

Chapter 1 : Probability Basics

Reference: Srinivasa Varadhan, *Probability Theory*

There is a space Ω of all possible outcomes and each individual outcome is represented as a point ω in that space. Subsets of Ω are **events** and each of them corresponds to a collection of outcomes. $\omega \in A \implies A$ has occurred.

1.1 Introduction

1.1.1 Definitions

Universal set S , Sigma field \mathcal{S} of subsets of S , Probability measure \mathbb{P}

Sample Space: (S, \mathcal{S}) ; Probability Space: $(S, \mathcal{S}, \mathbb{P})$

S is the set of outcomes, its power set is the Sample Space, whose subset is an event, collection of all of which forms a σ -field or Event Space.

Fields milder as their final criterion does not countability but only finiteness

A probability measure must satisfy Countable Additivity (alongside unit properties)

$\mathcal{S} \in 2^S =: \sigma\text{-field} \equiv$ inclusion of trivial subsets, complements and arbitrary union & thus intersection (for field, all is same except arbitrary becomes finite)

Intersection of σ -fields is also a σ -field, so $\sigma(\mathcal{C}) = \bigcap_{t \in \Lambda} \mathcal{S}_t$ is the minimal σ -field over \mathcal{C} . For the set of all intervals $I \subset [0, 1]$, $\sigma(I) = \mathcal{B}([0, 1])$, and $B \in \mathcal{B}$ is a borel set.

Probability satisfies the inclusion-exclusion principle. $\mathbb{P}(E) = \sum_{s_i \in E} p_i$

Discrete Probability Space: countable $p_i \geq 0$ adding upto 1 ($\implies \inf(p_i) = 0$).

Discrete Symmetric Probability: for finite S , $\mathcal{S} = \mathcal{S}_* = 2^S$, $\mathcal{P}(E) = \frac{|E|}{|S|}$

Bonferroni's identity: sum of events minus pairwise intersections \leq union \leq sum

1.1.2 Measure Theory

Universal inclusion, closure under finite union and complementation yield a field/**algebra**.

Extending it to countable union gives a σ -**field**, and a **probability measure** over it satisfies non-negativity, normality and disjoint countable additivity.

For σ -finite measure P over algebra \mathcal{A} , P 's countably additivity over \mathcal{A} is equivalent to: for any sequence $A_n \in \mathcal{A}$ with $A_n \downarrow \phi$, we have $P(A_n) \downarrow 0$.

Caratheodary's Extension Theorem: any probability measure P over algebra \mathcal{A} can be **extended uniquely** over the σ -field $\mathcal{F}(\mathcal{A} = \mathcal{F}$

Outer measure $P^*(A) = \inf\{\sum_{j=1}^{\infty} P(A_j) \mid A \subset \bigcup_{i=1}^{\infty} A_i \subset \mathcal{F}\}$ is countably subadditive

Class \mathcal{M} of measurable sets is $\mathcal{M} = \{A \in \mathcal{F} \mid \forall E \subset \Omega : P^*(E) = P^*(E \cap A) + P^*(E \cap A^c)\}$

\mathcal{M} is a σ -field containing \mathcal{A} , hence \mathcal{F} , and P^* extends P from \mathcal{A} to \mathcal{F}

For any algebra \mathcal{A} , $M(\mathcal{A}) = \mathcal{F}(\mathcal{A})$ where M denotes the monotone class

For each right continuous, non decreasing function $F(x)$ with $F(-\infty) = 0$ and $F(\infty) = 1$, there is a unique probability measure P on $\mathcal{B}(\mathbb{R})$ such that $F(x) = P(-\infty, x)$.

Lebesgue: P is countably additive if and only if F is right continuous

1.1.3 Fundamental Theorems

To find prob after n steps, consider total prob rule on step k after k-1 and induct

Bayes' theorem: $\mathbb{P}(G|E) = \mathbb{P}(E|G) \cdot \mathbb{P}(G) / \sum_{i=1}^N \mathbb{P}(E|G_i) \cdot \mathbb{P}(G_i)$

Independence (product for any event collection) implies pairwise independence [unidir]

Borel-Cantelli lemma: for countable events E_i , $\sum \mathbb{P}(E_i) < \infty \implies \mathbb{P}(\limsup E_i) = 0$

$\lim_{n \rightarrow \infty} \mathbb{P}(E) = \mathbb{P}(\bigcup_{n=1}^{\infty} E_n)$ for increasing sequence of E_i (standard set partial order)

Sup is union, inf is intersection. Taking lim just complementing the operation outside

Fatou's lemma: $P(\liminf) \leq \liminf(P) \leq \limsup(P) \leq P(\limsup)$

$\mathbb{P}(E|G) = \mathbb{P}(E \cap G) / \mathbb{P}(G) \implies \mathbb{P}(E) = \sum_{i=1}^N \mathbb{P}(E|G_i) \cdot \mathbb{P}(G_i)$ where G_i partition S

1.2 Random Events

1.2.1 Random Variables

A discrete real-valued random variable X is a function with domain as outcomes and a countable range $\{x_1, x_2, \dots\}$ of reals such that $\{\omega : X(\omega) = x_i\}$ is an event for all i

X is a random variable $\iff X^{-1}(B) = \{s \in S \mid X(s) \in B\} \in \mathcal{S} \quad \forall B \in \mathcal{B}(\mathbb{R})$

In other words, the pullback of a borel set must be an event (union of outcomes) in S

X is a continuous real RV $\iff \{s \in S \mid X(s) \leq x\} =: \{X \leq x\} \in \mathcal{S} \quad \forall x \in \mathbb{R}$

1.2.2 Random Vectors

The compilation of r real random variables forms a random vector

Density function for random vector \mathbf{X} is $f(x_1, \dots, x_r) = \mathbb{P}(X_1 = x_1, \dots, X_r = x_r)$

The set of vectors $\mathbf{x} \in \mathbb{R}^r$ such that $\mathbb{P}(\mathbf{x}) > 0$ is countable and sum is 1

For random vector \mathbf{X} , $f_{\mathbf{X}}$ is the joint density and f_{X_i} is marginal density

For $\mathbf{Z} = (X, Y)$ we have $f_X(x) = \sum_y f_{\mathbf{Z}}(x, y)$ and $f_Y(y) = \sum_x f_{\mathbf{Z}}(x, y)$

Various discrete probability distributions:

- Geometric: $f(x) = p(1-p)^x$; mean $(1-p)/p$ variance
- Poisson: $f(x) = \lambda^x / x! \cdot e^{-\lambda}$; mean λ variance λ
- Binomial: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; mean np variance $np(1-p)$
- Negative Binomial: $f(x) = \binom{x+r-1}{r-1} p^r (1-p)^x$; mean: $\frac{r(1-p)}{p}$, variance: $\frac{r(1-p)}{p^2}$
- Hypergeometric: $f(x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$; mean: $n \frac{K}{N}$, variance: $n \frac{K}{N} (1 - \frac{K}{N}) \frac{N-n}{N}$.

1.2.3 Expectation

Expected value of discrete RV X is its prob-weighted mean $\sum x \cdot f(x)$ over its range

For continuous random variables, the expectation is $E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)$

Even if the expected value by formula exists, it need not be defined. It's defined iff the mod version of the formula ($\sum |x|f(x)$) exists and is finite, not the customary one.

If $Z = \phi(\mathbf{x})$ is RV of random vector \mathbf{x} , then $E[Z] = \sum \phi(\mathbf{x})f(\mathbf{x})$ over sample space
 Independent random variables X, Y satisfy $\mathbb{P}(X = x, Y = y) = f_X(x)f_Y(y)$ and thus $E[XY] = E[X]E[Y]$ if X, Y are independent (not iff condition)
 Discrete expectation can also be computed using $\sum xf_X(x) = \sum \mathbb{P}(X \geq x)$

1.3 Analysis of Probability

1.3.1 Density-Distribution Functions

Distribution function $F : \mathbb{R} \rightarrow [0, 1]$ of real random variable X is $F(x) = \mathbb{P}(X \leq x)$

Discrete density function $f : \mathbb{R} \rightarrow [0, 1]$ is defined as $f(x) = \mathbb{P}(X = x) = F'(x)$ (if defined)

The set of possible events of X (so $f(x) > 0$) is countable and $\sum_i f(x_i) = 1$ over it
 A random variable is continuous, iff so is its distribution function (leading integral)
 Distribution function should be non-decreasing and one-side continuous, and they must asymptotically go from -1 to 1 , going left to right

Density function $f : \mathbb{R} \rightarrow [0, 1]$ of a real continuous RV X is a function satisfying $\int_{-\infty}^{\infty} f(x)dx = 1$ such that $F(x) = \int_{-\infty}^x f(y)dy \forall x \in \mathbb{R}$ (F absolutely continuous)

Density functions need not be bounded between 0 and 1 since continuous density is 0 at any given point, so $f(x) \neq \mathbb{P}(X = x) = 0$. The only fact is $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$

1.3.2 Discrete Moments

X has a moment of order r if X^r has finite expectation, and in that case we define it as $E[X^r]$. Analogously the central moment is given by $E[(X - \mu)^r]$

X has a moment of order $r \implies X$ has moments of all orders $k \leq r$

Existence of r^{th} moment of X_i yields the existence of that of their linear combinations

$$\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - E[X]^2$$

$E[(X - a)^2] = \text{Var}[X] + (\mu - a)^2$: proves the minimality of variance around the mean

Moment Generating Function of X : $\phi_X(t) = \sum_{x=0}^{\infty} f_X(x)t^x = E[t^x]$

$$\phi'(1) = E[X], \quad \phi''(1) + \phi'(1) - (\phi'(1))^2 = \text{Var}[X]$$

$$\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab \cdot \text{Cov}[X, Y] : \text{covariance of } X \text{ and } Y$$

$$\text{Cov}[X, Y] = E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y]$$

$$\text{Var}[\sum X_i] = \sum \text{Var}[X_i] + 2 \sum \sum \text{Cov}[X_i, X_j] \text{ (single pair counting)}$$

$$\text{Correlation Coefficient } \rho(X, Y) = \text{Cov}[X, Y] / \sqrt{\text{Var}[X] \cdot \text{Var}[Y]}$$

1.3.3 Continuous Random Variables

Various continuous probability distributions:

- Uniform: $f(x) = 1/(b - a)$ for $a \leq x \leq b$; mean $(a + b)/2$
- Normal: $\Phi(x; \mu, \sigma) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} / \sqrt{2\pi\sigma^2}$; mean μ variance σ^2

- Exponential: $f(x) = \lambda e^{-\lambda x}$; mean $1/\lambda$ variance $1/\lambda^2$
- Gamma: $f(x; \alpha, \lambda) = x^{\alpha-1} e^{-\lambda x} \cdot \lambda^\alpha / \Gamma(\alpha)$; mean α/λ variance α/λ^2

Let $\phi : I \rightarrow \mathbb{R}$ be a differentiable monotonic function. Let random variable X have PDF $f(x) = 0 \forall x \notin I$. Then $Y = \phi(X)$ has density given by $g(y) = 0 \forall y \notin \phi(I)$ and

$$g(y) = f(\phi^{-1}(y)) \left| \frac{d\phi^{-1}(y)}{dy} \right| \quad \forall y \in \phi(I)$$

In general you obtain the densities of related variables by converting to their distributions and then analysing

Cauchy Density: $f(x) = 1/\pi(1+x^2)$; $F(x) = 1/2 + \arctan(x)/\pi$ (Cauchy RV)

X is exponentially distributed $\iff \mathbb{P}(X > a+b \mid X > a) = \mathbb{P}(X > b)$

If X is normally distributed its square follows a gamma distribution

1.4 Convergence

Let X be a random variable defined on $(S, \mathcal{C}, \mathbb{P})$, and let $\{X_n\}$ also be defined on the same probability space. Further, let $A = \{s \in S \mid \lim_{n \rightarrow \infty} X_n(s) = X(s)\}$. Then, the sequence $\{X_n\}$ is said to converge to X **almost surely** if $\mathbb{P}(A) = 1$. We write $X_n \xrightarrow{\text{a.s.}} X$ or $X_n \rightarrow X$ a.s.

Let a sequence of random variables $\{X_n\}$ and a random variable X be defined on a probability space $(S, \mathcal{C}, \mathbb{P})$. Then X_n is said to **converge in probability** to X if, for each $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$. We write $X_n \xrightarrow{\mathbb{P}} X$ or $X_n \rightarrow X$ i.p.

Let $\{F_n\}$ be a sequence of distribution functions. If there exists a distribution function F such that, as $n \rightarrow \infty$, $F_n(x) \rightarrow F(x)$ at every point x at which F is continuous, we say that F_n converges in law or **converges weakly** to F and write $F_n \xrightarrow{w} F$.

Let $\{X_n\}$ be a sequence of random variables and let $\{F_n\}$ be the corresponding sequence of distribution functions. We say that X_n **converges in distribution** to X if $F_n \xrightarrow{w} F_X$ weakly, where F_X is the distribution function of X . In this case, we write $X_n \xrightarrow{d} X$.

Chapter 2 : Probability PhD

2.1 Random Variables & Convergence

A random variable or measurable function is a map $f : \Omega \rightarrow \mathbb{R}$ such that every Borel Set $B \in \mathcal{B}(\mathbb{R})$ is pulled back to a measurable subset of Ω .

f is measurable $\iff f^{-1}(B = (-\infty, x]) = \{\omega : f(\omega) \leq x\} \in \mathcal{F} \forall x \in \mathbb{R}$

A sequence of functions f_n converges to a function f everywhere or **pointwise** if $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega) \forall \omega \in \Omega$

f_n converges to f **almost surely** if the set on which it doesn't converge is measure 0

$S = \{\omega \in \Omega \mid \lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)\} \implies P(S^c) = 0$

f_n converges to f **in probability** if the set on which it nontrivially deviates is measure $\lim_{n \rightarrow \infty} P[\omega : |f_n(\omega) - f(\omega)| \geq \epsilon] = 0 \forall \epsilon \in \mathbb{R}^+$

Pointwise is strongest, followed by almost sure and then in probability.

Pointwise convergence \implies almost sure convergence \implies convergence in probability

If $f_n \rightarrow f$ in probability then there's a subsequence such that $f_{n_j} \rightarrow f$ almost surely

2.2 Multiple Events

2.2.1 Joint Density

For continuous RV X, Y over the same space: $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$

$\mathbb{P}(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

Marginal distribution functions: $F_X(z) = \mathbb{P}(X \leq z) = F(z, \infty)$

If there's a non-negative function f such that $F(x, y) = \int_{-\infty}^x (\int_{-\infty}^y f(u, v) dv) du$ then it's termed as the **joint density function**

For any area $A \in \mathbb{R}^2$: $\mathbb{P}((X, Y) \in A) = \int \int_A f(x, y) dx dy$

Marginal density functions: $f_X(z) = \int_{-\infty}^{\infty} f(z, y) dy$ and $F_X(z) = \int_{-\infty}^z f_X(x) dx$

By calculus: $\partial^2 F(x, y) / \partial x \partial y = f(x, y)$

X and Y are independent iff $\mathbb{P}(a < X \leq b, c < Y \leq d) = \mathbb{P}(a < X \leq b) \cdot \mathbb{P}(c < Y \leq d)$

Independence implies $F(x, y) = F_X(x) \cdot F_Y(y)$

2.2.2 Convolutional Form

For X, Y with joint density f and $Z = \phi(X, Y)$, define $A_z = \{(x, y) \mid \phi(x, y) \leq z\}$. Then if $g : \mathbb{R} \rightarrow \mathbb{R}^+$ satisfies the following equation $\forall z \in \mathbb{R}$ it is the density of Z

$$\iint_{A_z} f(x, y) dx dy = \int_{-\infty}^z g(u) du$$

Using the above, we get $f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x) dx \quad \forall z \in \mathbb{R}$

The RHS becomes a product if the RVs are independent, making it their **convolution**

If $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$ then $X + Y \sim \Gamma(\alpha + \beta, \lambda)$

If $X \sim n(\mu_1, \sigma_1^2)$, $Y \sim n(\mu_2, \sigma_2^2)$ for independent X, Y then $X + Y \sim n(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

If $X \sim n(\mu, \sigma^2)$, $Y = (X - \mu)^2$ follows gamma density $\Gamma(1/2, 1/2\sigma^2)$

2.2.3 Independence of Events

Random variables X_1, \dots, X_n are **independent** if their joint distribution function is the product of the marginals, $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$.

This implies that for independent variables, the probability of an intersection of events is multiplicative: $P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i)$.

This extends to expectations, $E[g_1(X_1) \dots g_n(X_n)] = \prod_{i=1}^n E[g_i(X_i)]$.

A collection of n random variables $\{X_1, X_2, \dots, X_n\}$ is said to be mutually independent (or simply independent) iff: $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \prod_{k=1}^n F_{X_k}(x_k)$ where F_{X_1, X_2, \dots, X_n} is the joint cumulative distribution function, and F_{X_k} is the marginal cumulative distribution function of X_k .

Discrete random variables X_1, X_2, \dots, X_r are mutually independent iff $f(\mathbf{X}) = f_1(x_1)f_2(x_2) \dots f_r(x_r)$

For mutually independent X_i and sets $A_i \in \mathbb{R}$, $\mathbb{P}(X_i \in A_i \forall i \in [r]) = \prod \mathbb{P}(X_i \in A_i)$

2.2.4 Conditional Joint Density

Conditional probability density of Y given X , i.e. $f_{Y|X}$ is given by $f(x, y)/f_X(x)$

Based on this, we define $\mathbb{P}(a \leq Y \leq b | X = x) = \int_a^b f(y|x) dy$

Consequently, $\mathbb{P}(Y \leq y | X = x) = \int_{-\infty}^y f(x, t)/f_X(x) dt$

Expectation of Y given $X = x$ is $E[Y|X = x] = \int_{-\infty}^{\infty} yf(y|x)dy$

Also note that $E[E[Y|X]] = E[Y]$

2.3 Moment Analysis

2.3.1 Moment Generating Functions

Moment Generating Function $M_X(t)$ of random variable X is a function $M_X : D \rightarrow \mathbb{R}$ defined by $M_X(t) = E[e^{tX}]$, where $D \subset \mathbb{R}$ is the set of all t which renders $M_X(t)$ finite

For $X \sim n(x; \mu, \sigma^2) : M_X(t) = e^{\mu t} e^{\sigma^2 t^2 / 2} \quad \forall t \in \mathbb{R}$

For $X \sim \Gamma(\alpha, \lambda) : M_X(t) = (\lambda / \lambda - t)^{-\alpha} \quad \forall t < \lambda$

For X with mean μ and PDF f , m^{th} moment is the expectation $E[X^m]$, also computed as $EX^m = \int_{-\infty}^{\infty} x^m f(x) dx$. Consequently variance $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx$

For normal distribution $n(x; \mu, \sigma^2)$, $Y = (X - \mu)^2$ follows gamma density $\Gamma(1/2, 1/2\sigma^2)$

For $X \sim n(x; \mu, \sigma^2) : E[X^{2n+1}] = 0$ and $E[X^{2n}] = \sigma^{2n} (2n)! / 2^n n!$

Independent variables $X, Y \implies M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

$M_X(t) = E[e^{tX}] = E \sum t^n X^n / n! = \sum E X^n \cdot t^n / n!$; Taylor: $M_X(t) = \sum t^n / n! \cdot M_X^{(n)}(t)$

Thus like an EGF we get: $\mathbf{E}[\mathbf{X}^n] = \left. \frac{d^n \mathbf{M}_X(t)}{dt^n} \right|_{t=0}$

2.3.2 Characteristic Functions

The characteristic function of a random variable X is defined as $\varphi_X(t) = \mathbb{E}[e^{itX}]$: $i^2 = -1$
 $\varphi_X(t)$ is finite for all random variables X and all real numbers t .

$$\varphi_X(0) = 1, |\varphi_X(t)| = |E[e^{itX}]| \leq 1 \quad \forall t \in \mathbb{R}$$

For a sum of independent random variables $S_n = X_1 + \dots + X_n$, the characteristic function of the sum is the product of the individual characteristic functions:

$$\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t)$$

$$\text{For } X \sim U(a, b) : \varphi_X(t) = te^{it(a+b)/2} \sin[(b-a)t/2]/2(b-a)$$

$$\text{For } X \sim \text{exp}(\lambda) = \Gamma(1, \lambda) : \varphi_X(t) = \lambda/(\lambda - it) \quad \forall t \in \mathbb{R}$$

$$\text{For } X \sim n(x; n, \sigma^2) : \varphi_X(t) = e^{itx} \cdot e^{-\sigma^2 t^2/2}$$

$$\varphi_X^{(n)}(0) = i^n \cdot E[X^n] : \varphi_X(t) = M_X(it)$$

Independent variables $X, Y \implies \varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \quad \forall t \in \mathbb{R}$

For discrete random variables: $\varphi_{\mathbf{X}}(\mathbf{t}) = \sum \mathbf{e}^{i\mathbf{t}\mathbf{j}} \mathbf{f}_{\mathbf{X}}(\mathbf{j})$

2.3.3 Inversion Formula

$$f_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi_X(t) dt \quad \forall k \in \mathbb{N}, t \in \mathbb{R}$$

For integer IIDs X_i , define $S_n = X_1 + X_2 + \dots + X_n$ then $\varphi_{S_n}(t) = \varphi_X(t)^n$ and thus:

$$f_{S_n}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} (\varphi_X(t))^n dt$$

If two random variables have same characteristic function, they have same distribution function.
 Distribution functions depend continuously on their characteristic functions

2.4 Inequalities

Schwarz Inequality: for finite second momented X, Y : $E[XY]^2 \leq E[X]^2 E[Y]^2$

This yields $Cov[X, Y]^2 \leq Var[X]Var[Y] \implies |\rho(X, Y)| \leq 1$

For any non-negative random variable X : $\mathbb{P}(X \geq t) \leq E[X]/t \quad \forall t \in \mathbb{R}^+$

2.4.1 Chebyshev's Inequality

For a random variable X with finite mean μ and finite variance σ^2 , Chebyshev's inequality gives an upper bound on the probability that X deviates from its mean by a certain amount. For any constant $a > 0$:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

This is a universal bound that applies to any probability distribution, making it a key tool in proving convergence results like the Weak Law of Large Numbers.

2.4.2 Weak Law of Large Numbers

Weak Law of Large Numbers: for IIDs X_i with finite mean μ , let $S_n = \sum X_i$. Then for any $L \in \mathbb{R}^+$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{S_n}{n} - \mu \right| \geq L \right) = 0$$

Standard integral: $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$

2.5 Kolmogorov's Theorems

2.5.1 Kolmogorov's 0-1 Law

A **tail event** for a sequence of random variables $\{X_n\}$ is an event whose occurrence is independent of any finite prefix of the sequence. The collection of all such events forms the tail σ -algebra $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$.

The 0-1 law states that for a sequence of **independent** random variables, any tail event $A \in \mathcal{T}$ must have a probability of either 0 or 1. For example, the event that the series $\sum X_n$ converges is a tail event.

2.5.2 Kolmogorov's Three-Series Theorem

This theorem provides necessary and sufficient conditions for the **almost sure convergence** of a sum of independent random variables $\sum X_n$. For a truncation constant $A > 0$, let $Y_n = X_n \mathbf{1}_{|X_n| \leq A}$. The series $\sum X_n$ converges almost surely if and only if the following three series all converge:

1. $\sum_{n=1}^{\infty} P(|X_n| > A)$
2. $\sum_{n=1}^{\infty} E[Y_n]$
3. $\sum_{n=1}^{\infty} \text{Var}(Y_n)$

2.6 Lévy-Cramér Continuity Theorem

This theorem provides an equivalence between the convergence of a sequence of characteristic functions and the convergence of the corresponding random variables in distribution.

Let $\{X_n\}$ be a sequence of random variables with characteristic functions $\{\phi_n(t)\}$. The sequence $\{X_n\}$ converges **in distribution** to a random variable X ($X_n \xrightarrow{d} X$) if and only if:

1. The sequence $\{\phi_n(t)\}$ converges pointwise to a function $\phi(t)$ for all $t \in \mathbb{R}$.
2. The limit function $\phi(t)$ is continuous at $t = 0$.

When these conditions hold, $\phi(t)$ is the characteristic function of the limiting random variable X .

2.7 Laws of Large Numbers

These theorems describe the long-term behavior of the sample mean of a sequence of random variables. Let X_1, X_2, \dots be a sequence of i.i.d. random variables with finite mean $E[X_i] = \mu$, and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ be the sample mean.

2.7.1 Weak Law of Large Numbers (WLLN)

The sample mean converges **in probability** to the true mean μ .

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0 \quad \forall \epsilon > 0$$

2.7.2 Strong Law of Large Numbers (SLLN)

The sample mean converges **almost surely** to the true mean μ .

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$$

The SLLN is a stronger result, as almost sure convergence implies convergence in probability.

2.8 Central Limit Theorem

For IIDs X_i with mean μ and variance σ^2 each, the random variable $S_n = \sum_1^n X_i = X_1 + \dots + X_n$ has mean $n\mu$ and variance $n\sigma^2$

Based on density function f of X_i , we recursively define $f_{S_n}(x) = \int f_{S_{n-1}}(y) \cdot f(x - y) dy$

The standard normal distribution is $\Phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ with mean 0 and variance 1

Central Limit Theorem: IIDs X_i with finite mean μ , variance σ^2 and $S_n = \sum X_i$:

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}$$

The power of this theorem lies in the sheer generality of its applicability: for ANY random variable with ANY distribution this result holds, even if no moments beyond order 2 exist. Discrete or continuous or neither, no concerns.

2.9 Martingales

Lyapunov's condition is strong and implies Lindeberg's condition which is general.

For some $\delta > 0, \forall \epsilon > 0$:

$$\text{Lyapunov: } \lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \int |x|^{2+\delta} d\alpha_j = 0$$

$$\text{Lindeberg: } \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i| \geq \epsilon s_n} x^2 d\alpha_i = 0$$