** For each n, let $X_{n,k}$, $1 \leq k \leq n$ be independent. Let $b_n > 0$ with $b_n \to \infty$, and let $\bar{X}_{n,k} = X_{n,k} 1_{\{|X_{n,k} \leq b_n|\}}$. Suppose that as $n \to \infty$,

  1. $\displaystyle\sum_{k=1}^{n} P(|X_{n,k} > b_n|) \to 0$, and
  2. $\displaystyle\frac{1}{b_n^2} \sum_{k=1}^{n} \mathbb{E}\bar{X}_{n,k}^2 \to 0$.

  If we let $S_n = X_{n,1} + \cdots + X_{n,n}$ and $a_n = \displaystyle\sum_{k=1}^{n} \mathbb{E}\bar{X}_{n,k}$, then $\displaystyle\frac{S_n - a_n}{b_n} \to 0$ in probability.
- (Durrett Thm 2.2.7) **Weak law of large numbers:** Let $X_1, X_2, \ldots$ be i.i.d. with $xP(|X_i| > x) \to 0$ as $x \to \infty$. If we let $S_n = X_1 + \cdots + X_n$ and $\mu_n = \mathbb{E}\left[X_1 1_{\{|X_1 \leq n|\}}\right]$, then $S_n/n - \mu_n \to 0$ in probability.
- **Kolmogorov Three Series Theorem:** Let $X_1, X_2, \ldots$ be independent random variables. The random series $\displaystyle\sum_{n=1}^{\infty} X_n$ converges almost surely in $\mathbb{R}$ iff the following conditions hold for some $A > 0$:

  1. $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq A)$ converges,
  2. Let $Y_n = X_n 1_{\{|X_n| \leq A\}}$, then $\displaystyle\sum_{n=1}^{\infty} \mathbb{E}Y_n$ converges, and
  3. $\displaystyle\sum_{n=1}^{\infty} \text{Var } Y_n$ converges.
- (Durrett Thm 2.5.3) **Special case of three series theorem:** In the set-up above, further assume that $\mathbb{E}X_n = 0$ for all n. If $\displaystyle\sum_{n=1}^{\infty} \text{Var } X_n < \infty$, then $\displaystyle\sum_{n=1}^{\infty} X_n$ converges almost surely.
- (Billingsley Prob 22.3, HW7) **Generalized Borel-Cantelli lemmas:** Suppose $\{X_n\}$ are non-negative. If $\sum \mathbb{E}X_n < \infty$, then $\sum X_n$ converges w.p. 1. If the X_n are independent and uniformly bounded, and $\sum \mathbb{E}X_n = \infty$, then $\sum X_n$ diverges w.p. 1.

Stein's method (Lec 14-15)

- Set-up:
  - $\{X_i\}_{i \in I}$ a collection of 0/1-valued random variables (I some index set),
  - $W := \sum_{i \in I} X_i$,
  - $p_i := \mathbb{E}X_i = P\{X_i = 1\}$,
  - $\lambda := \sum_{i \in I} p_i = \mathbb{E}W$.
  - A **graph** is an ordered pair (I, E), where I is the set of vertices and $E \subseteq I \times I$ is the set of edges. E must be symmetric (i.e. $(i,j) \in E \Leftrightarrow (j,i) \in E$) and has no loops (i.e. $(i,i) \notin E$ for all i).
  - A graph (I, E) is a **dependency graph** for $\{X_i\}_{i \in I}$ if for any two disjoint subsets $I_1, I_2 \subseteq I$ with no edges between them, $\{X_i\}_{i \in I_1}$, $\{X_j\}_{j \in I_2}$ are independent. (A dependency graph need not be unique.)
  - For a vertex $i \in I$, the **neighborhood** of i is $N_i := \{i\} \cup \{j : (i,j) \in E\}$.
- Let $P_\lambda(A) :=$ probability that value of Poisson(λ) random variable falls in A. For every $A \subseteq \mathbb{N}$, there is a unique function $f : \mathbb{N} \to \mathbb{R}$ such that

$$\lambda f(w+1) - w f(w) = \delta_A(w) - P_\lambda(A)$$

  for $w = 0, 1, 2, \ldots$. Moreover, $|f(w)| \leq 1.25$ and $|f(w+1) - f(w)| \leq \min(3, \lambda^{-1})$.
- **Stein's Equation**: $Z \sim$ Poisson(λ) if and only if for every $f : \mathbb{N} \to \mathbb{R}$ bounded, $\mathbb{E}\{\lambda f(Z+1) - Zf(Z)\} = 0$.
- Let $\{X_i\}_{i \in I}$ be a collection of 0/1-valued random variables, and let (I, E) be a dependency graph for $\{X_i\}_{i \in I}$. Let $p_{ij} = P(X_i = X_j = 1)$, and let P_W denote the probability distribution of $W = \sum X_i$. Then

$$\|P_W - \text{Poisson}(\lambda)\|_{TV} \leq \min(3, \lambda^{-1}) \left[\sum_{i \in I} \sum_{j \in N_i \setminus \{i\}} p_{ij} + \sum_{i \in I} \sum_{j \in N_i} p_i p_j \right].$$

Weak convergence & Central Limit Theorem (Lec 16-18)

- (Lec 18) **Slutsky's Theorem**: If $X_n \Rightarrow Z$ and $X_n - Y_n \Rightarrow 0$, then $Y_n \Rightarrow Z$.
- (Lec 16) Let $\mathcal{C}_b^\infty = \{f : f : \mathbb{R} \to \mathbb{R}, f$ has bounded derivatives of all order$\}$.

  If F_n, F are distribution functions on $\mathbb{R}$ such that $F_n \Rightarrow F$, then

$$\int_{-\infty}^{\infty} f dF_n \to \int_{-\infty}^{\infty} f dF$$

  for all $f \in \mathcal{C}_b^\infty$. The converse is true as well.
- (Lec 16) **Lindeberg's CLT**: Let $\{X_{ni}\}$ be a triangular array. Suppose that $\mathbb{E}(X_{ni}) = 0$ for all n and i, and that Var $X_{ni} = \sigma_{ni}^2 < \infty$. Let $S_n = \sum_{i=1}^{k_n} X_{ni}$, $s_n^2 = \sum_{i=1}^{k_n} \sigma_{ni}^2$ (i.e. sum of the n^{th} row and its variance.)

Suppose for every n, $\{X_{ni}\}_{i=1}^{k_n}$ is independent. Suppose also that the Lindeburg condition holds: i.e. for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{ni}|>\varepsilon s_n\}} X_{ni}^2 dP = 0.$$

Then for every $x \in \mathbb{R}$,

$$P\left\{\frac{S_n}{s_n} \le x\right\} \to \Phi(x).$$

- (Lec 17) **Lyapounov's CLT**: Let X_{ni} be a triangular array such that the X_{ni}'s have mean 0, $\mathbb{E}|X_{ni}|^{2+\delta} < \infty$ for some $\delta > 0$.

  If Lyapunov's condition holds:

  $$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{k_n} \mathbb{E}|X_{ni}|^{2+\delta} = 0,$$

  then $P\left\{\frac{S_n}{s_n} \le x\right\} \to \Phi(x)$.

- **CLT for single variable:** Assume X_i's are i.i.d. with finite mean μ and finite variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then $\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$. Equivalently, $\frac{S_n - n\mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$.

- In trying to prove that the condition for CLT holds (and a direct verification of the conditions doesn't work), consider using truncation or looking at characteristic functions.

- **Lindeberg-Feller**: Consider the set-up for Lindeberg CLT where instead of a triangular array we just have $X_1, X_2, \ldots$ independent. If this sequence satisfies $\max_{k=1,\ldots,n} \frac{\sigma_k^2}{s_n^2} \to 0$ as $n \to \infty$, then Lindeberg's condition is both necessary and sufficient.

- (Lec 17) **Berry-Esseen bounds for CLT approximation:** Assume $X_1, \ldots, X_n$ have mean 0, variance σ_i^2, and $\mathbb{E}|X_i|^3 = r_i$ are finite. Then

  $$\sup_{-\infty<x<\infty} \left| P\left\{\frac{S_n}{s_n} \le x\right\} - \Phi(x) \right| \le \frac{0.78 R_n}{s_n^3},$$

  where $R_n = \sum_{i=1}^n r_i$. If the X_i's are iid, the RHS is $\frac{0.78\mathbb{E}|X_1|^3}{\sigma^3\sqrt{n}}$.

Characteristic functions (Lec 18-20)

- (Lec 18) **Skorohod's Theorem:** Let F_n, F be distribution functions on $\mathbb{R}$ such that $F_n \Rightarrow F$. Then there exist $(\Omega, \mathcal{F}, P)$ and random variables $\{Y_n\}$, Y with distribution functions $\{F_n\}$, F such that $Y_n(\omega) \to Y(\omega)$ for all ω.

- (Lec 19) **Continuity Theorem:** $F_n \Rightarrow F$ iff $\phi_n(t) \to \phi(t)$ for all t. (The same is true for MGFs, if they exist.)

- (Lec 19) **Definition of tightness:** A family of probabilities $\{\mu_n\}$ on $\mathbb{R}$ is **tight** if for every $\varepsilon > 0$, there exist $a < b$ so that $\mu_n(a, b] > 1 - \varepsilon$ for all n. We say that $\{\mu_n\}$ is **almost compactly supported**.

- (Lec 19) $\{\mu_n\}$ is tight iff for every subsequence $\{n_k\}_{k=1}^\infty$, there exists a further subsequence $\{n_{k_i}\}_{i=1}^\infty$ and a probability μ so that $\mu_{n_{k_i}} \Rightarrow \mu$.

- (Lec 19) **Cantor's diagonal argument:** Let $\{X_{ij}\}_{i,j=1}^{\infty}$ be a 2D array of real numbers such that each row is bounded. Then there exists a subsequence $\{n_k\}_{k=1}^{\infty}$ of $\mathbb{N}$ and $\{l_r\}_{r=1}^{\infty}$ such that $x_{rn_k} \to l_r$ as $k \to \infty$ for all $r \in \mathbb{N}$.

- (Lec 19) **Helly selection theorem:** If $\{F_n\}_{n=1}^{\infty}$ are any distribution functions on $\mathbb{R}$, then there exist monotone, right-continuous F and a subsequence $n_k \nearrow \infty$ such that $F_{n_k}(x) \to F(x)$ for all points of continuity x of F. (If $\{F_n\}$ is tight, the subsequence can be chosen suc that limit F is a distribution function.)

- (Lec 20) **Inversion formula:** If $a < b$ with $\mu\{a\} = \mu\{b\} = 0$ (i.e. not atoms of the measure), then

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

- (Lec 20) If $\phi_\mu(t) = \phi_\nu(t)$ for all t, then $\mu = \nu$.

- (Lec 20) **Fourier transform:** If $\phi(t)$ is integrable, and if F is the corresponding distribution function, then F has density $f(x) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} e^{-itx} \phi(t) dt$.

- (Lec 20) A weighted average of characteristic functions is still a characteristic function.

- (Lec 20) **Pólya's criteria:** If ϕ is continuous, non-negative, even, $\phi(0) = 1$, ϕ convex on $(0, \infty)$ and $(-\infty, 0)$, then ϕ is a characteristic function.

- (Lec 20) 2 characteristic functions can agree in a neighborhood of 0 without having the same measure.

- $\phi_{X+Y} = \phi_X \phi_Y$ does **not** imply that X and Y are independent. (E.g. $X = Y =$ standard Cauchy.) However, if X and Y and $\mathbb{R}^d$-valued random variables, then

$$X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad \mathbb{E}\left[e^{t(X,Y)\cdot(\xi,\eta)}\right] = \mathbb{E}\left[e^{tX\cdot\xi} \cdot \mathbb{E}\left[e^{t(X,Y)\cdot(\xi,\eta)}\right]\right]$$