Chapter 1

Probability

§ Sample Space and Events.

- Sample Space (Ω) - Set of all possible outcomes of an experiment.

  $\omega \in \Omega$ are called sample outcomes/realization or elements.

- Event (A) - A subset of a sample space.

§ Probability

Definition (Probability)

$\mathbb{P}: \mathcal{P}(\Omega) \to \mathbb{R}$ is a probability distribution or a probability measure if

i) $\mathbb{P}(A) \geq 0 \quad \forall A \in \mathcal{P}(\Omega)$

ii) $\mathbb{P}(\Omega) = 1$

iii) If $A_1, A_2, \dots$ are disjoint, $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

↳ Countable additivity from measure theory ...?!

Not all $A \in \mathcal{P}(\Omega)$ are measurable ⇢ σ-algebra

§ View on Probability

- Frequentist.

- Bayesian

§ Independent Events

Definition (Independence of Event)

Two events A, B are independent if $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$

and we write $A \perp\!\!\!\perp B$

A set of events $\{A_i : i \in I\}$ is independent if $\mathbb{P}(\bigcap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$.

for all $J \subseteq I$, and denote

§ Conditional Probability

Definition (Conditional Probability)

If $\mathbb{P}(B) > 0$, the conditional probability of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$$

- $\mathbb{P}(\cdot|B)$ satisfies axiom of probability

A, B are independent $\Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$

$\mathbb{P}(A_1 \cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \cdots \mathbb{P}(A_n | A_1 \cdots A_{n-1})$

§ Bayes' Theorem

Theorem (The Law of the Probability)

Let $A_1, \cdots, A_k$ be a partition of Ω. Then for any event B,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

(pf) Let $C_j = BA_j$. Then, $C_1, \cdots, C_k$ are disjoint, $B = \bigcup_{j=1}^{k} C_j$

Hence, $\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{j=1}^{k} C_j\right) = \sum_{j=1}^{k} \mathbb{P}(C_j) = \sum_j \mathbb{P}(BA_j) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$.

Theorem (Bayes' Theorem)

Let $A_1, \cdots, A_k$ be a partition of Ω such that $\mathbb{P}(A_i) > 0$, $\mathbb{P}(B) > 0$.

Then, $$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

(pf) $$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_iB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{k} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

Remark $\mathbb{P}(A_i)$ – Prior probability of A_i.

$\mathbb{P}(A_i|B)$ – Posterior probability of A_i given B.

$$P(A|B) = \frac{0.7 \times 0.9}{0.7 \times 0.9 + 0.2 \times 0.01 + 0.1 \times 0.01}$$

$$= \frac{0.63}{0.633}$$

$$\frac{1}{211}$$

Random Variable

§ Random Variables.

Definition (Random Variable)

A random variable is a mapping $X : \Omega \to \mathbb{R}$ that assign a real nbr $X(\omega)$ for each outcome ω

- Given random variable X and $A \subseteq \mathbb{R}$, let $X^{-1}(A) := \{\omega \in \Omega \mid X(\omega) \in A\}$.
  Define $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$.
  $\mathbb{P}(X = x) = \mathbb{P}(X^{-1}(x))$

Example Flip a coin twice. X: nbr of heads.

$\mathbb{P}(X=0) = \mathbb{P}(\{TT\}) = 1/4$.

$\mathbb{P}(X=1) = \mathbb{P}(\{HT, TH\}) = 1/2$

$\mathbb{P}(X=2) = \mathbb{P}(\{HH\}) = 1/4$.

Definition (Cumulative Distribution Function (CDF))

The CDF of a random variable I is $F_X : \mathbb{R} \to [0,1]$ by

$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$

Example Flip a coin twice, X: nbr of heads.

$\mathbb{P}(X=0)=\frac{1}{4}$, $\mathbb{P}(X=1)=\frac{1}{2}$, $\mathbb{P}(X=2)=\frac{1}{4}$.

$$F_X(x)=\begin{cases}0 & \text{if } x<0\\ 1/4 & \text{if } 0\le x<1\\ 3/4 & \text{if } 1\le x<2\\ 1 & \text{if } x=2\end{cases}$$

Dynkin's theorem

$\mathcal{F}: \mathbb{R}\to[0,1]$

Theorem Let X have a CDF F and Y have a CDF G.

If $F(x)=G(x)$ $\forall x$, then $\mathbb{P}(X\in A)=\mathbb{P}(Y\in A)$ $\forall A$.

(pf) For $a<b$, $\mathbb{P}(X\in(a,b))=\mathbb{P}(X^{-1}(a,\infty))-\mathbb{P}(X^{-1}(b,\infty))$

Let $A_n=(a+\frac{1}{n},\infty)\Rightarrow \lim X^{-1}(A_n)=X^{-1}(-\infty,b)$.

$\lim Y^{-1}(A_n)=Y^{-1}(-\infty,b)$.

Thus, $\mathbb{P}(X^{-1}(a,\infty))=\lim \mathbb{P}(X^{-1}(A_n))=\lim \mathbb{P}(X>a)$

$=\lim 1-\mathbb{P}(X\le a)=\lim 1-\mathbb{P}(Y\le a)$

$=\lim \mathbb{P}(Y>a)=\lim \mathbb{P}(Y^{-1}(A_n))$

$=\mathbb{P}(Y^{-1}(a,\infty))$.

Similarly, $\mathbb{P}(X^{-1}(b,\infty))=\mathbb{P}(Y^{-1}(b,\infty))$. Hence, $\mathbb{P}(X\in(a,b))=\mathbb{P}(Y\in(a,b))$

For $A_n\in\mathcal{A}$, suppose $\mathbb{P}(X\in A_n)=\mathbb{P}(Y\in A_n)$ $\forall n$. Let $B=\bigcup A_n$, $C=\bigcap A_n$

$\mathbb{P}(X\in A^c)=1-\mathbb{P}(X\in A)=1-\mathbb{P}(Y\in A)=\mathbb{P}(Y\in A^c)$.

$\mathbb{P}(X\in B)=\sum \mathbb{P}(X\in B_n)=$

$A_n=\bigcup_{i<n} A_i$

For $A_i \in \mathcal{A}$ $(i \in \{1,2\})$ if $\mathbb{P}(X \in A_i) = \mathbb{P}(Y \in A_i)$ ☐

$\mathbb{P}(X \in A_1 \cap A_2) =$ ☐

For any open $U \subseteq \mathbb{R}$, $\mathbb{P}(X \in U) = \mathbb{P}(Y \in U)$.

Let $\mathcal{B}$ be collection of measurable sets $A \in \mathcal{A}$ s.t.

$\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$. (it suff. to show

$\mathcal{B}$ is a σ-algebra

$\mathcal{B}$ is closed under complement

Either closeness under countable intersection or union suff.

If closed under intersection btwn two sets, done.

$I := \{(a,b), [a,b), (a,b], [a,b] ; a < b\}$ closed under finite intersection

$I \subseteq \mathcal{D}$, $\mathcal{B} = \sigma(I) \subseteq \mathcal{D}$

Theorem A function $F : \mathbb{R} \to [0,1]$ is a CDF for some probability $\mathbb{P}$ if and only if F satisfies

(i) F is nondecreasing

(ii) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

(iii) F is right continuous.

(pf) ($\Rightarrow$) (i) $\mathbb{P}(X \le x_1) = \mathbb{P}(X^{-1}(-\infty, x_1)) \le \mathbb{P}(X^{-1}(-\infty, x_2)) = \mathbb{P}(X \le x_2)$

$\quad \| \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \|$

$\quad f_X(x_1) \qquad\qquad\qquad\qquad\qquad\qquad\quad f_X(x_2)$

$\forall x_1 \le x_2$

(ii) Consider $A_n = (-\infty, -n]$. $A_1 \supseteq A_2 \supseteq \cdots$

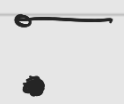
$\cap A_n = \phi$. Thus, $\lim \mathbb{P}(X \in A_n) = \mathbb{P}(X \in \phi) = 0$.

i.e. Given $\varepsilon > 0$, $\exists N$ s.t. $|\mathbb{P}(X \in A_n)| < \varepsilon$

$\Rightarrow 0 \le F(-N) = \mathbb{P}(X \in A_N) < \varepsilon. \Rightarrow |F(x)| < \varepsilon \ \forall x \le -N$

$\Rightarrow \lim_{x \to -\infty} F(x) = 0$.

Similarly, $\lim_{x \to \infty} F(x) = 1$.



(iii) Take $x_0 \in \mathbb{R}$. If $F(x_0) < 1$, let $2\varepsilon < 1 - F(x_0)$ be given.

Let $x_1 = \inf \{x \mid F(x) > F(x_0) + \varepsilon\}$. $x_1 \ge x_0$.

For any $(x_n) \to x_0$ with $x_n \ge x_0$, $A_n := (-\infty, x_n]$.

Then, $\mathbb{P}(X \in A_n) \to \mathbb{P}(X \in A)$. where $A = (-\infty, x_0)$

i.e. $F(x_n) \to F(x_0)$

Definition (Discrete RandomVar / Probability Mass Function)

X is discrete if it takes countably many values.

PMF for X is $f_X(x) = \mathbb{P}(X = x)$

- $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}, \quad \sum_i f_X(x_i) = 1$

  $F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$.

Definition (Continuous Random Variable / Probability distribution function)

A random variable X is continuous if $\exists$ a function f_X such that $f_X(x) \geq 0 \ \forall x$, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ and for $\forall a \leq b$, $\mathbb{P}(a < x < b) = \int_a^b f_X(x)\,dx$.

- We have $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$.

  and $f_X(x) = F_X'(x) \ \forall x$ at which F_X is diff.

Lemma Let F be the CDF for a random variable X.

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$
2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$
3. $\mathbb{P}(X > x) = 1 - F(x)$
4. If x is continuous, $F(b) - F(a) = \mathbb{P}(a < X \leq b)$

$\mathbb{P}(a \leq x \leq b)$

$\mathbb{P}(\overset{\uparrow}{a} < x \leq b)$

$\mathbb{P}(a \leq x \leq \overset{\uparrow}{b})$

Definition (Inverse CDF / Quantile Function)

X: random var. with CDF F.

$F^{-1}(q) = \inf\{x : F(x) \geq q\}$

$F^{-1}(q)$ named { first quantile, second, third } { if $q = 1/4$, if $q = 1/2$, if $q = 3/4$ }

Example (Discrete Random Variables)

1. Point mass distribution

$X \sim \delta_a$ if $F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$

PMF $f(x) = \begin{cases} 1 & x = a \\ 0 & \text{otherwise} \end{cases}$.

2. Discrete uniform distribution. $k > 1$.

$f(x) = \begin{cases} 1/k & \text{for } x = 1, \cdots, k \\ 0 & \text{otherwise.} \end{cases}$

3. Bernoulli Distribution

$\mathbb{P}(X=1) = p$. $\mathbb{P}(X=0) = 1-p$. for some $p \in [0,1]$

$X \sim \text{Bernoulli}(p)$.

$f(x) = p^x (1-p)^{1-x}$ for $x \in \{0,1\}$.

4. Binomial Distribution

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \dots, n \\ 0 & \text{otherwise.} \end{cases}$$

$f(x) = \mathbb{P}(X=x)$ be PMF. Then, $X \sim \text{Binomial}(n,p)$.

- If $X_1 \sim \text{Binomial}(n_1, p)$, $X_2 \sim \text{Binomial}(n_2, p)$,

then $X_1 + X_2 \sim \text{Binomial}(n_1+n_2, p)$

$$\because) \quad \mathbb{P}(X_1+X_2=x) = \sum_{y=0} \mathbb{P}(X_1+X_2=x \mid X_1=y)\, \mathbb{P}(X_1=y)$$

5. Geometric Distribution

X has a geometry distribution with $p \in (0,1)$,

denoted as $X \sim \text{Geom}(p)$ if

$\mathbb{P}(X=k) = p(1-p)^{k-1}$ $\forall k \geq 1$.

- $\sum \mathbb{P}(X=k) = \sum p(1-p)^{k-1} = \dfrac{p}{1-(1-p)} = 1.$

Means nbr of flips to the first heads

6. Poisson Distribution

X has a Poisson distribution with parameter λ, denoted as $X \sim \text{Poisson}(\lambda)$ if $P(X=x) = f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ $\forall x \geq 0,\ x \in \mathbb{Z}$.

- $\sum_{x=0}^{\infty} f(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$.
- If $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.
- $E[X] = \sum x f(x) = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$.

§2.4 Some Important Continuous Random Variable.

1. Uniform Distribution

$X \sim \text{Uniform}(a,b)$ if X has a PDF $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$

$\Rightarrow$ CDF $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$

PDF/CDF of Z is denoted as $\phi(z)$, $\Phi(z)$ respectively.

2. Normal (Gaussian) Distribution

$X \sim N(\mu, \sigma^2)$ if X has a PDF $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, $x \in \mathbb{R}$.

- X has a standard normal distribution if $\mu = 0, \sigma = 1$.

Standard Normal random variable is denoted by Z.

- μ: mean. σ: standard deviation.

Proposition (i) If $X \sim N(\mu, \sigma^2)$, then $Z = (X-\mu)/\sigma \sim N(0,1)$

(ii) If $Z \sim N(0,1)$, then $\mu + \sigma Z \sim N(\mu, \sigma^2)$

(iii) If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \cdots, n$ are independent, then

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

3. Exponential Distribution

$X \sim \text{Exp}(\beta)$ if $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$, $\forall x > 0$

$$\int_0^\infty f(x) = \left[e^{-\frac{x}{\beta}}\right]_0^\infty = 1.$$

$$\int_0^\infty x f(x) = \int_0^\infty \frac{1}{\beta} x e^{-\frac{x}{\beta}} dx = \left[x e^{-\frac{x}{\beta}} - \beta e^{-\frac{x}{\beta}}\right]_0^\infty$$

$$= \beta$$

4. Gamma Distribution

For $\alpha > 0$, $\Gamma(\alpha) := \int_0^\infty y^{\alpha-1} e^{-y} dy$

$X \sim \text{Gamma}(\alpha, \beta)$ if $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} x^{-\frac{x}{\beta}}$, $x > 0$

5. Beta Distribution

$X \sim \text{Beta}(\alpha, \beta)$ if $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$

6. t and Cauchy Distribution

$X \sim t_\nu$ if $f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{(1+\frac{x^2}{\nu})^{(\nu+1)/2}}$.

- Normal is t with $\nu = \infty$.

  Cauchy is t with $\nu = 1$. $f(x) = \frac{1}{\pi(1+x^2)}$.

7. χ^2 Distribution

$X \sim \chi^2_p$ if $f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}$, $x > 0$.

If $Z_1, \dots, Z_p$ are independent standard normal random variables, then $\sum_{i=1}^{p} Z_i^2 \sim \chi^2_p$

§2.5 Bivariate Distributions.

Definition (Joint CDF/Distribution).

- $F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y)$.
- Probability function Discrete: Joint mass fn $f(x,y) = \mathbb{P}(X=x, Y=y)$.

  Continuous: Joint PDF $\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$.

§2.6 Marginal Distribution.

Definition (Marginal Distribution)

- If (X,Y) have joint distribution with mass function $f_{X,Y}$,

  marginal mass function for X is $f_X(x) = \mathbb{P}(X=x) = \sum_y \mathbb{P}(X=x, Y=y) = \sum_y f(x,y)$

  Y is $f_Y(y) = \mathbb{P}(Y=y) = \sum_x \mathbb{P}(X=x, Y=y) = \sum_x f(x,y)$.

- For continuous r.v., the marginal densities are

  $f_X(x) = \int f(x,y)\,dy$, $f_Y(y) = \int f(x,y)\,dx$.

  Corresponding marginal distribution functions are denoted by F_X, F_Y.

§2.7 Independent Random Variables.

Definition (Independent RV).

Two RV X,Y are independent if $\forall A,B$,

$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)$.

and write $X \perp\!\!\!\perp Y$. Otherwise X,Y are dependent: $X \not\perp\!\!\!\perp Y$.

Theorem Let X, Y have joint PDF $f_{X,Y}$. Then, $X \perp\!\!\!\perp Y$ if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, y.$$

($\Rightarrow$) $A_n = \left[x - \frac{1}{n}, x + \frac{1}{n}\right]$, $B_n = \left[y - \frac{1}{n}, y + \frac{1}{n}\right]$.

$$\mathbb{P}(X \in A_n, Y \in B_n) = \iint_{A_n \times B_n} f_{X,Y}(x,y)\,dx\,dy.$$

$$\mathbb{P}(X \in A_n) = \int_{A_n} f_X(x)\,dx \qquad \mathbb{P}(Y \in B_n) = \int_{B_n} f_Y(y)\,dy.$$

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \text{a.e.} \quad \text{(starred)}$$

($\Leftarrow$) $\mathbb{P}(X \in A, Y \in B) = \iint_{A \times B} f_{X,Y}(x,y)\,dx\,dy$

$$= \iint_{A \times B} f_X(x) f_Y(y)\,dx\,dy = \int_A f_X(x)\,dx \int_B f_Y(y)\,dy$$

$$= \mathbb{P}(X \in A)\,\mathbb{P}(X \in B).$$

Theorem Suppose range of X, Y is a (possibly infinite) rectangle.

If $f(x,y) = g(x)h(y)$ for some g, h, then X, Y are independent.

(pf) $\iint_{\mathbb{R}^2} f\,dA = 1 \Rightarrow \int_{\mathbb{R}} g(x)\,dx \int_{\mathbb{R}} h(y)\,dy = 1.$

Let $P = \int_{\mathbb{R}} g(x)\,dx$. $Q = \int_{\mathbb{R}} h(y)\,dy$.

$f_X(x) = \int_y f(x,y)\,dy = \int_y g(x)h(y)\,dy = Q\,g(x)$

$f_Y(y) = P\,h(y)$ similarly.

$f(x,y) = g(x)h(y) = PQ\,g(x)h(y)$

$= f_X(x) f_Y(y).$

By thm 2.30, done.

§2.8 Conditional Distribution

Definition (Conditional Probability Mass Function / PDF).

$$f_{X|Y}(x|y) := \mathbb{P}(X=x, Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0.$$

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{assuming } f_Y(y) > 0$$

Then, $\mathbb{P}(X \in A \mid Y=y) = \int_A f_{X|Y}(x|y)\,dx$.

§2.9 Multivariate Distributions and IID Samples.

Definition (IID)

If $X_1, \dots, X_n$ are independent and has same marginal distribution with CDF F, we say $X_1, \dots, X_n$ are IID and write $X_1, \dots, X_n \sim F$.

If F has a density f, we also write $X_1, \dots, X_n \sim f$.

§ 2.10 Two Important Multivariate Distribution

Definition (Multinomial Distribution)

For $p=(p_1,\cdots,p_k)$, $p_j \geq 0$, $\sum_{j=1}^{k} p_j = 1$, $X=(X_1,\cdots,X_k)$ where X_j = # of color j appear from n draw of balls where probability of color j appearing is p_j on each draw.

We denote $X \sim \text{Multinomial}(n, p)$.

$$f(x) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} \text{ where } \binom{n}{x_1 \cdots x_k} = \frac{n!}{x_1! \cdots x_k!}$$

Lemma Suppose $X \sim \text{Multinomial}(n, p)$, $X=(X_1,\cdots,X_k)$, $p=(p_1,\cdots,p_k)$. The marginal distribution of X_j is $\text{Binomial}(n, p_j)$.

(pf) WLOG, $j=k$.

$$f_k(x_k) = \sum_{\substack{x_1,\cdots,x_{k-1} \\ x_1+\cdots+x_k=n}} f(x)$$, which is probability # of color $k = x_k$ after n draw with each draw probability of color $k = p_k$,

$$1 = 1^n = (\Sigma p_i)^n = \left(p_k + \sum_{i \neq k} p_k\right)^n$$

$$= (p + (1-p))^n.$$

Coeff. of $p_k^{x_k}$ coeff. of p^{x_k}

Definition (Multivariate Normal)

For vector μ, matrix $\Sigma \in S_+^n$, let $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$

$X \sim N(\mu, \Sigma)$ if $f(x;\mu,\Sigma) = \dfrac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

where $|\Sigma|$ is the determinant of Σ

$\mu = 0, \Sigma = I$ → standard Normal.

Theorem If $Z \sim N(0, I)$ and $X = \mu + \Sigma^{1/2} Z$, then $X \sim N(\mu, \Sigma)$

Conversely, if $X \sim N(\mu, \Sigma)$, then $\Sigma^{-1/2}(X-\mu) \sim N(0, I)$

$$S = r^{-1}(y) = \Sigma^{-1/2}(x-\mu) \qquad f_Y(y) = f_X(S(y)) \left|\frac{dS(y)}{dy}\right|$$

(pf) $X = \mu + \Sigma^{1/2} Z = r(Z)$

$$P(X \in A) = P(\mu + \Sigma^{1/2} Z \in A) = P(Z \in S(A))$$

$$= \int_{S(A)} f(z;0,I)\,dz = \int_{S(A)} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}z^T z}\,dz$$

$z =$

$$= \int_{z \in S(A)} \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}S(x)^T S(x)} \left|\frac{dS}{dx}\right| dx$$

$$= \int_A \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}\,dx$$

$\Rightarrow X \sim N(\mu; \Sigma)$

Similar for converse case

Theorem Let $X = (X_a, X_b)$, $\mu = (\mu_a, \mu_b)$, $\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$

Then, (1) Marginal distribution of X_a is $X_a \sim N(\mu_a, \Sigma_{aa})$

(2) Conditional distribution of X_b given $X_a = x_a$ is

$$X_b \mid X_a = x_a \sim N(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a),\ \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$$

(3) If a is a vector, then $a^T X \sim N(a^T\mu, a^T\Sigma a)$

(4) $V = (X-\mu)^T\Sigma^{-1}(X-\mu) \sim \chi^2_k$

(pf) (2) $f_{X_b|X_a}(x_b|x_a) = \dfrac{f_X(x)}{f_{X_a}(x_a)}$

$$\overset{(1)}{=} \frac{\frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}{\frac{1}{(2\pi)^{k_a/2}|\Sigma_a|^{1/2}} e^{-\frac{1}{2}(x_a-\mu_a)^T\Sigma_a^{-1}(x_a-\mu_a)}}$$

§2.11 Transformations of Random Variable.

Observation Let $Y = r(X)$, let $A_y := \{x \mid r(x) \le y\}$.

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(r(X) \le y)$$
$$= \mathbb{P}(\{x \mid r(x) \le y\}) = \int_{A_y} f_X(x)\,dx.$$

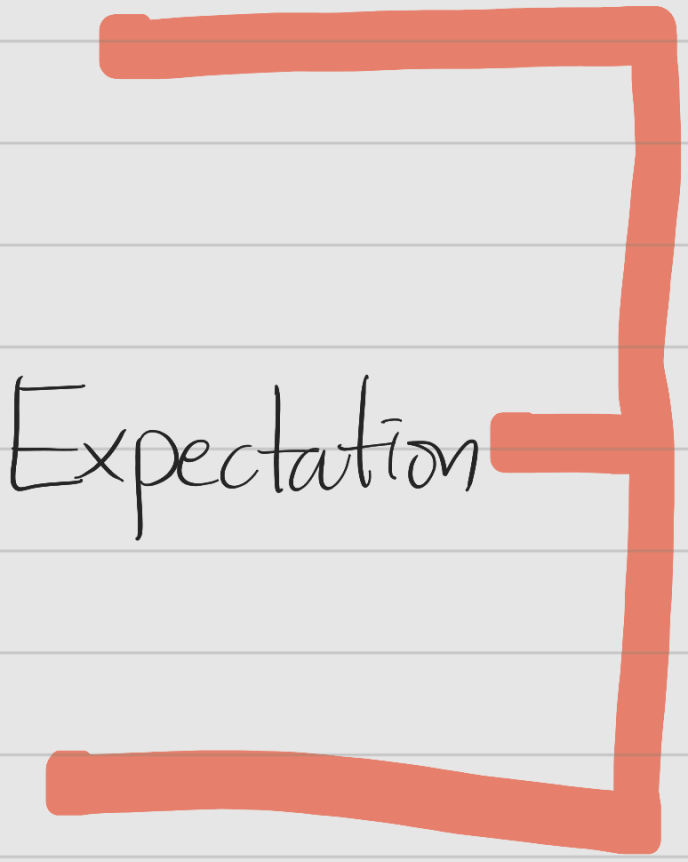
If r is strictly monotone, $f_Y(y) = f_X(s(y)) \left|\frac{ds(y)}{dy}\right|$

§2.12 Transformations of Several Random Variables

Observation $Z = r(X, Y)$, let $A_z = \{(x, y) \mid r(x, y) \le z\}$.

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}(r(X, Y) \le z)$$
$$= \mathbb{P}(\{(x, y) \mid r(x, y) \le z\}) = \iint_{A_z} f_{X,Y}(x, y)\,dx\,dy$$

$$f_Z(z) = F_Z'(z)$$


Expectation

§ 3.1 Expectation of a Random Variable.

Definition (Expectation Value)

The expectation of X is

$$\mathbb{E}(X) = \int x\, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

assuming the sum (or integral) is well defined.

We denote $\mathbb{E}(X) = \mathbb{E}X = \int x\, dF(x) = \mu = \mu_X$.

k^{th} central moment: $\mathbb{E}((X-\mu)^k)$

Example 1. $X \sim \text{Bernoulli}(p)$. $\mathbb{E}(X) = \sum_{x=0}^{1} x f(x) = 0 \times (1-p) + 1 \times p = p$

2. Flip a fair coin twice, X = # of heads.

$$\mathbb{E}(X) = \sum_x x f_X(x) = 0 \times f(0) + (1 \times f(1) + 2 \times f(2)$$

$$= 0 \times \tfrac{1}{4} + (1 \times \tfrac{1}{2} + 2 \times \tfrac{1}{4} = 1.$$

3. $X \sim \text{Uniform}(-1, 3)$. $\mathbb{E}(X) = \int x\, dF_X(x) = \int x f_X(x)\, dx$

$$= \frac{1}{4}\int_{-1}^{3} x\, dx = 1$$

Theorem (Linearity of the Expectation)

If $X_1, \cdots, X_n$ are RV, $a_1, \cdots, a_n$ are constants,

$\mathbb{E}(\sum a_i X_i) = \sum a_i \mathbb{E}(X_i)$

(pf) By linearity of integration.

Theorem (Multiplicity of the Expectation)

Let $X_1, \cdots, X_n$ be independent. Then,

$\mathbb{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathbb{E}(X_i)$

(pf) $\mathbb{E}(\prod X_i) = \int \prod x_i \, f_X(x_1, \cdots, x_n) \, dx$ $\quad x = (x_1, \cdots, x_n)$, $y = (x_1, \cdots, x_n)$

$= \int \prod x_i \prod f_{X_i}(x_i) \, dx$

$= \int \prod x_i f_{X_i}(x_i) \, dx = \prod \int x_i f_{X_i}(x_i) \, dx_i$.

Theorem (Expectation of Transformed RV)

Let $Y = r(X)$. Then, $\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x) \, dF_X(x)$

(pf) $\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int y f_Y(y) \, dy = \int r(x) \left| \frac{ds(y)}{dy} \right| f_X(s(y)) \, dy$

$= \int r(x) f_X(x) \, dx$.

§3.3 Variance and Covariance.

Definition (Variance)

Let X be a RV with mean μ. The variance of X denoted by σ^2 or σ_X^2 or $\mathbb{V}(X)$ or $\mathbb{V}X$ is

$\sigma^2 := \mathbb{E}(X-\mu)^2 = \int (x-\mu)^2 dF(x)$, assuming the expectation exists.

The standard deviation is $sd(X) := \sqrt{\mathbb{V}(X)}$ denoted by σ and σ_X

Theorem Assuming variance is well defined,

1. $\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2$
2. If a, b are constants, then $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$
3. If $X_1, \cdots, X_n$ are independant, $a_1, \cdots, a_n$ are constants, then $\mathbb{V}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \mathbb{V}(X_i)$

(pf) 1. $\mathbb{V}(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2$

$= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 = \mathbb{E}(X^2) - \mu^2$

2. $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b = a\mu + b$

$\mathbb{V}(aX+b) = \mathbb{E}\left((aX+b) - (a\mu+b)\right)^2 = \mathbb{E}(a^2(X-\mu)^2)$

$= a^2\mathbb{E}(X-\mu)^2 = a^2\mathbb{V}(X)$

3. $\mathbb{V}(\sum X_i) = \mathbb{E}(\sum X_i - \sum \mu_i)^2 = \mathbb{E}\left((\sum X_i - \mu_i)^2\right)$

$= \sum \mathbb{V}(X_i)$ by induction on n

Use (2).

Example $X \sim \text{Binomial}(n,p)$, $X = \sum X_i$, $X_i = \begin{cases} 1 & \text{if toss } i = \text{head} \\ 0 & \text{otherwise} \end{cases}$.

$X = \sum X_i$, and X_i are independent. $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1-p$.

$\mathbb{E}(X_i) = p$.

$\mathbb{E}(X_i^2) = p \times 1^2 + (1-p) \times 0^2 = p$.

Thus, $\mathbb{V}(X_i) = \mathbb{E}(X_i^2) - p^2 = p - p^2 = p(1-p)$.

$\mathbb{V}(X) = \mathbb{V}(\sum X_i) = \sum p(1-p) = np(1-p)$.

Definition (Sample Mean / Sample Variance)

If $X_1, \cdots, X_n$ are random variables, then

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ is the sample variance

↳ To ensure unbiased estimator (following theorem)

Theorem Let $X_1, \cdots, X_n$ be IID, $\mu = \mathbb{E}(X_i)$, $\sigma^2 = \mathbb{V}(X_i)$. Then

$\mathbb{E}(\bar{X}_n) = \mu$, $\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n}$, $\mathbb{E}(S_n^2) = \sigma^2$

pf) $\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu$

$\mathbb{V}(\bar{X}_n) = \mathbb{V}\left(\frac{1}{n}\sum X_i\right) = \sum \frac{1}{n^2} \mathbb{V}(X_i) = n \times \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$

$\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1}\sum (X_i - \bar{X}_n)^2\right) = \frac{1}{n-1}\sum \mathbb{E}(X_i - \bar{X}_n)^2$

$X_n - \bar{X}_n = \frac{n-1}{n} X_n - \frac{1}{n} \sum_{i=1}^{n-1} X_i$, $\mathbb{E}(X_n - \bar{X}_n) = 0$.

$$\Rightarrow \mathbb{E}(X_n-\bar{X}_n)^2 = \mathbb{V}(X_n-\bar{X}_n)^2 = \mathbb{V}\left(\frac{n-1}{n}X_n - \frac{1}{n}\sum_{i=1}^{n-1}X_i\right)$$

$$= \mathbb{V}\left(\frac{n-1}{n}X_n - \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\right)$$

$$= \left(\frac{n-1}{n}\right)^2\left(\mathbb{V}(X_n) + \mathbb{V}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\right)\right)$$

$$= \left(\frac{n-1}{n}\right)^2\left(\sigma^2 + \frac{\sigma^2}{n-1}\right) = \frac{n-1}{n}\sigma^2.$$

$$\mathbb{E}(S_n^2) = \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i-\bar{X}_n)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Definition (Covariance)

For random variables X, Y with mean μ_X, μ_Y, and standard deviation σ_X, σ_Y respectively, the covariance btwn X and Y is

$$Cov(X,Y) = \mathbb{E}((X-\mu_X)(Y-\mu_Y))$$

and the correlation is

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

Theorem $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

$$-1 \le \rho(X,Y) \le 1.$$

(pf). $Cov(X,Y) = \mathbb{E}(X-\mu_X)(Y-\mu_Y)$

$$= \mathbb{E}(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)$$
$$= \mathbb{E}(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y$$
$$= \mathbb{E}(XY) - \mu_X \mu_Y = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

$$\mathbb{E}|(X-\mu_X)(Y-\mu_Y)| = \int |(x-\mu_X)(y-\mu_Y)| f_{X,Y}(x,y)\,dx\,dy$$
$$\leq \left(\int (x-\mu_X)^2 dx\right)^{1/2} \left(\int (y-\mu_Y)^2 dy\right)^{1/2} \text{ by C-S ineq.}$$
$$= \sigma_X \sigma_Y.$$

Thus, $|\mathrm{Cov}(X,Y)| \leq \sigma_X \sigma_Y$. i.e., $-1 \leq \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \leq 1$.

- If $Y = aX + b$, $\rho(X,Y) = 1$ if $a > 0$
  $\rho(Y,Y) = -1$ if $a < 0$.
- X, Y are independent $\Rightarrow$ $\mathrm{Cov}(X,Y) = \rho = 0$.

Theorem $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathrm{Cov}(X,Y)$.

More generally $\mathbb{V}\left(\sum a_i X_i\right) = \sum_i a_i^2\,\mathbb{V}(X_i) + 2\sum_{i<j} a_i a_j\,\mathrm{Cov}(X_i, X_j)$

(pf) $\mathbb{V}(aX+bY) = \mathbb{E}\left(aX+bY-(a\mu_X+b\mu_Y)\right)^2$

$= \mathbb{E}\left(a(X-\mu_X)+b(Y-\mu_Y)\right)^2$

$= a^2\mathbb{E}(X-\mu_X)^2 + b^2\mathbb{E}(Y-\mu_Y)^2 + 2ab\,\mathbb{E}(X-\mu_X)(Y-\mu_Y)$

$= a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\,\mathrm{cov}(X,Y)$.

Use induction with fact that

$\mathrm{Cov}(X, Y+Z) = \mathbb{E}(X-\mu_X)(Y+Z-(\mu_Y+\mu_Z))$

$= \mathbb{E}(X-\mu_X)(Y-\mu_Y) + \mathbb{E}(X-\mu_X)(Z-\mu_Z)$

$= \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$

§3.5 Conditional Expectation

Definition (Conditional Expectation)

The conditional expectation of X given $Y=y$ is

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum x f_{X|Y}(x|y)\,dx & \text{(discrete)} \\ \int x f_{X|Y}(x|y)\,dx & \text{(continuous)} \end{cases}$$

If $r(x,y)$ is a fn of x,y,

$$\mathbb{E}(r(X,Y)|Y=y) = \begin{cases} \sum r(x,y) f_{X|Y}(x,y)\,dx & \text{(discrete)} \\ \int r(x,y) f_{X|Y}(x,y)\,dx & \text{(continuous)} \end{cases}$$

- $y \mapsto \mathbb{E}(X|Y=y)$ is a fn of y.

$\mathbb{E}(X|Y)$ is the random variable whose value is $\mathbb{E}(X|Y=y)$ when $Y=y$.

R.V. $X: \Omega \to E$, $\mathbb{P}: \mathcal{F}(\Omega) \to \mathbb{R}$

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$

Theorem (The Rule of Iterated Expectations)

For R.V. X,Y, assuming the expectations exists,

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y), \quad \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X).$$

More generally

$$\mathbb{E}(\mathbb{E}(r(X,Y)|X)) = \mathbb{E}(r(X,Y))$$

(pf) $\mathbb{E}(\mathbb{E}(Y|X)) = \int \mathbb{E}(Y|X=x) f_X(x)\,dx$

$= \int\int y f_{Y|X}(x|y)\,dy\, f_X(x)\,dx$

$= \iint y f_{Y|X}(x|y) f_X(x)\,dy\,dx$

$= \iint y f(x,y)\,dy\,dx = \mathbb{E}(Y).$

Definition (Conditional Variance)

The conditional variance is

$$\mathbb{V}(Y|X=x) = \int (y-\mu(x))^2 f(y|x)\,dy,$$

where $\mu(x) = \mathbb{E}(Y|X=x)$

Theorem $\mathbb{V}(Y) = \mathbb{E}\,\mathbb{V}(Y|X) + \mathbb{V}\,\mathbb{E}(Y|X)$

(pf) $\mathbb{E}\,\mathbb{V}(Y|X) = \mathbb{E}\int (y-\mu(X))^2 f(y|X)\,dy$

$= \int\int (y-\mu(x))^2 f(y|x)\,dy\, f_X(x)\,dx$

$= \iint (y-\mu(x))^2 f(x,y)\,dx\,dy$ $\qquad (y-\mu_y)^2$

$\mathbb{V}\,\mathbb{E}(Y|X) = \mathbb{V}\int y f(y|x)\,dy$ $\qquad \mathbb{E}\,\mathbb{E}(Y|X) = \mathbb{E}(Y)$

$$= \int \left(\int y f(y|x) dy - \mu_y \right)^2 f_x(x) dx$$

$$= \int \left(\mu(x) - \mu_y \right)^2 f_x(x) dx$$

$$= \int \left(\int (\mu(x) - \mu_y) f(y|x) dy \right) f_x(x) dx$$

§3.6 Moment Generating Functions

Definition (Moment Generating Function)

The moment generating function (MGF) or Laplace transform, of X is $\psi_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx}\, dF(x)$.

where $t \in \mathbb{R}$

Observation $\psi'(0) = \left[\frac{d}{dt}\mathbb{E}e^{tx}\right]_{t=0} = \mathbb{E}[Xe^{tx}]_{t=0} = \mathbb{E}(X)$.

$\psi^{(k)}(0) = \mathbb{E}(X^k)$ (k^{th} moment).

Lemma (Properties of the MGF)

(1) If $Y = aX + b$, then $\psi_Y(t) = e^{bt}\psi_X(at)$

(2) If $X_1, \cdots, X_n$ are independent and $Y = \sum X_i$, then $\psi_Y(t) = \prod \psi_i(t)$

where ψ_i is the MGF of X_i.

(A) (1) $\psi_Y(t) = \mathbb{E}\, e^{tY} = \mathbb{E}\, e^{t(aX+b)} = e^{bt}\, \mathbb{E}\, e^{atX} = e^{bt}\psi_X(at)$

(2) $\psi_Y(t) = \mathbb{E}(e^{t\sum X_i}) = \mathbb{E}(\prod e^{tX_i}) = \prod \mathbb{E}(e^{tX_i}) = \prod \psi_{X_i}(t)$

Example $X \sim \text{Binomial}(n,p)$, $X = \sum_{i=1}^{n} X_i$, where $\mathbb{P}(X_i=1)=p$, $\mathbb{P}(X_i=0)=1-p$.

$\psi_i(t) = \mathbb{E}\, e^{X_i t} = pxe^t + (1-p)x1 = pe^t + q$ where $q = 1-p$.

$\psi_X(t) = \pi \psi_i(t) = (pe^t + q)^n$.

Theorem Let X, Y be R.V. If $\psi_X(t) = \psi_Y(t)$ for $\forall t$ in some open interval around 0, then $X \overset{d}{=} Y$ (i.e. they have the same distrib.)

Example. $X_1 \sim \text{Binomial}(n_1, p)$, $X_2 \sim \text{Binomial}(n_2, p)$, $Y = X_1 + X_2$.

Then $\psi_Y(t) = \psi_{X_1}(t)\psi_{X_2}(t) = (pe^t + q)^{n_1}(pe^t + q)^{n_2} = (pe^t + q)^{n_1+n_2}$.

Y has MGF of $\text{Binomial}(n_1 + n_2, p)$.

Hence by thm, $Y \sim \text{Binomial}(n_1 + n_2, p)$

Note (Moment Generation Functions for Some Common Distributions)

Distribution	MGF $\psi(t)$	
Bernoulli (p)	$pe^t + (1-p)$	
Binomial (n, p)	$(pe^t + (1-p))^n$	
Poisson (λ)	$e^{\lambda(e^t - 1)}$	
Normal (μ, σ)	$e^{\mu t + \frac{\sigma^2 t^2}{2}}$	
Gamma (α, β)	$\left(\frac{1}{1-\beta t}\right)^\alpha$	for $t < \frac{1}{\beta}$.

$\because$) Poisson (λ):

4.1 Inequalities

§ 4.1 Probability Inequality.

Theorem (Markov's Inequality)

Let X be a nonnegative R.V. and $\exists\, \mathbb{E}(X)$. For any $t>0$,
$\mathbb{P}(X>t) \le \frac{\mathbb{E}(X)}{t}$.

(pf) $\mathbb{E}(X) = \int_0^\infty x f(x)\,dx = \int_0^t x f(x)\,dx + \int_t^\infty x f(x)\,dx$
$\ge \int_t^\infty x f(x)\,dx \ge t\int_t^\infty f(x)\,dx = t\,\mathbb{P}(X>t)$.

Theorem (Chernoff's Inequality)

Let X be a R.V. then, $\mathbb{P}(X>\varepsilon) \le \inf_{t\ge 0} e^{-t\varepsilon}\mathbb{E}(e^{tx})$.

pf) $t>0$ $\mathbb{P}(X>\varepsilon) = \mathbb{P}(e^{tx}>e^{t\varepsilon}) \le e^{-t\varepsilon}\mathbb{E}(e^{tx})$ (by Markov's ineq).
Thus, $\mathbb{P}(X\ge\varepsilon) \le \inf_{t>0} e^{-t\varepsilon}\mathbb{E}(e^{tx})$.

Theorem (Chebyshev's Inequality)

Let $\mu = \mathbb{E}(X)$, $\sigma^2 = \mathbb{V}(X)$. Then
$\mathbb{P}(|X-\mu|\ge t) \le \frac{\sigma^2}{t^2}$ and $\mathbb{P}(|Z|\ge k) \le \frac{1}{k^2}$, where $Z = \frac{X-\mu}{\sigma}$.

(pf) Second ineq. follows from the first by setting $t=k\sigma$.

$$P(|X-\mu|\geq t) = P((X-\mu)^2 \geq t^2) \leq \frac{E(X-\mu)^2}{t^2} = \frac{\sigma^2}{t^2}.$$

↳ Markov's ineq.

Example $V(\bar{X}_n) = \frac{V(X_1)}{n} = \frac{p(1-p)}{n}$.

By Chebyshev, $P(|\bar{X}_n - p| > \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$.

↳ $p(1-p) \leq \frac{1}{4}$ bc p

Theorem (Hoeffding's Inequality) ← Used for bounded random variable.

Let $Y_1, \cdots, Y_n$ be independent such that $E(Y_i)=0$, $a_i \leq Y \leq b_i$.

Let $\varepsilon > 0$. Then for any $t>0$,

$$P\left(\sum_{i=1}^n Y_i \geq \varepsilon\right) \leq e^{-t\varepsilon} \prod_{i=1}^n e^{t^2(b_i-a_i)^2/8}$$

In particular, $P\left(\sum_{i=1}^n Y_i \geq \varepsilon\right) \leq \inf_{t>0} e^{-t\varepsilon} \prod_{i=1}^n e^{t^2(b_i-a_i)^2/8}$

(pf) $P(\sum Y_i \geq \varepsilon) \leq e^{-t\varepsilon} E(e^{t\sum Y_i})$ (Chernoff ineq.)

$$= e^{-t\varepsilon} \prod_{i=1}^n E e^{tY_i}.$$

Let $Y_i = \alpha b_i + (1-\alpha) a_i$, $\alpha = \frac{Y_i - a_i}{b_i - a_i}$.

$$e^{tY_i} = e^{t(\alpha b_i + (1-\alpha)a_i)} \leq \alpha e^{tb_i} + (1-\alpha) e^{ta_i} = \frac{Y_i - a_i}{b_i - a_i} e^{tb_i} + \frac{b_i - Y_i}{b_i - a_i} e^{ta_i}$$

$$E(e^{tY_i}) \leq \frac{-a_i}{b_i-a_i} e^{tb_i} + \frac{b_i}{b_i-a_i} e^{ta_i} = e^{g(u)}, \text{ where}$$

$$u = t(b_i - a_i),\ g(u) = -\gamma u + \log(1-\gamma+\gamma e^u),\ \gamma = -\frac{a_i}{b_i-a_i}.$$

$g(0) = g'(0) = 0$. $g''(u) \leq \frac{1}{4}$ $\forall u > 0$. By Taylor, $\exists \xi \in (0,u)$ s.t.

$$g(u) = \frac{u^2}{2} g''(\xi) \le \frac{u^2}{8} = \frac{t^2(b_i-a_i)^2}{8}.$$

Thus, $\mathbb{E}(e^{tY_i}) \le e^{g(u)} = e^{t^2(b_i-a_i)^2/8}$

Theorem Let $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Then for $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - p| > \varepsilon) \le 2e^{-2n\varepsilon^2},$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

(pf) $\mathbb{P}(\bar{X}_n - p > \varepsilon) = \mathbb{P}(\sum Y_i > n\varepsilon)$, where $Y_i = X_i - \mu_{X_i}$ has $\mathbb{E}Y_i = 0$
$= X_i - p$

$-p \le Y_i \le 1-p$.

By Hoeffding, $\mathbb{P}(\bar{X}_n - p > \varepsilon) \le e^{-t n\varepsilon} \prod e^{t^2/8}$

$$\le e^{-4n\varepsilon^2} \cdot e^{2n\varepsilon^2} = e^{-2n\varepsilon^2} \quad (t = 4\varepsilon)$$

Similarly, $\mathbb{P}(\bar{X}_n - p < \varepsilon) = e^{-2n\varepsilon^2}$ so that $\mathbb{P}(|\bar{X}_n - p| > \varepsilon) \le 2e^{-2n\varepsilon^2}$.

Theorem (Mill's Inequality)

Let $Z \sim N(0,1)$. Then $\mathbb{P}(|Z| > t) \le \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$

(pf) $p(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $p'(z) = -z p(z)$.

$$p(z > t) = \int_t^\infty p(z)dz = \int_t^\infty -\frac{p'(z)}{z} dz \le -\frac{1}{t}\int_t^\infty p'(z)dz$$

$$= -\frac{1}{t} p(z)\Big|_t^\infty = \frac{1}{t} p(t) = \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t} = P(z < -t).$$

Thus, $P(|z| > t) = P(z > t) + P(z < -t) = \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$.

§4.2 Inequalities For Expectations

Theorem (Cauchy-Schwartz Inequality)

If X, Y have finite variance, then

$$\mathbb{E}|XY| \le \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

(pf) $\mathbb{E}(t|X|+|Y|)^2 = t^2\mathbb{E}(X^2) + 2t\mathbb{E}(|XY|) + \mathbb{E}(Y)^2 \ge 0 \ \forall t$.

Thus, $(\mathbb{E}(|XY|))^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$

$\Rightarrow \mathbb{E}|XY| \le \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$

Theorem (Jensen's Inequality)

If g is convex, then $\mathbb{E}\,g(X) \ge g(\mathbb{E}(X))$

If g is concave, then $\mathbb{E}\,g(X) \le g(\mathbb{E}(X))$

(pf) Let g be cvx. $L(x) = a+bx$ be a line tangent to $g(x)$ at $\mathbb{E}(X)$.

Then, $\mathbb{E}\,g(X) \ge \mathbb{E}\,L(X) = \mathbb{E}(a+bX) = a+b\mathbb{E}(X) = L(\mathbb{E}(X)) = g(\mathbb{E}(X))$.

Corollary $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$. $|\mathbb{E}X| \leq \mathbb{E}|X|$.

If $X > 0$, $\mathbb{E}(1/X) \geq 1/\mathbb{E}(X)$

$\mathbb{E}(\log X) \leq \log \mathbb{E}(X)$.

Definition (Kullback-Leibler (KL) divergence)

KL divergence btwn densities p, q is

$$D(p,q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

Convergence of Random Variables

§5.2 Types of Convergence

Definition (Types of Convergence)

Let $X_1, X_2, \ldots$ be a sequence of random variables and let X denote another R.V. Let F_n denote the CDF of X_n, F be CDF of X.

1. X_n converges to X in probability, written $X_n \xrightarrow{P} X$, if
   $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$
2. X_n converges to X in distribution, written $X_n \rightsquigarrow X$, if
   $\lim_{n\to\infty} F_n(t) = F(t)$
   at all t for which F is continuous

Definition (Convergence in Quadratic Mean / L_2 / as)

- X_n converges to X in quadratic mean (convergence in L_2), written $X_n \xrightarrow{qm} X$ if $\mathbb{E}(X_n - X)^2 \to 0$
- X_n converges to X in L_1, written $X_n \xrightarrow{L_1} X$ if
  $\mathbb{E}|X_n - X| \to 0$ as $n \to \infty$
- X_n converges almost surely to X, written $X_n \xrightarrow{as} X$ if
  $\mathbb{P}(\{s : X_n(s) \to X(s)\}) = 1$.

Theorem The followings hold:

(1) $X_n \xrightarrow{\text{a.s.}} X$ implies $X_n \xrightarrow{P} X$

(2) $X_n \xrightarrow{P} X$ implies $X_n \rightsquigarrow X$

(3) If $X_n \rightsquigarrow X$ and $\mathbb{P}(X=c)=1$ for some c, then $X_n \xrightarrow{P} X$.

(pf) (1) $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n-X|^2 > \varepsilon^2) \le \frac{1}{\varepsilon^2}\mathbb{E}(X_n-X)^2 \to 0$

(2) Let F be conti. at x. Fix x.

$$F_n(x) = \mathbb{P}(X_n \le x) = \mathbb{P}(X_n \le x, X \le x+\varepsilon) + \mathbb{P}(X_n \le x, X > x+\varepsilon)$$
$$\le \mathbb{P}(X \le x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$
$$= F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

$$F(x-\varepsilon) = \mathbb{P}(X \le x-\varepsilon) = \mathbb{P}(X \le x-\varepsilon, X_n \le x) + \mathbb{P}(X \le x-\varepsilon, X_n > x)$$
$$\le \mathbb{P}(X_n \le x) + \mathbb{P}(|X_n - X| > \varepsilon)$$
$$= F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon)$$

Thus, $F(x-\varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \le F_n(x) \le F(x+\varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$.

$n \to \infty$ gives $F(x-\varepsilon) \le \liminf_n F_n(x) \le \limsup_n F_n(x) \le F(x+\varepsilon)$.

By $\varepsilon \to 0$, $\lim_{n\to\infty} F_n(x) = F(x)$.

(3) $F(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \ge c. \end{cases}$ Thus, $\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x > c. \end{cases}$

Let $\varepsilon > 0$. $\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n > c+\varepsilon) + \mathbb{P}(X_n < c-\varepsilon)$

$$\le 1 - F_n(c+\varepsilon) + F_n(c-\varepsilon)$$
$$\to 1 - F(c+\varepsilon) + F(c-\varepsilon) = 1 - 1 + 0 = 0.$$

Theorem (1) $X_n \xrightarrow{as} X$ implies $X_n \xrightarrow{P} X$.

(2) $X_n \xrightarrow{qm} X$ implies $X_n \xrightarrow{L_1} X$

(3) $X_n \xrightarrow{L_1} X$ implies $X_n \xrightarrow{P} X$.

(pf) (1) $P(|X_n - X| > \varepsilon) = 1 - P(|X_n - X| < \varepsilon)$

Let $A_{N,\varepsilon} = \{s \mid |X_n(s) - X(s)| < \varepsilon \quad \forall n \geq N\}$. Then, $A_{1,\varepsilon} \subseteq A_{2,\varepsilon} \subseteq \cdots$

$A_\varepsilon = \bigcup A_{N,\varepsilon} = \{s \mid \exists N \text{ s.t. } |X_n(s) - X(s)| < \varepsilon \ \forall n \geq N\}$.

$A = \bigcap A_\varepsilon$.

By $X_n \xrightarrow{a.e.} X$, $P(A) = 1$.

$$P(|X_N - X| > \varepsilon) = 1 - P(|X_N - X| < \varepsilon)$$
$$\leq 1 - P(|X_n - X| < \varepsilon \ \forall n \geq N)$$
$$= 1 - P(A_{\varepsilon,N})$$

$\lim_N P(A_{\varepsilon,N}) = P(A_\varepsilon) \geq P(A) = 1$. (↳ theorem)

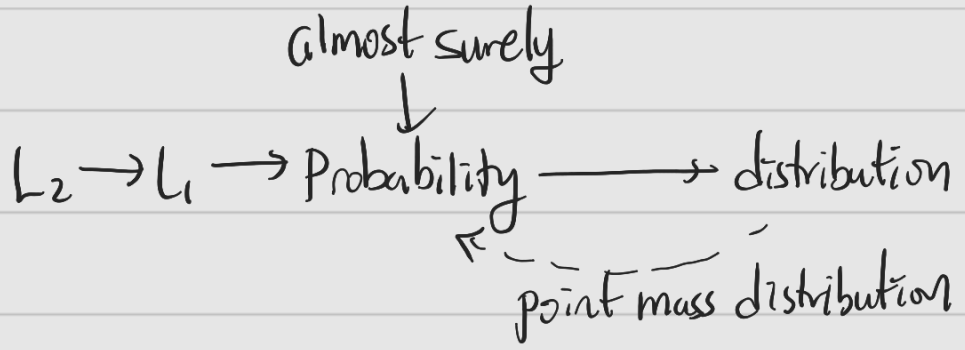
Thus, $\lim_N P(|X_N - X| > \varepsilon) = 0$.

(2) $E(X_n - X)^2 \to 0$.

$(E|X_n - X|)^2 \leq E(X_n - X)^2 \to 0$. Thus, $E|X_n - X| \to 0$.

(3) $P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon} E|X_n - X| \to 0$. Thus, $X_n \xrightarrow{P} X$.

Note



Counter example for direction with no arrows.

Theorem (Convergence Rules)

Let X_n, X, Y_n, Y be distribution, g be continuous.

(1) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n+Y_n \xrightarrow{P} X+Y$

(2) If $X_n \xrightarrow{gm} X$ and $Y_n \xrightarrow{gm} Y$, then $X_n+Y_n \xrightarrow{gm} X+Y$

(3) If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow C$, then $X_n+Y_n \rightsquigarrow X+C$

(4) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_nY_n \xrightarrow{P} XY$

(5) If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow C$, then $X_nY_n \rightsquigarrow CX$

(6) If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$

(7) If $X_n \rightsquigarrow X$, then $g(X_n) \rightsquigarrow g(X)$.

(pf) (1), (2), (4), follows if convergence in p, gm are convergence in topological space induced by

$$d(X,Y) = \sup_{\varepsilon>0} \mathbb{P}(|X-Y| > \varepsilon)$$

$$d'(X,Y) = \mathbb{E}(X-Y)^2$$ respectively.

↳ Have to check it is norm.

§5.3 The Law of Large Numbers

Theorem (The Weak Law of Large Numbers)

If $X_1, \cdots, X_n$ are IID, then $\overline{X}_n \xrightarrow{P} \mu$. ($\mu = \mathbb{E}(X_1)$).

(pf) Assume $\sigma < \infty$. (Just to simplify proof)

By Chebyshev, $\mathbb{P}(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\mathbb{V}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$ as $n \to \infty$.

Theorem (The Strong Law of Large Numbers)

Let $X_1, \cdots, X_n$ be IID. If $\mu = \mathbb{E}|X_1| < \infty$, then $Y_n \xrightarrow{as} \mu$

(pf)