一、基本术语 basic terminology

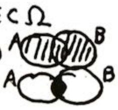
1. sample space Ω: the set of all outcomes of the experiment

$\Omega=\{1,2,3,4,5,6\}$ $E_0=\{6\}$, $E_1=\{1,3,5\}$

event E: subset of the sample space, $E\subset\Omega$

empty set ϕ, union $E=A\cup B$ (A or B)

intersection $E=A\cap B$ (或AB) (A and B)



2. $\{E_i\}_i$ are called mutually exclusive $E_i\cap E_j=\phi$ for all $i\neq j$

↳ form a partition of Ω if ①mutually exclusive ② $\cup_i E_i=\Omega$

3. complement $E^c=\{w: w\in\Omega \text{ and } w\notin E\}$

De Morgan's Law: $(A\cup B)^c=A^c\cap B^c$

4. probability space (Ω,F,P): Ω: sample space F: collections of events P: probability measure on F

$|\Omega|$: the cardinality of Ω (Ω元素数量); 2^Ω: power set (the collection of all subsets of Ω)

[event axioms]: F满足 ① $\Omega\in F$ ②若$A\in F$, $A^c\in F$ ③若$A,B\in F$, $A\cup B\in F$

⇒引理: $\phi\in F$, 若$A,B\in F$, $AB\in F$

[probability axioms]: P满足 ① $P(A)\geq 0$对$\forall A\in F$成立

②对$\forall A,B\in F$, 若$AB=\phi$, $P(A\cup B)=P(A)+P(B)$ ③ $P(\Omega)=1$

⇒引理: $P(A^c)=1-P(A)$; $P(A)=P(AB)+P(AB^c)$;

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

5. experiment with equiprobable events: $P(A)=\frac{|A|}{|\Omega|}$

6. permutation排列: $n!$ 有序 $A_n^k=\frac{n!}{k!}$

binomial coefficient $\binom{n}{k}=\frac{n!}{(n-k)!k!}$ 无序

7. K-map

	B	B^c
A		
A^c		

	BC	BC^c	B^cC^c	B^cC
A				
A^c				

8. countably infinite: 可以用0,1,2…自然数编号

二、离散随机变量 discrete-type random variables

1. 定义: 在(Ω,F,p)中, $X:\Omega\to R$, $X(w)\in R$是X结果(w)的值, X为random variable

pmf: probability mass function $P_X(x):=P(X=x)$

↳多变量: $P_{XY}(x,y)=P\{w: X(w)=x, Y(w)=y\}$

2. mean/expectation期望: $E(X)=\sum_{x\in\Omega}x\cdot P(X=x)$ $E(aX+b)=aE(X)+b$

[LOTUS]: 有关于X的函数$G(X)$, $E(G)=\sum_{x\in\Omega}G(x)\cdot P(X=x)$

(直接把x换成G(x), 乘上对应概率)

3. variance方差: $Var(X)=E[(X-E(X))^2]$ standard deviation标准差: $\sigma_X=\sqrt{Var}$

$Var(aX+b)=a^2Var(X)$; $Var(X)=\underbrace{E(X^2)}_{\text{second moment of }X}-(E(X))^2$

4. conditional probability条件概率: $P(B|A)=\begin{cases}\frac{P(AB)}{P(A)} & P(A)>0\\ \text{undefined} & P(A)=0\end{cases}$

independent独立: $P(AB)=P(A)\cdot P(B)$ 或 $P(B|A)=P(B)$

⇒ pairwise independent (A,B,C): $P(AB)=P(A)\cdot P(B)$, $P(AC)=P(A)\cdot P(C)$, $P(BC)=P(B)\cdot P(C)$

$P(AB)=0$ ↳ mutually exclusive 互不相容!

A,B,C independent: $P(ABC)=P(A)\cdot P(B)\cdot P(C)$ 且A,B,C pairwise independent

[性质]: $E[XY]=E[X]E[Y]$; $E[X+Y]=E[X]+E[Y]$, $Var[X+Y]=Var[X]+Var[Y]$ (X,Y相互独立)

4. the bernoulli distribution伯努力分布 X with range $\{0,1\}$

⇒ $p_X(1)=p$, $p_X(0)=(1-p)$ ⇐ $E(X)=p$, $Var(X)=p(1-p)$

the binomial distribution二项分布 n次0-1分布 order does not matter

pmf: $p_X(k)=\binom{n}{k}p^k(1-p)^{n-k}$ for $k=0\sim n$ ⇐ $E(X)=np$, $Var(X)=np(1-p)$

the geometric distribution几何分布 L: 直到出现"1"为止的次数

$p_L(k)=(1-p)^{k-1}\cdot p$, $P(L>k)=(1-p)^k$ ⇐ $E[L]=\frac{1}{p}$, $Var(L)=\frac{1-p}{p^2}$

[证]: I: 第1次是0, 直到出现"1"为止的次数 ⇒ $E[L]=E[I]$

又$E[L]=p+(1-p)\cdot E[1+I]$ ⇒ 证毕 各独立事件性质!

memoryless: $P(L>k+n|L>n)=\frac{P(L>k+n)}{P(L>n)}=(1-p)^k=P(L>k)$

Bernoulli Process伯努力过程: a sequence of Bernoulli random variable $X_1,X_2\cdots$

$P(X_k=1)=p$, $P(X_k=0)=1-p$

⇒① C_j: cumulative variable: $C_j=\sum_{k=1}^{j}X_k$ ↖ binomial distribution

⇒② L_j: the number of trails between count j-1 and j geometric dis~

⇒③ negative binomial distribution: $S_j=\sum_{i=1}^{r}L_i$ 前k-1个中有j-1个count, 第k个是第j个count

(考虑r个相互独立的几何分布 ⇒ 加起来就是负二项分布)

$p(n)=C_{n-1}^{r-1}p^r(1-p)^{n-r}$ for $n\geq r$ $E[S_r]=\frac{r}{p}$, $Var(S_r)=rVar(L_1)=\frac{r(1-p)}{p^2}$

the Poisson distribution泊松分布 $p(k)=\frac{e^{-\lambda}\cdot\lambda^k}{k!}$ $E(Y)=\lambda$, $Var(Y)=\lambda$

当n很大, p很小时, 可以用来估计binomial distribution! ⇒ $\lambda=np$

[eg]. an event with the poisson distribution, happen 4 times per minute

对于每min: $\lambda=4$! 对于每3分钟: $\lambda=3\times4=12$!

5. maximum likelihood parameter estimation极大似然参数估计

考虑随机变量X, X满足pmf: p_θ, θ为参数, 进行试验, 观测值k, the maximum likelihood estimate of θ for observation k, $\hat{\theta}_{ML}(k)$就是最大化出现已发生事件的概率值时的θ即为所求

[做法]: ①写出X的pmf(二项? 泊松? 几何? …) ②对观测值$p_X(k)$进行求导, 极大值时的θ即为所求 ⇐不好求导的值就θ的邻值相比 ($\frac{p_X(\theta+1)}{p_X(\theta)}\leq1$, $\frac{p_X(\theta-1)}{p_X(\theta)}\leq1$)

6. 几个不等式和置信区间

① Markov inequality X是随机变量, $c>0$, 有 $P\{X\geq c\}\leq\frac{E[X]}{c}$

[证]: $E[X]=\sum_i u_i p_X(u_i)\geq\underbrace{\sum_{i:u_i<c}0\cdot p_X(u_i)}_{\text{小于}c:=0}+\underbrace{\sum_{i:u_i\geq c}c\cdot p_X(u_i)}_{\text{大于}c:=c}=c\cdot P\{X\geq c\}$.

② Chebychev inequality X是随机变量, mean $\mu=E[X]$, variance σ^2, 有:

$P\{|X-\mu|\geq d\}\leq\frac{\sigma^2}{d^2}\xrightarrow{d=a\sigma}P\{|X-\mu|\geq a\sigma\}\leq\frac{1}{a^2}$

[证]: 令$Y=(X-\mu)^2$, $E[Y]=\sigma^2$, 代入Markov inequality可得$(c=d^2)$

③ confidence Intervals置信区间: 表示在大样本中取小样本估计时的可信度

$\hat{p}=\frac{X}{n}$, X是小样本中发生事件数, n是小样本总数, $\hat{p}$为我们所得概率, p为真实

由chebychev inequality, $P\{|X-\mu|<a\sigma\}>1-\frac{1}{a^2}$ ⇒ $P\{X\in(\mu-\sigma a,\mu+\sigma a)\}>1-\frac{1}{a^2}$

考虑0-1事件Y: 在区间为1, 不在为0 ⇒ $\sigma=\sqrt{np(1-p)}$

$P\{p\in(\hat{p}-a\sqrt{\frac{p(1-p)}{n}},\hat{p}+a\sqrt{\frac{p(1-p)}{n}})\}\geq1-\frac{1}{a^2}$, 又$p(1-p)\leq\frac{1}{4}$, 有:

interval estimator $P\{p\in(\hat{p}-\frac{a}{2\sqrt{n}},\hat{p}+\frac{a}{2\sqrt{n}})\}\geq1-\frac{1}{a^2}$

a	$1-\frac{1}{a^2}$
2	75%
5	96%
10	99%

the confidence interval with confidence level $1-\frac{1}{a^2}$ is $(\hat{p}-\frac{a}{2\sqrt{n}},\hat{p}+\frac{a}{2\sqrt{n}})$.

7. The law of total probability全概率公式, Bayes formula贝叶斯公式

$E_1\sim E_k$ forms a partition of Ω 如果 ① $E_1\sim E_k$ mutually exclusive ② $\Omega=E_1\cup\cdots\cup E_k$

⇒全概率公式: $P(A)=P(AE_1)+\cdots+P(AE_k)$

Bayes formula: $P(E_i|A)=\frac{P(AE_i)}{P(A)}=\frac{P(A|E_i)\cdot P(E_i)}{P(A)}=\frac{P(A|E_i)\cdot P(E_i)}{P(A|E_1)\cdot P(E_1)+\cdots+P(A|E_k)\cdot P(E_k)}$

可以用来计算mean! 离散随机变量X, 事件A, ← LOTUS

$E[X|A]=\sum_i u_iP(X=u_i|A)$; $E[g(X)|A]=\sum_i g(u_i)P(X=u_i|A)$

8. binary hypothesis testing

H_1, H_0 → system $\xrightarrow{X}$ decision rule → H_1 or H_0

① likelihood matrix:

	X=0	X=1	X=2	X=3
H_1	0	0.1	<u>0.3</u>	<u>0.6</u>
H_0	<u>0.4</u>	<u>0.3</u>	0.2	0.1

"underlie" means the decision rule used (如出现X=0, 认为H_0是对的)

false alarm: H_0 is true, but H_1 is declared $P_{false\ alarm}=P(\text{declare } H_1 \text{ true}|H_0)$

数值上等于H_0行中没被underlie的.

miss: H_1 is true, but H_0 is declared $P_{miss}=P(\text{declar } H_0 \text{ true}|H_1)$

数值上等于H_1行中没被underlie的

② Maximum likelihood (ML) decision rule 如上图所示, 对于每一列, 选最可能发生的hypothesis作为undelie的(相同就随意) 看要求! 如min P_{miss}…

⇒ likelihood ratio test (LRT): $\Lambda(k)=\frac{p_1(k)}{p_0(k)}$

$\Lambda(X)>1$: declare H_1 is true | $\Lambda(X)<1$: declare H_0 is true.

with threshold τ:

$\Lambda(X)>\tau$: declare H_1 is true | $\Lambda(X)<\tau$: declare H_0 is true.

③ MAP decision rule 给 $\pi_0=P(H_0)$, $\pi_1=P(H_1)$, H_0,H_1有prior probability.

⇒ joint probability matrix: 表中每一项是$P(H_i, X=k)$; $\pi_1=\pi_0$ ⇒ uniform

	X=0	X=1	X=2	X=3	
H_1	0	0.02	0.06	<u>0.12</u>	⇐乘π_1
H_0	<u>0.32</u>	<u>0.24</u>	<u>0.16</u>	0.08	⇐乘π_0

posteriori probability: $P(H_1|X=2)=\frac{P(H_1,X=2)}{P(X=2)}=\frac{0.06}{0.06+0.16}$

还是underlie每一列中最大的项 → 用$\Lambda(X)$来计算, 相当于threshold $\tau=\frac{\pi_0}{\pi_1}$

⇒ average error probability: $p_e=\pi_0 P_{false\ alarm}+\pi_1 P_{miss}$

the sum of all not underlied numbers, 要看用什么rule (MAP还是ML!)

一、[binary hypothesis testing]: $H_1, H_0 \to$ system $\xrightarrow{X}$ decision rule $\to H_1$ or H_0

1. likelihood matrix

	X=0	X=1	X=2	X=3
H_1	0	0.1	0.3	0.6
H_0	0.4	0.3	0.2	0.1

"underlie" means the decision rule used (X=0，则认为H_0 true)

false alarm: H_0 is true, but H_1 is declared $P_{false\ alarm} = P(\text{declare } H_1 \mid H_0)$
数值上等于H_0行中没被underlie的和

miss: H_1 is true, but H_0 is declared $P_{miss} = P(\text{declare } H_0 \mid H_1)$
数值上等于H_1行中没被underlie的和 ← π_0, π_1是H_0, H_1的概率

average error probability: $p_e = \pi_0 P_{false\ alarm} + \pi_1 \cdot P_{miss}$

2. Maximum likelihood (ML) decision rule
如上图所示，对于每一列，选择最有可能发生的hypothesis作为underlie的(相同就看要求)

likelihood ratio test (LRT): $\Lambda(k) = \frac{p_1(k)}{p_0(k)}$

$\Lambda(x) > 1$: declare H_1 is true | $\Lambda(x) < 1$: declare H_0 is true

with threshold τ:

$\Lambda(x) > \tau$: declare H_1 is true | $\Lambda(x) < \tau$: declare H_0 is true

3. MAP decision rule
H_0, H_1有prior probability $\pi_0 = P(H_0)$, $\pi_1 = P(H_1)$

joint probability matrix: 表中每一项是$P(H_i, X=k)$

	X=0	X=1	X=2	X=3	
H_1	0	0.02	0.06	0.12	乘π_1
H_0	0.32	0.24	0.16	0.08	乘π_0

posteriori probability: $P(H_1 \mid X=2) = \frac{P(H_1, X=2)}{P(X=2)} = \frac{0.06}{0.06+0.16}$

还是underlie每列最大项→用$\Lambda(x)$计算，相当于$\tau = \frac{\pi_0}{\pi_1}$

$p_e = \pi_0 P_{false\ alarm} + \pi_1 P_{miss}$ ⇐ MAP最小化p_e.
相当于没被underlie的项的和 ☆要是joint probability matrix!

二、[Reliability]

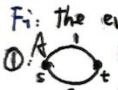
1. Union bound
$P(A \cup B) \le P(A) + P(B)$

$P(A_1 \cup A_2 \cdots \cup A_m) \le P(A_1) + P(A_2) + \cdots + P(A_m)$ ☆

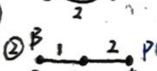
2. Network outage probability
s-t network ⇒ source node s, terminal node t.
each link i fails with probability p_i

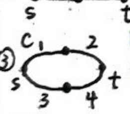
F_i: the event link i fails Network outage: s→t 无path

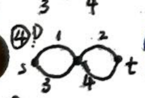
union bound

①  $P(F) = p_1 p_2$ 并联!

串联!

②  $P(F) = p_1 + p_2 - p_1 p_2$ | $P(F) = P(F_1 \cup F_2) \le P(F_1) + P(F_2)$

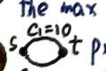
③  $P(F) = P(F_1 F_3 \cup F_1 F_4 \cup F_2 F_3 \cup F_2 F_4)$
$P(F) = (p_1 + p_2 - p_1 p_2)(p_3 + p_4 - p_3 p_4)$

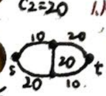
④  $P(F) = P(F_1 F_3 \cup F_2 F_4)$
$P(F) = p_1 p_3 + p_2 p_4 - p_1 p_2 p_3 p_4$

⑤  $P(E) = (F_1 F_3 \cup F_2 F_4 \cup F_1 F_5 F_4 \cup F_3 F_5 F_2)$
$P(F) = p_1 p_3 + p_2 p_4 - p_1 p_2 p_3 p_4 + p_1 q_2 q_3 p_4 p_5 + q_1 p_2 p_3 q_4 p_5$ $q_i = 1 - p_i$
考虑5有无: $P(E) = P(C, 5无) + P(D, 5有)$

☆写成互斥的几个事件的和!

5. distribution of the capacity of a flow network
the max flow rate from s to t. $q_i = 1 - p_i$

 $p_X(0) = p_1 p_2$, $p_X(10) = q_1 p_2$, $p_X(20) = p_1 q_2$, $p_X(30) = q_1 q_2$
从s流出，每条路最多C_i的flow.

 $Y = 30, 20, 10, 0$

三、[continuous-type random variable]

1. cmmulative distribution function (CDF)
$F_X(c) = P\{w : X(w) \le c\} = P\{X \le c\}$

left limit: $F(x-) = \lim_{y \to x, y < x} F(y)$ right limit: $F(x+) = \lim_{y \to x, y > x} F(y)$

$\Delta F_X(x) = F_X(x) - F_X(x-) \Rightarrow P\{X < c\} = F_X(c-)$, $P\{X = c\} = \Delta F_X(c)$

① F is the CDF of some random variable if and only if:
Ⅰ. F is nondecreasing Ⅱ. $\lim_{c \to \infty} F(c) = 1$ and $\lim_{c \to -\infty} F(c) = 0$
Ⅲ. F is right continuous ($F_X(c) = F_X(c+)$)
对于离散变量的CDF，断点为$P(X=k)$有值的点

2.
对于离散: $F_X(c) = \sum_{u: u \le c} p_X(u)$ (pmf) 对于连续: $F_X(c) = \int_{-\infty}^{c} f_X(u)\,du$ (probability density function)

[continuous-type random variable]: a random variable X is a continuous-type random variable if exist f_X: $F_X(c) = \int_{-\infty}^{c} f_X(u)\,du$ for all $c \in R$.

☆ $P\{a < X \le b\} = F_X(b) - F_X(a) = \int_a^b f_X(u)\,du$

mean: $\mu_X = E[X] = \int_{-\infty}^{\infty} u f_X(u)\,du \xrightarrow{LOTUS} E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u)\,du$

variance: $Var(X) = E[(X - \mu_X)^2] = E[X^2] - (E[X])^2$

↳ standard deviation $\sigma = \sqrt{Var(X)}$, standardized random variable $\frac{X - \mu_X}{\sigma_X}$ ← mean 0, variance 1

☆ $1 = \int_{-\infty}^{\infty} f_X(u)\,du$ 概率和为1.

3. 几种分布

① uniform distribution: $f_X(u) = \begin{cases} \frac{1}{b-a} & a \le u \le b \\ 0 & \text{else} \end{cases}$

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

for $a=0, b=1$: $E[X^k] = \int_0^1 u^k du = \frac{1}{k+1}$, $Var = \frac{1}{12}$

② exponential distribution: $f_T(t) = \begin{cases} \lambda e^{-\lambda t} & t \ge 0 \\ 0 & \text{else} \end{cases}$ | $F_T(t) = \begin{cases} 1 - e^{-\lambda t} & t \ge 0 \\ 0 & t < 0 \end{cases}$

Ⅰ. mean: $E[T^n] = \int_0^{\infty} t^n \cdot \lambda e^{-\lambda t} dt$
$= -t^n e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} n \cdot t^{n-1} e^{-\lambda t} dt \Rightarrow E[T] = \frac{1}{\lambda}$, $E[T^n] = \frac{n!}{\lambda^n}$
$= \frac{n}{\lambda} E[T^{n-1}]$ $E[T^2] = \frac{2}{\lambda^2}$

Ⅱ. $Var(T) = \frac{1}{\lambda^2}$, $\sigma(X) = \frac{1}{\lambda}$

Ⅲ. memoryless property: $P(T > s+t \mid T > s) = e^{-\lambda t} = P\{T > t\}$

the exponential distribution is the limit of scaled geometric distributions

[eg]. 若一个灯只在h, 2h, 3h… 灭，$p = \lambda h$. L_h是灭时经过的h数量

L_h: geometric distribution $E[L_h] = \frac{1}{p} = \frac{1}{\lambda h}$ T_h是灭时经过的时间 $T_h = h \cdot L_h$

$P\{T_h > c\} = P\{L_h \cdot h > c\} = P\{L_h > \frac{c}{h}\} = (1 - \lambda h)^{\lfloor \frac{c}{h} \rfloor}$ $(1 - \frac{\lambda}{n})^n \to e^{-\lambda}, n \to \infty$

$\therefore P\{T_h > c\} = (1 - \lambda h)^{\lfloor \frac{c}{h} \rfloor} \to e^{-\lambda c} = P\{T > c\}$

4. Poisson process ← limit of scaled Bernoulli processes
描述一个counting process关于experiment: a collection of random variable

$N = (N_t : t > 0)$ $N_t = \sum_{n=1}^{\infty} I_{\{t \ge T_n\}}$ 时间t内出现的count次数

$T_n = \min\{t : N_t \ge n\}$ 第n次count出现的时间 $T_n = U_1 + \cdots + U_n$ ← 第n-1次与第n次出现的间隔

①定义: a Poisson process with rate λ is a random counting process $N = \{N_t : t \ge 0\}$ such that:

Ⅰ. N has independant increments: if $0 \le t_0 \le t_1 \le \cdots \le t_n$, $N_{t_1} - N_{t_0} \cdots N_{t_n} - N_{t_{n-1}}$ (the increments) are independant 每次count之间独立，用泊松分布估计每一段

Ⅱ. The increment $N_t - N_s$ has the $Poi(\lambda(t-s))$ distribution for $t \ge s$

(或等价说法: $U_1 \cdots U_n$ are mutually independent, exponentially distributed random variables with parameter λ) 时间间隔符合exponential分布!

$P\{N_t - N_s = k\} = \frac{[\lambda(t-s)]^k}{k!} e^{-\lambda(t-s)}$ $E[N_t - N_s] = \lambda(t-s)$

更一般的: $P\{N_t = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ $Var[N_t - N_s] = \lambda(t-s)$

5. The Erlang Distribution
令$T_r = U_1 + \cdots + U_r$, the distribution of T_r is the Erlang distribution.

☆ $P\{T_r > t\} = P\{N_t \le r-1\} = \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

the CDF: $F_{T_r}(t) = 1 - P\{T_r > t\}$

$\Rightarrow f_{T_r}(t) = -\frac{dP\{T_r > t\}}{dt} = e^{-\lambda t}\left[\sum_{k=0}^{r-1} \lambda \cdot \frac{(\lambda t)^k}{k!} - \sum_{k=0}^{r-1} \frac{k \cdot \lambda^k}{k!} t^{k-1}\right]$
pdf

$= e^{-\lambda t}\left(\sum_{k=0}^{r-1} \frac{\lambda^{k+1} t^k}{k!} - \sum_{k=1}^{r-1} \frac{\lambda^k t^{k-1}}{(k-1)!}\right) = e^{-\lambda t}\left(\sum_{k=0}^{r-1} \frac{\lambda^{k+1} t^k}{k!} - \sum_{k=0}^{r-2} \frac{\lambda^{k+1} t^k}{k!}\right)$

$= \frac{e^{-\lambda t} \cdot \lambda^r \cdot t^{r-1}}{(r-1)!}$ mean: $E[T_r] = E[U_1] + \cdots + E[U_r] = \frac{r}{\lambda}$

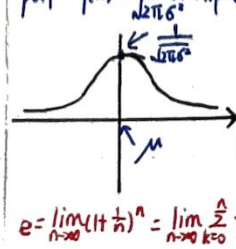
var: $Var(T_r) = Var(U_1) + \cdots + Var(U_r) = \frac{r}{\lambda^2}$

6. linear scaling of pdfs
X is random variable, $Y = aX + b$, $a > 0$:

$Y = aX + b \Rightarrow f_Y(u) = f_X\left(\frac{u-b}{a}\right) \cdot \frac{1}{a}$ $\frac{u-b}{a}$用的是X的范围

$E[Y] = aE[X] + b$ $Var(Y) = a^2 Var(X)$ $\sigma_Y = a\sigma_X$ $F_Y(u) = F_X\left(\frac{u-b}{a}\right)$
遇到正态分布情况算E(Y)和Var(Y)

7. The Gaussian (normal) distribution
pdf: $f(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)$, 又叫$N(\mu, \sigma^2)$ distribution.



Ⅰ. pdf is symmetric around the value μ.
Ⅱ. the peak height: $\frac{1}{\sqrt{2\pi\sigma^2}}$
Ⅲ. $[\mu - \sigma, \mu + \sigma]$: 68.3%
$[\mu - 2\sigma, \mu + 2\sigma]$: 95.44%

$e = \lim_{n \to \infty}(1 + \frac{1}{n})^n = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{1}{k!}$; $\lim_{x \to 0} \frac{e^x - 1}{x} = 1$

$\Phi(0)=\frac{1}{2}$

① standard normal distribution: $\mu=0, \sigma^2=1$ $N(0,1)$ distribution

CDF: $\Phi(u)=\int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}\exp(-\frac{v^2}{2})dv$

complementary CDF: $Q(u)=\int_{u}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{v^2}{2})dv = 1-\Phi(u)=\Phi(-u)$

u大于3或4时用Q(u), u小时用Φ(u)

② let $I=\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du$

$I^2=\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du\int_{-\infty}^{\infty}e^{-\frac{v^2}{2}}dv=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{u^2+v^2}{2}}du\,dv \xrightarrow{\text{极坐标}} \int_0^{2\pi}\int_0^{\infty}e^{-\frac{r^2}{2}}r\,dr\,d\theta$

$=-2\pi e^{-\frac{r^2}{2}}\Big|_0^{\infty}=2\pi$ → standard normal distribution pdf integrates to 1.

X: N(0,1) distribution $E[X]=0$.

$E[X^2]=\int_{-\infty}^{\infty}\frac{u^2}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})du=\int_{-\infty}^{\infty}\frac{u}{\sqrt{2\pi}}\cdot u\exp(-\frac{u^2}{2})du$

$=-\frac{u}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})du=0+1=1$, $Var[X]=1$

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

[eg]. X有N(10,16) distribution

$P(X\ge 15)=P\{\frac{X-10}{\sqrt{16}}\ge\frac{15-10}{4}\}=Q(\frac{15-10}{4})=Q(1.25)=1-\Phi(1.25)$

$P(X\le 5)=P\{\frac{X-10}{4}\le\frac{5-10}{4}\}=\Phi(\frac{5-10}{4})=\Phi(-1.25)=Q(1.25)$

化成标准正态分布,然后用Φ( ), Q( )查表

8. CLT: central limit theorem and Gaussian approximation

① Gaussian approximation 前提: 多个independent random variables加起来,

如 X是n个Bernoulli随机变量的和 mean: np var: $np(1-p)$ $\{np, n(1-p)\}$都很大

$\tilde{X}$: Gaussian random variable, mean和var 相同 ⇒ 适用于 估计的不接近0, 1

$P(X\le v)\approx P(\tilde{X}\le v)$

② Gaussian approximation with the continuity correction:

$P(X\le k)\approx P(\tilde{X}\le k+0.5)$ $P(X\ge k)\approx P(\tilde{X}\ge k-0.5)$ $P(X=k)=P(k-\frac{1}{2}\le\tilde{X}\le k+\frac{1}{2})$

DeMoivre-Laplace limit theorem: $S_{n,p}$是binomial random variable,

$\lim_{n\to\infty}P\{\frac{S_{n,p}-np}{\sqrt{np(1-p)}}\le c\}=\Phi(c)$

[eg]. 60个学生参加考试, $p=\frac{5}{6}$, 用Gaussian approximation with the continuity correction估计$P\{X\le 52\}$

先算 mean, var: $\mu_X=np=50$, $\sigma_X=\sqrt{np(1-p)}=2.887$

再套估计公式,化简 $P\{X\le 52\}=P\{X\le 52.5\}=P\{\frac{X-50}{2.887}\le\frac{52.5-50}{2.887}\}$

成标准正态 $=\Phi(0.866)$

[eg]. p是包中DVD比 检查n个, X是表示DVD的数量, $\hat{p}=\frac{X}{n}$

用Gaussian approximation 估计 $P\{|\hat{p}-p|\le\delta\}$:

$P\{|\hat{p}-p|\le\delta\}=P\{|\frac{X}{n}-p|\le\delta\}=P\{|X-np|\le n\delta\}=P\{|\frac{X-np}{\sqrt{np(1-p)}}|\le\delta\sqrt{\frac{n}{p(1-p)}}\}$

$\approx\Phi(\delta\sqrt{\frac{n}{p(1-p)}})-\Phi(-\delta\sqrt{\frac{n}{p(1-p)}})=2\Phi(\delta\sqrt{\frac{n}{p(1-p)}})-1$

$p=0.5, n=1000$, 找δ使 $P\{|\hat{p}-p|\le\delta\}\approx 0.99$ ← 表述成: find the half-width δ of the confident interval so we have 99% confidence that ...

找δ使 $2\Phi(\delta\sqrt{\frac{1000}{p(1-p)}})-1=0.99$ ← 不好直接求导时试试两边取ln( ).

9. ML parameter estimation for continuous-type variables

[eg]. r.v T是exponential distribution, 参数λ, observed that $T=t$, 找ML estimate $\hat{\lambda}_{ML}(t)$. 对要观测事件求导(多个时概率相乘)

$P(T=t)=\lambda\cdot e^{-\lambda t}$ $\frac{d(\lambda\cdot e^{-\lambda t})}{d\lambda}=(1-\lambda t)e^{-\lambda t}\Rightarrow\hat{\lambda}_{ML}(t)=\frac{1}{t}$

[eg]. X在[0,b]上uniform distribution, observed $X=u$, 求ML estimate of b

pdf: $f_b(u)=\frac{1}{b}I_{\{0\le u\le b\}}$, 其中 $I_{\{0\le u\le b\}}=\begin{cases}1 & 0\le u\le b\\ 0 & \text{else}\end{cases}$

↳ $b<u$: $f_b(u)=0$; $b\ge u$: $f_b(u)=\frac{1}{b}$

$\Rightarrow\hat{b}=u$

coin, head概率为X, 有 $f_X(p)=\cdots$; 求5次扔, 4次head的概率 Y: head次数; $P_{Y|X}(v|p)\sim B(5,p)$

$P(Y=4)=\int_0^1 f_X(p)P_{Y|X}(v|p)dp=\int_0^1 f_X(p)\cdot C_5^4\cdot p^4\cdot(1-p)dp$

10. functions of a random variable

① the distribution of a function of a random variable. $Y=g(X)$

I. 分析X的pdf 和函数g( ), 注意Y是离散还是连续

连续: 求Y的CDF: $F_Y(c)=P\{Y\le c\}=P\{g(X)\le c\}$ ↳ 注意g(X)增减性

再求导: $\frac{dF_Y(c)}{dc}=f_Y(c)$

离散: 直接求pmf: $p_Y(c)=P\{Y=c\}=P\{g(X)=c\}$

[eg]. $Y=X^2$, X: $f_X(u)=\frac{e^{-|u|}}{2}$ for $u\in R$, 求pdf, mean和var of Y

先算cdf: $P\{Y\le c\}=P\{X^2\le c\}=P\{-\sqrt{c}\le X\le\sqrt{c}\}=\int_{-\sqrt{c}}^{\sqrt{c}}f_X(u)du=1-e^{-\sqrt{c}}$

再求导: $f_Y(c)=\begin{cases}0 & c\le 0\\ \frac{e^{-\sqrt{c}}}{2\sqrt{c}} & c>0\end{cases}$ 如 $\int_0^{\infty}u^n e^{-u}=n!$ for $n\in Z^+$

mean: $E[Y]=\int_{-\infty}^{\infty}g(u)f_X(u)du=\int_{-\infty}^{\infty}\frac{u^2e^{-|u|}}{2}du=2!=2$

var: $E[Y^2]=\int_{-\infty}^{\infty}[g(u)]^2f_X(u)=\int_{-\infty}^{\infty}\frac{u^4e^{-|u|}}{2}du=4!=24$

$Var[Y]=E[Y^2]-(E[Y])^2=24-4=20$

[eg]. X is uniformly distributed on [0,3], $Y=(X-1)^2$ 分段讨论

$X\in[0,3]$, $Y\in[0,4]$, $P\{(X-1)^2\le c\}=P\{-\sqrt{c}+1\le X\le\sqrt{c}+1\}$

① $0\le c\le 1$: 取得到整个区间: $F_Y(c)=\frac{2\sqrt{c}}{3}$

② $1\le c\le 4$, 只得取$[0,1+\sqrt{c}]$: $F_Y(c)=\frac{1+\sqrt{c}}{3}$

$F_Y(c)=\begin{cases}0 & c<0\\ \frac{2\sqrt{c}}{3} & 0\le c<1\\ \frac{1+\sqrt{c}}{3} & 1\le c<4\\ 1 & c\ge 4\end{cases}\xrightarrow{\text{求导}} f_Y(c)=\begin{cases}\frac{1}{3\sqrt{c}} & 0\le c<1\\ \frac{1}{6\sqrt{c}} & 1\le c<4\\ 0 & \text{else}\end{cases}$

mean: $E[Y]=E[(X-1)^2]=\int_0^3(u-1)^2\cdot\frac{1}{3}du=1$

11. generate a random variable with a specified distribution

考虑对uniformly distributed [0,1]的X进行变换g( ).

如 $F^{-1}(u)=\min\{c: F(c)\ge u\}$ $P\{F_X^{-1}(U)\le c\}=P\{U\le F_X(c)\}=F_X(c)=P(X\le c)$ $(\lambda=1)$

[eg]. U: uniformly distributed [0,1], 求g(U)使其exponentially distributed

先看$F(c)=1-e^{-c}$ for $c\ge 0$ $F(c)=u$ ⇒ 用u表示c ⇒ g(U)

再求反函数 $g(u)=F^{-1}(c)$ ⇒ $1-e^{-c}=u\Rightarrow c=-\ln(1-u)$

↳ $g(u)=F^{-1}(u)=-\ln(1-u)$ 如 $P(U<c)=c$ $(c\in(0,1))$

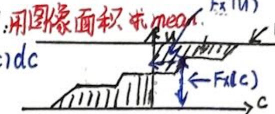
[eg]. U, 使其g(U)有与6面骰子相同的分布

$g(u)=i$ for $\frac{i-1}{6}<u\le\frac{i}{6}$ 处理离散时看图像.

[the area rule for expectation]: 用图像面积求mean.

$E[X]=\int_0^{\infty}[1-F_X(c)]dc-\int_{-\infty}^{0}F_X(c)dc$

$=\int_0^1 F_X^{-1}(u)du$



如 注意只适用于从U[0,1]出发的情况.

[failure rate functions]: T表示lifetime of an item ($T\ge 0$)

failure rate function: $h(t)=\lim_{\epsilon\to 0}\frac{P(t<T\le t+\epsilon|T>t)}{\epsilon}=\lim_{\epsilon\to 0}\frac{P\{t<T\le t+\epsilon\}}{P\{T>t\}\cdot\epsilon}$

$=\lim_{\epsilon\to 0}\frac{F_T(t+\epsilon)-F_T(t)}{\epsilon}\cdot\frac{1}{1-F_T(t)}=\frac{f_T(t)}{1-F_T(t)}$ ← 因为T非负

① 给定$h(t)$: $F(t)=1-e^{-\int_0^t h(s)ds}$, $E[T]=\int_0^{\infty}(1-F(t))dt$

② the failure rate function for the $\min\{T_1,T_2\}$ (T_1, T_2 independant) is the sum of their failure rate functions.

[Binary hypothesis testing with continuous-type observation]:

f_1, f_0是两个已知pdf, $f_i(u)$是the likelihood of $X=u$ if H_i is true hypothesis

the likelihood ratio $\Lambda(u)=\frac{f_1(u)}{f_0(u)}$ LRT (likelihood ratio test) with τ:

$\Lambda(X)\begin{cases}>\tau & \text{declare } H_1 \text{ true}\\ <\tau & \text{declare } H_0 \text{ true}\end{cases}$ 如 $\ln\Lambda(X)\begin{cases}>\ln\tau & \text{declare } H_1 \text{ true}\\ <\ln\tau & \text{declare } H_0 \text{ true}\end{cases}$

如果$f_1(u)$, $f_0(u)$中, u的取值范围不重合,

图.补充 直接写策略执行. probability of error是0.

1. 1min 接到2次电话, N_t表示t时刻前接到的电话次数, $P(N_t=i)=\frac{e^{-2t}(2t)^i}{i!}$

求第三次电话到达时间的pdf

先求cdf: $F(t)=1-P(N_t\le 2)=1-e^{-2t}(1+2t+\frac{(2t)^2}{2})$, 对t求导 $f(t)=e^{-2t}(2t)^2$

2.

X	10	9	8
P	0.5	0.3	0.2

100次后, 总分$P\{915\le X\le 945\}$? $X=X_1+\cdots+X_{100}$

$E[X]=9.3$ $Var[X]=0.61$, $\frac{\sum X_i-930}{\sqrt{61}}\sim N(0,1)$

$P=\{915\le X\le 945\}=P\{\frac{915-930}{\sqrt{61}}\le\frac{X-930}{\sqrt{61}}\le\frac{945-930}{\sqrt{61}}\}=\Phi(1.92)-\Phi(-1.92)$

3. 求导: $\arctan(x)'=\frac{1}{1+x^2}$ $\arccos(x)'=\frac{-1}{\sqrt{1-x^2}}$ $\arcsin(x)'=\frac{1}{\sqrt{1-x^2}}$

4. 求exponential distribution with $\lambda=0.02$的median. 设为m: $P\{T\ge m\}=P\{T\le m\}=\frac{1}{2}$

$P\{T\le m\}=1-e^{-\lambda m}=\frac{1}{2}\Rightarrow e^{-\lambda m}=\frac{1}{2}\Rightarrow m=50\ln 2$

5. X指数分布, λ=2; $Y\sim N(2,9)$; U: [0,1]上均匀分布

(1) $g(X)=U$, 求g $F_U(c)=P\{U\le c\}=c=P\{g(X)\le c\}=P\{X\le g^{-1}(c)\}=F_X(g^{-1}(c))$

$\Rightarrow F_X(g^{-1}(c))=c$, $g^{-1}(c)=F_X^{-1}(c)\Rightarrow g(c)=F_X(c)$, $g(X)=1-e^{-2X}$

(2) $k(X)=Y$, 求k $F_Y(c)=\Phi(\frac{c-2}{3})=P\{Y\le c\}=P\{X\le k^{-1}(c)\}=F_X(k^{-1}(c))=1-e^{-2k^{-1}(c)}$

$\Rightarrow k^{-1}(c)=u=-\frac{1}{2}\ln(1-\Phi(\frac{c-2}{3}))$, $k(u)=c=3\Phi^{-1}(1-e^{-2u})+2$

6. T指数分布, λ=2, 求$E(T|T>1)$: 指数分布 memoryless property

$E(T|T>1)=1+E(T)=1.5$; X在[0,2]上均匀分布, 求$E(X|X>1)$: $X|X>1$在[1,2]上仍均匀分布. $E(X|X>1)=\frac{1+2}{2}=1.5$ 通解 令$Y=T|T>1$, $u\le 1$: $F_Y(u)=0$

$u>1$: $F_Y(u)=P(Y\le u)=P(T\le u|T>1)=\frac{P(T\le u, T>1)}{P(T>1)}=\frac{P(T\le u)-P(T<1)}{1-P(T<1)}$

7. 判断independant: $P(AB)=P(A)\cdot P(B)$ 或 $P(B|A)=P(B)$

8. Poission distribution $p(k)=\frac{e^{-\lambda}\lambda^k}{k!}$ (n很大, p很小时估计bionomial distribution)

9. 分部积分: $\int uv'dx=uv-\int u'vdx$ ← $E(L)=\frac{1}{p}$, $Var(L)=\frac{1-p}{p^2}$

10. geometric distribution: $p_L(k)=(1-p)^{k-1}p$; 直到出现"1"为止的次数为L

11. $F_X(u)=\frac{1}{1+\exp(-\lfloor u\rfloor)}$ 离散!, $\lfloor u\rfloor$: 下界函数, 求$P\{|X|\le 3\}$.

$P\{|X|\le 3\}=P\{-3\le X\le 3\}=P\{X\le 3\}-P\{X<-3\}$ 相当于$P\{X\le -4\}$

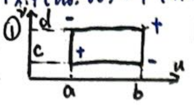
12. 换底公式: $\log_a b=\frac{\ln b}{\ln a}$, $(\log_a x)'=\frac{1}{x\ln a}$

一. Jointly Distributed Random variables

1. joint cumulative distribution function (CDF)

X, Y 是同一个 probability space $(\Omega, \mathcal{F}, P)$ 的 r.v. CDF 是：

$F_{X,Y}(u_0, v_0) = P\{X \le u_0, Y \le v_0\}$ for any $(u_0, v_0) \in \mathbb{R}^2$

① 

$P\{(X,Y) \in R\} = F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{X,Y}(a,d) + F_{X,Y}(a,c)$

② X, Y 有 CDF $F_{X,Y}$，对 u, v 有 $\lim\limits_{v \to \infty} F_{X,Y}(u,v) = F_{X,Y}(u, +\infty)$ 存在.

$F_X(u) = F_{X,Y}(u, \infty)$；$F_Y(v) = F_{X,Y}(\infty, v)$

③ CDF 性质：(F 就是 $F_{X,Y}$) 仅全部满足是 CDF 的充要条件.

I. $0 \le F(u,v) \le 1$ for all $(u,v) \in \mathbb{R}^2$ II. $F(u,v)$ is nondecreasing in u/v

III. $F(u,v)$ is right-continuous in u/v.

IV. 若 $a<b$, $c<d$, 有 $F(b,d) - F(b,c) - F(a,d) + F(a,c) \ge 0$

V. 对所有 v, $\lim\limits_{u \to -\infty} F(u,v) = 0$；对所有 u: $\lim\limits_{v \to -\infty} F(u,v) = 0$

VI. $\lim\limits_{u,v \to +\infty} F(u,v) = 1$

2. joint probability mass functions (joint pmf)

$p_{X,Y}(u,v) = P\{X=u, Y=v\}$ 考虑 X 取值 $u_1, u_2 \cdots$ Y 取值 $v_1, v_2 \cdots$

① marginal pmf: $p_X(u) = P\{X=u\} = \sum_j p_{X,Y}(u, v_j)$, $p_Y(v) = \sum_i p_{X,Y}(u_i, v)$

② conditional pmf: $p_{Y|X}(v|u_0) = P(Y=v \mid X=u_0) = \frac{p_{X,Y}(u_0,v)}{p_X(u_0)}$

若 $p_X(u_0) = 0$, $p_{Y|X}(v|u_0)$ 无定义

③ 性质 (全部加起来是 pmf 充要条件)

I. p 非负 II. $\sum_i \sum_j p(u_i, v_j) = 1$ 或 countably finite set

III. 有限集 $\{u_1, u_2 \cdots\}$, $\{v_1, v_2 \cdots\}$，对 $u, v \notin$ 集合，有 $p(u,v) = 0$

3. joint probability density functions (joint pdf)

① X, Y 是 joint continuous-type 若存在 $f_{X,Y}$ (pdf)，有

$F_{X,Y}(u_0, v_0) = \int_{-\infty}^{u_0} \int_{-\infty}^{v_0} f_{X,Y}(u,v)\,du\,dv$ 或对于矩形区域 $R\ (a,b] \times (c,d]$ 有：

$P\{(X,Y) \in R\} = \iint_R f_{X,Y}(u,v)\,du\,dv$ 实际上只要 piecewise differentiable boundary

② LOTUS: $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u,v) f_{X,Y}(u,v)\,du\,dv \to$ linear of expectation.

$E[aX + bY + c] = aE[X] + bE[Y] + c$

③ 性质 (全部加起来是 pdf 充要条件)

I. for any $(u,v) \in \mathbb{R}^2$, $f(u,v) \ge 0$ II. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(u,v)\,du\,dv = 1$

④ support of a joint pdf: u, v 平面上，$f(u,v)$ 不为 0 的区域

marginal pdf: $f_X(u) = \int_{-\infty}^{+\infty} f_{X,Y}(u,v)\,dv$；$f_Y(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u,v)\,du$

期望: $E[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \cdot f_{X,Y}(u,v)\,du\,dv = \int_{-\infty}^{\infty} u f_X(u)\,du$

condition pdf: $f_{Y|X}(v|u_0) = \frac{f_{X,Y}(u_0,v)}{f_X(u_0)}$ $-\infty < v < +\infty$ $f_X(u_0) = 0$ 时无定义

$\Rightarrow f_{X,Y}(u,v) = f_{Y|X}(v|u) f_X(u)$, $f_Y(v) = \int_{-\infty}^{\infty} f_{Y|X}(v|u) f_X(u)\,du$

conditional mean: $E[Y|X=u] = \int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,dv \to$ 只与 u 有关 (deterministic function of u)

⑤ uniform joint pdf: $f_{X,Y}(u,v) = \begin{cases} \frac{1}{\text{area of } S} & \text{if } (u,v) \in S \\ 0 & \text{else} \end{cases}$

$P\{(X,Y) \in A\} = \frac{\text{area of } A \cap S}{\text{area of } S}$

4. independence of r.v.

① r.v. X, Y is independent if: 对于 $\{X \in A\}$ $\{Y \in B\}$,

$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$ 或 $A: \{u \le u_0\}$; $B: \{v \le v_0\}$：用于计算

$F_{X,Y}(u_0, v_0) = F_X(u_0) F_Y(v_0) \leftarrow$ 用于 check 是否独立 (充要条件)

$\Rightarrow$ discrete-type r.v. $p_{X,Y}(u,v) = p_X(u) \cdot p_Y(v)$

$\Rightarrow$ continuous-type r.v. $f_{X,Y}(u,v) = f_X(u) \cdot f_Y(v)$

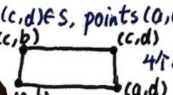
② 从 pdf 判断 independence:

I. X, Y is independent 当且仅当对 all $u \in \mathbb{R}$, $f_X(u) = 0$ 或 $f_{Y|X}(v|u) = f_Y(v)$ for all $v \in \mathbb{R}$

[product set]: $A \times B = \{(u,v): u \in A, v \in B\}$; total area: $|A \times B| = |A| \times |B|$

$|A|$: denote the sum of the lengths of all intervals making up A.

II. a subset S in $\mathbb{R}^2$ has the swap property if for any two points $(a,b) \in S$, $(c,d) \in S$, points (a,d), (c,b) also in S.



4 个顶点都在 S 中 $\to$ holds for a product set $A \times B$.

考虑顶点上的两个点作矩形

III. $S \in \mathbb{R}^2$. S is product set 当且仅当 S 有 swap property

IV. X, Y independent jointly continuous-type r.v. 的 necessary condition (必要条件): the support of $f_{X,Y}$ is a produced set.

$\Rightarrow$ 推论：X, Y is uniformly distributed over a set S in the plane, X, Y independent 当且仅当 S 是 product set.

[eg]. $f_{X,Y}(u,v) = \begin{cases} Cu^2v^2 & u,v \ge 0, u+v \le 1 \\ 0 & \text{else} \end{cases}$ support is not product set! $\to$ not independent

$f_{X,Y}(u,v) = \begin{cases} u+v & u,v \in [0,1] \\ 0 & \text{else} \end{cases}$ 假设 $u+v = f_X(u) \cdot f_Y(v)$, $0<u_1<u_2<1$, $0<v_1<v_2<1$

有 $(u_1+v_1)(u_2+v_2) = (u_1+v_2)(u_2+v_1)$ $f_{X,Y}(u_1,v_2) \cdot f_{X,Y}(u_2,v_1)$

$\Rightarrow (u_2-u_1)(v_2-v_1) = 0$ 矛盾 $\Rightarrow$ not independent: $f_{X,Y}(u_1,v_1) \cdot f_{X,Y}(u_2,v_2) =$

$f_{X,Y}(u,v) = \begin{cases} 9u^2v^2 & u,v \in [0,1] \\ 0 & \text{else} \end{cases}$ yes $\Rightarrow \begin{cases} f_X(u) = 3u^2 \\ f_Y(v) = 3v^2 \end{cases}$

5. distribution of sums of r.v. 算 g(X,Y) 的通法和之前算 g(X) 一样，这里先考虑特殊情况：S = X + Y

① integer-valued r.v. 的和：X, Y 只能取整数.

$p_S(k) = P\{X+Y=k\} = \sum_j P\{X=j, Y=k-j\} = \sum_j p_{X,Y}(j, k-j) \overset{X,Y\ \text{independent}}{=} \sum_j p_X(j) p_Y(k-j) = p_X * p_Y$ 卷积

[eg]. $X \sim B(m,p)$, $Y \sim B(n,p)$, $S = X+Y \to S \sim B(m+n, p)$ $\swarrow = \binom{m+n}{k}$

$p_S(k) = \sum_{j=0}^{k} \binom{m}{j} p^j (1-p)^{m-j} \binom{n}{k-j} p^{k-j} (1-p)^{n-k+j} = \left(\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}\right) p^k (1-p)^{m+n-k}$

$= \binom{m+n}{k} p^k (1-p)^{m+n-k}$

[eg]. $X \sim Poi(\lambda_1)$, $Y \sim Poi(\lambda_2)$, $S = X+Y \to S \sim Poi(\lambda_1 + \lambda_2)$

$p_{X+Y}(k) = \sum_{j=0}^{k} \frac{\lambda_1^j e^{-\lambda_1}}{j!} \frac{\lambda_2^{k-j} e^{-\lambda_2}}{(k-j)!} = e^{-(\lambda_1+\lambda_2)} \cdot \left(\sum_{j=0}^{k} \frac{\lambda_1^j \lambda_2^{k-j}}{j!(k-j)!}\right) = \left(\sum_{j=0}^{k} \binom{k}{j} p^j (1-p)^{k-j}\right) \frac{\lambda^k e^{-\lambda}}{k!}$

$= \frac{\lambda^k e^{-\lambda}}{k!}$ 其中 $\lambda = \lambda_1 + \lambda_2$, $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

② sum of joint continuous-type r.v.: 先算 CDF 再求导得到 pdf

$F_S(c) = P\{S \le c\} = \int_{-\infty}^{\infty} du \cdot \int_{-\infty}^{c-u} f_{X,Y}(u,v)\,dv$

$f_S(c) = \frac{dF_S(u)}{dc} = \int_{-\infty}^{\infty} du \cdot \frac{d}{dc}\left(\int_{-\infty}^{c-u} f_{X,Y}(u,v)\,dv\right) = \int_{-\infty}^{\infty} f_{X,Y}(u, c-u)\,du$

X, Y independent: $f_S(c) = \int_{-\infty}^{\infty} f_X(u) \cdot f_Y(c-u)\,du$

[解法]：I. 要考虑 X, Y 区间 $\Rightarrow$ 画图，分类来做 (重叠地方面积) II. 不考虑区间，套公式

[eg]. $W = \max\{X,Y\}$：$F_W(t) = P\{\max\{X,Y\} \le t\} = P\{X \le t\} \cdot P\{Y \le t\} = F_X(t) \cdot F_Y(t)$

$f_W(t) = f_X(t) F_Y(t) + f_Y(t) F_X(t) \leftarrow$ 两边求导 $W = \min\{X,Y\}$ 同理：用 $P\{W \ge t\}$！

$P\{W \ge t\} = P\{\min\{X,Y\} \ge t\} = P\{X \ge t\} P\{Y \ge t\} \Rightarrow 1 - F_W(t) = [1 - F_X(t)][1 - F_Y(t)]$

6. joint pdf of functions of r.v. 考虑 W, Z 均为 f(X,Y) 函数时的联合分布

① linear mapping: $W = aX + bY$, $Z = cX + dY$, W, Z 在 α, β 平面上，X, Y 在 u, v 平面上

$\binom{W}{Z} = A\binom{X}{Y}$, $\binom{\alpha}{\beta} = A\binom{u}{v}$, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow \binom{u}{v} = A^{-1}\binom{\alpha}{\beta}$, $A^{-1} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$\det(A) = |A| = ad - bc$ 仅 $f_{W,Z}(\alpha,\beta) = \frac{1}{|\det A|} f_{X,Y}\left(A^{-1}\binom{\alpha}{\beta}\right)$，用了一个引理：

if R 是 u-v 平面的 set, S 是 R 映射到 α-β 平面的 image, 有 $area(S) = |\det A| \cdot area(R)$

② one-to one mapping: $\binom{W}{Z} = g\left(\binom{X}{Y}\right)$, $\alpha = g_1(u,v)$, $\beta = g_2(u,v)$, $\binom{\alpha}{\beta} = g\left(\binom{u}{v}\right)$

$J = J(u,v) = \begin{pmatrix} \frac{\partial g_1(u,v)}{\partial u} & \frac{\partial g_1(u,v)}{\partial v} \\ \frac{\partial g_2(u,v)}{\partial u} & \frac{\partial g_2(u,v)}{\partial v} \end{pmatrix}$ Jacobian 矩阵

给定点 (u_0, v_0)，邻近值用 linear func 估计：$g\left(\binom{u}{v}\right) = g\left(\binom{u_0}{v_0}\right) + A\left(\binom{u}{v} - \binom{u_0}{v_0}\right)$, $A = J(u_0, v_0)$

$f_{W,Z}(\alpha,\beta) = \frac{1}{|\det J|} f_{X,Y}\left(g^{-1}\binom{\alpha}{\beta}\right)$

[eg]. $f_{X,Y} = \begin{cases} u+v & (u,v) \in [0,1]^2 \\ 0 & \text{else} \end{cases}$ $W = X^2$, $Z = X(1+Y)$，求 pdf: $f_{W,Z}$

先算 $f_{W,Z}(\alpha,\beta)$ support: $\begin{cases} \alpha = u^2 \in [0,1] \\ \beta = u(1+v) \in [\sqrt{u}, 2\sqrt{u}] \end{cases}$

用 α, β 表示 u, v: $u = \sqrt{\alpha}$, $v = \frac{\beta}{\sqrt{\alpha}} - 1$ ← one-to-one!

$|\det J| = \left|\det\begin{pmatrix} 2u & 0 \\ 1+v & u \end{pmatrix}\right| = 2u^2 = 2\alpha$

$f_{W,Z}(\alpha,\beta) = \begin{cases} \frac{\sqrt{\alpha} + \frac{\beta}{\sqrt{\alpha}} - 1}{2\alpha}, & (\alpha,\beta) \in A \\ 0 & \text{else} \end{cases}$

$A = \{(\alpha,\beta): 0 < \alpha \le 1, \sqrt{\alpha} \le \beta \le 2\sqrt{\alpha}\}$

③ many-to-one mapping: 把同一个 (α, β) 对应的 (u, v) 全部加起来

[eg]. $W = \min\{X,Y\}$, $Z = \max\{X,Y\}$，用 $f_{X,Y}$ 表示 $f_{W,Z}$ 显然 (X,Y), (Y,X) 等价：2-to-1

$f_{W,Z}(\alpha,\beta)$ support: $\{(\alpha,\beta): \alpha \le \beta\} \to$ 实际上，$\alpha = \beta$，即 $u = v$ 时，$P\{W=Z\} = P\{X=Y\} =$

$\int_{-\infty}^{\infty} \int_{v}^{v} f_{X,Y}(u,v)\,du\,dv = 0 \to$ So 不重要 $\Rightarrow$ support: $U = \{(\alpha,\beta): \alpha < \beta\}$

考虑 subset $A \in U$: $\{(W,Z) \in A\} = \{(X,Y) \in A\} \cup \{(Y,X) \in A\}$, 对 $f_{Y,X}(u,v) = f_{X,Y}(v,u)$

$f_{W,Z}(\alpha,\beta) = \begin{cases} f_{X,Y}(\alpha,\beta) + f_{X,Y}(\beta,\alpha) & \alpha < \beta \\ 0 & \alpha \ge \beta \end{cases}$

7. correlation and covariance X, Y 是两个 r.v.

correlation: $E[XY]$ covariance $Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$

correlation coefficient: $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ 相关系数

① covariance: $Cov(X,Y) = E[X(Y - E[Y])] = E[(X - E[X])Y] = E[XY] - E[X]E[Y]$

linear: $Cov(X+Y, U+V) = Cov(X,U) + Cov(X,V) + Cov(Y,U) + Cov(Y,V)$

$Cov(aX+b, cY+d) = ac\,Cov(X,Y)$

如果 X, Y is uncorrelated: $Var(X+Y) = Cov(X+Y, X+Y) = Cov(X,X) + Cov(Y,Y) + 2Cov(X,Y)$

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

Consider $S_n = X_1 + \cdots + X_n$,

$X_1 \sim X_n$ uncorrelated, $E[X_i] = \mu$, $Var[X_i] = \sigma^2$: $E[S_n] = n\mu$

$Var(S_n) = Cov(S_n, S_n) = Cov\left(\sum_i X_i, \sum_j X_j\right) = n\sigma^2 \to$ standardized version: $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$

② X and Y uncorrelated: $Cov(X,Y) = 0$ X and Y positively correlated: $Cov(X,Y) > 0$

X and Y negatively uncorrelated: $Cov(X,Y) < 0$

仅 X, Y independent $\xrightarrow{\quad}$ X, Y uncorrelated (反之不亦然！)

多个 r.v. uncorrelated 当 pairwise (两两) uncorrelated.

③ $Cov(X,Y)$与X, Y单位有关，但$\rho_{X,Y}$无关

$\Rightarrow Cov\left(\frac{X-E[X]}{\sigma_X}, \frac{Y-E[Y]}{\sigma_Y}\right) = Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = \rho_{X,Y}$ ☆ $\rho_{aX+b,cY+d} = \rho_{X,Y}$ $(a,c>0)$

④ Schwarz's inequality: X, Y两个r.v. $|E[XY]| \le \sqrt{E[X^2]E[Y^2]}$

$\Leftrightarrow (a_1b_1+\cdots+a_nb_n)^2 \le (a_1^2+\cdots+a_n^2)(b_1^2+\cdots+b_n^2)$，若$E[X^2]\ne0$，取等当且仅当

$P\{Y=cX\}=1$ for some constant c.

$\Rightarrow$推论：$|Cov(X,Y)| \le \sqrt{Var(X)Var(Y)}$ 若$Var(X)\ne0$，取等当且仅当$Y=aX+b$

[相关系数correlation coefficient $\rho_{X,Y}$]：$|\rho_{X,Y}|\le1$，

$\rho_{X,Y}=1$当且仅当$Y=aX+b, a>0$；$\rho_{X,Y}=-1$当且仅当$Y=aX+b, a<0$

[covariance matrix]：$\begin{pmatrix}X_1\\X_2\\X_3\end{pmatrix}$ $\begin{pmatrix}5&2&0\\2&5&2\\0&2&5\end{pmatrix}$ $Cov(X_i,X_j)$：i^{th} from the top，j^{th} from the left.

[eg]. $X_1 \sim X_n$独立，mean μ，variance σ^2，用sample mean $\hat{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$和sample variance $\hat{\sigma}^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\hat{X})^2$估计$\mu$和$\sigma^2$，an estimator is called unbiased 若estimator的mean和parameter的mean相当.

(a) sample mean: $E[\hat{X}]=\frac{1}{n}\sum_{k=1}^{n}E[X_k]=\frac{1}{n}\sum_{k=1}^{n}\mu=\mu \Rightarrow$ unbiased.

(b) find the mean square error $E[(\mu-\hat{X})^2]$ for 用sample mean估计mean:

$E[(\mu-\hat{X})^2]=Var(\hat{X})=\frac{1}{n^2}Var\left(\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}Var(X_k)=\frac{\sigma^2}{n}$

(c) sample variance: $E[\hat{\sigma}^2]=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\hat{X})^2=\frac{n}{n-1}E[(X_1-\hat{X})^2]$ (由symmetry)，又

$E[X_1-\hat{X}]=\mu-\mu=0$，$E[(X_1-\hat{X})^2]=Var(X_1-\hat{X})=Var\left(\frac{n-1}{n}X_1-\sum_{k=2}^{n}\frac{X_k}{n}\right)=\left[\left(\frac{n-1}{n}\right)^2+\sum_{k=2}^{n}\frac{1}{n^2}\right]\sigma^2$

$=\frac{n-1}{n}\sigma^2$，$E[\hat{\sigma}^2]=\sigma^2 \Rightarrow$ unbiased

8. minimum mean square error estimation | MSE: mean square error

① constant estimator：用常数δ去估计Y (not observed)：

MSE: $E[(Y-\delta)^2]=\int_{-\infty}^{\infty}(y-\delta)^2f_Y(y)dy$，在$\delta^*=E[Y]$取到最小值，为$Var(Y)$

[证]：$E[(Y-\delta)^2]=E[Y^2]-2\delta\cdot E[Y]+\delta^2$ 二次曲线

② unconstrained estimator：用函数$g(X)$去估计Y (observation X) 在已知joint pdf下

eg. $X=10$，$g(X)$为常数，转化为constant estimator $\Rightarrow E[Y|X=10]=\int_{-\infty}^{+\infty}v\cdot f_{Y|X}(v|10)dv$

MSE: $E[(Y-E[Y|X=10])^2]=E[Y^2|X=10]-(E[Y|X=10])^2$

optimal estimator: $g^*(u)=E[Y|X=u]=\int_{-\infty}^{\infty}vf_{Y|X}(v|u)dv$，用$E[Y|X]$表示$g^*(X)$

MSE: $E[(Y-E[Y|X])^2]=E[Y^2]-E[(E[Y|X])^2]$

[证]：$MSE=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}[v-g^*(u)]^2f_{Y|X}(v|u)dv\right)f_X(u)du=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(v^2-g^*(u)^2)f_{Y|X}(v|u)dv\right)f_X(u)du$

③ linear estimator，②中$g(X)$转化成$L(X)=aX+b$

MSE: $E[(Y-(aX+b))^2]=E[((Y-aX)-b)^2] \Rightarrow$ 转化成constant estimator，

$b=E[Y-aX]=\mu_Y-a\mu_X$，代入$MSE=E[(Y-\mu_Y-a(X-\mu_X))^2]=Var(Y-aX)=$

$Var(Y)-2aCov(X,Y)+a^2Var(X) \Rightarrow a^*=\frac{Cov(Y,X)}{Var X}$

minimum MSE linear estimator: $L^*(X)=\hat{E}[Y|X]=\mu_Y+\left(\frac{Cov(Y,X)}{Var(X)}\right)(X-\mu_X)$ 相等 $\mu_Y+\sigma_Y\cdot\rho_{X,Y}\left(\frac{X-\mu_X}{\sigma_X}\right)$

MSE: $\sigma_Y^2-\frac{Cov(Y,X)^2}{Var(X)}=\sigma_Y^2(1-\rho_{X,Y}^2)=\sigma_Y^2-Var(\hat{E}[Y|X])=E[Y^2]-E[\hat{E}[Y|X]^2]$

$\hat{E}[Y|X]$又叫wide sense condition expectation of Y given X，将X, Y理解为两个r.v.

[若X, Y是standard r.v.] ($\mu=0$, var $=1$)：$\hat{E}[Y|X]=\rho_{X,Y}\cdot X$；MSE: $1-\rho_{X,Y}^2$

④ 总结：I. order: $E[(Y-g^*(X))^2] \le \sigma_Y^2(1-\rho_{X,Y}^2) \le \sigma_Y^2 \leftarrow$ MSE for $\delta^*=E[Y]$

MSE for $g^*(X)=E[Y|X]$ MSE for $L^*(X)=\hat{E}[Y|X]$

II. linear: $E[aX+bY+c]=aE[X]+bE[Y]+c$

$E[aY+bZ+c|X]=aE[Y|X]+bE[Z|X]+c$；$\hat{E}[aY+bZ+c|X]=a\hat{E}[Y|X]+b\hat{E}[Z|X]+c$

9. law of large number：关于sample average of n个r.v.即$\frac{S_n}{n}$，$S_n=X_1+\cdots+X_n$，

$X_1\sim X_n$相互独立，mean μ，var $\le c$，有：(对$\forall\delta>0$)

$P\left\{\left|\frac{S_n}{n}-\mu\right|\ge\delta\right\}\le\frac{c}{n\cdot\delta^2}\xrightarrow{n\to\infty}0$ 用Chebychev不等式证，算$E[\frac{S_n}{n}]$和$Var(\frac{S_n}{n})$

[eg]. $U_1\sim U_n$独立，指数分布(λ)，(a) observation $u_1\sim u_n$，求MLestimator $\hat{\lambda}$

$f_T(u_1,\ldots,u_n)=\lambda e^{-\lambda u_1}\cdots\lambda e^{-\lambda u_n}=\lambda^n e^{-\lambda S_n}$，$S_n=u_1+\cdots+u_n$，求得有$\hat{\lambda}_{ML}=\frac{n}{S_n}$

(b) 求足够大的n，使$[0.9\hat{\lambda}_{ML}, 1.1\hat{\lambda}_{ML}]$ is a confidence interval for estimation of λ with confident level 96%. (用Chebychev)

$P\left\{\frac{0.9n}{S_n}\le\lambda\le\frac{(1.1)n}{S_n}\right\}\ge0.96 \Rightarrow P\left\{0.9\le\lambda\frac{S_n}{n}\le1.1\right\}\ge0.96$，$E[\lambda\cdot\frac{S_n}{n}]=E[\lambda U_1]=1$

$Var\left(\frac{\lambda S_n}{n}\right)=\frac{\lambda^2}{n^2}Var(S_n)=\frac{\lambda^2}{n}\cdot Var(U_1)=\frac{1}{n} \Rightarrow P\left\{\left|\frac{\lambda S_n}{n}-1\right|\ge\delta\right\}\le\frac{1}{n\delta^2}$，

令$\frac{1}{n\delta^2}=0.04$，$\delta=0.1 \Rightarrow n=2500$

10. Central limit theorem $X_1\sim X_n$独立，identically distributed r.v. (mean μ, var σ^2)

$S_n=X_1+\cdots+X_n$，有：$\lim_{n\to\infty}P\left\{\frac{S_n-n\mu}{\sqrt{n\sigma^2}}\le c\right\}=\Phi(c)$，或：

$S_n-\frac{n}{2}$是高斯分布，mean 0，var: $\frac{n}{12}$

11. Joint Gaussian distribution

[想法]：有$(U_1,V_1)\cdots(U_n,V_n)$独立，相同分布r.v.，mean $=0$，当$n\to\infty$，limit分布是joint distribution of 两个r.v.

$\left(\frac{U_1+\cdots+U_n}{\sqrt{n}}, \frac{V_1+\cdots+V_n}{\sqrt{n}}\right)$是limiting bivariate distribution，由CLT，若X, Y有这个性质，$aX+bY$也有 $\Rightarrow$

[定义]：r.v. X, Y是joint Gaussian当所有线性组合$aX+bY$是Gaussian r.v.

(包括退化情况：常数，高斯分布，var $=0$；X, Y线性相关：$X=aY+b$或$Y=aX+b$)

① 公式：bivariate normal pdf (5个参数：$\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$)

$$f_{X,Y}(u,v)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left(-\frac{\left(\frac{u-\mu_X}{\sigma_X}\right)^2+\left(\frac{v-\mu_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{u-\mu_X}{\sigma_X}\right)\left(\frac{v-\mu_Y}{\sigma_Y}\right)}{2(1-\rho^2)}\right)$$

☆判断pdf是否bivariate normal：$f_{X,Y}(u,v)=C\cdot\exp(-P(u,v))$，其中

$P(u,v)=au^2+buv+cv^2+du+ev+f$ (f实际上可以归到常数C中，仍5个参数)

其中$a>0, c>0, b^2-4ac<0$

② 从standard 2-d normal到general：W, Z是独立，standard normal r.v.

相关系数$\rho=0$，有$f_{W,Z}(\alpha,\beta)=\left(\frac{e^{-\frac{\alpha^2}{2}}}{\sqrt{2\pi}}\right)\left(\frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}}\right)=\frac{e^{-\frac{\alpha^2+\beta^2}{2}}}{2\pi}$ $\Leftarrow$ standard bivariate normal pdf

有线性变换$\begin{pmatrix}X\\Y\end{pmatrix}=A\begin{pmatrix}W\\Z\end{pmatrix}+\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}$，

$A=\begin{pmatrix}\sqrt{\frac{\sigma_X^2(1+\rho)}{2}} & -\sqrt{\frac{\sigma_X^2(1-\rho)}{2}}\\ \sqrt{\frac{\sigma_Y^2(1+\rho)}{2}} & -\sqrt{\frac{\sigma_Y^2(1-\rho)}{2}}\end{pmatrix}$，则X, Y有上面公式.

③ 图象：I. 对于标准(W, Z)：bell shape that rotationally symmetric about origin，level set是围着原点的circles，half-peak level set: $\{(\alpha,\beta):\alpha^2+\beta^2=2\ln2\}$

II. 对于一般(X, Y)，half-peak level set：μ_X, μ_Y表示圆心，σ_X, σ_Y表示长短轴的长度，ρ表示旋转角度

$\left\{(u,v):\frac{\left(\frac{u-\mu_X}{\sigma_X}\right)^2+\left(\frac{v-\mu_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{u-\mu_X}{\sigma_X}\right)\left(\frac{v-\mu_Y}{\sigma_Y}\right)}{1-\rho^2}=2\ln2=(1.18)^2\right\}$ 是个椭圆ellipse!

④ 性质：I. X有$N(\mu_X,\sigma_X^2)$分布，Y有$N(\mu_Y,\sigma_Y^2)$分布

II. 线性组合$aX+bY$是高斯r.v. (或：X, Y是joint Gaussian)

III. ρ是correlation coefficient between X, Y ($\rho_{X,Y}=\rho$) $\Leftarrow$ X, Y独立当且仅当$\rho=0$

IV. for estimation of Y from X，$L^*(X)=g^*(X)$ (或：$E[Y|X]=\hat{E}[Y|X]$)：the best unconstrained estimator $g^*(X)$ is linear.

V. the conditional distribution of Y given $X=u$ is $N(\hat{E}[Y|X=u],\sigma_e^2)$，$\sigma_e^2$是$\hat{E}[Y|X]$的MSE.

⑤ higher dimensional joint Gaussian distribution：n dimensional Gaussian density $X=\begin{pmatrix}X_1\\\vdots\\X_n\end{pmatrix}$，$\mu=\begin{pmatrix}\mu_1\\\vdots\\\mu_n\end{pmatrix}$，$u=\begin{pmatrix}u_1\\\vdots\\u_n\end{pmatrix}$

n×n covirance matrix Σ和$f_X(u)=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma)}}\cdot\exp\left(-\frac{(u-\mu)^T\Sigma^{-1}(u-\mu)}{2}\right)$

Σ中i行$j$th是$Cov(X_i,X_j)$，i^{th} diagonal element是$Var(X_i)$

[对于n=2]：$\mu=\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}$，$\Sigma=\begin{pmatrix}\sigma_X^2&\rho\sigma_X\sigma_Y\\\rho\sigma_X\sigma_Y&\sigma_Y^2\end{pmatrix}$，$\det(\Sigma)=\sigma_X^2\sigma_Y^2(1-\rho)$.

$\Sigma^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\frac{1}{\sigma_X^2}&-\frac{\rho}{\sigma_X\sigma_Y}\\-\frac{\rho}{\sigma_X\sigma_Y}&\frac{1}{\sigma_Y^2}\end{pmatrix}$

12. 补充

① 关于MSE：consider r.v. Y，估计$\hat{Y}=g(X)=E[Y|X]$，error(e)：$Y-\hat{Y}$

MSE: $E[e^2]=Var(e)+E[e]^2$ 一直等于0 $\Leftarrow$ 解释joint Gaussian distribution性质

$E[Y|X]=\int E[Y|X=u]f_X(u)du \Rightarrow$ 就算给出了X的值，也不能用$E[Y|X=u]$来算MSE!

② die扔n次，$S_n=X_1+\cdots+X_n$，$E[X_1]=\frac{7}{2}$，$Var(X_1)=\frac{35}{12}$，决定n使$\frac{S_n}{n}$ is within 10% of the μ_X is greater than 0.92

I. 用law of large number：$P\left\{\left|\frac{S_n}{n}-\mu\right|\ge\delta\right\}\le\frac{\sigma^2}{n\cdot\delta^2}$，又$P\left\{\left|\frac{S_n}{n}-\mu\right|\le0.1\mu\right\}\le0.08$，

$\delta=0.1\mu=\frac{7}{20} \Rightarrow \frac{Var(\frac{S_n}{n})}{n\cdot\delta^2}\le0.08$，$n\ge\frac{1}{0.08\cdot0.1^2}\cdot\frac{4}{49}\cdot\frac{35}{12}=297.6$

II. 用CLT：$\frac{S_n-n\mu}{\sqrt{n\sigma^2}}$是高斯分布，$\mu=\frac{7}{2}n$，$\sigma^2=\frac{35}{12}n$，有

$P\{|S_n-\mu|\le\frac{7}{20}n\}=P\left\{\left|\frac{S_n-\mu}{\sqrt{\frac{35}{12}n}}\right|\le\frac{\frac{7}{20}n}{\sqrt{\frac{35}{12}n}}\right\}=1-2Q\left(\frac{\frac{7}{20}n}{\sqrt{\frac{35}{12}n}}\right)=2\Phi\left(\frac{\frac{7}{20}n}{\sqrt{\frac{35}{12}n}}\right)-1\ge0.92\cdots$