概率论与随机过程.

1. Event and Their Probilities
2. Random Variable
3. Random Vectors
4. Sequences of Random Variables.
5. Introduce to Stochastic Processes.
6. Stationary Processes.
7. Finite Markov Chain
8. Independent - Increment Process

一、Event and Their Probabilites.

1、集合 $A-B$，A, but no B.

$A-B = A-AB = A\bar{B}$, $A\cup B = A\cup \bar{A}B$.

2. ① Classical Probability

有限样本空间，等概率，$P(A) = \frac{\#A}{\#\Omega}$ → 表示数量.

有限可加性 $P(\bigcup_i A_i) = \sum_i P(A_i)$，这三种都满足 互斥, disjoint / mutually exclusive.

如果事件不同，顺序取 n，$\binom{N}{n} n!$ 为所有排列数.

② Geometric Probability

$P(A) = \frac{L(A)}{L(\Omega)}$ → 测度，蒲丰投针，互相等待.

③ Frepuency

$F_n(A) = \frac{f_n(A)}{n}$，频率可以当成概率.

3、Probability space.

σ-algebra → $\begin{cases} \Omega \in \mathcal{F} \\ F \in \mathcal{F} \Rightarrow \bar{F} \in \mathcal{F} \\ \bigcup_n F_n \in \mathcal{F} \end{cases}$ 则 $\mathcal{F}$ 最小的 $\{\emptyset, \Omega\}$.

probability space $(\Omega, \mathcal{F}, P)$.

运算性质：

$P(A\cup B) = P(A) + P(B) - P(AB)$.

$P(A\cup B\cup C) = P(A) + P(B) + P(C) - [P(AB) + P(AC) + P(BC)] + P(ABC)$ 双数减单数加.

$P(\bigcup_i A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) - \sum_{i<j<k<l} P(A_i A_j A_k A_l) + \dots + (-1)^{n+1} P(A_1 A_2 \cdots A_n)$

4. Conditional Probabilities.

条件概率: $P(A|B)=\dfrac{P(AB)}{P(B)}$，条件概率也是概率.

乘法公式: $P(A_1A_2\ldots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1})$

全概率公式: B_i 互斥, $\bigcup_i B_i=\Omega \Rightarrow P(A)=\sum_i P(B_i)P(A|B_i)$.

贝叶斯公式: $P(B_i|A)=\dfrac{P(B_iA)}{P(A)}=\dfrac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)}$ → 条件概率乘法公式 → 全概率公式.

5. Independence of event.

$P(AB)=P(A)P(B) \Leftrightarrow P(A|B)=P(A) \Leftrightarrow P(B|A)=P(B)$

互斥必不独立(因为已有互斥关系了).

A,B 独立，$\bar{A},A,\bar{B},B$，之间(A类和B类)的之间都独立.

① 多事件独立.

mutually independent: 两两独立并且 $P(ABC\ldots)=P(A)P(B)P(C)\ldots$

pairwise independent: 只有两两独立.

说多变量(事件) independent(不是两两 independent)，那么 $P(A_{i_1}A_{i_2}\ldots A_{i_j})=P(A_{i_1})\ldots P(A_{i_j})$.

对于所有可能的联合概率，都可拆分。反之，所有组合均可拆分，这些事件 independent.

② Bernoulli Trials.

两结果，互相独立. $P_n(k)=\binom{n}{k}p^k(1-p)^{n-k}$

二、Random Variable.

1. 分布. cumulative distribution function $F(x)=P(X\le x),\ \forall x\in R$. (d.f.).

$F(x)$:
- 不严格递增, $x_1>x_2$, $F(x_1)\ge F(x_2)$.
- $\lim_{x\to-\infty}F(x)=0$, $\lim_{x\to\infty}F(x)=1$.
- 右连续, $F(x)=F(x^+)$, $\forall x\in R$.

$P(X>x)=1-F(x)$.

$P(x_1<X\le x_2)=F(x_2)-F(x_1)$.

$P(X<x)=F(x^-)$.

$P(X=x)=F(x)-F(x^-)$

$P(X\ge x)=1-F(x^-)$.

① 离散型 → Probability function $p(x)$ (p.f.). 画分布表格呈现. d.f. 就是求和

② 连续型 → probability density function (p.d.f) $P(a\le X\le b)=\int_a^b f(x)\,dx$.

d.f. $F(x)=P(X\le x)=\int_{-\infty}^{x}f(x)\,dx$, $f(x)=\dfrac{d}{dx}F(x)$, $\int_{-\infty}^{+\infty}f(x)\,dx=F(+\infty)=1$

因为连续 所以 $F(x^{+-})$ 的 +, − 或者 $P(a<X)$ $P(a\le X)$ 不论在哪等号都一样.

<多布随机变量的分布函数求法>. $Y=g(X)$

离散情况: 画表格, 直接计算, 有公式 $p_Y(y)=P(Y=y)=P(g(X)=y)=P(X=g^{-1}(y))=p_X(g^{-1}(y))$.

连续情况: 给定 $f_X(x)$.

注意判定新 r.v. 的作用域.

$F_Y(y)=P(Y\le y)=P(g(X)\le y)=P(X\le g^{-1}(y))$.

$=F_X[g^{-1}(y)]$

$f_Y(y)=\dfrac{d}{dy}F_Y(y)=\dfrac{d}{dy}F_X[g^{-1}(y)]=\dfrac{d}{dg^{-1}(y)}F_X[g^{-1}(y)]\dfrac{dg^{-1}(y)}{dy}=f_X[g^{-1}(y)]\dfrac{dg^{-1}(y)}{dy}$

2、Mathematical Expectation and Variance.

① 期望 Expectation.（绝对收敛时 $\int |x| f(x)dx < \infty$ 时）存在）.

$$\mu := E(X) = \begin{cases} \sum_{i=1}^{\infty} x_i p(x_i), & \text{discrete} \\ \int_{-\infty}^{+\infty} x f(x) dx = \underline{\int_{-\infty}^{+\infty} x \, dF(x)}, & \text{continuous} \end{cases}$$

→ 当 X 整数且非负, $E(X) = \sum_{x=0}^{\infty} P(X > x)$.

$$E[f(X)] = \begin{cases} \sum_i f(x_i) p(x_i) \\ \int_{-\infty}^{+\infty} f(x) f(x) dx = \underline{\int_{-\infty}^{+\infty} f(x)\, dF(x)} \end{cases}$$

线性算子.

$E[af(X)+b]$
$= aE[f(X)]+b$.

如果 $f(\cdot)$ 是线性算子, 那么 $E[f(X)] = f[E(X)]$, 反之, 不等.

② 方差 Variance.

$$\sigma^2 = E[(X-\mu(X))^2] = Var(X) = \begin{cases} \sum_i (x_i-\mu)^2 p(x_i) \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx. \end{cases}$$

$Var[af(X)+b]$
$= a^2 Var[f(X)]$.

$Var(X) = E(X^2) - E^2(X)$, $\sigma = SD[X] = \sqrt{Var(X)}$.

方差为0, X 为 constant.

原点矩, 中心矩, 高阶矩存在, 低阶矩一定存在.

③ 数字特征的应用.

a) $X \geq 0 \Rightarrow \underline{P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}}$, Markov's inequality. 令 $X' = (X-\mu)^2$, $\varepsilon' = \varepsilon^2$

b) $\underline{P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}}$, Chebyshev's inequality.

3. Discrete Random Variables.

① Binomial Distribution. $X \sim B(n,p)$, $x = 0, 1, 2, \dots$,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad E(X) = np, \quad Var(X) = np(1-p).$$

当 n 非常大, p 非常小时.

$$\lim_{n \to +\infty} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}, \text{ where } \lim_{n\to\infty} np = \lambda$$

与泊松分布的概率会相对应.

② Geometric Distribution. $X \sim G(p)$, $\underline{x = 1, 2, \dots}$,

前 x-1 次均不成, 最后成. $p(x) = (1-p)^{x-1} p$.

$E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$, $F(x) = 1-(1-p)^x$.

无记忆性. $P(X > x+y \mid X > x) = P(X > y)$.

③ Poisson Distribution $X \sim Poi(\lambda)$, $x = 0, 1, \dots$

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \dots$$

$E(X) = Var(X) = \lambda$.

4. Continuous Random Variables

① Uniform Distribution. $X \sim U(a,b)$.

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases} \Rightarrow a=0,\ b=1 \Rightarrow \text{standard uniform distribution}$$

$$E(X) = \frac{a+b}{2}. \quad Var(X) = \frac{(b-a)^2}{12}.$$

如果 $Y = F_X(x)$，那么变量 Y 必然在 $[0,1]$ 上均匀分布.

反过来也成立，一个均匀分布 $Y \sim U(0,1)$，通过构造 $X = F_X^{-1}(y)$，可使得 X 的累积分布函数满足 $P(X \le x) = F_X(x)$.

② Exponential Distribution. $X \sim Exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

$E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$，无记忆性 $P(X > t+s \mid X > s) = P(X > t)$.

③ Normal Distribution. $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty.$$

重要极限 $\rightarrow \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$. $\mu = 0,\ \sigma = 1 \Rightarrow$ standard normal distribution.

对于标准正态分布 $Z \sim N(0,1)$，$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$，$-\infty < z < \infty$. 且 $\Phi(z) = P(Z \le z)$ 可查表

可以使用正态分布来近似 Poisson 和 binomial 分布，从离散到连续，一般加一个 0.5 校正，如.

$$\underset{(\text{离散})}{P(X=20)} = \underset{(\text{连续})}{P(19.5 < Y < 20.5)}.$$

总结:
- B, Poi → count
- G, Exp → waiting time.
- N → sums.

三、Random Vectors. 相同样本空间上的不同随机变量.

1. Random Vectors and Joint Distributions.

Joint d.f. $F(x,y) = P(X \le x, Y \le y)$, $-\infty < x, y < +\infty$.

$P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) \ge 0$

① 单调
② $F(-\infty, y) = 0$, $F(x, -\infty) = 0$
$F(\infty, \infty) = 1$.
③ $\lim_{x \to x_0^+} F(x,y) = F(x_0, y)$.
$\lim_{y \to y_0^+} F(x,y) = F(x, y_0)$.
右连续.

marginal distribution: $F_Y(y) = \lim_{x \to +\infty} F(x,y)$，使用画图方式理解.

① Discrete Random Vectors.

$p(x,y) = P(X=x, Y=y)$, $p_X(x) = \sum_y p(x,y)$, $F(x,y) = \sum_{x_i \le x} \sum_{y_k \le y} p(x_i, y_k)$，使用表呈现.

② Continuous Random Vectors.

$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\,du\,dv$. $P[(X,Y) \in A] = \iint_{(x,y) \in A} f(x,y)\,dx\,dy$.

$F_X(x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(x,y)\,dy \right) dx \Rightarrow F_X(x) = \lim_{y \to \infty} F(x,y)$, $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$.

③ $f(x,y) = \begin{cases} \dfrac{\partial^2 F}{\partial x \partial y}, & \text{where exists} \\ 0 & \text{orthermise.} \end{cases}$ ⇒ 但是一般很少有这种类型的题目，好吧，这是描述本风格，要记住，不过这些很多会被注意.

Uniform distribution $(0<x<a, 0<y<b.)$ $P((X,Y)\in A) = \dfrac{L(A)}{ab}.$

2维 normal distribution. $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$.

并不是所有联合事件都拥有 Joint p.d.f. 边缘分布是两个正态分布 $N(\mu_1,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$.

2、Independence of Random Variables.

$$F(x,y) = F_X(x)F_Y(y) \Rightarrow X, Y \text{ independent}, X \perp Y.$$

$$X \perp Y \Leftarrow \begin{cases} P(x,y) = P_X(x)P_Y(y), & \text{discrete} \\ f(x,y) = f_X(x)f_Y(y), & \text{continuous.} \end{cases}$$

$\boxed{\text{独立的充要条件：}\ \Omega = \{(x,y),\ \text{scope为矩形}\},\ f(x,y) \xrightarrow{\text{可写为}} f(x)\,h(x)g(y)}$ 分离变量.

X, Y 独立，$g(X)$，$h(Y)$ 也独立.

3. Conditional Distributions.（条件概率也是概率）

①离散型.

$$P_{X|Y}(x|y) = P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P(x,y)}{P_Y(y)}.$$

$$F_{X|Y}(x|y) = \sum_{x_i \le x} P_{X|Y}(x_i|y)$$

在Poission分布的基础上，将Poission视为样本空间，再进行二项分布，得到的分布依旧为Poission分布，唯一变化的是强度变为 λp，仍也符合Poission定义.

②连续型.

$$f_{X|Y} = \frac{f(x,y)}{f_X(x)} \quad \text{if } 0<f_X(x)<+\infty, \text{ for } x\in I, \text{ and } 0 \text{ orthermise.}$$

只有 $f_X(x) > 0$ 时，条件概率才有意义，条件概率是单变量概率，但是 x 的 scope I 与 y（可能）有关，注意讨论存在和范围.

利用连续条件概率，可以求 $P(Y \ge y \mid X = x_0)$ 这样的概率，常方法是做不了的，因为没有考虑极限.

另外，二维正态分布的边缘概率和条件概率都是正态分布.

4. One Function of Two Random Variables.

①离散. 可以通过表格呈现，对角线 → $X+Y$，max(X,Y)，min(X,Y)

$$P_{X+Y}(n) = \sum_x P_X(x)P_Y(n-x) = P_X * P_Y$$

注意求和的范围，只有 n 个离散分布从1开始，$\sum_{x=1}^{n-1}$，其余都从0开始 $\sum_{x=0}^{n}$.

ex1. 对于独立 $X\sim B(n_1,p)$，$Y\sim B(n_2,p)$，由若积公式和 $\sum_{x_1=0}^{k}\binom{n_1}{x_1}\binom{n_2}{k-x_1}=\binom{n_1+n_2}{k}$

有 $X+Y\sim B(n_1+n_2,p)$。依然为二项分布。

ex2. 对于独立 Poisson，$X\sim Poi(\lambda)$，$Y\sim Poi(\mu)$，$X+Y\sim Poi(\lambda+\mu)$，强度相加。

则这两种（以及后面的 Gauss），为稳定特征，B ⇒ 总次数为 n_1+n_2，Poi ⇒ 总强度为 $\lambda+\mu$。

② 连续型

a) 正常情况组合：每一类也和离散分。

$$F_Z(z)=P(Z\le z)=P(\varphi(X,Y)\le z)=P((X,Y)\in A)=\iint_A f(x,y)\,dx\,dy$$

求区域的过程中，将 z 看作一个 constant.

$$f_{\varphi(X,Y)}(z)=\frac{d}{dz}F_Z(z)=\frac{d}{dz}\iint_A f(x,y)\,dx\,dy$$

正规化方法都要使用积分，但有一个不成熟的技巧。

$$\frac{d}{dz}\int_{-\infty}^{+\infty}dx\int_{-\infty}^{z-x}f(x,y)\,dy=\int_{-\infty}^{+\infty}f(x,z-x)\,dx.$$

→ 对于积分的上下限的变换次序，并不影响结果。

只需记住 $f_{X+Y}=f_X * f_Y$，$f_{X+Y}(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\,dx$ 要求只即可，其余不用记住。

对于这个方法的要求熟悉，反转＋平移位即可，剩下不必要麻烦。

关于两个独立 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$ 相加，$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$ 依然为正态分布。

b) $\max(X,Y)$

利用概率转换。$F_Z(z)=P(\max(X,Y)\le z)=P(X\le z,Y\le z)=F(z,z)$.

加上独立条件便可更简洁。

c) $\min(X,Y)$.

概率转换，
$$\begin{aligned}F_Z(z)=P(\min(X,Y)\le z)&=1-P(\min(X,Y)>z)\\&=1-P(X>z,Y>z)\\&=1-(1-F_X(z)-F_Y(z)+F(z,z))\\&=F_X(z)+F_Y(z)-F(z,z).\end{aligned}$$

五. Numerical Characteristic of Random Vectors.

① 期望 Expectation.

级数求和 ⇒ $E[h(X,Y)]=\begin{cases}\sum\sum h(x,y)\,p(x,y).\\ \iint_{x\,y} f(x,y)\,h(x,y)\,dx\,dy\end{cases}$，必须有 $h(x,y)$，将 X, Y 联系起来

$E(X+Y)=E(X)+E(Y)\to E(aX+bY+c)=aE(X)+bE(Y)+c$.

$|E(X)|\le E(|X|)$，独立 → $E[g(X)h(Y)]=E[g(X)]E[h(Y)]$.

Cauchy-schwarz inequality $(E(XY))^2\le E(X^2)E(Y^2)$.

$Var(X+Y) = Var(X) + Var(Y) + \cancel{2Cov(X,Y)}\ 2Cov(X,Y)$.

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

若 $X \perp Y$，$Var(X+Y) = Var(X) + Var(Y)$

$Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y)$.

② Covariance and Correlation.

协方差. $Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$

$= E(XY) - E(X)E(Y)$.

$Cov(aX+c, bY+d) = ab\,Cov(X,Y)$

$Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$

$\hookrightarrow \begin{cases} \sum_x\sum_y (X-E(X))(Y-E(Y))\, P(x,y). \\ \iint_{x\,y} (X-E(X))(Y-E(Y)) f(x,y)\,dx\,dy \end{cases} \to$ 不如直接计算 $f(x,y)$ 的均值.

correlation coefficient.

$\rho(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$ 消除量纲, 归一化线性关系. $\rho(aX+b, cY+d) = \rho(X,Y)$.

独立 ⇒ 不相关, 反之不成立

等价表达 ⇒ uncorrelation

$Cov(X,Y)=0$

$E(XY) = E(X)E(Y)$

$Var(X+Y) = Var(X) + Var(Y)$.

, 但是, 对于 Normal 分布是特例!

对于正态分布 $(X,Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ 来说, $X \perp Y \Leftrightarrow \rho = 0$.

四、Sequences of Random Variables.

1. 分布族于数字特征.

$\{X_n(w), w \in \Omega\}_{n=1}^{\infty} \to \{X_n\}_{n=1}^{\infty}$

分布函数族: $\{F(x_1, \ldots x_k; n_1 \ldots n_k), \forall k, \forall n_i\}$.

IID ~ independent and identically distribution 独立同分布 ⇒ 最平稳随机过程.

$\mu_X(n) = E(X_n)$, $\underbrace{\Psi_X^2(n)}_{\uparrow\ \text{power function}} = E(X_n^2)$, $\sigma_X^2(n)$, $\underbrace{R_X(n,m)}_{\uparrow\ \text{自相关}} = E(X_n X_m)$

$C_X(n,m) = Cov(X_n, X_m)$ 自协方差. $\{\underbrace{C_{XY}(n,m)}_{\text{互协方差}} = \underbrace{R_{XY}(n,m)}_{\text{互相关}} - \mu_X(n)\mu_Y(m)\}$ (cross-).

2. 收敛类型

① almost surely. $X_n \xrightarrow{a.s.} X$ if $\{w \in \Omega: X_n(w) \to X(w)$ as $n \to \infty\}$ 概率为1.

就极限意义而言, 可能有有限个点不收敛, 但只要整体概率已为1, 不影响.

② in probability $X_n \xrightarrow{P} X$, if $\underbrace{P(|X_n - X| > \varepsilon)}_{} \to 0$ as $n \to \infty$ for all $\varepsilon > 0$

使用切比雪夫不等式进行证明. $P(|X_n - \mu| > \varepsilon) \le \dfrac{Var(X_n)}{\varepsilon^2}$

③ in rth mean, $X_n \xrightarrow{r} X$, $\underbrace{E(|X_n - X|^r)}_{} \to 0$ as $n \to \infty$, $r=2$ 时, 叫均方收敛.

a.s. 收敛 + $|X_n| < \infty$ ⇒ 均方收敛, 使用 $E(|X_n - X|^r)$ 进行验证.

④ $X_n \to X$ in distribution, $X_n \xrightarrow{D} X$.

$$P(X_n \le x) \to P(X \le x),\ n \to \infty.$$

分布一致，即依分布收敛.

$$\begin{array}{ccc} (X_n \xrightarrow{a.s.} X) & \searrow & \\ r\ge 1 \Uparrow & & (X_n \xrightarrow{P} X) \longrightarrow (X_n \xrightarrow{D} X) \\ (X_n \xrightarrow{r} X) & \nearrow & \end{array}$$

3、The Law of Large Numbers. 大数定律 (LLN)

$$\overline{X}_n = \frac{S_n}{n} = \frac{\sum_1^n X_i}{n},\ S_n = \sum_1^n X_i$$

WLLN: $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \ge \varepsilon\right) = 0$ 或 $\lim_{n\to\infty} P(|\overline{X}_n - E(\overline{X}_n)| \ge \varepsilon) = 0$, $\underline{P}$.

SLLN: $P(\lim_{n\to\infty} |\overline{X}_n - E(\overline{X}_n)| = 0) = 1$, $\underline{a.s.}$

Chebyshev's LLN.

均值存在，方差有界 $\Rightarrow P\left(\left|\overline{X}_n - \frac{\sum_{i=1}^n \mu_i}{n}\right| \ge \varepsilon\right) \to 0$, as $n \to \infty$

证明 $P\left(\left|\overline{X}_n - \frac{\sum \mu_i}{n}\right| \ge \varepsilon\right) \le$? 切比雪夫不等式.

Bernoulli's LLN.

$P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) \to 1$, as $n \to \infty$., 频率的稳定值为概率.

4. The Central Limit theorem 中心极限定理. (CLT)

对于 $S_n = \sum_1^n X_i$, 我们可以用卷积公式，但，对于 Poission, Normal, binomial, 最简便的方法是通过 $(\lambda)(\mu, \sigma^2)(n, p)$ 确定，而不是卷积公式（虽然也是由卷积证明的）.

无论 Levy-Linderberg CLT, De Moivre-Laplace CLT, Liapunov CLT. 都表明.

$$S_n^* = \frac{S_n - E(S_n)}{\sigma_{S_n}},\quad S_n = \sum_{i=1}^n X_i$$

$$F_n(x) = P(S_n^* \le x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}} ds.$$

即可以用正态分布去近似任何独立重复实验，只需将 $S \to S^*$ 标准化

大数定律让我们知道 n 足够大时，序列的均值可收敛到事件的概率均值.

而中心极限定理告诉我们，

$$P(|\overline{X}_n - E(\overline{X}_n)| \ge \varepsilon) = 1 - P(|\overline{X}_n - E(\overline{X}_n)| < \varepsilon).$$

$$= 1 - P\left(\frac{|\overline{X}_n - E(\overline{X}_n)|}{\sqrt{Var(\overline{X}_n)}} < \frac{\varepsilon}{\sqrt{Var(\overline{X}_n)}}\right)$$

$$= 2\left(1 - \Phi\left(\frac{\varepsilon}{\sqrt{Var(\overline{X}_n)}}\right)\right).$$

序列如何逼近这个均值.

五、Introduction to Stochastic Processes.

1、定义与分类.

2、The Distribution Family and the moment Functions

3、The Moments of the Stochastic Processes.

本节所求的，矩都存在.

① Mean, Autocorrelation, Autocovariance.

$$\mu_X(t),\ \sigma_X^2(t),\ R_X(t_1,t_2)=\cancel{E(X_{t_1}X_{t_2})}\ E(X_{t_1}X_{t_2}),\ C_X(t_1,t_2)=\operatorname{Cov}(X_{t_1},X_{t_2})$$

玩随机过程，一定分清哪个是随机变量，哪个不是随机变量.

$E(X_{t_1}X_{t_2})$ 是一种内积，交换，非负，三角不等式都满足.

分布 $\begin{cases} \text{first-order distribution of } X_t \to \cancel{\mu_X(t),\ \sigma_X^2(t)} \to \text{确定 } \mu_X(t),\ \sigma_X^2(t). \\ \text{second-order distribution of } X_t \Rightarrow \to \text{确定 } \mu_X(t)_{(2个)},\ \sigma_X^2(t)_{2个},\ \rho. \end{cases}$

$\mu\ \Sigma$.

② Cross-correlation and Cross-covariance.

$$R_{XY}(t_1,t_2)=E(X_{t_1}Y_{t_2})$$

$C_{XY}(t_1,t_2)=0 \Rightarrow$ uncorrelated.，独立必不相关.

$R_{XY}(t_1,t_2)=0 \Rightarrow$ orthogonal.

4、Stochastic Analysis.

< continuous > < differentiable >.

连续 ~~收敛~~ 一指均方收敛，它的均值也连续.

$$\lim_{\varepsilon\to 0} E\left[(X_{t_0+\varepsilon}-X_{t_0})^2\right]=0.$$

$\Updownarrow$ 等价于 (X_t 连续).

$R_X(t_1,t_2)$ 连续.

$$R_{X'}(s,t)=E(X'_sX'_t)=\frac{\partial^2 R_X(s,t)}{\partial s\,\partial t}.$$

积分与求导与普通解析处理无异.

六、Stationary Processes.

0. 和差化积与积化和差.

一般不用.

$$\sin\alpha+\sin\beta=2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

$$\cos\alpha+\cos\beta=2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

$$\sin\alpha\cos\beta=\frac{1}{2}\left(\sin(\alpha+\beta)+\sin(\alpha-\beta)\right)$$

$$\cos\alpha\cos\beta=\frac{1}{2}\left(\cos(\alpha+\beta)+\cos(\alpha-\beta)\right)$$

$$\sin\alpha\sin\beta=-\frac{1}{2}\left(\cos(\alpha+\beta)-\cos(\alpha-\beta)\right)$$

1. Strict Stationary Process. (SSS).

$F_X(x_1,\dots,x_n;t_1,\dots,t_n)=F_X(x_1,\dots,x_n;t_1+h,\dots,t_n+h)$.

IID 序列 $F(x_1,x_2,\dots,x_k)=\prod_{i=1}^{k}F(x_k)$, 必然 SSS

求 $E(X_nX_m)$ 时 注意讨论 $m=n$ 的情况.

$E(X_t^2)<\infty \Rightarrow$ second order process.

SSS + second order $\Rightarrow$ WSS. { $\mu_X(t)$ 为 constant. / $R_X(\tau)$ 只与差值有关 } 一阶和二阶分布矩 不随着时间变化发生变化.

2. Wide Stationary Processes (~~…~~ WSS).

secondc order + μ_X 是 constant + $R_X(\tau)$ 只依靠差值 $\Rightarrow$ SSS. (这并不意味着WSS可以推SSS)

$|R_X(t,s)|\le\sqrt{R_X(s,s)R_X(t,t)}\le|R_X(0)|$, 零点为正且最大.

实偶函数. $R_X(t,s)=R_X(\tau)=R_X(s,t)=R_X(-\tau)$. 三角不等式.

在针对抽象的随机过程证明 WSS 时, { 均值直接求 = constant. / R_X 的求法不直接套求 $R_X(t,t+\tau)$ 可剩去一部分换元 }

<随机电报过程>.

$E(X_tX_{t+\tau})=I^2\sum_{n=0}^{\infty}P(N_t=2n)-I^2\sum_{n=0}^{\infty}P(N_t=2n+1)$

I^2 或 $-I^2$ 　　后使用 ~~…~~ 泰勒展开证明.

注意讨论 $t=0$ 未定义, $t'=t+\tau$ 讨论全部定义域 $(-\infty,+\infty)$, 偶函数, 都应只与 τ 有关.

$E(X_tX_{t+\tau})=E(X_{t'}X_{t'-\tau})=E(X_{t'}X_{t'+|\tau|})$

X_t 周期T, R_τ 同样周期T; $R_X(0)=R_X(T)\Rightarrow R_X(\tau)=R_X(\tau+T)$ 即有相当反之一定正确.

$[R_X(t_j-t_k)]$ 非负正定, $aSa^T\ge0$.

对于 WSS, $R_X(\tau)$ 在0处连续, 随机序列均方连续.

3. Joint Stationary Process.

$\{X_t\}$ 和 $\{Y_t\}$ ~~分别平稳,且~~ $R_{XY}(t,t+\tau)=R_X(\tau)$

那么 $\{X_t\},\{Y_t\}$ jointly stationary process.

4. Ergodicity of Stationary Process.

每个时间段都代表全部时间段 $\rightarrow$ ergodic.

ensemble average, ~~…~~ ensemble correlation. 集平均, 集相关, 与概率论中定义一致

time average, 时平均 $\langle X_t\rangle=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X_t\,dt$

time correlation, 时相关 $\langle X_tX_{t+\tau}\rangle=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}X_tX_{t+\tau}\,dt$.

一般给的时性 来代相等判别. 均方为0 { $\lim_{T\to\infty}E\left[\left(\frac{1}{2T}\int_{-T}^{T}X_t\,dt-\mu_X\right)^2\right]=E\left[(\langle X_t\rangle-\mu_X)^2\right]=0\Rightarrow$ ergodic in mean.

$\lim_{T\to\infty}E\left[\left(\frac{1}{2T}\int_{-T}^{T}X_tX_{t+\tau}\,dt-R_X(\tau)\right)^2\right]=E\left[(\langle X_tX_{t+\tau}\rangle-R_X(\tau))^2\right]=0\Rightarrow$ ergodic in correlation.

WSS + 时间遍历性（均值 + 相关）= ergodic process.

对于随机相位过程式遍历性，很多时候对答案影响不太大，当然，很多设有区别，都也较容易

均值遍历 iff. $\lim_{T\to+\infty}\frac{1}{T}\int_0^{2T}(1-\frac{\tau}{2T})[R_X(\tau)-\mu_X^2]d\tau=0$.

<随机电信过程> 利用上述 $\frac{1}{T}\int_0^{2T}(1-\frac{\tau}{2T})[R_X(\tau)-\mu_X^2]d\tau=0$ 证明均值 ergodic.

5. Power Spectral Density of stationary Processes.

① 信号总能量 $=\int_{-\infty}^{+\infty}x^2(t)dt=\int_{-\infty}^{+\infty}\frac{1}{2\pi}|F(w)|^2dw$

信号功率 $=\frac{1}{2T}\int_{-T}^{T}x^2(t)dt=\frac{1}{(4\pi T)}\int_{-\infty}^{+\infty}|F(w)|^2dw$

(Q) 平均功率 $=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-\infty}^{+\infty}\underbrace{\frac{1}{2T}|F(w)|^2}_{\text{功率谱密度 (PSD)}}dw \Rightarrow S_X(w)=\lim_{T\to\infty}\frac{1}{2T}|F_T(w)|^2$.

对功率谱密度积分是平均功率.

$Q=\lim_{T\to\infty}E\left(\frac{1}{2T}\int_{-T}^{T}x_T^2dt\right)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\underbrace{E(X_T^2)}_{\tau=0}dt$

$=\frac{1}{2T}\int_{-T}^{T}R_X(0)dt=R_X(0)$.

$R_X(\tau)\xleftrightarrow{FT}S(w)$, $Q=\int_{-\infty}^{+\infty}\frac{S(w)}{2\pi}dw=\frac{1}{2\pi}\int_{-\infty}^{+\infty}S(w)e^{jw0}dw=R_X(0)$.

($S(w)\geq0$, $\int_{-\infty}^{+\infty}S(w)dw$ 为有限，归一化，便可称其为密度)

② Power Spectral Density and Autocorrelation Function.

<Wiener-Khintchine theorem> $R_X(\tau)\xleftrightarrow{FT}S_X(w)$，将功率按 f 展开.

~~$e^{-\beta|\tau|}\longleftrightarrow\frac{2\beta}{\beta^2+w^2}$~~ 时域相乘 ⟺ 频域卷积 $\cdot\frac{1}{2\pi}$.

$e^{-\beta|\tau|}\longleftrightarrow\frac{2\beta}{\beta^2+w^2}$

$e^{jw_0t}\longleftrightarrow2\pi\delta(w-w_0)\Rightarrow1\longleftrightarrow2\pi\delta(w)$ } 直流 ⟷ 冲激.

$\delta(t-t_0)\longleftrightarrow e^{-jwt_0}$

$f(x)h(x)\longleftrightarrow\frac{1}{2\pi}G(w)*H(w)$.

③ Cross-Power Spectral Density

针对于 X Y 的联合平稳序列.

$S_{XY}(w)\longleftrightarrow R_{XY}(\tau)$.

依然满足维纳辛钦定理

6. Finite Markov Chains.

1. Basic Concepts.

markov property → 遗忘 ⇒ 有马尔可夫性即是马尔可夫链

时齐 $P(X_{m+1}=j \mid X_m=i) = P(X_{n+1}=j \mid X_n=i) \Rightarrow \underline{P_{ij}(n, n+1)}$ 时间转移

一步转移.

转移矩阵 → stochastic matrix.

2. Higher Order Transition Probabilities and Distributions.

$p_i(0) = P(X_0=i)$ $p_{ij}(n) = P(X_{m+n}=j \mid X_m=i)$.

<Chapman-Kolmogorov equation (C-K)>.

$P_{ij}(n+m) = \sum_k p_{ik}(n)p_{kj}(m) \Rightarrow P(n) = P^n,\ p(n) = p(0)P(n)$.

由已知条件求转移事件.

当转移时概率内找不到可用时写为 P_{ij}.

使用乘法公式和马氏性不断拆分，再求转移概率，最后得出结果.

$p(0)$ 和 P 可以决定这个马氏链的所有特性.

观察题目选择最合适的解法，有些算矩阵乘法，有些展开代数乘法.

3. Invariant Distribution and Ergodic markov Chain.

$P^n = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$ 随机过程退化为一个随机变量.

$\pi = \pi P$, π → invariant distribution.

$\exists n, s.t.\ P(n)$ 为正矩阵 ⇒ ergodic ⇒ 极限存在.

⇕

即 $P_{ij}(n) > 0$ for $\forall i, j \in S$.

初态为 π 的马尔可夫链是平稳过程.

↳ $P\{X_{n_1}=i_1, \ldots, X_{n_k}=i_k\} = P\{X_{n_1+h}=i_1, \ldots, X_{n_k+h}=i_k\}$.

求解 π $\begin{cases} \pi = \pi P \\ \sum \pi_i = 1. \end{cases}$，解线性方程组

证明遍历性 $\xrightarrow{\text{极限存在}}$ 是否可写为 $\begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$ $\xrightarrow{\text{是}}$ 有遍历性

是否可写为 $\begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$ $\xrightarrow{\text{否}}$ 无遍历性

证明遍历性 $\xrightarrow{\text{极限不存在（振荡）}}$ 无遍历性

⇔ 极限存在 即矩阵是正则矩阵，即使我们可以求出 P^n，如果 $P^n \neq \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}$，也说 P^n 极限不存在.

八、Independent - Increment Processes.

1、独立增量过程引入。

独立增量、平稳增量。

$X_0=0$ 的独立增量过程可完全由 X_t-X_s 决定。

$X_0=0$ 且独立增量。

$$C_X(t,s)=E[(X_t-\mu_X(t))(X_s-\mu_X(s))]$$

$$\underbrace{\text{令 } X_t-\mu_X(t)=Y_t}\ \Big|=E(Y_tY_s) \qquad \text{假设 } 0<s<t$$

Y_t 也为独立增量

$$\text{独立增量}\ =E[(Y_t-Y_s)(Y_s-Y_0)]+E(Y_s^2)$$

$$=E(Y_t-Y_s)E(Y_s-Y_0)+E(Y_s^2)$$

$$=E(Y_s^2)=E[(X_s-\mu_X(s))^2]=Var_X(s).$$

即如果 $X_0=0$ 且独立增量 $\underbrace{C_X(t,s)=Var_X[\min(s,t)]}$

另外 $X_0=0$ 的独立增量过程也是 Markov 过程。

2、Poission Processes

本课本中定义的 Poission 过程为平稳、独立增量过程，并且初值为0，是正常的计数过程。

$$P(N_t=n)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

仍然需要知道的，还是独立增量过程的共同特点 $C_X(t,s)=Var_X[\min(t,s)]$。

当条件概率的条件前后两次独立时，条件概率就是非条件概率。

Poission Process 为计数过程，事件之间的时间间隔为指数分布。

3. Gaussian Processes.

高斯过程不是独立增量过程。

高斯过程在任何时点的线性组合都为 Gauss 过程。

$$\text{对于平稳高斯过程}\begin{cases}\mu_X(t)=\mu_X=\text{constant.}\\ C_X(t,t+h)=C_X(h),\ \text{或理解为自相关和协方差都只与差值有关,}\end{cases}$$

因为 $C_X(t,t+h)=R_X(t,t+h)-\mu_X^2(t)$

$\mu_X(t)$ 为常数。

高斯过程2阶矩一定存在，对于 Gauss 过程 SSS $\Leftrightarrow$ WSS.

<证明一个过程为 Gauss 过程>.

即证明 $\sum_{i=1}^{n}a_iX_{t_i}$ 为正态分布，for all $t_1,t_2,\ldots t_n\in T$, for all $a_i\in R$, all $n>0$.

<写出 Gauss 分布的概率密度函数(p.d.f)>.

一阶，$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

高阶，$X\sim N(\mu,\Sigma)$ $\qquad f_X(x_1,x_2\ldots;t_1,t_2\ldots)=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

均值向量. 协方差矩阵.

参数向量.

4. Brownian Motion and Wiener Processes.

Wiener process $\begin{cases} \text{独立增量} \\ \text{平稳增量 } W_t - W_s = W_{t-s}, \text{ where } 0 \le s \le t. \\ W_0 = 0 \\ W_t \sim N(0, \sigma^2 t) \end{cases}$

$\mu_W(t) = 0$, $C_W(s,t) = R_W(s,t) = \sigma^2 \min(s,t)$.

$\sum_{k=1}^{n} a_k W_{t_k} = \sum_{k=1}^{n} b_k (W_{t_k} - W_{t_{k-1}}) \rightarrow$ Gauss 过程.

$\min(s,t) = \dfrac{s+t-|s-t|}{2}$ (中点减距离的一半).

Wiener Process 是独立增量过程，也是 Gauss 过程.