Section Notes for PSTAT 213B

Haosheng Zhou

Jan, 2025

Contents

Week 1

Elements of Measure Theory

The measure-based probability theory is established on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. It consists of the sample space Ω, the sigma field $\mathscr{F}$ (definition?) on Ω, and the probability measure $\mathbb{P} : \mathscr{F} \to [0,1]$. The probability measure satisfies the axioms of probability (definition?), which contains the essential "countable additivity". Within this framework, a random variable $X : \Omega \to \mathbb{R}$ is just a mapping $\omega \mapsto X(\omega)$ that is $\mathscr{F}$-measurable, i.e., $\{\omega : X(\omega) \in B\} \in \mathscr{F}, \forall B \subset \mathbb{R}$ Borel.

The Borel sigma field, denoted $\mathscr{B}_{\mathbb{R}}$, is the collection of all Borel measurable subsets of $\mathbb{R}$. Mathematically speaking, it is defined as $\mathscr{B}_{\mathbb{R}} := \sigma(\mathscr{Q})$, where $\mathscr{Q} := \{(a, b] : a, b \in \mathbb{R}\}$ is a π-system. That is the reason the CDF is defined as $F_X(x) = \mathbb{P}(X \leq x)$. To connect the CDF with the distribution/law of the r.v., we define the law of X, denoted μ_X, as the probability measure on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ induced by X: $\mu_X(B) = \mathbb{P}(X \in B)$ for any Borel set $B \in \mathscr{B}_{\mathbb{R}}$. The following exercise proves that the CDF characterizes the law of a random variable, which is a standard application of the $\pi - \lambda$ theorem.

Exercise 1. *Given μ, ν as two probability measures on $(\Omega, \mathscr{F})$, and $\mathscr{Q}$ is a π-system such that $\sigma(\mathscr{Q}) = \mathscr{F}$. If $\mu_1(A) = \mu_2(A)$ for $\forall A \in \mathscr{Q}$, prove that $\mu_1(A) = \mu_2(A)$ for $\forall A \in \mathscr{F}$.*

Consequently, prove that if μ_X, μ_Y are the laws of r.v. X and Y, then $F_X \equiv F_Y$ iff $\mu_X(B) = \mu_Y(B)$ for $\forall B \in \mathscr{B}$.

Hint. Consider $\mathscr{H} := \{A \in \mathscr{F} : \mu_1(A) = \mu_2(A)\}$ and prove that it is a λ-system. □

The next core topics are integration and the convergence theorems. By definition, the expectation $\mathbb{E}X := \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$ is nothing else but the Lebesgue integral of the r.v. (measurable function) X. By a simple change of variable $x = X(\omega)$, we recover the important identity $\mathbb{E}X = \int_{\mathbb{R}} x\, d\mu_X(x)$ (prove it on your own). The key insight from the Lebesgue integration is that the behavior of the function on any zero-Lebesgue-measure set can be neglected, which naturally motivates the definition of "almost sure" and "almost everywhere" in measure theory. Any pointwise properties that are required to prove some relationships of Lebesgue integrals can be reduced to almost sure/almost everywhere properties without any costs.

One of the key problems w.r.t. the integration is when one has a sequence of random variables X_n and $X_n \overset{a.s.}{\to} X(\omega)$ $(n \to \infty)$ (pointwise can be reduced to almost sure convergence since neglected by the Lebesgue integral). We hope to understand under what conditions $\mathbb{E}X_n \to \mathbb{E}X$ $(n \to \infty)$ holds, i.e. the interchange of the limit and the integration is allowed. Three most important convergence theorems provide the sufficient conditions: (please check)

- Monotone convergence theorem (MCT): $X_n \geq 0$ $a.s.$ for $\forall n,\ \omega \in \Omega$, and X_n is $a.s.$ increasing in n.
- Dominated convergence theorem (DCT): $\sup_n |X_n| \leq Y$ $a.s.$, and $\mathbb{E}Y < \infty$.
- Bounded convergence theorem (BCT): $\sup_n |X_n| \leq M$ $a.s.$, where $M \in \mathbb{R}$.

If one has a weaker condition (e.g., only non-negativity but no monotonicity), it is possible to derive a weaker conclusion, stated by the Fatou's lemma (please check).

The next important concept is the Radon-Nikodym derivative, which often comes up when studying conditional expectation. The problem of interest is that, when having two probability measures $\mathbb{P}, \mathbb{Q}$ on the same measurable space $(\Omega, \mathscr{F})$, if the following representation exists:

$$\mathbb{P}(A) = \int_A f(\omega)\, d\mathbb{Q}(\omega),\ \forall A \in \mathscr{F}, \tag{1}$$

for some measurable function $f : \Omega \to \mathbb{R}$. As a necessary condition, for any $A \in \mathscr{F}$ such that $\mathbb{Q}(A) = 0$, it holds that $\mathbb{P}(A) = 0$. We call this property the absolute continuity of $\mathbb{P}$ w.r.t. $\mathbb{Q}$, denoted $\mathbb{P} << \mathbb{Q}$. Amazingly, this is also the sufficient condition! The Radon-Nikodym theorem states that, if $\mathbb{P} << \mathbb{Q}$ for two probability measures (actually σ-finite, so it applies for λ, the Lebesgue measure on $\mathbb{R}$), then such measurable function f must exist and is almost surely unique such that equation (1) holds. We call such a function f the Radon-Nikodym derivative and denote it by $\frac{d\mathbb{P}}{d\mathbb{Q}}$. The following exercise states some fundamental properties of the RN-derivative.

Exercise 2. *Argue that $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is actually a r.v. on the probability space $(\Omega, \mathscr{F}, \mathbb{Q})$. Check that $\mathbb{E}_{\mathbb{Q}} \frac{d\mathbb{P}}{d\mathbb{Q}} = 1$, where $\mathbb{E}_{\mathbb{Q}}$ denotes the expectation under measure $\mathbb{Q}$.*

If $\frac{d\mathbb{P}}{d\mathbb{Q}} > 0,\ \mathbb{Q} - a.s.$, prove that $\frac{d\mathbb{Q}}{d\mathbb{P}}$ exists and that $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{1}{\frac{d\mathbb{P}}{d\mathbb{Q}}}$ under both $\mathbb{P} - a.s.$ sense and $\mathbb{Q} - a.s.$ sense.

Hint. To identify $\frac{d\mathbb{Q}}{d\mathbb{P}}$, use the almost sure uniqueness of the RN-derivative. □

Why does the notation of the RN derivative look like a differential between two measures? The following exercise provides the intuition.

Exercise 3. *For disjoint sets $A \in \mathscr{F}$ and $\Delta A \in \mathscr{F}$ (understood as the perturbation in A), calculate $\frac{\mathbb{P}(A \cup \Delta A) - \mathbb{P}(A)}{\mathbb{Q}(A \cup \Delta A) - \mathbb{Q}(A)}$. Prove that this difference quotient is equal to $f(\omega_0)$ if $\Delta A = \{\omega_0\}$ and $\mathbb{Q}(\{\omega_0\}) > 0$.*

Argue intuitively that $\frac{\mathbb{P}(A \cup \Delta A) - \mathbb{P}(A)}{\mathbb{Q}(A \cup \Delta A) - \mathbb{Q}(A)} \to f(\Delta A)$ as $\mathbb{Q}(\Delta A) \to 0$, which is interpreted as the relative rate of change in the measure when A receives an infinitesimal perturbation.

Hint. Definition. □

The Lebesgue decomposition theorem implies that for the law μ_X of r.v. X, there exists a unique decomposition of the law $\mu_X = \mu + \nu$ such that $\mu << \lambda$ (absolute continuous w.r.t. Lebesgue measure) and $\nu \perp \lambda$ (singular w.r.t. Lebesgue measure). We say $\nu \perp \lambda$ if there exists disjoint$A, B \in \mathscr{B}$ such that $A \cup B = \mathbb{R}$, but $\nu(A) = \lambda(B) = 0$. This provides the classification of random variables into discrete, continuous, and singular r.v. In some sense, that is why probability densities and probability mass functions are both called "likelihood" in statistics.

Exercise 4. *Consider the probability space $(\Omega = [0, 1], \mathscr{F} = \mathscr{B}_{[0,1]}, \mathbb{P} = \lambda)$, where λ is the Lebesgue measure.*

If $X(\omega) = \begin{cases} 0 & \text{if } \omega < \frac{1}{2} \\ 1 & \text{if } \omega \geq \frac{1}{2} \end{cases}$, find out μ, ν in the Lebesgue decomposition of μ_X (discrete).

If $X(\omega) = \omega$, find out μ, ν in the Lebesgue decomposition of μ_X (continuous).

Show that there exists a Borel measurable set C which is uncountable, but has zero Lebesgue measure. Consider r.v. X that is supported on C. Prove that $\mu \equiv 0$ in the Lebesgue decomposition of μ_X. This is neither a discrete r.v. nor a continuous r.v. (singular).

Hint. Consider Cantor set C. □

After clearing up those concepts for a single r.v., we start to talk about random vectors, most importantly, the concept of independence between two r.v. X and Y. The independence of X and Y is defined as the independence between sigma fields $\sigma(X)$ and $\sigma(Y)$. Two sigma fields $\mathscr{A}, \mathscr{B}$ are defined to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B),\ \forall A \in \mathscr{A}, B \in \mathscr{B}. \tag{2}$$

The following theorem is again an application of the $\pi - \lambda$ theorem.

Exercise 5. *$\mathscr{Q}, \mathscr{R}$ are π-systems and subsets of $\mathscr{F}$. If $\mathscr{Q}, \mathscr{R}$ are independent, then $\sigma(\mathscr{Q}), \sigma(\mathscr{R})$ are independent. Consequently, prove that X, Y are independent iff $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\,\mathbb{P}(Y \leq y)$, $\forall x, y \in \mathbb{R}$.*

Hint. Fix $A \in \mathscr{Q}$, consider $\mathscr{H}_A := \{B \in \sigma(\mathscr{R}) : \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)\}$ and prove $\mathscr{Q}, \sigma(\mathscr{R})$ are independent. Then fix $B \in \mathscr{R}$ and perform the similar argument once more. □

To build a larger probability space where two (finitely many) random variables on $(\Omega_1, \mathscr{F}_1, \mathbb{P})$ and $(\Omega_2, \mathscr{F}_2, \mathbb{Q})$ can live, one sets the sample space as $\Omega_1 \times \Omega_2$ and wishes to build a sigma field $\mathscr{F}_1 \otimes \mathscr{F}_2$ on $\Omega_1 \times \Omega_2$. The procedure is similar to that in one dimension, except setting $\mathscr{F}_1 \otimes \mathscr{F}_2 := \sigma(\mathscr{Q})$, where the π-system $\mathscr{Q} := \{A \times B : A \in \mathscr{F}_1, B \in \mathscr{F}_2\}$ is the collection of measurable rectangles. $\mathscr{F}_1 \otimes \mathscr{F}_2$ is called the product sigma field, which is typically larger than $\mathscr{F}_1 \times \mathscr{F}_2$. Concerning the product probability measure $\mathbb{P} \otimes \mathbb{Q} : \mathscr{F}_1 \otimes \mathscr{F}_2 \to [0, 1]$, it is also only defined on $\mathscr{Q}$ through $(\mathbb{P} \otimes \mathbb{Q})(A \times B) := \mathbb{P}(A) \cdot \mathbb{Q}(B)$ and later extended to the whole $\mathscr{F}_1 \otimes \mathscr{F}_2$. That is the construction of the product probability space $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mathbb{P} \otimes \mathbb{Q})$. Actually, Kolmogorov's extension lemma guarantees the conclusion for uncountably many r.v., provided the consistency condition.

As the last topic, we shall talk about Fubini's theorem when it comes to double integration/summation. If $\mathbb{P}, \mathbb{Q}$ are both sigma finite measures, the key condition is that either the integrand f is a.s. non-negative or integrable, i.e., $\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)|\, d(\mathbb{P} \otimes \mathbb{Q})(\omega_1, \omega_2) < \infty$. Under this, one can interchange the order of integration, i.e.

$$\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2)\, d(\mathbb{P} \otimes \mathbb{Q})(\omega_1, \omega_2) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2)\, d\mathbb{P}(\omega_1)\, d\mathbb{Q}(\omega_2) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, d\mathbb{Q}(\omega_2)\, d\mathbb{P}(\omega_1). \tag{3}$$

The following exercise is one of the most important implications of Fubini's theorem to keep in mind.

Exercise 6. *Prove $\mathbb{E}|X|^p = \int_0^\infty p y^{p-1} \mathbb{P}(|X| > y)\, dy$ for $\forall p > 0$.*

Hint. Apply Fubini's theorem using $|x|^p = \int_0^{|x|} p y^{p-1}\, dy$. □

Week 2

Almost Sure and Convergence in Probability

The following exercise tests the understanding on almost sure and convergence in probability.

Exercise 7. *Consider the probability space* $([0,1], \mathscr{B}_{[0,1]}, \lambda)$*, on which there exists a sequence of random variables* $X_n(\omega) := \mathbb{I}_{(0,\frac{1}{n})}(\omega)$*. Judge if this sequence of r.v. converges a.s./in probability, and identify the limit.*

Repeat the exercise for a sequence of independent r.v. Y_n *on the same probability space such that* $Y_n \overset{d}{=} X_n$*,* $\forall n$*.*

Hint. $\{X_n\}$ converges almost surely but not $\{Y_n\}$. Independence matters. $\square$

Exercise 8 (Convergence for i.i.d. r.v.)**.** *Refer to Lemma 14 (1), (3) in 2024 notes.*

Week 3

Convergence Mode, Uniform Integrability

Exercise 9. *Let* $\{X_n\}$ *be i.i.d. random variables following* $\mathcal{E}(1)$*, and* $M_n := \max_{1\leq i\leq n}\{X_i\}$*, show that*

$$\frac{M_n}{\log n} \xrightarrow{a.s.} 1 \ (n\to\infty). \tag{4}$$

Note that we have already proved $\limsup_{n\to\infty}\frac{X_n}{\log n} = 1$ *a.s. in the homework.*

Hint. Prove the following conclusions under the almost sure sense: $\limsup_{n\to\infty}\frac{M_n}{\log n}\geq 1$, $\limsup_{n\to\infty}\frac{M_n}{\log n}\leq 1$, $\liminf_{n\to\infty}\frac{M_n}{\log n}\geq 1$, among which the first one is obvious.

To prove $\limsup_{n\to\infty}\frac{M_n}{\log n}\leq 1$, split the maximum w.r.t. n terms into the first N terms (finitely many) and the tail $n-N$ terms (infinitely many). Use $\limsup_{n\to\infty}\frac{X_n}{\log n}=1$ to bound the tail part.

To prove $\liminf_{n\to\infty}\frac{M_n}{\log n}\geq 1$, apply Borel-Cantelli and prove $\forall\varepsilon\in(0,1)$, $\sum_{n=1}^\infty \mathbb{P}\left(\frac{M_n}{\log n}<1-\varepsilon\right)<\infty$. □

Exercise 10. *Show that*

$$d(X,Y) := \mathbb{E}\frac{|X-Y|}{1+|X-Y|} \tag{5}$$

is a metric on the space of random variables (the equality is under almost sure sense). Show that $X_n \xrightarrow{p} X$ $(n\to\infty)$ *iff* $d(X_n,X)\to 0$*, which shows that convergence in probability can be embedded into a metric space.*

Hint. The triangle inequality of d follows from the fact that $x\mapsto\frac{x}{1+x}$ is increasing. The equivalence in convergence follows from BCT and Markov inequality. □

Exercise 11. *Show that for* $1\leq p<q\leq\infty$*,* L^q *convergence implies* L^p *convergence.*

Hint. Holder's inequality. □

Exercise 12. *Check that for* $X_n\equiv X$*,* $\{X_n\}$ *is U.I. iff* $X\in L^1$*. This provides the motivation of the def of U.I.*

If $X\in L^1$*, prove that* $\forall\varepsilon>0$*,* $\exists\delta>0$*, such that for any* $A\in\mathscr{F}, \mathbb{P}(A)<\delta$*,* $\mathbb{E}|X|\mathbb{I}_{|X|\in A}<\varepsilon$ *holds.*

Hint. By Cauchy principle, $\forall\varepsilon>0,\ \exists M>0$ such that $\int_{|X(\omega)|\geq M}|X(\omega)|\,d\mathbb{P}(\omega)<\frac{\varepsilon}{2}$. □

Exercise 13. *If* $\{X_n\}$ *and* $\{Y_n\}$ *are U.I., show that* $\{X_n+Y_n\}$ *is U.I.*

Hint. Definition. □

Exercise 14. *Show that* $\{X_n\}$ *is U.I., if one of the following conditions holds:*

(1): (moment) exists $\varepsilon>0$*, such that* $\sup_n\mathbb{E}|X_n|^{1+\varepsilon}<\infty$*.*

(2): (dominated) exists $Y\in L^1$*, such that* $\sup_n|X_n|\leq Y$ *a.s..*

Explain why condition (1) cannot be weakened to $\sup_n\mathbb{E}|X_n|<\infty$*.*

Hint. Part (1): consider $\mathbb{E}\frac{|X_n|}{|X_n|^{1+\varepsilon}}|X_n|^{1+\varepsilon}\mathbb{I}_{|X_n|\geq\lambda}$. Use the fact that $\frac{x}{x^{1+\varepsilon}}\to 0$ $(x\to+\infty)$ to connect with the moment condition. Part (2): by definition.

Counterexample: $X_n(\omega)=n\mathbb{I}_{(0,\frac{1}{n})}(\omega)$. Contradiction follows from Vitali's convergence theorem. □

Week 4

WLLN

Exercise 15. *Let $\{X_{n,k}\}_{1\leq k\leq n}$ be a lower triangular array of r.v. such that each row consists of independent r.v. If there exists a sequence of positive $\{b_n\}$ such that:*

1. $\sum_{k=1}^n \mathbb{P}\left(|X_{n,k} > b_n|\right) \to 0$ *as* $n\to\infty$,
2. $\frac{\sum_{k=1}^n \mathbb{E}X_{n,k}^2 \mathbb{I}_{|X_{n,k}|\leq b_n}}{b_n^2} \to 0$ *as* $n\to\infty$,

then

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0 \ (n\to\infty), \tag{6}$$

where $S_n := \sum_{k=1}^n X_{n,k}$ *and* $a_n := \sum_{k=1}^n \mathbb{E}X_{n,k}\mathbb{I}_{|X_{n,k}|\leq b_n}$.

Hint. Set the truncation $Y_{n,k} := X_{n,k}\mathbb{I}_{|X_{n,k}|\leq b_n}$ and consider $T_n := \sum_{k=1}^n Y_{n,k}$. First prove $\mathbb{P}(S_n \neq T_n) \to 0$ as $n\to\infty$. Then apply Chebyshev's inequality for T_n. □

Exercise 16. $\{X_n\}$ *is a sequence of i.i.d. r.v. following the distribution*

$$\mathbb{P}\left(X_1 = 2^j\right) = 2^{-j}, \ \forall j \geq 1. \tag{7}$$

Let $S_n := \sum_{i=1}^n X_i$. *Prove that* $\limsup_{n\to\infty} \frac{X_n}{n\log_2 n} = \infty$ *a.s. while* $\frac{S_n}{n\log_2 n} \xrightarrow{p} 1$ *as* $n\to\infty$. *This example illustrates the difference between almost sure convergence and convergence in probability.*

Hint. The almost sure part follows from a standard Borel-Cantelli argument.

The convergence in probability part follows from the WLLN above. Consider the form $b_n := 2^{m_n}$ such that $\frac{n}{b_n} \to 0$ as $n\to\infty$ (check two conditions above). One example would be $b_n = n\log_2 n$ and $a_n = nm_n = n\log_2 n + n\log_2\log_2 n$. □

Exercise 17 (An unfair "fair" game)**.** $\{X_n\}$ *is a sequence of i.i.d. r.v. following the distribution*

$$\mathbb{P}\left(X_1 = -1\right) = p_0, \quad \mathbb{P}\left(X_1 = 2^k - 1\right) = p_k, \ \forall k \geq 1, \tag{8}$$

where $p_k := \frac{1}{2^k k(k+1)}$, $\forall k\geq 1$ *and* $p_0 := 1 - \sum_{k\geq 1} p_k$. *Let* $S_n := \sum_{i=1}^n X_i$, *prove that*

$$\frac{S_n}{\frac{n}{\log_2 n}} \xrightarrow{p} -1 \ (n\to\infty). \tag{9}$$

Compare with the conclusion you can draw by directly applying SLLN. What can you observe?

Hint. Apply the WLLN above. Consider $b_n := 2^{m_n}$ where $m_n := \inf\left\{m\in\mathbb{N} : 2^{-m}m^{-\frac{3}{2}} \leq \frac{1}{n}\right\}$. □

Week 5

Applications of SLLN

Exercise 18 (Empirical CDF)**.** *Given i.i.d. samples $\{X_n\}$. Assume X_1 has CDF F. Define*

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{X_i \le x} \tag{10}$$

as the empirical CDF. Show that for any fixed $x \in \mathbb{R}$, $F_n(x) \xrightarrow{a.s.} F(x)$ $(n \to \infty)$. Calculate $Var(F_n(x))$. Based on those conclusions, construct a 95%-pointwise confidence band (interval) for $F_n(x)$ when $n \to \infty$. Look at numerical experiments in Figure 1, what do you think is a possible problem for this confidence band?

Hints. Apply SLLN, $\text{Var}(F_n(x)) = \frac{F(x)[1-F(x)]}{n}$. By CLT, $\frac{F_n(x)-F(x)}{\sqrt{\frac{F(x)[1-F(x)]}{n}}} \xrightarrow{d} N(0,1)$ $(n \to \infty)$ for any fixed $x \in \mathbb{R}$. By Slutsky's theorem, $\frac{F_n(x)-F(x)}{\sqrt{\frac{F_n(x)[1-F_n(x)]}{n}}} \xrightarrow{d} N(0,1)$ $(n \to \infty)$ for any fixed $x \in \mathbb{R}$ (the true F is unknown). Hence the 95%-confidence band is $F_n(x) \pm Z_{0.025}\sqrt{\frac{F_n(x)[1-F_n(x)]}{n}}$. □

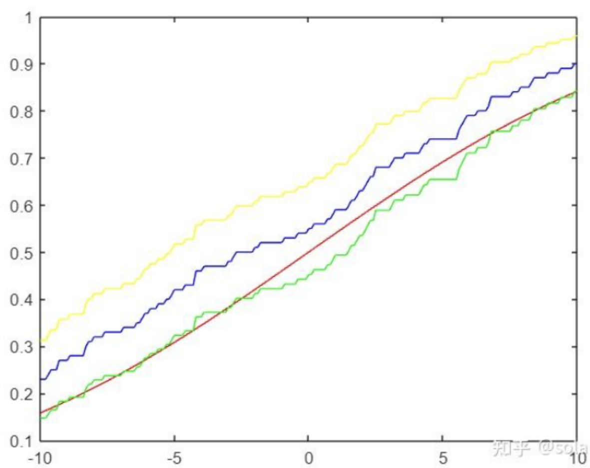

Figure 1: Confidence Band of Empirical CDF
Red line denotes F, the CDF of $N(0,100)$. Blue line denotes a sample of F_n. Yellow line denotes the corresponding 95%-upper confidence band of F_n. Green line denotes the corresponding 95%-lower confidence band of F_n.

Exercise 19 (Dvoretzky-Kiefer-Wolfowitz)**.** $\mathbb{P}\left(\sup_x |F_n(x) - F(x)| > \varepsilon\right) \le 2e^{-2n\varepsilon^2}, \forall \varepsilon > 0$.

Hints. This result is highly non-trivial, don't try to prove it on your own! Check this if interested. □

Exercise 20 (Glivenko-Cantelli)**.** *Prove* $\sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0$ $(n \to \infty)$*. What's the difference between this and the first conclusion? Rebuild the 95%-confidence band according to DKW. Look at numerical experiments in Figure 2,* *is that better?*

Hints. Borel-Cantelli. This conclusion is uniform in x. Set $0.05 = 2e^{-2n\varepsilon^2}$ so that $\varepsilon = \sqrt{-\frac{1}{2n}\log 0.025}$, the 95%-uniform confidence band is $\left[0 \vee F_n(x) - \sqrt{-\frac{1}{2n}\log 0.025}, 1 \wedge F_n(x) + \sqrt{-\frac{1}{2n}\log 0.025}\right]$. □

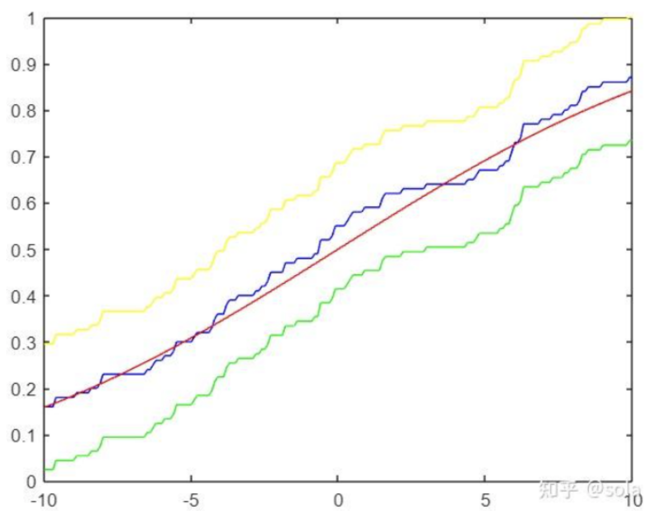

Figure 2: Uniform Confidence Band of Empirical CDF
Red line denotes F, the CDF of $N(0, 100)$. Blue line denotes a sample of F_n. Yellow line denotes the corresponding 95%-upper uniform confidence band of F_n. Green line denotes the corresponding 95%-lower uniform confidence band of F_n.

Exercise 21 (Entropy)**.** *Let $\{X_n\}$ be i.i.d. sequence of letters taking values in $\{1, 2, ..., r\}$, $p_k := \mathbb{P}(X_1 = k) > 0$. For a randomly typed letter sequence $X_1...X_n$, let $\pi_n := \prod_{i=1}^n p_{X_i}$ denote the probability of observing such a sequence. Prove $-\frac{\log \pi_n}{n} \xrightarrow{a.s.} H := -\sum_{k=1}^r p_k \log p_k$ $(n \to \infty)$, the limit is defined as the entropy of the distribution, measuring how chaotic/random the distribution is.*

Hints. SLLN. □

Exercise 22. *Calculate the discrete distribution supported on $\{1, 2, ..., n\}$ that has the maximum entropy.*

Hints. Use Lagrange multiplier. The uniform distribution. □

Exercise 23. *Calculate the discrete distribution supported on $\{1, 2, ..., n\}$ that has the maximum entropy and also exhibits the conservation of energy, i.e., $\sum_{i=1}^n p_i E_i = U$.*

Hints. Use Lagrange multiplier. The Boltzmann distribution w.r.t. energy levels $E_1, ..., E_n$. □

Exercise 24. *Calculate the continuous distribution that has the maximum entropy, subject to a zero mean and a unit variance.*

Hints. Require the calculus of variation. $N(0, 1)$. □

Week 6

Weak Convergence

Exercise 25. *Prove that the Levy metric between two CDFs F and G*

$$d(F,G) := \inf \{\delta > 0 : F(x-\delta) - \delta \leq G(x) \leq F(x+\delta) + \delta,\ \forall x \in \mathbb{R}\} \tag{11}$$

is indeed a metric.

Hints. First prove symmetricity. Then use symmetricity and right-continuity of CDF to prove positivity. Finally prove triangle inequality by using the definition of the infimum in Levy metric. □

Exercise 26. *Prove that $F_n \xrightarrow{d} F$ iff $d(F_n, F) \to 0$ as $n \to \infty$.*

Hints. One direction is obvious by setting $n \to \infty$. For the other direction, cut the real line at points $x_1 < ... < x_k \in C(F)$ such that the spacing is less than ε and the probability mass left outside $[x_1, x_k]$ is less then ε. Prove that in this case $d(F_n, F) < 2\varepsilon$. □

Exercise 27. *Discuss for different convergence modes if $X_n \to X$ implies that $\{X_n\}$ and X live in the same probability space? Try to construct an example if the answer is negative. How shall we understand the Skorohod theorem?*

Hints. Negative for weak convergence. □

Exercise 28 (Slutsky)**.** *If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c \in \mathbb{R}$, prove $(X_n, Y_n) \xrightarrow{d} (X, c)$ as $n \to \infty$. Use this to show that the T-statistic $T_n := \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ for i.i.d. samples $X_1, ..., X_n$ with mean μ and variance σ^2 is asymptotically Gaussian, i.e., $T_n \xrightarrow{d} N(0,1)$ as $n \to \infty$.*

Hints. Definition of weak convergence. □

Week 8

Conditional Expectation

Keep in mind the definition of conditional expectation of $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, where $\mathscr{G} \subset \mathscr{F}$:

1. $\mathbb{E}(X|\mathscr{G}) \in \mathscr{G}$
2. $\forall A \in \mathscr{G}, \mathbb{E}(X\mathbb{I}_A) = \mathbb{E}(\mathbb{E}(X|\mathscr{G}) \cdot \mathbb{I}_A)$

Those two properties actually **characterize** CE, due to the almost sure uniqueness of CE as a consequence of Radon-Nikodym theorem. This is crucial in proving the properties of CE.

Exercise 29. *Check that $\mu(A) := \mathbb{E}(X\mathbb{I}_A)$ is a finite signed measure on the measurable space $(\Omega, \mathscr{G})$ if $X \in L^1$. Check that $\mu << \mathbb{P}$ on $(\Omega, \mathscr{G})$ and show that $\mathbb{E}(X|\mathscr{G}) = \frac{d\mu}{d\mathbb{P}}\Big|_{\mathscr{G}}$. CE is almost surely unique under what sense?*

Hints. The countable additivity comes from DCT. CE is $\mathbb{P}|_{\mathscr{G}} - a.s.$ unique. □

Exercise 30. *Let $Y \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, and X_1, X_2 are independent r.v.s such that X_2 is independent of Y. Prove that $\mathbb{E}(Y|X_1, X_2) = \mathbb{E}(Y|X_1)$ under the almost sure sense.*

Hints. Firstly, $\mathbb{E}(Y|X_1) \in \sigma(X_1, X_2)$, only need to prove $\forall A \in \sigma(X_1, X_2), \mathbb{E}(Y\mathbb{I}_A) = \mathbb{E}(\mathbb{E}(Y|X_1)\mathbb{I}_A)$ by definition.

First prove this property holds for measurable rectangles $A = B \times C \in \mathscr{R}$, where $B \in \sigma(X_1), C \in \sigma(X_2)$.

Then notice that $\sigma(X_1, X_2) = \sigma(\mathscr{R})$ and apply $\pi - \lambda$ theorem to conclude the proof. □

Exercise 31 (CE for a given value)**.** *Show that there exists a Borel measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}(Y|X) = h(X)$. Define $\mathbb{E}(Y|X = x) := \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)}\, dy$ if the density functions exist, $Y \in L^1$ and $f_X(x) \neq 0$. Show that $\mathbb{E}(Y|X = x) = h(x)$. What does this proposition tell us?*

Hints. The first result follows directly from $\mathbb{E}(Y|X) \in \sigma(X)$.

Denote $g(x) := \int y \frac{f_{(X,Y)}(x,y)}{f_X(x)}\, dy$ and it suffices to show that $\mathbb{E}(Y|X) = g(X)$. Use the definition of CE to conclude the proof. □

Exercise 32 (Borel Paradox)**.** *Consider X_1, X_2 i.i.d. having exponential distribution with mean θ. If X_2 is observed, what is the maximum likelihood estimator (MLE) for θ?*

Now consider $Z := \frac{X_2 - 1}{X_1}$, calculate the joint density of (X_1, Z) and the marginal density of Z.

Suppose one observes $X_2 = 1$, and notice that $\{X_2 = 1\} = \{Z = 0\}$. Compare the two MLEs for θ derived from the observation of X_2 and Z respectively. What do you observe? Why is this happening?

Hints. $\hat{\theta} = X_2$ if X_2 is observed. The joint density $f_{(X_1,Z)}(x, z) = \frac{x}{\theta^2} e^{-\frac{1+x+xz}{\theta}}$ $(x > 0, 1 + xz > 0)$ and the marginal density is

$$f_Z(z) = \begin{cases} (1+z)^{-2} e^{-\frac{1}{\theta}} & z \geq 0 \\ \frac{(1+z-\theta z)e^{\frac{1+z}{\theta z}} + \theta z}{\theta z (1+z)^2} e^{-\frac{1}{\theta}} & z < 0 \end{cases}. \quad (12)$$

When $X_2 = 1$ is observed, the first MLE gives $\hat{\theta} = 1$ while the second MLE through Z gives $\hat{\theta} = +\infty$ even if $\{X_2 = 1\} = \{Z = 0\}$. This is because the zero measure event $\{X_2 = 1\}$ can be approximated in different ways. □