Chapter 1 Probspace

- Sample Space: $\omega \in \Omega$ is a sample or elementary event.
- probability mass function (pmf) $p: \Omega \to [0,1]$ satisfies $\sum_{\omega\in\Omega} p(\omega) = 1$

$$\Rightarrow \Pr(A) = \sum_{\omega \in A} p(\omega)$$

- Event: a subset $A \subseteq \Omega$ of the sample space
- σ-Algebra: A family $\Sigma \subseteq 2^{\Omega}$ (Ω的幂集) of subsets of Ω, if:
  ① $\phi \in \Sigma$
  ② $A \in \Sigma \Rightarrow A^c \in \Sigma$
  ③ $A_1, A_2, \cdots \in \Sigma \Rightarrow \bigcup_i A_i \in \Sigma$
- Probability Space: $\Sigma \subseteq 2^{\Omega}$ is a σ-Algebre, a probability measure, also called probability law, is a function $\Pr: \Sigma \to [0,1]$ satisfying:
  ① 归一化: $\Pr(\Omega) = 1$
  ② σ-可加性: for 不相交的 $A_i \in \Sigma$: $\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$.
  三元组 $(\Omega, \Sigma, \Pr)$ 称做一个概率空间.
- Basic Properties of Probability:
  $\Pr(A^c) = 1 - \Pr(A)$ $\quad$ $\Pr(\phi) = 0$ $\quad$ $\Pr(A \setminus B) = \Pr(A) - \Pr(A \cap B)$
  $A \subseteq B \Rightarrow \Pr(A) \le \Pr(B)$ $\quad$ $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
- Union Bound: for $A_1, A_2, \cdots A_n \in \Sigma$:

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \Pr(A_i)$$

- 容斥原理: for $A_1, A_2, \cdots A_n \in \Sigma$:

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) - \cdots$$

$$= \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \phi}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right)$$

- Boole-Bonferroni Inequality: for $A_1, A_2, \cdots A_n \in \Sigma$, for any $k \ge 0$:

$$\sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ 1 \le |S| \le 2k}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right) \le \Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ 1 \le |S| \le 2k+1}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right).$$

Chapter 2 Conditional Probability

- For events A, B, $\Pr(B) > 0$, the conditional probability that A occurs given that B occurs is: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$

eg. Von Neumann's Bernoulli factory, boy and girl paradox.

· Chain rule: $Pr(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} Pr(A_i \mid \bigcap_{j<i} A_j)$

· Law of Total Probability (全概率定理): Let events $B_1, B_2, \dots B_n$ be a partition of Ω such that $Pr(B_i)>0$ for all i, then:

$$Pr(A) = \sum_{i=1}^{n} Pr(A \cap B_i) = \sum_{i=1}^{n} Pr(A|B_i) Pr(B_i)$$

eg: Monty Hall Problem, Gambler's Ruin

· Bayes' Law. Let events $B_1, B_2, \dots B_n$ be a partition of Ω such that $Pr(B_i)>0$ for all i, if event A has $Pr(A)>0$, then:

$$Pr(B_i|A) = \frac{Pr(B_i) Pr(A|B_i)}{Pr(A)} = \frac{Pr(B_i) Pr(A|B_i)}{Pr(A|B_1) Pr(B_1) + \cdots + Pr(A|B_n) Pr(B_n)}$$

Chapter 3 Independence

· $Pr(A \cap B) = Pr(A) Pr(B) \Rightarrow Pr(A \cap B) = Pr(A|B) Pr(B) = Pr(A) Pr(B)$

$\Rightarrow Pr(A|B) = Pr(A)$

即A的概率不受B的影响.

· Several Events: mutually independent:

$$Pr(\bigcap_{i \in J} A_i) = \prod_{i \in J} Pr(A_i), \text{ where } J \subseteq I, \{A_i \mid i \in I\}.$$

An event A is called mutually independent of a family $\{B_i \mid i \in I\}$ of events if for all disjoint finite subsets $J^+, J^- \subseteq I$:

$$Pr(A) = Pr\left(A \mid \bigcap_{i \in J^+} B_i \cap \bigcap_{i \in J^-} B_i^c\right)$$

· pairwise independent: 两两独立.

pairwise $\not\rightleftarrows$ mutually

eg. Network Reliability

· Conditional independence: $\Pr(A \cap B \mid C) = \Pr(A \mid C)\Pr(B \mid C)$.

Chapter 4 Random Variable

· Definition: Given $(\Omega, \Sigma, \Pr)$, a random variable is a function $X: \Omega \to \mathbb{R}$ satisfying that $\forall x \in \mathbb{R}$, $\{\omega \in \Omega \mid X(\omega) \le x\} \in \Sigma$ (Σ-measurable).

· Distribution (分布): CDF (累积分布函数) of a RV X is the F_X: $\mathbb{R} \to [0,1]$ given by $F_X(x) = \Pr(X \le x)$

① 单调性: $\forall x, y \in \mathbb{R}$, if $x \le y$ then $F_X(x) \le F_X(y)$

② 有界性: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$

· Discrete Random Variable: 离散的.

pmf: $p_X: \mathbb{R} \to [0,1]$ is given by: $p_X(x) = \Pr(X = x)$

☆ CDF F_X satisfies: $F_X(y) = \sum_{x \le y} p_X(x)$ (A kind of sum?)

· Continuous Random Variable: 连续的.

CDF can be expressed as: $F_X(x) = \Pr(X \le y) = \int_{-\infty}^{y} f_X(x)\,dx$

pdf: (概率密度函数) f_X

· Independence (mutually, pairwise).

· joint CDF (联合累积分布函数) $F_X: \mathbb{R}^n \to [0,1]$

$$F_X(x_1, \cdots, x_n) = \Pr(X_1 \le x_1 \cap \cdots \cap X_n \le x_n)$$

· joint mass function (联合质量函数):

$$p_X(x_1, \cdots, x_n) = \Pr(X_1 = x_1 \cap \cdots \cap X_n = x_n).$$

· marginal distribution of X_i in $(X_1, \cdots, X_n)$ is given by:

$$p_{X_i}(x_i) = \sum_{x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n} p_{(X_1, \cdots, X_n)}(x_1, \cdots, x_n)$$

Chapter 5 Discrete RV

- pmf
- Some basic discrete probability distributions:

① Bernoulli Trial
$$\begin{cases} X\in\{0,1\},\ p_X(k)=\Pr(X=k)=\begin{cases} p & \text{if } k=1 \\ 1-p & \text{if } k=0\end{cases} \\ E[X]=0\cdot(1-p)+1\cdot p=p \qquad Var[X]=p(1-p)\end{cases}$$

☆ Indicator: the indicator X of event A is a random variable:

$$X=I(A)=\begin{cases}1 & \text{if } A \text{ occurs} \\ 0 & \text{o.w.}\end{cases} \qquad E[X]=\Pr(A) \qquad Var[X]=\Pr(A)\Pr(A^c)$$

② Binomial Distribution
$$\begin{cases} X\in[n],\ p_X(k)=\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}. \\ E[X]=np \qquad Var[X]=np(1-p)\end{cases}$$

(consider n i.i.d. Bernoulli trials)

$X=\sum_{k\geq 1} I_k,\ I_k\in\{0,1\}$

$E[X]=\sum_{k\geq 1} E[I_k]$

$=\sum_{k\geq 1}(1-p)^{k-1}$

$=\frac{1}{p}.$

③ Geometric Distribution
$$\begin{cases} X\in\{1,2,\dots\},\ p_X(k)=\Pr(X=k)=(1-p)^{k-1}p \\ E[X]=\frac{1}{p} \qquad Var[X]=(1-p)/p^2\end{cases}$$

☆ memoryless: $\Pr(X=k+n \mid X>n)=\Pr(X=k)$

☆ Sum of Independent RVs:

$$p_{X+Y}(z)=\Pr(X+Y=z)=\sum_x \Pr(X=x\cap Y=z-x)$$

$$=\sum_x p_X(x)\,p_Y(z-x)=\sum_y p_X(z-y)\,p_Y(y)$$

Convolution (卷积): $p_{X+Y}=p_X * p_Y$.

④ Negative Binomal Distribution:

X: number of failures in a sequence of i.i.d. Bernoulli trials before r successes

$$X\in\{0,1,2,\dots\},\ p_X(k)=\Pr(X=k)=\binom{k+r-1}{k}(1-p)^k p^r$$

$$=(-1)^k\binom{-r}{k}(1-p)^k p^r$$

Actually, $X=(X_1-1)+(X_2-1)+\dots+(X_r-1)$ for i.i.d. $X_i\sim Geo(p)$

$E[X] = \frac{r(1-p)}{p}$ $Var(X) = \frac{r(1-p)}{p^2}$

(Consider r 次 几何分布再减 r)

⑤ Hypergeometric Distribution:

X: number of successes in n draws, without replacement, from a finite population of N objects, including exactly M ones, drawing of whom are counted as successes.

$X \in \{0,1,\cdots,n\}$, $p_X(k) = Pr(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$

$E[X] = \frac{nM}{N}$

⑥ Multinomial Distribution:

n balls are thrown into m bins, where each ball is thrown independently such that the ith bin receives the ball with probability p_i, where $p_1 + \cdots + p_m = 1$ is given.

$(X_1, X_2, \cdots, X_m)$: the ith bin receives exactly X_i balls.

$(X_1, X_2, \cdots, X_m)$ takes values $(k_1, \cdots k_m) \in \{0,1,\cdots,n\}^m$ that $k_1 + \cdots + k_m = n$.

$$p_{(X_1,\cdots,X_m)}(k_1,\cdots,k_m) = Pr\left(\bigcap_{i=1}^{m}(X_i = k_i)\right) = \frac{n!}{k_1!k_2!\cdots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}.$$

☆ $X_i \sim Bin(n, p_i)$, the marginal distribution of X_i is $Bin(n, p_i)$.

⑦ Poisson Distribution: $X \in \{0,1,2,\cdots\}$.

$p_X(k) = Pr(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ $E[X] = \lambda$ $Var(X) = \lambda$

☆ Sum: Independent $X \sim Bin(n_1, p)$, $Y \sim Bin(n_2, p) \Rightarrow X+Y \sim Bin(n_1+n_2, p)$.

$\sim$ $X \sim Pois(\lambda_1)$, $Y \sim Pois(\lambda_2) \Rightarrow X+Y \sim Pois(\lambda_1+\lambda_2)$.

· Some Famous Examples:

Balls into Bins Pattern Matching Random Walk

Random Graph Coupon Collector

Random Tree. Maximum Cut

· Expectation: $E[X] = \sum_x x\, p_X(x)$

☆ LOTUS: For $f: \mathbb{R} \to \mathbb{R}$, for discrete X and $X = (X_1, \cdots, X_n)$

· $E[f(X)] = \sum_x f(x) p_X(x)$

· $E[f(X_1, \cdots, X_n)] = \sum_{(x_1, \cdots, x_n)} f(x_1 \cdots x_n) p_X(x_1, \cdots, x_n)$

· Linearity of Expectation: $E[aX+b] = a E[X] + b$

$E[X+Y] = E[X] + E[Y]$.

☆ · Double Counting: 对于非负的 RV $X \in \{0, 1, 2, \cdots\}$ 有:

$$E[X] = \sum_{k=0}^{\infty} \Pr[X > k]$$

Pf: $E[X] = \sum_{x \geq 0} x \Pr[X=x] = \sum_{x \geq 0} \sum_{k=0}^{x-1} \Pr[X=x] = \sum_{k \geq 0} \sum_{x > k} \Pr[X=x] = \sum_{k \geq 0} \Pr[X > k]$.

Or proof by indicator.

??? · A random number of random variables: $X_1, X_2, \cdots, X_N$ for random N:

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]\, E[X_1].$$

· 条件期望: $E[X \mid A] = \sum_x x \Pr(X = x \mid A)$

· 全期望定律: $E[X] = \sum_{i=1}^{n} E[X \mid B_i] \Pr(B_i)$, B_i is a partition of Ω.

· $E[E[X \mid Y]] = \sum_y E[X \mid Y=y] \Pr(Y=y) = E[X]$

· Jensen's Inequality.

· Monotonicity of Expectation:

For RVs X, Y, for $c \in \mathbb{R}$:

① if $X \leq Y$ a.s. then $E[X] \leq E[Y]$.

② if $X \leq c$ a.s. then $E[X] \leq c$.

③ $E[|X|] \geq |E[X]| \geq 0$

· Averaging Principle

$$\begin{cases} Pr(X \geq E[X]) > 0 & \Leftarrow \text{ if } Pr(X<c)=1 \text{ then } E[X]<c \\ Pr(X \leq E[X]) > 0 & \Leftarrow \text{ if } Pr(X>c)=1 \text{ then } E[X]>c \end{cases}$$

Chapter 6 Moments and Deviations

· Markov 不等式：X 是一个非负值的 RV，对于 $\forall a>0$，$Pr(X \geq a) \leq \frac{E[X]}{a}$.

Pf：（用 indicator）令 $I = I(X \geq a)$，因为 $X \geq 0$ 且 $a>0$，有：

$$I = 1(X \geq a) \leq \left\lfloor \frac{X}{a} \right\rfloor \leq \frac{X}{a}$$

所以 $Pr(X \geq a) = E[I] \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}$.

· 推论：for any $c>1$，$Pr(X \geq cE[X]) \leq 1/c$.

· Generalized Markov's Inequality：X 是一个 RV 且 $f: \mathbb{R} \to \mathbb{R}_{\geq 0}$ 是一个非负值函数，则对于 $\forall a>0$，$Pr(f(X) \geq a) \leq \frac{E[f(X)]}{a}$.

· RV X 的 k 阶矩是 $E[X^k]$，k 阶中心矩是 $E[(X-E[X])^k]$.

如果 $E[X]=0$，则称 X 是中心化的，一个 RV X 可以用 $Y = X - E[X]$ 进行中心化

$Var[X] = E[(X-E[X])^2]$

· Chebyshev 不等式：X 是一个 RV，对于 $\forall a>0$，$Pr(|X-E[X]| \geq a) \leq \frac{Var[X]}{a^2}$

· 推论：对于标准差 $\sigma = \sqrt{Var[X]}$，对于 $\forall k \geq 1$，$Pr(|X-E[X]| \geq k\sigma) \leq 1/k^2$.

· median 中位数：RV X 的 median 是 any value m s.t.：

$$Pr(X \leq m) \geq 1/2 \quad \text{and} \quad Pr(X \geq m) \geq 1/2$$

· The expectation $\mu = E[X]$ is the value that minimizes: $E[(X-\mu)^2]$

· The median m is the value that minimizes：$E[|X-m|]$

· If X is a RV with finite expectation μ, median m, and standard deviation σ, then: $|\mu - m| \leq \sigma$.

Pf: $|\mu - m| = |E[X]-m| = |E[X-m]| \leq E[|X-m|] \leq E[|X-\mu|]$

$$= E\left[\sqrt{(X-\mu)^2}\right] \leq \sqrt{E[(X-\mu)^2]} = \sigma.$$

· Variance：$Var[X] = E[(X-E[X])]^2$

$$= E[X^2 - 2E[X]X + E[X]^2]$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2$$

$$= E[X^2] - E[X]^2$$

· Variance of Linear Function:

For RV X, Y and $a \in \mathbb{R}$：

① $Var[a] = 0$

② $Var[X+a] = Var[X]$

③ $Var[aX] = a^2 Var[X]$

④ $Var[X+Y] = Var[X] + Var[Y] + 2(E[XY] - E[X]E[Y])$

· Covariance (协方差): $Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$.

Properties:

① $Var[X] = Cov(X,X)$

② Symmetric: $Cov(X,Y) = Cov(Y,X)$

③ Distributive: $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$

$Cov(aX,Y) = a\,Cov(X,Y)$

④ if X,Y are independent then: $Cov(X,Y) = E[XY] - E[X]E[Y] = 0$

$\Rightarrow E[XY] = E[X]E[Y]$

· Expectation of Product:

(Cauchy-Schwarz): $E[XY]^2 \le E[X^2]E[Y^2]$

(Hölder): for $\forall p,q>0$ satisfying $\frac{1}{p}+\frac{1}{q}=1$:

$$E[XY] \le E[|X|^p]^{1/p} E[|Y|^q]^{1/q}$$

· Correlation 相关性: $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]\cdot Var[Y]}} \in [-1,1]$

X,Y are uncorrelated if $Cov(X,Y)=0$, means: $E[XY] = E[X]E[Y]$.

$Var[X+Y] = Var[X] + Var[Y]$.

· Variance of Sum:

① $Var[X+Y] = Var[X] + Var[Y] + 2Cov(X,Y)$.

② $Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i] + \sum_{i\neq j} Cov(X_i,X_j)$

③ for pairwise independent $X_1, X_2 \dots X_n$:

$Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} Var[X_i]$

· Variance of Discrete Uniform Distribution:

For integers $a \le b$, let X be chosen from $[a,b]$ u.a.r.

① $E[X] = \sum_{k=a}^{b} \frac{k}{b-a+1} = \frac{a+b}{2}$

② $E[X^2] = \sum_{k=a}^{b} \frac{k^2}{b-a+1} = \frac{2b^2+2ab+2a^2+b-a}{6}$

③ $Var[X] = E[X^2] - E[X]^2 = \frac{(b-a)(b-a+2)}{12}$

- Examples: Two-point Sampling. Cliques in Random Graph. (Erdős - Rényi)

  A Threshold Behavior in RG.

- The moment Problem:

  moment generating function (MGF):

$$M_X(t) = E[e^{tX}] = \sum_{k \geq 0} \frac{t^k E[X^k]}{k!}$$

Chapter 7 Continuous Random Variable

- A random variable $X: \Omega \to \mathbb{R}$ is called continuous, if its CDF can be expressed as:

$$F_X(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(u)du$$

for some integrable function $f: \mathbb{R} \to [0, \infty)$

f is called pdf of X.

- $f_X(x) = F_X'(x)$

- $Pr(X = x) = 0$ for all $x \in \mathbb{R}$.
- The density value $f_X(x) \geq 0$ is NOT a probability, but a proportion.

  $Pr(x < X \leq x + \Delta x) = F_X(x + \Delta x) - F_X(x) \approx f_X(x)\Delta x$

- PDF: f_X is the pdf of a CRV $X: \Omega \to \mathbb{R}$, iff

  ① $\int_{-\infty}^{\infty} f_X(x)dx = Pr(-\infty < X < \infty) = 1$

  ② $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.

- Joint Distribution: $F_{X,Y}: \mathbb{R} \to [0,1]$. given by:

$$F_{X,Y}(x, y) = Pr(X \leq x \cap Y \leq y)$$

joint pdf: $f_{X,Y}: \mathbb{R}^2 \to [0, \infty)$ for all $x, y \in \mathbb{R}$.

$$F_{X,Y}(x,y) = \int_{v=-\infty}^{y}\int_{u=-\infty}^{x} f_{X,Y}(x,y)\,du\,dv$$

$$\Rightarrow f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

· Marginal Distribution:

$$F_X(x) = Pr(X \le x) = F_{X,Y}(x,\infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{u,Y}(u,y)\,dy\,du$$

$$F_Y(y) = Pr(Y \le y) = F_{X,Y}(\infty,y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,V}(x,v)\,dx\,dv$$

marginal pdf: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

· Independence: $F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$

$$\Leftrightarrow f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$$

· Conditional Distribution:

① event A: $F_{X|A}(x) = Pr(X \le x \mid A) = \int_{-\infty}^{x} f_{X|A}(u)\,du$

全概率定理: For partition $B_1, \cdots, B_n$ of Ω with $Pr(B_i) > 0$ for all i: $f_X(x) = \sum_{i=1}^{n} Pr(B_i)\, f_{X|B_i}(x)$.

② RVs X, Y: $F_{X|Y}(x|y) = Pr(X \le x \mid Y = y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u,y)}{f_Y(y)}\,du$

Pf: $Pr(X \le x \mid y \le Y \le y+dy) = \frac{Pr((X \le x) \cap (y \le Y \le y+dy))}{Pr(y \le Y \le y+dy)}$

$$= \frac{\int_{u=-\infty}^{x} f_{X,Y}(u,y)\,dy\,du}{f_Y(y)\,dy} = \int_{-\infty}^{x} \frac{f_{X,Y}(u,y)}{f_Y(y)}\,dy$$

· Expectation: $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} x\,d F_X(x).$

k-th moment: $E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \int_{-\infty}^{\infty} x^k\,d F_X(x)$

· Double Counting: if CRV X takes on only nonnegative values (i.e. the pdf $f_X(x)=0$ if $x<0$), then:

$$E[X]=\int_0^\infty [1-F_X(x)]\,dx=\int_0^\infty \Pr(X>x)\,dx$$

Pf: $\int_0^\infty [1-F_X(x)]dx=\int_0^\infty \Pr(X>x)\,dx=\int_0^\infty\left(\int_x^\infty f_X(u)\,du\right)dx$

$$=\int_{u=0}^\infty f_X(u)\int_{x=0}^u 1\,dx\,du=\int_0^\infty u f_X(u)\,du=E[X].$$

- LOTUS: $E[g(X)]=\int_{-\infty}^\infty g(x)f_X(x)\,dx$
- Monotonicity of Expectation:
  ① If $X\geq 0$, then $E[X]\geq 0$
  ② If $X\geq Y$, then $E[X]\geq E[Y]$.
- Total Expectation: events $B_1,\dots,B_n$ be a partition of Ω such that $\Pr(B_i)>0$ for all i. $E[X]=\sum_{i=1}^n E[X|B_i]\Pr(B_i)$.

- $E[E[X|Y]]=E[X]$

Pf: $E[E[X|Y]]=\int_{-\infty}^\infty E[X|Y=y]f_Y(y)\,dy=\int_{-\infty}^\infty f_Y(y)\int_{-\infty}^\infty x\frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx\,dy$

$$=\int_{-\infty}^\infty x f_X(x)\,dx=E[X].$$

- Continuous Uniform Distribution:

RV X is uniform on $[a,b]$.

pdf $f(x)=\begin{cases}\frac{1}{b-a} & \text{if } a<x<b\\ 0 & \text{o.w.}\end{cases}$ and CDF $F(x)=\begin{cases}0 & \text{if } x\leq a\\ \frac{x-a}{b-a} & \text{if } a<x\leq b\\ 1 & \text{if } x>b\end{cases}$

$E[X]=\int_a^b \frac{x}{b-a}\,dx=\frac{a+b}{2}$

$E[X^2]=\int_a^b \frac{x^2}{b-a}\,dx=\frac{a^2+b^2+ab}{3}$

$\Rightarrow Var[X]=\frac{(b-a)^2}{12}$

- Exponential Distribution:

pdf $f(x)=\lambda e^{-\lambda x}$ and CDF $F(x)=1-e^{-\lambda x}$, $x\geq 0$

$E[X]=\int_0^\infty x\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda}$ $\quad E[X^2]=\frac{2}{\lambda^2}$

$Var[X] = \frac{1}{\lambda^2}$.

① Continuous limit of geometric distribution: an i.i.d. Bernoulli trial (with $p=\lambda\delta$) being performed after every δ time elapse. let X be the time of the first success.

$$Pr(X>x) = (1-p)^{x/\delta} = (1-\lambda\delta)^{x/\delta} \to e^{-\lambda x} \text{ as } \delta \downarrow 0$$

② memoryless: for $s,t \geq 0$:

$$Pr(X>s+t \mid X>t) = Pr(X>s)$$

Pf: $Pr(X>s+t \mid X>t) = \dfrac{Pr(X>s+t)}{Pr(X>t)} = \dfrac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s}$

$$= Pr(X>s).$$

③ if $X_1, \dots, X_n$ are independent exponential RVs with params $\lambda_1, \dots, \lambda_n$, respectively, then $\min_{1\le i\le n} X_i$ is exponential with param $\sum_{i=1}^{n}\lambda_i$.

Pf: $Pr(\min_{1\le i\le n} X_i > x) = Pr(\bigcap_{1\le i\le n}(X_i>x)) = \prod_{i=1}^{n} Pr(X_i>x)$

$$= \prod_{i=1}^{n} e^{-\lambda_i x} = e^{-x\sum_{i=1}^{n}\lambda_i}$$

· Poisson Point Process.

· Normal Distribution(正态分布): $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

$\mu=0, \sigma=1$, standard normal distribution $N(0,1)$.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$E[X] = \mu$, $Var[X] = \sigma^2$

① $Y = aX+b \sim N(a\mu+b, a^2\sigma^2)$.

$X \sim N(\mu, \sigma^2)$, then $\dfrac{X-\mu}{\sigma} \sim N(0,1)$

$X \sim N(0,1)$, then $\sigma X+\mu \sim N(\mu, \sigma^2)$.

· Convolution (卷积):

$$f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy.$$

if CRVs X, Y are independent, then:

$$f_{X+Y} = f_X * f_Y.$$

· Standard Normal Distribution:

pdf: $\phi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$

CDF: $\Phi(z) = Pr(X \le z) = \int_{-\infty}^{z} \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

没有闭式解.

· MGF of ND: $X \sim N(0,1)$: $M_X(t) = E[e^{tX}] = e^{t^2/2}$

· Bivariate ND: the joint density of standard bivariate normal RVs (X, Y) with param $-1<\rho<1$, is given by:

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

① the marginal distributions of X and Y are $N(0,1)$. And:

$Cov(X,Y) = E[XY] - E[X]E[Y] = \rho$.

② $\rho = 0 \Rightarrow f_{X,Y}(x,y) = \phi(x)\phi(y)$.

· General bivariate normal RVs (X,Y) with means μ_1, μ_2, variances σ_1^2, σ_2^2, and correlation $-1<\rho<1$, is given by:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)}$$

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Marginally $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $Cov(X,Y) = \sigma_1\sigma_2\rho$.

Chapter 8 Limit Theorems

· Modes of Convergence:

Let $X, X_1, X_2, \cdots : \Omega \to \mathbb{R}$ be RVs on probspace (Ω, Σ, Pr).

① $\{X_n\}$ converges in distribution (依分布) to X, denoted $X_n \xrightarrow{D} X$, if

weak $F_{X_n}(x) = Pr(X_n \le x) \to F_X(x) = Pr(X \le x)$ as $n \to \infty$

for all $x \in \mathbb{R}$ which $F_X(x)$ is continuous.

② $\{X_n\}$ converges in probability (依概率) to X, denoted $X_n \xrightarrow{P} X$, if

strong $Pr(|X_n - X| > \varepsilon) = 0$ as $n \to \infty$ for all $\varepsilon > 0$

③ $\{X_n\}$ converges almost surely to X, denoted $X_n \xrightarrow{a.s.} X$, if $\exists A \in \Sigma$ such that: $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in A$ and $Pr(A) = 1$

Some properties:

① $X_n \xrightarrow{D} X$ and $F_X = F_Y \Rightarrow X_n \xrightarrow{D} Y$

② $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$

③ if $X_n \xrightarrow{D} c$, where $c \in \mathbb{R}$ is a constant, then $X_n \xrightarrow{P} c$

④ $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X$

Other modes:

① $X_n \xrightarrow{1} X$ (convergence in mean) if $\lim_{n\to\infty} E[|X_n - X|] = 0$

② $X_n \xrightarrow{r} X$ (convergence in rth mean) if $\lim_{n\to\infty} E[|X_n - X|^r] = 0$

Sthrength of Convergence:

$$(X_n \xrightarrow{a.s} X) \Rightarrow (X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{D} X)$$

$$\Uparrow$$

$$(X_n \xrightarrow{s} X) \Rightarrow (X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{1} X)$$

for $s > r \ge 1$

· LLN: Let $X_1, X_2, \ldots$ be i.i.d. RVs with finite mean $E[X_1] = \mu$. And let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean.

① Weak LLN: $\bar{X}_n \xrightarrow{P} \mu$ as $n \to \infty$

② Strong LLN: $\bar{X}_n \xrightarrow{a.s.} \mu$ as $n \to \infty$

- De Moivre-Laplace Theorem:

Let $p \in (0,1)$ and $X_n \sim B(n,p)$. Then its standardization

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty$$

- CLT: Let $X_1, X_2, \ldots$ be i.i.d. RVs with $E[X_1] = \mu$ and $Var[X_1] = \sigma^2$. And Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean. Then:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty$$

Another Form: let $S_n = \sum_{i=1}^{n} X_i$, then:

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} N(0,1).$$

- Berry-Esseen Theorem: Let $X_1, X_2 \ldots$ be i.i.d. RVs with $E[X_1] = \mu$, $Var[X_1] = \sigma^2$, and $\rho = E[|X_1 - \mu|^3]$, And let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. There is an absolute constant C, such that for any z:

$$\left| Pr\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z\right) - \Phi(z) \right| \le \frac{C\rho}{\sigma^3\sqrt{n}}$$

where Φ stands for the CDF for SND $N(0,1)$.

Chapter 9 Concentration of Measure

- Chernoff Bound:

Let $X_1 \ldots X_n \in \{0,1\}$ be independent trials $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$.

① For any $\delta > 0$.

$$Pr(X \ge (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \le \begin{cases} e^{-\frac{\mu\delta^2}{3}} & 0 < \delta < 1 \\ 2^{-(1+\delta)\mu} & (1+\delta) \ge 2e \end{cases}$$

② For any $0 < \delta < 1$.

$$\Pr(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \le e^{-\frac{\mu\delta^2}{2}}$$

- Chernoff-Hoeffding bound:

  if $X_1, \dots, X_n \in \{0,1\}$ are independent and $S_n = \sum_{i=1}^{n} X_i$, then for any $t > 0$,
  $$\Pr(|S_n - E[S_n]| \ge t) \le 2e^{-\frac{2t^2}{n}}$$

- Hoeffding's Lemma:

  if $X \in [a,b]$ a.s. and $E[X] = 0$, then its MGF:
  $$M_X(\lambda) = E[e^{\lambda X}] \le e^{\lambda^2(b-a)^2/8}$$

- Hoeffding's inequality:

  if $S_n = \sum_{i=1}^{n} X_i$ and $X_i \in [a_i, b_i]$ are independent, then:
  $$\forall t > 0, \quad \Pr(|S_n - E[S_n]| \ge t) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

- Large Deviation (Concentration) Bound:

  if $X \sim N(\mu, \sigma^2)$, then for any $a > 0$:
  $$\Pr\left(\left|\frac{X-\mu}{\sigma}\right| \ge a\right) \le 2e^{-a^2/2}$$

- Example Problems: Controlling a Fair Voting, Sub-Gaussian tail. Error Reduction (two-sided case)
- The Method of Bounded Differences:

① McDiarmid's Inequality:

Let $X_1, \dots, X_n$ be independent random variables, where $X_i \in \mathcal{X}_i$ for all i. If $f: \mathcal{X}_1 \times \dots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the bounded differences property:
$$\forall i: \sup_{x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n, x_i' \in \mathcal{X}_i} |f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)| \le c_i$$

then for any $t > 0$,
$$\Pr(|f(X_1, \dots, X_n) - E[f(X_1, \dots, X_n)]| \ge t) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

Chapter 10 Martingale

- Doob Sequence: $Y_0, Y_1, \dots, Y_n$ of n-variate function $f: \mathbb{R}^n \to \mathbb{R}$ on RVs $X_1, \dots, X_n$.

is given by $\forall\, 0 \le i \le n$: $Y_i = E[f(X_1, \dots, X_n) \mid X_1, \dots, X_n]$

- Martingale (鞅): A sequence $\{Y_n : n \ge 0\}$ of RVs is a martingale with respect to another sequence $\{X_n : n \ge 0\}$ if, for all $n \ge 0$.
  ① $E[|Y_n|] < \infty$
  ② $E[Y_{n+1} \mid X_0, X_1, \dots, X_n] = Y_n$
- Super-martingale, sub-martingale
- filtration
- Examples of Martingale:
  ① Doob martingale
  ② Capital in a fair gambling game.
  ③ Unbiased 1D Random Walk
  ④ De Moivre's martingale
  ⑤ Polya's urn.

- Azuma's inequality: If a sequence $\{Y_n : n \ge 0\}$ is a martingale (w.r.t some sequence $\{X_n : n \ge 0\}$), and for all $n \ge 1$.
$$|Y_n - Y_{n-1}| \le c_n$$
then any $n \ge 1$ and any $t > 0$:
$$Pr(|Y_n - Y_0| \ge t) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right).$$

- Fair Gambling Game.
  If $\{Y_n : n \ge 0\}$ is a martingale w.r.t. $\{X_n : n \ge 0\}$, then $\forall n \ge 0$
$$E[Y_n] = E[Y_0]$$

Pf: By total expectation: $E[Y_n] = E[E[Y_n \mid X_0, X_1, \cdots X_{n-1}]]$
As a martingale, $E[Y_n \mid X_0, X_1, \dots X_{n-1}] = Y_{n-1}$
$\Rightarrow E[Y_n] = E[Y_{n-1}] = \cdots = E[Y_0]$

- OST: Let $\{Y_t : t \ge 0\}$ be a martingale and T be a stopping time, both w.r.t. $\{X_t : t \ge 0\}$. Then:
$$E[Y_T] = E[Y_0]$$

if any one of the following conditions holds:

① bounded time: there is a finite N such that $T < N$.

② bounded range: $T < \infty$ a.s., and there is a finite c s.t. $|Y_t| < c$ for all t

③ bounded differences: $E[T] < \infty$ and there is a finite c such that: $E[|Y_{t+1} - Y_t| \mid X_0, X_1, \ldots, X_t] < c$ for all $t \geq 0$.

· Wald's Equation:

Let $X_1, X_2, \ldots$ be i.i.d. RVs with $\mu = E[X_i] < \infty$. Let T be a stopping time with respect to $X_1, X_2, \ldots$ if $E[T] < \infty$, then:

$$E\left[\sum_{i=1}^{T} X_i\right] = E[T] \cdot \mu.$$

Chapter 11 Random Process (*)

· Markov Chain

Chapter 12 Statistics. (*)

· Parameter Estimation (无偏, 相合, 一致)

· Hypothesis Test

· Bayes Estimation

· Variance Analysis

· Relation and Regression Analysis