概率论

casio911教程

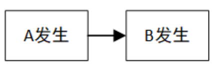
$s^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$	**样本方差**
$\sigma^2=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-\mu)^2$	总体方差
$\int_0^{+\infty}e^{-x^2}dx=\frac{\sqrt{\pi}}{2}(\text{Gamma})$	$\displaystyle\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$ $\displaystyle\int_{-\infty}^{+\infty}e^{-t^2}dt=\sqrt{\pi}$

双变量计算可算$\sum x_iy_i,\sum x_i^2,\sum y_i^2$，回归计算的$a$是$\hat{b}$，$b$是$\hat{a}$，$r$是相关度

概率

基本概念

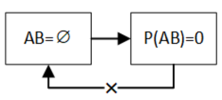
- 基本事件：必发生一个且仅发生一个的最简单事件　基本事件互不相容
- 随机实验的定义：
    - (1) 在相同的条件下可以重复进行;
    - (2) 试验的结果是可以预知的;
    - (3) 试验完成前无法知道试验的最终结果
- 在随机试验中必然发生的事件称为必然事件,用符号Ω表示
- 在随机试验中必然不发生的事件称为不可能事件,用符号∅表示
- 随机事件的关系：（事件A与事件B）

　$\star$ 若 

称为 $A\subset B$ (包含)

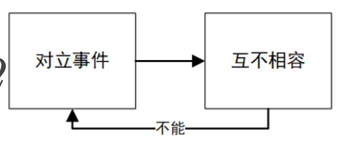
　$\star$ 若$A\subset B$且$B\subset A$，记为 $A=B$.(相等)

　$\star$ 若 A 与 B 不能在**同一次试验中同时发生**称 A 与 B 是互斥(互不相容)的　$A\cap B=\varnothing$,

- 

可能$P(A)$或$P(B)=0$，**概率为0不能代表是空集**

　$\star$ 若 A 发生当且仅当 B 不发生，则称事件 B 是事件 A 的对立事件，记为 $\bar{A}$.

　$A\cap B=\varnothing$且$A+B=\Omega$ 

基本性质

性质 1　$P(\varnothing)=0$；

性质 2　$0\le P(A)\le 1$；

性质 3 设 $A_1,A_2,\cdots,A_n$ 为两两互斥的事件，$P\left(\bigcup\limits_{i=1}^{n}A_i\right)=\sum\limits_{i=1}^{n}P\left(A_i\right)$ (有限可加性);

性质 5 若 $A\subset B$，则 $P(A)\le P(B)$；$P(AB)=P(A)$

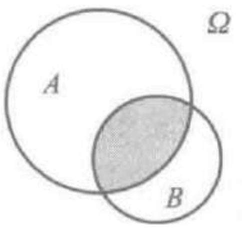
$P(AB)\le\frac{P(A)+P(B)}{2}$　　$(P(AB)\le P(A),P(AB)\le P(B))$

性质 6 $P(A\cup B)=P(A)+P(B)-P(AB)$ (加法公式);

三并：$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P()$

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

性质7 设 A,B 为任意事件，则 $P(A-B)=P(A\bar{B})=P(A)-P(AB)$ （减法公式）;

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

$P(A),P(AB),P(A\cup B),P(A)+P(B)$关联：

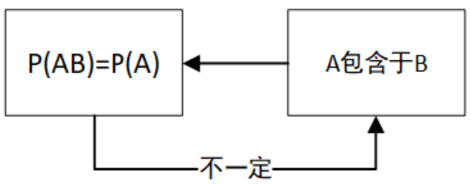
$$P(AB)\le P(A)\le P(A\cup B)\le P(A)+P(B)$$

等号条件：1：$A\subset B$ 2：$A\cup B=A$　3：$A\cap B=\varnothing$

基本关系式

2. 分配律 $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$　　$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$.
3. 对偶律 $\overline{A\cup B}=\bar{A}\cap\bar{B};\overline{A\cap B}=\bar{A}\cup\bar{B}$.
4. 差事件：$A-B=A\bar{B}=A-AB$
5. 条件概率：$P(A|B)=\frac{P(AB)}{P(B)}$ （**数量少的可以直接列举**）

$$P(A|B)=1,\ P(AB)=P(B)\rightarrow\begin{cases}P(A\cup B)=P(A)+P(B)-P(AB)=P(A)\\P(B-A)=P(B)-P(AB)=0\end{cases}$$



反例：$X\sim U(0,1)$，事件A：$X<0$，事件B：$X>0$，则$P(AB)=P(A)=0$，但A不包含于B

基本公式

乘法公式：$P(AB)=P(B)P(A\mid B)$

两两独立：$P(AB)=P(A)P(B)$ $P(A|\bar{B})=P(A|B)$可证　**对立事件也满足两两独立**

相互独立：$P(ABC)=P(A)P(B)P(C)$

共有$C_n^2+C_n^3+C_n^4+\cdots+C_n^n=(C_n^0+C_n^1+\cdots+C_n^n)-C_n^0-C_n^1=2^n-n-1$个式子

两两独立的例子：

- A={第一次掷骰子为偶}，B={第一次掷骰子为奇}，C={两次同时为奇或为偶}

  注意：$A\not\subset C,B\not\subset C$，C的子事件不是他们，例如：第一次为1，第二次为1

全概率公式：

$$P(B)=\sum_{i=1}^{n}P\left(A_i\right)P\left(B\mid A_i\right)$$

贝叶斯公式：

$$P\left(A_i\mid B\right)=\frac{P\left(A_i\right)P\left(B\mid A_i\right)}{\sum\limits_{i=1}^{n}P\left(A_i\right)P\left(B\mid A_i\right)}=\frac{P(A_iB)}{P(B)}$$

题

例题：假设A与B同时发生的时候C必发生，满足条件：

可知：$AB \subset C$,所以$P(C) \geqslant P(AB) = P(A) + P(B) - P(A \cup B) \geqslant P(A) + P(B) - 1$

例题：

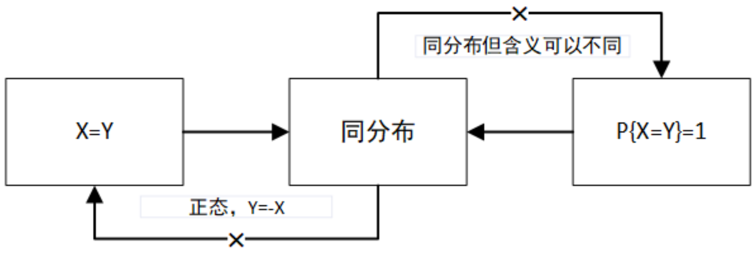
若$0 < P(A) < 1, 0 < P(B) < 1$，则$P(A|B) > P(A|\bar{B})$的充分必要条件是?

$P(B|A) > P(B|\bar{A}) \rightarrow P(AB) - P(B)P(A) > 0$

关键：$P(A\bar{B}) = P(A - AB) = P(A) - P(AB)$

随机变量及其分布

基本概念



例：在人群中随机选取一个人，记其性别为随机变量X，$P(X=0)=1/2$；$P(X=1)=1/2$。

令随机变量$Y=1-X$；易知Y与X的分布相同，但Y与X是不同的随机变量

分布函数$F(X)$

设X为随机变量，对于任意实数x，$F(x) = P(X \leq x), -\infty < x < +\infty$

对任意的两个常数 $a<b$ ，有$P(a < X \leq b) = F(b) - F(a)$

分布函数的性质（**判定是不是分布函数的条件**）

1. $$0 \leq F(x) \leq 1, \quad \lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$$
2. $F(x)$ 单调不减，即当 $x_1 < x_2$ 时，有 $F(x_1) \leq F(x_2)$；
3. $F(x)$ 是 x 的右连续函数，即 $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$

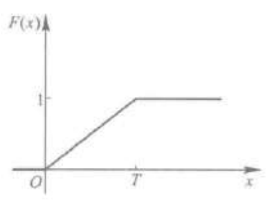
判断连续还是离散：如果$F(x)$部分连续则不为离散，如果某处$F(x+0) \neq F(x-0)$则不连续，**有可能既不离散也不连续**；

书本例：

某系统的寿命为 T, 系统在 t 时刻尚正常运行的条件下, 其失效率总保持为某个常数 $\lambda(>0)$, 即有

$$\lim_{h \rightarrow 0^+} \frac{1}{h} P\{t < T \leqslant t+h \mid T > t\} = \lambda,$$

试写出 T 的分布函数.



解：当 $t<0$ 时, $F(t)=0$，当 $t>0, h>0$ 时, (**把一件事情拆解为两件事件的交集，然后可以使用条件概率的公式**):
$\{t < T \leqslant t+h\} = \{t < T \leqslant t+h\} \cap \{t < T\}$,

$$P\{t < T \leqslant t+h \mid T > t\} = \frac{P\{t < T \leqslant t+h\}}{P\{T > t\}} = \frac{F(t+h) - F(t)}{1 - F(t)}.$$

由题设条件可知，有

$$\frac{\lim\limits_{h\to0^+}\frac{F(t+h)-F(t)}{h}}{1-F(t)}=\lambda,$$

得到关于 $F(t)$ 的微分方程

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t}=\lambda[1-F(t)]\rightarrow F(t)=1-\mathrm{e}^{-\lambda t}\rightarrow F(t)=\begin{cases}1-\mathrm{e}^{-\lambda t}, & t\geqslant 0\\ 0, & t<0\end{cases}$$

分布律$P(X=x_i)$

离散型随机变量

一维离散型随机变量的分布律可用下表来表示：

X	x_1	x_2	$\cdots$	x_n	$\cdots$
概率	p_1	p_2	$\cdots$	p_n	$\cdots$

满足：

> 非负性 $p_i\geq 0,\quad i=1,2,\ldots;$
>
> 规范性$\sum\limits_{i=1}^{+\infty}p_i=1$

概率密度函数$f(x)$

连续性随机变量

$f(x)$满足$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$，在$x$处连续则$F'(x)=f(x)$

满足：

> 非负性 $f(x)\geq 0,-\infty<x<+\infty$
>
> 规范性 $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1.$
>
> $P\{x_1<X<x_2\}=\int_{x_1}^{x_2}f(x)dx$**(不等关系有无等号结果相同)**
>
> $P\{X=x_0\}=0$

> 推导：$0\leqslant P\{X=x_0\}\leqslant P\{x_0-\Delta x<X\leqslant x_0\}=F(x_0)-F(x_0-\Delta x)$
>
> $\Delta x\rightarrow 0^+$，则右式趋于0

判断是不是概率密度函数：积分为1

1. $f_1(x)F_2(x)+f_2(x)F_1(x)$是概率密度函数
2. $2f_1(x)-f_2(x),x\in R$不一定，因为在某些x处可以是负数，违背非负性
3. $2f_1(x)f_1(y),-\infty<x\leq y<+\infty$ 轮换对称性 是概率密度函数

分布函数类型

离散型：

几何分布：在n次伯努利试验中，试验k次才得到第一次成功的机率。（前k-1次皆失败，第k次成功的概率）

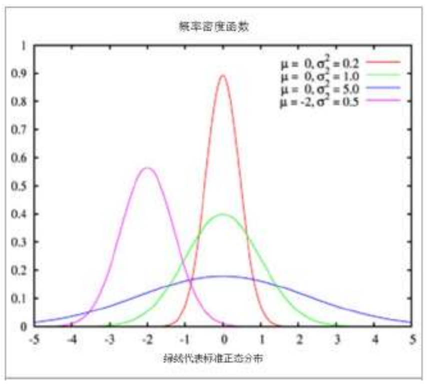
名称	记号	取值	分布律	期望	方差
二项分布	$X\sim B(n,p)$	$0,1,2,\ldots,\mathrm{n}$	$P(X=k)=C_n^kp^k(1-p)^{n-k}$	np	$np(1-p)$
泊松分布	$X\sim P(\lambda)$	$0,1,2,\ldots..$	$P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\lambda>0$	λ	λ

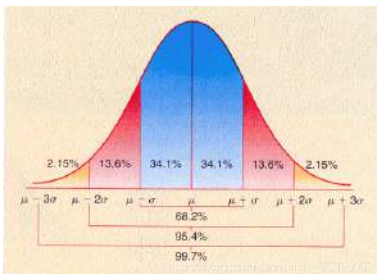
连续性：

名称	记号	取值	概率密度	期望	方差
均匀分布	$X\sim U(a,b)$	$\mathrm{a}<\mathrm{x}<\mathrm{b}$	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}.\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布	$X\sim E(\lambda)$	$\mathrm{x}>0$	$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, & x>0,\\ 0, & \end{cases}$	λ^{-1}	λ^{-2}
正态分布	$X\sim N\left(\mu,\sigma^2\right)$	R	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$	μ	σ^2
标准正态分布	$X\sim N(0,1)$	R	$f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$	0	1

指数分布无记忆性、指数分布的分布函数：$F(x)=\begin{cases}0, & x<0\\ 1-\mathrm{e}^{-\lambda x}, & x\geqslant 0\end{cases}$

正态分布





一维正态：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

标准正态分布：$P(a<X\leqslant b)=\Phi(b)-\Phi(a)$

$f(0)=\frac{1}{\sqrt{2\pi}\sigma}$ σ越大标准正态分布曲线越平缓，σ影响最大值

当x小于0时：$\Phi(x)=1-\Phi(-x)$ $P(|x|<a)=2\Phi(a)-1$

3σ原理：$P(\mu-3\sigma<x<\mu+3\sigma)=0.9974$

落在$(\mu-3\sigma$，$\mu+3\sigma)$以外的概率小于千分之三，在实际问题中常认为相应的事件是不会发生

非标准正态：$X\sim N\left(\mu,\sigma^2\right)$：$P(a<X\leqslant b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$

二维正态：

标准二维正态$(X,Y)\sim N(0,1;0,1;\rho)$：（**二维的$2\pi$没有根号**）

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

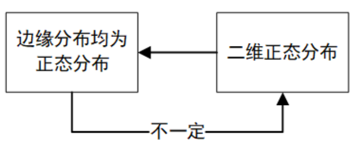
例题：

判断σ_1,σ_2的大小关系

$X\sim N\left(\mu_1,\sigma_1^2\right),Y\sim N\left(\mu_2,\sigma_2^2\right)$, 且 $P\left(\left|X-\mu_1\right|\leqslant 1\right)>P\left(\left|Y-\mu_2\right|\leqslant 1\right)\rightarrow P(|X-\mu|\leqslant a)=2\Phi(\frac{a}{\sigma})-1\rightarrow\sigma_1<\sigma_2$

性质：

1. 边缘分布仍为正态分布



若 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$,则$X\sim N\left(\mu_1,\sigma_1^2\right),Y\sim N\left(\mu_2,\sigma_2^2\right)$

2. 可加性

   若相互独立 $X \sim N\left(\mu_1, \sigma_1^2\right), Y \sim N\left(\mu_2, \sigma_2^2, \rho\right)$，则 $X+Y \sim N\left(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2\right)$

3. 线性不变性

   若$X \sim N\left(\mu, \sigma^2\right)$,则 $kX+c \sim N\left(k\mu+c, k^2\sigma^2\right)$

4. 相互独立的充要条件 $\boxed{\rho=0} \longleftrightarrow \boxed{X、Y相互独立}$ (仅针对正态分布)

可加性

X与Y服从同类型的分布，且相互独立，则$X+Y$也服从同类型的分布

二项分布： 若 $X \sim B(m, {\color{red}p}), Y \sim B(n, {\color{red}p})$，则$X+Y \sim B(m+n, p)$

泊松分布：若 $X \sim P\left(\lambda_1\right), Y \sim P\left(\lambda_2\right)$，则 $X+Y \sim P\left(\lambda_1+\lambda_2\right)$

正态分布：若 $X \sim N\left(\mu_1, \sigma_1^2\right), Y \sim N\left(\mu_2, \sigma_2^2, \rho\right)$，则 $X+Y \sim N\left(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2\right)$

例：

已知运载火箭在飞行中,进入它的仪器舱的宇 宙粒子数服从参数为 λ 的泊松分布. 而进入仪器舱的 每个粒子落到仪器的重要部位的概率等于 p ,试求恰 有 k 个粒子落到仪器重要部位的概率.
解：从第一个试验入手，**划分样本空间。**

设 X 表示宇宙粒子进入仪器舱的个数 $X \sim P(\lambda)$ $P\{X=m\}=\frac{\lambda^m}{m!} e^{-\lambda}, m=0,1,2, \ldots$

显然 $\{X=m\},(m=0,1,2, . .)$ 构成样本空间的一个划分

设Y表示落到重要部位的粒子数, 由题意知

$P\{Y=k \mid X=m\}=C_m^k p^k(1-p)^{m-k}, k=0,1,2, \ldots . . m$

由全概率公式得所求概率为

$$
\begin{aligned}
P\{Y=k\} & =\sum_{m=0}^{+\infty} P\{X=m\} P\{Y=k \mid X=m\}=\sum_{m=0}^{+\infty} \frac{\lambda^m e^{-\lambda}}{m!} C_m^k p^k(1-p)^{m-k} \quad(m \geq k) \\
& =\frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{m=k}^{+\infty} \frac{\lambda^{m-k}(1-p)^{m-k}}{(m-k)!} \xrightarrow{n=m-k} \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=0}^{+\infty} \frac{[\lambda(1-p)]^n}{n!} \\
& =\frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)}=\frac{(\lambda p)^k}{k!} e^{-\lambda p} \quad k=0,1,2, \ldots .
\end{aligned}
$$

落到仪器重要部位的粒子数服从参数为λp 的泊松分布

求解

连续型随机变量：

1. 由随机变量X的取值范围Ω_X 确定随机变量Y的取值范围 Ω_Y ;
2. 对任意一个 $y \in \Omega_Y$ ，求出：

$$
F(y)=P(Y \leq y)=P(g(X) \leq y)=P\left\{X \in G_y\right\}=\int_{G_y} f(x) \mathrm{d} x
$$

3. 按分布函数的定义写出 $F_Y(y)$ $f_Y(y)=F_Y^{\prime}(y)$，$-\infty<y<+\infty$.

*设连续型随机变量 X 的密度函数为 $f_X(x)$，$Y=g(X)$ 是连续型随机变量

*若 $y=g(x)$ 为严格单调函数，$x=h(y)$为相应的反函数，且为可导函数，则 $Y=g(X)$ 的密度函数为
$f_Y(y)=f_X(h(y)) \cdot {\color{red}\left|\left[h(y)^{\prime}\right]\right|}$ 少用

二维随机变量及其分布

设 $F(x, y)$是二维随机变量 (X, Y)的联合分布函数。

性质：（验证是不是联合分布函数）

1. $0 \leq F(x,y) \leq 1$;
2. 当固定y值时，$F(x,y)$是变量x的非减函数；

   当固定x值时，$F(x,y)$是变量y的非减函数；
3. 求解未知数用：

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

4. 当固定$y(x)$值时，$F(x,y)$是变量$x(y)$的右连续函数；
5. $P\left(x_1<X\leq x_2, y_1<Y\leq y_2\right)=F\left(x_2,y_2\right)-F\left(x_2,y_1\right)-F\left(x_1,y_2\right)+F\left(x_1,y_1\right)$。

在$f(x,y)$的连续处有：$f(x,y)=\frac{\partial F(x,y)}{\partial x \partial y}$

边缘：

边缘分布函数：$F_X(x)=\lim_{y\to\infty} F(x,y) \qquad F_Y(x)=\lim_{x\to\infty} F(x,y)$

边缘分布律：$P\left(X=x_i\right)=\sum_{j=1}^{\infty} P\left(X=x_i, Y=y_j\right)=\sum_{j=1}^{\infty} p_{ij}, i=1,2,\cdots$

边缘密度函数：

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y, -\infty<x<+\infty$$

条件概率：

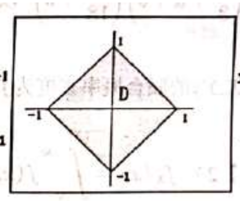
	离散	连续
有值条件	$x_i \in \Omega_{X\mid Y=y_j}$	$x \in \Omega_{X\mid Y=y}$
条件概率	$P\left(X=x_i \mid Y=y_j\right)=\frac{p_{ij}}{p_{\cdot j}}$	$f_{X\mid Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}$
归一性	$\sum_{i=1}^{\infty} P\left(X=x_i \mid Y=y_j\right)=\sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}}=1.$	$\int_{-\infty}^{+\infty} f_{X\mid Y}(x\mid y)\mathrm{d}x=\int_{-\infty}^{+\infty} \frac{f(x,y)}{f_Y(y)}\mathrm{d}x=\frac{\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x}{f_Y(y)}=1.$

> 注意，要是在$Y=y$的条件下x**不存在**则可以直接写0不用计算
>
> 相除后要写上X，Y两者的约束条件($X|Y$就是X关于Y的约束条件)

例：

设二维随机变量 (X,Y) 在 $(1,0),(0,1),(-1,0),(0,-1)$ 四点构成的正方形上服从均匀分布，求条件概率密度

$f_{Y\mid X}(y\mid x)$;



$$f(x,y)=\begin{cases}\frac{1}{2},(x,y)\in D\\ 0,(x,y)\notin D\end{cases} \quad f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases}1+x, -1\leqslant x<0\\ 1-x, \quad 0\leqslant x\leqslant 1\\ \quad 0, \quad 其他\end{cases}$$

当$-1<x<0$时 $f_{Y\mid X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{1}{2(1+x)}, -x-1<y<x+1\\ \quad 0, \quad 其他\end{cases}$

当$0<x<1$时，$f_{Y\mid X}(y\mid x)=\dfrac{f(x,y)}{f_X(x)}=\begin{cases}\dfrac{1}{2(1-x)}, x-1<y<-x+1\\ \quad 0, \quad 其他\end{cases}$

当 $x\leqslant -1$或$x\geqslant 1$时$f_{Y\mid X}(y\mid x)$ 不存在

相互独立的充要条件

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

离散：$p_{ij}=p_i\cdot p_j$ 连续：$f(x,y)=f_X(x)\cdot f_Y(y)$

n个的情况：

$F\left(x_1,x_2,\cdots,x_n\right)=\prod_{i=1}^{n}F_{X_i}\left(x_i\right)$

$P\left(X_1=x_1,X_2=x_2,\cdots,X_n=x_n\right)=\prod_{i=1}^{n}P\left(X_i=x_i\right)$

$f\left(x_1,x_2,\cdots,x_n\right)=\prod_{i=1}^{n}f_{X_i}\left(x_i\right)$

n个相互独立其中任意k个也相互独立

> 1. x_1^2,x_2^2,x_3^2独立
> 2. x_1+x_2,x_3独立
> 3. x_1-x_2,x_1+x_2不一定独立

$g(X,Y)$分布计算方法

离散型：利用联合分布律表格，在表格上标出$g(X,Y)$的取值，对应概率相加，得到$P(Z=z)$

连续型

$$F_z(z)=P(g(X,Y)\leq z)=\iint_{g(x,y)\leq z}f(x,y)\mathrm{d}x\,\mathrm{d}y\rightarrow f_Z(z)=F_Z'(z)$$

卷积

需满足：X**与**Y**相互独立**

$$f_z(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)\mathrm{d}x\quad f_z(z)=\int_{-x}^{+\infty}f_X(z-y)f_Y(y)\mathrm{d}y$$

商的分布

$$z=X/Y\rightarrow f_Z(z)=\int_{-\infty}^{+\infty}|y|f(zy,y)\mathrm{d}y.$$

例题

直接计算型：

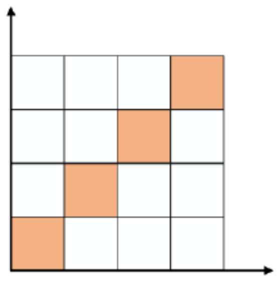
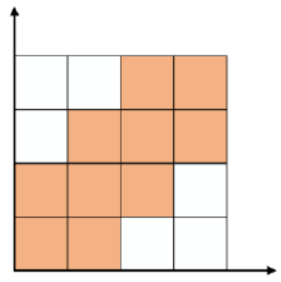
甲乙两人约定在下午 1 时到 2 时之间的任何时刻到达某车站乘公共汽车, 并且分别独立到达车站。这段时间内有四班公共汽车, 他们的开车时刻分别为 $1:15,1:30,1:45,2:00$, 如果他们约定:

（1）见车就乘；（2）最多等一辆车。求他们同乘一辆车的概率。

设甲乙两人到车站的时间分别为 x,y

（1）甲乙两人见车就乘则：

$$P\left\{\frac{1}{4}(n-1)\leqslant x\leqslant\frac{1}{4}n,\frac{1}{4}(n-1)\leqslant y\leqslant\frac{1}{4}n,n=1,2,3,4\right\}=\frac{4\times\left(\frac{1}{4}\right)^2}{1}=\frac{1}{4}$$

第(1)题图: 　　第二题图: 

(2) 最多等一辆车则:

$$P\left\{\frac{1}{4}(n-1)\leqslant x\leqslant\frac{1}{4}(n+1),\frac{1}{4}(n-1)\leqslant y\leqslant\frac{1}{4}(n+1),n=1,2,3\right\}=\frac{1-\frac{1}{4}\times\frac{1}{4}\times 6}{1}=\frac{5}{8}$$

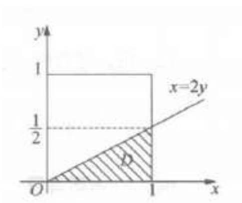
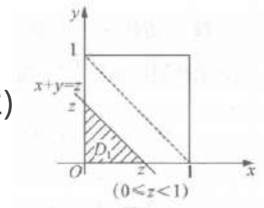
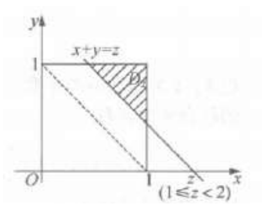
非卷积: （非相互独立）

eg：设二维随机变量（X,，Y）的联合密度函数为

$$f(x,y)=\begin{cases}2-x-y, & 0<x<1,0<y<1,\\ 0, & \text{其他}.\end{cases}$$

(1) 求 $P(X>2Y)$；

(2) 求 $Z=X+Y$ 的密度函数 $f_Z(z)$.

（1）由已知得: 　　(2)  

$$P(X>2Y)=\iint_{x>2y}f(x,y)\mathrm{d}x\,\mathrm{d}y=\int_0^{\frac{1}{2}}\mathrm{d}y\int_{2y}^{1}(2-x-y)\mathrm{d}x=\frac{7}{24}$$

（2）（题目没说相互独立，不能老想着卷积）

先求 Z 的分布函数:

$$F_Z(z)=P(X+Y\leqslant z)=\iint_{x+y\leqslant z}f(x,y)\mathrm{d}x\,\mathrm{d}y$$

$\Omega_Z=(0,2)$, 所以 当 $0\leqslant z<1$ 时,

$$F_Z(z)=\iint_{D_1}f(x,y)\mathrm{d}x\,\mathrm{d}y=\int_0^z\mathrm{d}y\int_0^{z-y}(2-x-y)\mathrm{d}x=z^2-\frac{1}{3}z^3$$

当 $1\leqslant z<2$ 时,

$$F_Z(z)=1-\iint_{D_2}f(x,y)\mathrm{d}x\,\mathrm{d}y=1-\int_{z-1}^{1}\mathrm{d}y\int_{z-y}^{1}(2-x-y)\mathrm{d}x=1-\frac{1}{3}(2-z)^3$$

故 $Z=X+Y$ 的密度函数为

$$f_Z(z)=\begin{cases}2z-z^2,0<z<1,\\(2-z)^2,1<z<2,\\0, \quad others\end{cases}$$

卷积

(必须满足X，Y相互独立才可以，而且X，Y得是连续的)

例题：设随机变量 X, Y **相互独立**，均服从区间（0,1）上的均匀分布, 求: $Z=X+Y$ 的概率密度 $f_Z(z)$

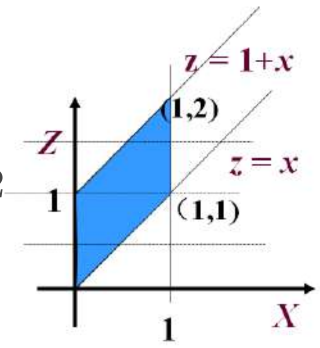
解:

1. X, Y 相互独立, 所以有:

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx$$

2. $f_X(x)f_Y(z-x)$ 非零的区域为:$0 \le x \le 1 \quad 0 \le z-x \le 1$

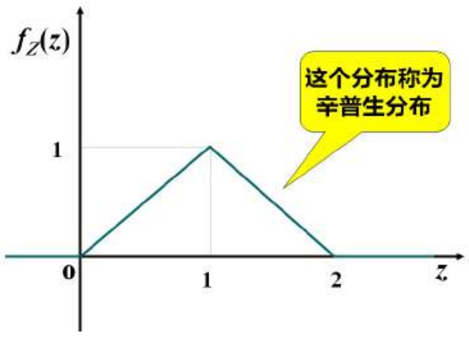
3. $f_Z(z)$ 的非零区域为: $0 \le z \le 2$



4. 在不同的区间段积分的上下限是不相同的(分段求解)

当 $z \le 0$ 或 $z > 2$ 时 $f_Z(z) = 0$
当 $0 < z \le 1$ 时$f_Z(z) = \int_0^z 1dx$
当 $1 < z < 2$ 时$f_Z(z) = \int_{z-1}^1 1dx = 2-z$

综上 $f_Z(z) = \begin{cases} z & 0 < z \le 1 \\ 2-z & 1 < z \le 2 \\ 0 & \text{其他} \end{cases}$



*最大最小值分布

最大最小分布(**条件是相互独立**)

> 最大: $F_U(u) = P(X \le u, Y \le u) = F_X(u)F_Y(u)$
>
> 最小: $F_U(u) = 1 - P(X > u, Y > u) = 1 - (1 - F_X(u))(1 - F_Y(u))$

随机变量的数字特征

期望的定义:

	离散	连续
X	$\sum\limits_{i=1}^{\infty} x_i p_i$	$\int\limits_{-\infty}^{+\infty} x f(x) \mathrm{d}x$
$g(X)$	$\sum\limits_{i=1}^{\infty} g(x_i) p_i$	$\int\limits_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x$
$g(X, Y)$	$\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty} g(x_i, y_j) p_{ij}$	$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x, y) f(x, y) \mathrm{d}x\mathrm{d}y$

期望的性质:

> 1. $E(c) = c$
> 2. $E(kX + c) = kE(X) + c$
> 3. $E(X + Y) = E(X) + E(Y)$(**不考虑独不独立**)
> 4. X, Y为**相互独立**的随机变量$E(XY) = E(X)E(Y)$

例:

民航机场的送客汽车载有 20 名乘客, 从机场出发, 乘客可以在 10 个车站下车, 如果达到某一站点时无顾客下车, 则在该车站不停车, 设随机变量 X 表示停车次数, 假定每个乘客在各车站下车是等可能的, 求平均停车次数。

设 x_i 为汽车在第 i 站的停车次数, $x_i = \begin{cases} 1 \\ 0 \end{cases}, i = 0, 1, 2, \ldots 10$

如果要停车则至少有一人需要下车, **不停车就是每个人都不下车(这个概率好算)**

$P\{x_i=1\}=P\{A_1\cup A_2\cup\ldots\cup A_{20}\}=1-P\left(\overline{A_1}\,\overline{A_2}\cdots\overline{A_{20}}\right)=1-\left(1-\frac{1}{10}\right)^{20}=1-\left(\frac{9}{10}\right)^{20}$

$E(X)=\sum_{i=1}^{10}E\left(X_i\right)=\sum_{i=1}^{10}\left[1-\left(\frac{9}{10}\right)^{20}\right]=10\left[1-\left(\frac{9}{10}\right)^{20}\right]$

标准化随机变量

(类比正态分布)

$$X^*=\frac{X-E(X)}{\sqrt{D(X)}}\Rightarrow E(X^*)=0,D(X^*)=1\qquad \frac{X-\mu}{\sigma}\sim N(0,1)(\text{对比形式})$$

方差

$D(X)=E\left\{[X-E(X)]^2\right\}=E\left(X^2\right)-[E(X)]^2$

算方差直接算出$E(X)$后算$E(X^2)$

性质:

> [1] $D(X)=0\leftrightarrow P(X=c)=1,c=E(X)$
>
> [2] $D(kX+c)=k^2D(X)$
>
> [3] **非相互独立时**
>
> $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)=D(X)+D(Y)\pm 2\rho_{xy}\sqrt{D(X)D(Y)}$
>
> [4] 当X与Y **相互独立**时$D(X\pm Y)=D(X)\pm D(Y)$ (相互独立$\rho=0\to\operatorname{cov}=0$)

协方差

$\operatorname{cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}=E(XY)-E(X)E(Y)$

计算cov **可以把两个参数拆开成已知的参量，一步步慢慢拆开**

关系式：$\operatorname{cov}(X,Y)=\rho_{xy}\sqrt{D(X)D(Y)}$

性质:

> [1]$\operatorname{cov}(X,c)=0$
>
> [2]$\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$
>
> [3]$\operatorname{cov}(kX,lY)=kl\operatorname{cov}(X,Y)$
>
> [4]$\operatorname{cov}\left(X_1+X_2,Y\right)=\operatorname{cov}\left(X_1,Y\right)+\operatorname{cov}\left(X_2,Y\right)$
>
> [5]$\operatorname{cov}(X,X)=D(X)$ (根据定义)

协方差矩阵:

设 n维随机变量 $(X_1,X_2,\cdots,X_n)$ 的协方差均存在，称矩阵

$$\boldsymbol{C}=\begin{bmatrix}C_{11}&C_{12}&\cdots&C_{1n}\\C_{21}&C_{22}&\cdots&C_{2n}\\\vdots&\vdots&&\vdots\\C_{n1}&C_{n2}&\cdots&C_{nn}\end{bmatrix}\text{为}n\text{维随机变量 }(X_1,X_2,\cdots,X_n)\text{的协方差矩阵.}$$

协方差矩阵 $\boldsymbol{C}=(C_{ij})_{n\times n}$ 满足:

$(1)C_{ii}=D\left(X_i\right),\quad i=1,2,\cdots,n;$

$(2)C_{ij}=C_{ji},\quad i\neq j,\quad i,j=1,2,\cdots,n.$

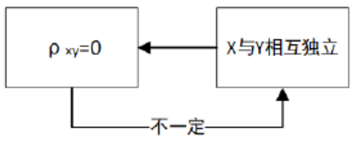
协方差矩阵是一个对称矩阵.

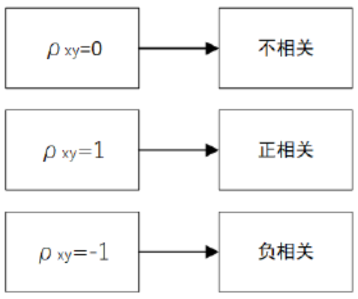
(X,Y)的协方差矩阵：$C=\begin{bmatrix}C_{xx} & C_{xy}\\ C_{yx} & C_{yy}\end{bmatrix}=\begin{bmatrix}D(X) & \rho_{xy}\sqrt{D(X)D(Y)}\\ \rho_{xy}\sqrt{D(X)D(Y)} & D(Y)\end{bmatrix}$

相关系数

$$\rho(X,Y)=E\left[\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{Y-E(Y)}{\sqrt{D(Y)}}\right]=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}=E(X^*Y^*)=\operatorname{cov}(X^*,Y^*)$$

性质：

> [1]$|\rho_{xy}|\le 1$
>
> [2]$|\rho_{XY}|=1\Leftrightarrow P(Y=aX+b)=1$
>
> [3] 
>
> 需要满足相关系数存在的条件

若ρ_{xy}存在 

方差，协方差，相关系数

相关系数跟协方差有关，影响随机变量相加后的方差，**均值不影响**

$D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)=D(X)+D(Y)\pm 2\rho_{xy}\sqrt{D(X)D(Y)}$

例：

设随机变量 (X,Y) 服从二维正态分布, 其中 $X\sim N(1,1), Y\sim N(-2,1)$, 并且 X 和 Y 的相关系数为 0.5, 令 $Z=2X+Y$, 试求: $E(Z), D(Z)$ 和概率密度 $f_Z(z)$

$E(Z)=E(2X+Y)=2E(X)+E(Y)=2\times 1-2=0$(均值不影响)

$$\begin{aligned}D(Z)&=D(2X+Y)=4D(X)+D(Y)+2\times 2\operatorname{cov}(X,Y)\\&=4D(X)+D(Y)+4\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}=4\times 1+1+4\times 0.5=7\end{aligned}$$

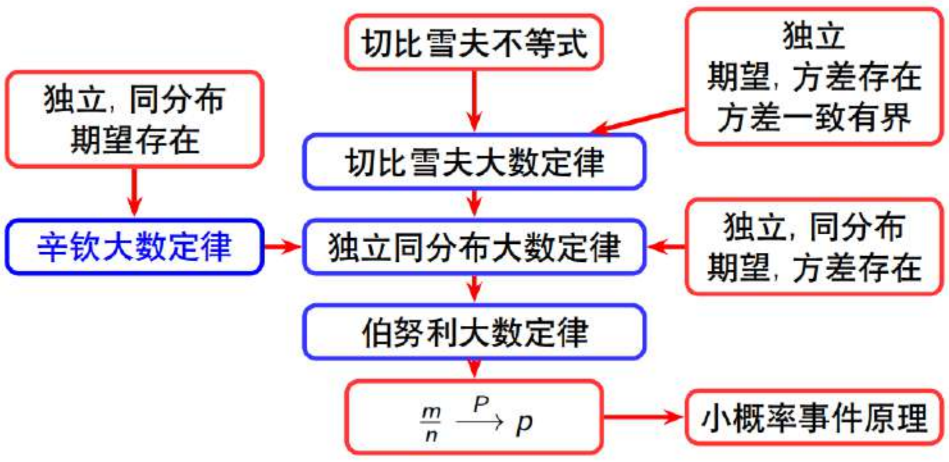
(X,Y) 服从二维正态分布, 线性组合 $Z=2X+Y$ 也服从正态分布

$Z\sim N(0,7)\to f_Z(z)=\frac{1}{\sqrt{2\pi}\sqrt{7}}e^{-\frac{z^2}{14}}, z\in R$

*矩

k阶原点矩	$\gamma_k=E\left(X^k\right)$
k阶绝对原点矩.	$\alpha_k=E\left(\lvert X\rvert^k\right)$
k阶中心矩	$\mu_k=E\left\{[X-E(X)]^k\right\}(\mu_2=\gamma_2-\gamma_1^2)$
k阶绝对中心矩	$\beta_k=E\left\{\lvert X-E(X)\rvert^k\right\}$

大数定理及中心极限定理



依概率收敛 $\lim_{n\to\infty} P\left(|X_n - c| < \varepsilon\right) = 1 \Longrightarrow X_n \xrightarrow{P} c$

切比雪夫不等式

$$P(|X - E(X)| \leq \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$$

没提到用切比雪夫的一般用中心极限定理

例题：

设噪声电压 $X_1, X_2, \cdots X_{100}$ 相互独立且都服从区间(0,6)上的均匀分布，用切比雪夫不等式估计叠加后的总噪声电压 $Y = \sum^{100} X_k$ 在 260 到 340 之间的概率。

$E_i(X), D_i(X)$算出，得到$E(Y), D(Y)$，其中$E(X)$恰好为300

$$P\{260 < Y < 340\} = P\{|Y - E(Y)| < 40\} \geq 1 - \frac{300}{40^2}$$

计算方法：先算个体X_i的期望和方差，确认$E(X)$、ε，然后代入求解

大数定律

条件：**随机变量序列$X_1, X_2, \cdots X_n, \cdots$的期望均存在**

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right\} = 1 \rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) \xrightarrow{P} 0$$

算术平均值法则、用频率估计概率、小概率原理(3σ) 68.3 95.4 99.7

切比雪夫大数定律

$X_1, X_2, \cdots$相互独立且**方差一致有界**　存在C使得$D(X_k) < C, k = 1, 2, \ldots$

独立同分布大数定理

$X_1, X_2, \cdots$ 独立同分布且期望、方差存在

对任意$\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1 \rightarrow \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - EX^2\right| < \varepsilon\right\} = 1$$

伯努利大数定理

$X_1, X_2, \cdots$**独立同分布**

对任意$\varepsilon > 0$，m为n重伯努利实验中发生的次数，满足$\lim_{n\to\infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1$

中心极限定理

定义:

独立随机变量之和的极限分布为**正态分布**(n很大的条件下)

设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，服从中心极限定理

则$\sum\limits_{k=1}^{n} X_k$的标准化随机变量的极限分布为标准正态分布;

独立同分布中心极限定理

$X_1, X_2, \cdots$ **独立同分布**且期望和方差存在，$E(X_k) = \mu, D(X_k) = \sigma^2$

则: $\sum\limits_{i=1}^{n} X_i$近似服从$N\left(n\mu, n\sigma^2\right)$ $\lim\limits_{n \to \infty} P\left(\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x).$

例:

某商店负责供应某地区 10000 人的商品, 某种商品在一段时间内每人购买一件的概率为 0.6, 假定在这一段时间内各人购买与否相互独立, 问商店至少预备多少件这种商品, 才能以 99% 的概率保证不会脱销。(假定该商品在某一段时间内每人最多可以购买一件) $\Phi(2.33) = 0.99$

解: (题目套话)

> 每个人是否购买X_i相互独立且同(0-1)分布，服从中心极限定理，设$X = \sum X_i$表示该段时间购买商品的人数，满足 $X \sim B(10000, 0.6)$
>
> (写出$E(X_i), D(X_i), E(X), D(X)$)
>
> 假设应备预N件商品，由中心极限定理可得:

$$P\{X \leqslant N\} = P\left\{\frac{X - 10000 \times 0.6}{\sqrt{10000 \times 0.6 \times 0.4}} \leqslant \frac{N - 10000 \times 0.6}{\sqrt{10000 \times 0.6 \times 0.4}}\right\} \approx \Phi\left(\frac{N - 10000 \times 0.6}{\sqrt{10000 \times 0.6 \times 0.4}}\right) \geqslant 0.99$$

因$\Phi(2.33) = 0.99 \Rightarrow \frac{N - 10000 \times 0.6}{\sqrt{10000 \times 0.6 \times 0.4}} \geqslant 2.33 \Rightarrow N \geqslant 6114.1$, 故应预备 6115 件商品.

流程:

> 1. **先计算一次的**$E(X_i), D(X_i)$
> 2. 满足独立同分布然后累加得到$\sum X_i$的$E(X)$和$D(X)$
> 3. 满足中心极限定理，代入公式
> 4. 然后整理右边的常数项，得到$P = \Phi(x)$

数理统计的基本概念

基本概念:

- 总体 : 研究对象的单位元素所组成的集合。 (是随机变量)
- 个体: 组成总体的每个单位元素。
- 样本: 等概率从总体中抽出的一部分个体，**与总体同分布且相互独立**

  样本总体同分布，各样本之间相互独立
- 若总体 X 的分布函数为$F(x)$, 则样本 $X_1, X_2, \cdots, X_n$的联合分布函数为:

$$F\left(x_1, x_2, \cdots, x_n\right) = P\left\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\right\} = \prod_{i=1}^{n} P\left\{X_i \leqslant x_i\right\}$$

  联合概率密度: $f\left(x_1, x_2, \cdots, x_n\right) = \prod\limits_{i=1}^{n} f\left(x_i\right)$

例题:

设电子元件的寿命(小时)服从参数λ=0.0015的指数分布，今测试6个元件，记录下它们各自失效的时间。问:这里的总体和样本分别是什么?

总体为电子元件的寿命X，**样本**是测试的六个原件的寿命$X_1, X_2 \cdots X_6$

- **总体分布**是指数量指标X的分布
- **统计量**：不包含未知参数的样本的函数

  $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个样本，若样本函数 $g(X_1, X_2, \cdots, X_n)$ 中不含任何未知参数，则称 $g(X_1, X_2, \cdots, X_n)$ 为一个统计量

常用统计量

> **样本均值**：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$
>
> 样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$，$S^2$ 是样本 $X_1, X_2, \cdots, X_n$ 的分散程度的合理的刻画
>
> 样本 k 阶原点矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad A_1=\bar{X}$
>
> 样本 k 阶中心矩.：$M_k=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^k \quad M_2=\frac{n-1}{n}S^2=A_2-A_1^2$

A_k、M_k统称样本矩(**样本矩是随机变量, 总体矩是数值**)

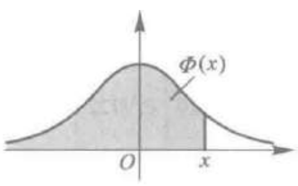
$\bar{X}, S^2, A_k, M_k \longleftrightarrow \bar{x}, s^2, a_k, m_k$

大写统计量，小写统计值（将样本观测值$x_1, x_2, \ldots, x_n$代入统计量的公式，得到统计值）

四种常用的分布

标准正态分布

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}, x\in R$$

上侧分位数 $u_\alpha(0<\alpha<1)$ 满足 

$$P\{X>u_\alpha\}=\int_{u_\alpha}^{+\infty}f(x)dx=\alpha \qquad \boldsymbol{\Phi(u_\alpha)=1-\alpha, u_{1-\alpha}=-u_\alpha}$$

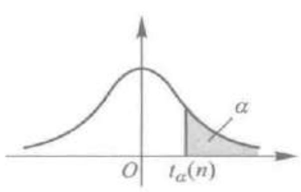
Gamma函数

$$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx, \alpha>0$$

主要性质：$\Gamma(1)=1, \quad \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}, \quad \Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$

抽样分布（三种）

t分布



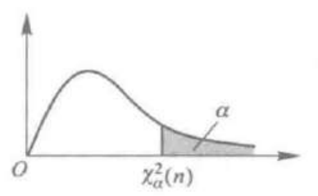
定义：设随机变量 X, Y 相互独立, $X\sim N(0,1)$, $Y\sim\chi^2(n)$, 则（n为自由度）

$$T=\frac{X}{\sqrt{Y/n}} \sim t(n) \qquad P\{T>t_\alpha\}=\alpha$$

性质:

> 1. 关于纵轴对称: $t_{1-\alpha}(n)=-t_\alpha(n)$
> 2. n较大时 $(n>30)$, $t_\alpha(n)\approx u_\alpha$

χ^2 （卡方）分布



定义：设n个相互独立并且都服从正态分布 $N(0,1)$ 的随机变量 $X_1,X_2,\cdots,X_n$, 记

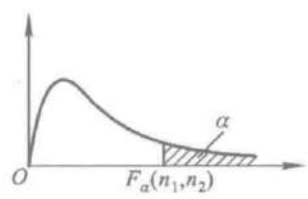
$$\chi^2=\sum_{i=1}^{n}X_i^2 \sim \chi^2(n) \qquad P\left\{\chi^2>\chi^2_\alpha\right\}=\alpha$$

则称随机变量 $\mathcal{X}^2$ 服从自由度为n的 χ^2 分布.自由度是右端所包含的独立变量的个数.

性质:

> 1. (数字特征) 设 $\chi^2\sim\chi^2(n)$, 则有$E\left(\chi^2\right)=n, \quad D\left(\chi^2\right)=2n$
> 2. (可加性)设 Y_1、Y_2相互独立,且 $Y_1\sim\chi^2\left(n_1\right), Y_1\sim\chi^2\left(n_2\right)$, 则$Y_1+Y_2\sim\chi^2\left(n_1+n_2\right)$.
> 3. 当n足够大时$\chi^2_\alpha(n)\approx n+u_\alpha\sqrt{2n}$

F分布



定义：设随机变量 X,Y 相互独立, $X\sim\chi^2\left(n_1\right), Y\sim\chi^2\left(n_2\right)$, 则

$$F=\frac{X/n_1}{Y/n_2}\sim F\left(n_1,n_2\right) \qquad P\left\{F>F_\alpha\right\}=\alpha$$

即随机变量 F 服从第一自由度为 n_1, 第二自由度为 n_2 的 F 分布.

推论

> $\frac{1}{F(n_1,n_2)}\sim F\left(n_2,n_1\right)$
>
> $F_{1-\alpha}\left(n_1,n_2\right)=\frac{1}{F_\alpha(n_2,n_1)}$

证明(例):

X,Y 相互独立, 且 $X\sim\chi^2(12), Y\sim\chi^2(9)$, 则 $F=\frac{X/12}{Y/9}\sim F(12,9)$,

令 $F'=\frac{1}{F}=\frac{Y/9}{X/12}\sim F(9,12)$,$P(F>\partial)=P\left\{\frac{1}{F}<\frac{1}{\partial}\right\}=P\left(F'<\frac{1}{\partial}\right)=0.95$,

则$P\left(F'>\frac{1}{\partial}\right)=0.05$, 那么有 $\frac{1}{\partial}=F_{0.05}(9,12)$, 则 $\partial=\frac{1}{F_{0.05}(9,12)}=\frac{1}{2.8}$

*自由度的解释

(总体均值已知) （1）式: $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

(总体均值未知) （2）式: $\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)}{\sigma^2}=\frac{(n-1)}{\sigma^2}S^2\sim\chi^2(n-1)$

问题: (1) 式与 (2) 式的自由度为何不同?

(1) 式是由 n 个相互独立的随机变量 $\frac{X_i-\mu}{\sigma}$ 的平方和构成, 因含有 n 个相互独立的随机变量, 故自由度为 n;

(2) 式虽然也是 n 个随机变量 $\frac{X_i-\bar{X}}{\sigma}$ 的平方和, 但这 n 个随机变量中**含有一个线性关系:**

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)=\sum_{i=1}^{n}X_i-n\bar{X}=n\bar{X}-n\bar{X}=0$$

即 n 个随机变量受到 $X_1+X_2+\cdots+X_n=n\bar{X}$ 这一条件的约束，因而只有 $n-1$ 个随机变量可以独立自由变动，所以其自由度为 $n-1$.

参数估计

矩估计法

基本思想：用**样本k阶矩作为总体k阶矩**的估计量,建立含有待估参数的方程，从而解出待估参数

样本矩是随机变量，而总体矩是数值（为估计量）$E(X^2)=E^2(X)+D(X)$

步骤：

求出总体的期望$E(X)$ $\rightarrow$ 求出样本的平均值$\bar{X}$ $\rightarrow$ 用$\bar{X}$代替$E(X)$ $\rightarrow$估计参数的矩估计值

例：设总体$X\sim N(\mu,\sigma^2)$，X的密度函数为$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in R$，其中,$\mu,\sigma$未知，$X_1,X_2,\cdots,X_n$为来自$X$ 的样本

(1)求σ^2的矩估计　　(2)求$E(X^2)$的极大似然估计

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

$E\left(X_i^2\right)=D\left(X_i\right)+\left[E\left(X_i\right)\right]^2=\sigma^2+\mu^2,E\left(\bar{X}^2\right)=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2$

故 $E\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)=\frac{n}{n-1}\left(\sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2\right)=\sigma^2$

故 σ^2 的一个矩估计为 $\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}$

(2)欲求$E(X^2)$的极大似然估计，因为$E(X^2)=E^2(X)+D(X)$，$E(X)$已知为μ，需求$D(X)$的极大似然估计值进而得到$E(X^2)$的

极大似然估计法

定义

设总体分布形式已知, 含未知参数 $\theta_1,\theta_2,\cdots,\theta_m$,$X_1,X_2,\cdots,X_n$是来自该总体的样本,

$x_1,x_2,\cdots,x_n$是相应的样本观测值, 相当于事件$\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$已经发

生.把已经发生的事件, 看成最可能出现的事件, 即认为它具有最大的概率.

满足

$$L\left(x_1,x_2,\cdots,x_n;\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m\right)=\max_{\theta_1,\theta_2,\cdots,\theta_m}L\left(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_m\right),$$

则称 $\hat{\theta}_i$ 为 θ_i的**最大似然估计值**，相应的估计量 $\hat{\theta}=\hat{\theta}\left(X_1,X_2,\ldots,X_n\right)$称为参数$\theta$的**最大似然估计量.**

步骤:

写出似然函数$L\left(x_1,x_2,\ldots,x_n;\theta\right)=\prod_{i=1}^{n}p\left(x_i;\theta_1,\theta_2,\ldots,\theta_m\right)$ $\rightarrow$取对数 $\ln L$ $\rightarrow$ 求导(对待估计的量)$\rightarrow$ 令导数为0 得到待估计量的表达式 $\rightarrow$ 极大似然估计为$\hat{\theta}=$

例题

从一批产品中随机抽取 n 个进行检测, 发现其中次品个数为 m 个, 试用极大似然估计法估计该批产品的次品率.(表示方法)

设次品率 p, 则 $0<p<1$, $X_i=\begin{cases}1, & \text{第 } i \text{ 次取得次品} \\ 0, & \text{第 } i \text{ 次取得合格品}\end{cases}, i=1,2,\ldots,n$

则 $P(X_i=1)=p, P(X_i=0)=1-p$

设$x_1,x_2,\cdots x_n$为样品$X_1,X_2,\cdots X_n$的观测值(这句话必须要写)，似然函数为:

$$L(p)=\prod_{i=1}^{n} P\{X_i=x_i\}=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}=p^m(1-p)^{n-m}$$

$$\ln L(p)=m\ln p+(n-m)\ln(1-p)$$

似然方程为:

$$\frac{d\ln L(p)}{dp}=\frac{m}{p}-\frac{n-m}{1-p}=\frac{m-np}{p(1-p)}=0$$

例题：（不能直接求出0）

设某种元件的使用寿命总体 X 的概率密度为

$$f(x,\theta)=\begin{cases}2\mathrm{e}^{-2(x-\theta)}, & x>\theta, \\ 0, & x\leqslant\theta,\end{cases}$$

其中 $\theta>0$ 为未知参数. 又设 $x_1,x_2,\cdots,x_n$ 是 X 的一组样本观测值, 求参数 θ 的极大似然**估计值.**

似然函数 $L(\theta)=\begin{cases}2^n\exp\left\{-2\sum\limits_{i=1}^{n}(x_i-\theta)\right\}, & x_{(1)}>\theta, \\ 0, & x_{(1)}\leqslant\theta,\end{cases}$

其中 $x_{(1)}=\min_{1\leqslant i\leqslant n} x_i$, 此处与前例不同, $L(\theta)$ 在 $\theta=x_{(1)}$ **处间断**, 因此只能直接求函数 $L(\theta)$ 的最大值点.

注意 $L(\theta)\geqslant 0$，且当 $\theta<x_{(1)}$ 时, $L(\theta)=2^n\exp\left\{-2\sum\limits_{i=1}^{n}(x_i-\theta)\right\}$ 随 θ 递增而递增,

($\theta=x_{(1)}$时指数项为1，1为可获得的最大值)，因而当 $\theta=x_{(1)}$ 时, $L(\theta)$ 达到最大.

区分量和值: $\hat{\theta}=X_{(1)}$ 是 θ 的极大似然估计量, $\hat{\theta}=x_{(1)}$ 是 θ 的极大似然估计值.

估计量的优良性准则

$\bar{X},S^2$是μ,σ^2的无偏、有效、相合估计(**μ未知的情况下**)

无偏性

定义：设$\hat{\theta}(X_1,X_2,\ldots,X_n)$是未知参数$\theta$的估计量(样本)，若$E(\hat{\theta})=\theta$，则称$\hat{\theta}$ 为θ的无偏估计.

X_1为样本，则$E(X_1)=E(X),D(X_1)=D(X)$，**样本分布和总体分布相同**

$E(\bar{X})=\mu \quad D(\bar{X})=\frac{1}{n^2}D(\sum\limits_{i=1}^{n}X_i)=\frac{D(X)}{n}, \qquad E^2(\bar{X})=E^2(X)$

> 若μ未知:
>
> $$E(S^2)=E(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2)=\sigma^2, E(M_2)=E(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2)=\frac{n-1}{n}E(S^2)\neq\sigma^2$$
>
> 已知μ: $E(\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2)=\sigma^2$

给出证明: $E(S^2)=\sigma^2$

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

$E\left(X_i^2\right)=D\left(X_i\right)+\left[E\left(X_i\right)\right]^2=\sigma^2+\mu^2 E\left(\bar{X}^2\right)=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2$

故 $E\left(\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}\right)^2\right)=\frac{n}{n-1}\left(\sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2\right)=\sigma^2$

例题:

设总体 $X \sim N\left(\mu, \sigma^2\right), X_1, X_2, \cdots, X_n$ 为其样本, 试求常数 C 使 $C \sum_{i=1}^{n-1}\left(X_{i+1}-X_i\right)^2$ 为 σ 的 无偏估计量:

$X_1, X_2, \cdots, X_n$**相互独立且同分布** ，有$E(X_i)=\mu, D(X_i)=\sigma^2$，$E(X_i^2)=\mu^2+\sigma^2$

$E[(X_{i+1}-X_i)^2]=E(X_{i+1}^2)+E(X_i^2)-2E(X_iX_{i+1})=2\sigma^2$

则$E[C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2]=2C(n-1)\sigma^2$，若等于$\sigma^2, C=\frac{1}{2(n-1)}$

有效性

定义：设 $\hat{\theta}_1\left(X_1, X_2, \ldots, X_n\right)$ 和$\hat{\theta}_2\left(X_1, X_2, \ldots, X_n\right)$ 是未知参数θ的两个无偏估计量,

若对θ的所有可能取值都有$D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$则称 $\hat{\theta}_1$ 比$\hat{\theta}_2$有效(优效).

设$\hat{\theta}_0$是θ的无偏估计，如果对θ的任何一个无偏估计量$\hat{\theta}$都有$D\left(\hat{\theta}_0\right) \leq D\left(\hat{\theta}\right)$

则称$\hat{\theta}_0$为θ的**最小方差无偏估计**

例题：假设总体服从正态分布 $N\left(\mu, \sigma^2\right)$, 抽样后构造统计量对 μ 进行估计, 下面哪些估计量是无偏的, 哪个估计量最有效, 为什么? $(n>2)$

(1) $\bar{X}$ (2) X_n (3) $\frac{X_1+X_2}{2}$ (4) X_1+X_2

$E(\bar{X})=E(X_n)=E(\frac{X_1+X_2}{2})=\mu$，(1),(2),(3)为无偏估计

$D(\bar{X})=\frac{\sigma^2}{n} \quad D(X_n)=\sigma^2 \quad D(\frac{X_1+X_2}{2})=\frac{\sigma^2}{2}$，所以(1)最有效

抽样分布定理

设$X_1, X_2, \cdots, X_n$是正态总体$X\sim N(\mu, \sigma^2)$的样本则:

> 1. $\bar{X}$ **与** S^2 **相互独立**;
> 2. $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$； $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$
> 3. $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$;
> 4. $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$
> 5. $\frac{1}{\sigma^2}\sum_{i=1}^n(X_i-\mu)^2 \sim \chi^2(n)$

区间估计

$\hat{\theta}_1=\hat{\theta}_1\left(X_1, \ldots, X_n\right), \quad \hat{\theta}_2=\hat{\theta}_2\left(X_1, \ldots, X_n\right)$对于给定的实数$\alpha(0<\alpha<1)$, 满足

$$P\left\{\hat{\theta}_1\left(X_1, \ldots, X_n\right) \leq \theta \leq \hat{\theta}_2\left(X_1, \ldots, X_n\right)\right\}=1-\alpha$$

称随机区间 $\left[\hat{\theta}_1, \hat{\theta}_2\right]$ 为θ 的**置信度**为 $1-\alpha$的置信区间 $1-\alpha$又称**置信水平或置信概率**

优良性准则： $\bar{X} \to \mu$， $S^2 \to \sigma^2$

区间估计的枢轴变量法

步骤：

1. 选取待估参数θ的估计量
2. 建立枢轴变量W
3. 确定W的分布.W 通常具有经典分布
4. 根据W的分布，建立概率等式：$P\left\{w_{1-\alpha/2} \le W \le w_{\alpha/2}\right\} = 1-\alpha$
5. 改写不等式$P\{A \le \theta \le B\} = 1-\alpha$

一个正态总体参数的置信区间

抽样分布定理　　看评估啥留啥在不等式中间　**置信度为**$1-\alpha$！！！

被估参数	条件	枢轴变量	满足条件
μ	σ^2 已知	$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$P\{\lvert U\rvert \le u_{\alpha/2}\}=1-\alpha$
μ	σ^2 未知	$T=\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$P\{\lvert T\rvert \le t_{\alpha/2}(n)\}=1-\alpha$
σ^2	μ 已知	$\chi^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$	$P\{\chi^2_{1-\alpha/2}(n) \le \chi^2 \le \chi^2_{\alpha/2}(n)\}=1-\alpha$
σ^2	μ 未知	$\chi^2=\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$P\{\chi^2_{1-\alpha/2}(n-1) \le \chi^2 \le \chi^2_{\alpha/2}(n-1)\}=1-\alpha$

两个正态总体的区间估计

置信度为$1-\alpha$！！！　若$\mu_1-\mu_2$的置信下限大于零,则可认为$\mu_1>\mu_2$；

被估参数	条件	枢轴变量
$\mu_1-\mu_2$	σ_1^2 已知 σ_2^2 已知	$U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
$\mu_1-\mu_2$	σ_1^2 未知 σ_2^2 未知 $\sigma_1^2=\sigma_2^2$	$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t\left(n_1+n_2-2\right)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1 未知 μ_2 未知	$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2} \sim F\left(n_1-1,n_2-1\right)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1 已知 μ_2 已知	$F=\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2/\sigma_1^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i-\mu_2)^2/\sigma_2^2} \sim F\left(n_1,n_2\right)$

接受域：$F_{1-\alpha/2}(n_1-1,n_2-1) \le F \le F_{\alpha/2}(n_1-1,n_2-1)$

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2,\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2\right) \quad V=\frac{1}{\sigma^2}\left[(n_1-1)S_1^2+(n_2-1)S_2^2\right] \sim \chi^2\left(n_1+n_2-2\right)$$

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t\left(n_1+n_2-2\right) \qquad S_w=\sqrt{\frac{1}{m+n-2}\left[(m-1)S_X^2+(n-1)S_Y^2\right]}$$

假设检验

基本概念：

1. **原假设H_0 和对立假设(备择假设)H_1**

   备择假设的内容与原假设对立，在假设检验问题中，否定原假设后则选择对立假设的结论.

2. **检验统计量**

   要求：检验统计量的取值范围和变化情况，能反映H_0与 H_1 所描述的内容, 并且当H_0成立时,

   能够确定检验统计量的概率分布

3. **显著性水平**

   小概率事件的标准α称为假设检验的显著性水平，通常, 取α= 0.05 或 0.01。

4. **接受域和拒绝域**

   使原假设H_0 得以接受的检验统计量的取值区域称为检验的接受域；

   使原假设H_0 被拒绝的检验统计量取值的区域称为检验的拒绝域。

两类错误：

第一类错误概率 (弃真概率) : 原假设 H_0成立, 而最终错误地拒绝H_0的概率,（就是置信度α）

第二类错误概率 (采伪概率) : 原假设 H_0不成立, 而错误地接受它的概率（不考）β

(即认为在接受域内，但实际不在接受域内)

例题:

在正态总体 $N(\mu, 1)$ 中抽取了 100 个样品, 计算得 $\bar{x}=5.2$.

(1)试检验假设 $H_0: \mu=5 \leftrightarrow H_1: \mu<5($ 取显著性水平 $\alpha=0.01)$;

(2) 计算上述检验在 $\mu=4.8$ 时犯第二类错误的概率.

解

(1) 这是一个单正态总体 $\sigma^2=1$ 已知，关于均值的左侧检验问题, 取显著性水平 $\alpha=0.01$, 在原假设成立的情况下，拒绝域:

$$W=\left\{\frac{\left(\bar{X}-5\right)}{\sigma / \sqrt{n}}<-u_\alpha\right\}=\left\{10(\bar{X}-5)<-u_{0.01}\right\},$$

计算可得 $10(\bar{x}-5)=10 \times(5.2-5)=2>-u_{0.01}=-2.326$, 即样本观测值没有落在拒绝域内, 因此不能拒绝原假设.

(2) 犯第二类错误的概率为当原假设 H_0 不成立而 H_1 是成立时,

$$\begin{aligned} P\left(10(\bar{X}-5) \geqslant-u_{0.01} \mid \mu=4.8\right) & =P\left(10(\bar{X}-4.8) \geqslant-u_{0.01}+2 \mid \mu=4.8\right) \\ & \approx 1-\Phi(-0.326) \approx 0.6278 .\end{aligned}$$

基本步骤

(1)根据实际问题提出原假设H_0和备择假设H_1

(2)建立**检验统计量**:

当H_0成立时，检验统计量为：（能够确定检验统计量的概率分布）;

参数的假设形式：（**一般都是双侧检验**）

1. $\mathrm{H}_0: \theta=\theta_0; \quad \mathrm{H}_1: \theta \neq \theta_0$，**双侧检验**
2. $\mathrm{H}_0: \theta \leq \theta_0; \quad \mathrm{H}_1: \theta>\theta_0$，单侧检验
3. $\mathrm{H}_0: \theta \geq \theta_0; \quad \mathrm{H}_1: \theta<\theta_0$，单侧检验

形如 2、3 的单侧假设检验问题需要作一个转换

2^*：$H_0: \theta=\theta_0; \quad H_1: \theta>\theta_0$ 3^*：$H_0: \theta=\theta_0; \quad H_1: \theta<\theta_0$ 与2、3具有相同的拒绝域

(3)写出H的**拒绝域:**

选定显著性水平α，依据检验统计量的分布和H_0的内容，确定H_0的拒绝域;

(4)决策:

根据检验统计量的**观测值**是否落在拒绝域，确定拒绝或接受H_0

正态总体参数的假设检验

如果备择假设是单边的要根据大于还是小于还有检验假设函数来写拒绝域，看变量分子（分母）偏大还是偏小，检验假设函数的值会偏大还是偏小

(备择假设>，那么拒绝域就是>)

单正态总体参数的假设检验

根据备择假设写拒绝域，判断在不在拒绝域

检测统计量不能写成枢轴变量！！！

显著性水平为α！！！，当H_0成立时

检验参数	条件	原假设与备择假设	检验统计量	拒绝域
μ	σ^2 已知	$H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$ $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu>\mu_0$ $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu<\mu_0$	$U=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$	$\lvert U\rvert>u_{\frac{\alpha}{2}}$ $U>u_\alpha$ $U<-u_\alpha$
μ	σ^2 未知	$H_0:\mu=\mu_0 \leftrightarrow H_1:\mu\neq\mu_0$ $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu>\mu_0$ $H_0:\mu=\mu_0 \leftrightarrow H_1:\mu<\mu_0$	$T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\lvert t\rvert>t_{\frac{\alpha}{2}}(n-1)$ $t>t_\alpha(n-1)$ $t<-t_\alpha(n-1)$
σ^2	μ 已知	$H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2\neq\sigma_0^2$ $H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2>\sigma_0^2$ $H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2<\sigma_0^2$	$\chi^2=\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2}\sim\chi^2(n)$	$\chi^2>\chi^2_{\alpha/2}(n)$ 或$\chi^2<\chi^2_{1-\alpha/2}(n)$ $\chi^2>\chi^2_\alpha(n)$ $\chi^2<\chi^2_{1-\alpha}(n)$
σ^2	μ 未知	$H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2\neq\sigma_0^2$ $H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2>\sigma_0^2$ $H_0:\sigma^2=\sigma_0^2 \leftrightarrow H_1:\sigma^2<\sigma_0^2$	$\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$	$\chi^2>\chi^2_{\alpha/2}(n-1)$ 或$\chi^2<\chi^2_{1-\alpha/2}(n-1)$ $\chi^2>\chi^2_\alpha(n-1)$ $\chi^2<\chi^2_{1-\alpha}(n-1)$

两个正态总体参数的假设检验

检验参数	条件	原假设与备择假设	检验统计量
$\mu_1-\mu_2$	σ_1^2 已知 σ_2^2 已知	$H_0:\mu_1=\mu_2 \leftrightarrow H_1:\mu_1\neq\mu_2$	$U=\frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$
$\mu_1-\mu_2$	σ_1^2 未知 σ_2^2 未知 $\sigma_1^2=\sigma_2^2$	$H_0:\mu_1=\mu_2 \leftrightarrow H_1:\mu_1\neq\mu_2$	$T=\frac{(\bar{X}-\bar{Y})}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1 未知 μ_2 未知	$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$	$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1 已知 μ_2 已知	$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$	$F=\frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}\sim F(n_1,n_2)$

U、T检验法的拒绝域写法同上

F检验法拒绝域：检验统计量观测值$f>F_{\alpha/2}$或$f<F_{1-\alpha/2}$

$$S_w=\sqrt{\frac{1}{m+n-2}\left[(m-1)S_X^2+(n-1)S_Y^2\right]}$$

区间估计和假设检验的区别

	区间估计	**假设检验**
函数	枢轴函数	检验统计量
方法核心	求待定量的区间	假设待定量已知看是否满足条件
	注意区分**置信区间**和**单侧置信上限**	注意区分**单双侧检验**
区别α大小	题目给**置信度**（0.95），一般用$1-\alpha$=置信度	题目给**显著性水平**α（0.05/0.01）
	导致用的α是小数字(0.05/0.01)	

考虑代入的假设比真实的偏大还是偏小，然后想拒绝域（出来的统计量是高于上限还是低于下限）

区间估计

置信区间

1.设某地区男、女身高 X, Y 均**服从正态分布且方差相等**, 随机抽取成人男、女各 100 名, 测量 并计算得男子身高 $\bar{x}=1.71\ \mathrm{m}, s_1=0.035\ \mathrm{m}$,女子身高 $\bar{y}=1.67\ \mathrm{m}, s_2=0.038\ \mathrm{m}$。求男、女平 均身高之差的**置信度 0.95 的置信区间。**

解

设 $X\sim N\left(\mu_1,\sigma^2\right), Y\sim N\left(\mu_2,\sigma^2\right)$, X, Y 相互独立，要估计$\mu_1-\mu_2$, **两总体方差未知但相等**,

选取枢轴变量$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$

其中$S_w=\sqrt{\frac{1}{n_1+n_2-2}\left[(n_1-1)S_1^2+(n_2-1)S_2^2\right]}$

令 $P\left\{|T|\leqslant t_{\frac{\alpha}{2}}\left(n_1+n_2-2\right)\right\}=1-\alpha$ **(这里$\alpha=0.05$)**

得$\mu_1-\mu_2$的置信区间

$$\left[\bar{X}-\bar{Y}-t_{\frac{\alpha}{2}}\left(n_1+n_2-2\right)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \bar{X}-\bar{Y}+t_{\frac{\alpha}{2}}\left(n_1+n_2-2\right)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

单侧置信上（下）限

2.从某型号的一批电子管中抽出容量为 10 的样本做寿命试验, 算得 $s=45$(小时), 设整批电子管的寿命**服从正态分布**, 试求这批电子管寿命标准差的**单侧置信上限**（置信度为 0.95）。

设电子管寿命$X\sim N\left(\mu,\sigma^2\right)$, 要估计$\sigma^2$, μ未知,

选取枢轴变量$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

令 $P\left\{\chi^2\geqslant\chi^2_{1-\alpha}(n-1)\right\}=1-\alpha$ **(这里$\alpha=0.05$)**

解得 σ^2的单侧置信上限 $\frac{(n-1)S^2}{\chi^2{}_{1-\alpha}(n-1)}$

然后代入计算就行

假设检验

双侧检验

1.目前我国男子平均身高为 169 厘米.现对某地 40 名男子进行身高测量，计算得身高平均数为 171.2 厘米, **标准差**（是已知S^2，不是方差σ^2)为 9.2 厘米，问：是否可认为该地区男子的平均身高为全国平均水平?

$(\alpha=0.05，\ t_{0.05}(40)=1.6839, t_{0.05}(39)=1.6849, t_{0.025}(40)=2.0211, t_{0.025}(39)=2.0227)$

设原假设$H_0:\mu=169$，备择假设$H_1:\mu\neq169$

已知样本方差，总体均值和总体方差未知，在原假设H_0成立的条件下，

选取**检验统计量** $T=\frac{\bar{X}-169}{S/\sqrt{n}}\sim t(n-1)$

因为**显著性水平**$\alpha=0.05$，故拒绝域为$|t|>t_{\alpha/2}(39)$

计算统计量t与拒绝域对比得结果

故接受原假设，即认为该地区男子平均身高为全国平均水平.

单侧检验

2.从甲，乙两个灯泡厂分别抽取 30 个灯泡，测试寿命，甲厂灯泡平均寿命为 $\bar{x}=1500$， 样本标准差为 $S_1=80$， 乙厂灯泡平均寿命为 $\bar{y}=1450$， 样本标准差为 $S_2=90$， 假设甲厂灯泡寿命$X\sim N\left(\mu_1,\sigma^2\right)$， 乙厂灯泡寿命 $Y\sim N\left(\mu_2,\sigma^2\right)$， 其中参数未知。试问在显著水平 $\alpha=0.05$ 下，能否断定甲厂灯泡比乙厂灯泡好?

解：设原假设和备则假设 $H_0:\mu_1=\mu_2, H_1:\mu_1>\mu_2$

已知样本方差，总体均值和总体方差未知，在原假设H_0成立的条件下，

选取检验统计量$T=\frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$

其中 $S_w=\sqrt{\frac{1}{n_1+n_2-2}\left[(n_1-1)S_1^2+(n_2-1)S_2^2\right]}$

代入$n_1=n_2=30$

因为**显著性水平**$\alpha=0.05$，故拒绝域为$t>t_\alpha(58)$

（原本分子该更小一点，假设相等导致分子偏大）

计算统计量t与拒绝域对比得结果

故拒绝原假设，可以断定甲厂产品比乙厂产品好

回归分析

假定自变量 $X_1,X_2,\cdots,X_k$在试验中的取值为 $x_1,x_2,\cdots,x_k$， 归结为随机误差的那一部分记为 ε ，则有

$$Y=\mu\left(x_1,x_2,\cdots,x_k\right)+\varepsilon$$

Y 关于$X_1,X_2,\cdots,X_k$的回归模型为:

$$Y=\mu\left(x_1,x_2,\cdots,x_k\right)+\varepsilon\quad E(\varepsilon)=0,D(\varepsilon)=\sigma^2$$

回归函数：$\mu\left(x_1,x_2,\cdots,x_k\right)$

回归方程：$y=\mu\left(x_1,x_2,\cdots,x_k\right)$

误差方差σ^2的意义：反映回归模型的有效性，σ^2越小越有效

一元线性回归模型

$Y=a+bX+\varepsilon,\quad \varepsilon\sim N(0,\sigma^2),\quad E(\varepsilon)=0,D(\varepsilon)=\sigma2$

a、b为常系数, b称为Y对X的**回归系数**，a称为**回归常数**

y_i的估计值$\hat{y}=a+bx_i$称为回归值

应使偏差$y_i-\hat{y}$都尽可能小$\rightarrow$使残差的平方和$\sum\left(y_i-\hat{y}_i\right)^2$最小.

最小二乘法

残差平方和满足:

$$Q(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

要使Q最小

结果:

$$\begin{cases}\hat{b}=\dfrac{l_{xy}}{l_{xx}}\\ \hat{a}=\bar{y}-\hat{b}\bar{x}\end{cases}\quad 其中\begin{cases}l_{xy}=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}=\sum_{i=1}^{n}x_iy_i-\dfrac{1}{n}\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i\\ l_{xx}=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2=\sum_{i=1}^{n}x_i^2-\dfrac{1}{n}(\sum_{i=1}^{n}x_i)^2\\ l_{yy}=\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2=\sum_{i=1}^{n}y_i^2-\dfrac{1}{n}(\sum_{i=1}^{n}y_i)^2\end{cases}$$

σ^2的无偏估计

$$\hat{\sigma}^2=\frac{1}{n-2}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\frac{1}{n-2}\left(l_{yy}-\hat{b}^2l_{xx}\right)$$

显著性检验

样本相关系数检验法

$$\hat{\rho}=R=\frac{l_{xy}}{\sqrt{l_{xx}}\sqrt{l_{yy}}}$$

给定显著性水平α, 计算样本相关系数R的值

如果$|R|>R_a(n-2)$，则认为Y与X之间的线性相关关系显著

考卷

1. 某产品年需求量为随机变量X，且已知X~U(2000, 4000)，每售出1吨，获利3w;未售出的存储费用为1万元/t，设本年度该公司准备了t吨该商品，试计算平均收益，并计算每年准备多少吨获益最大；

$X \sim U(2000,4000) \quad f_X(x)=\begin{cases}\frac{1}{2000}, & 2000 \leqslant x \leqslant 4000 \\ 0, & \text{其他}\end{cases}$

收益 $Y=\begin{cases}3x+(-1)(t-x)=4x-t, & x \leqslant t \\ 3t, & x>t\end{cases}$

$$E(Y)=\int_{2000}^{t}(4x-t)\frac{1}{2000}dx+\int_{t}^{4000}\frac{3t}{2000}dx=\frac{9t^2-(18000-t)^2}{16000}+\frac{12000t-3t^2}{2000}$$

令 $\frac{dE(Y)}{dt}=7-\frac{4t}{2000}=0 \Rightarrow t=3500$, 即每年准备 3500 吨, 平均收益最大.

2. 设 $X_1,X_2,\ldots,X_n$ 为取自总体 $X \sim N(0,\theta)$ 的样本, 其中 $\theta>0$ 为方差, 令 $T=\left|\sum_{i=1}^{n}X_i\right|$, 确定常数 b, 使 bT 为 $\sqrt{\theta}$ 的无偏估计

思路：找$E(T)$

记 $Y=\sum_{i=1}^{n}X_i \sim N(0,n\theta)$, 从而 $\frac{Y}{\sqrt{n\theta}} \sim N(0,1)$, 于是

$$E(T)=E(|Y|)=\sqrt{n\theta}E\left|\frac{Y}{\sqrt{n\theta}}\right|=\sqrt{n\theta}\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{\sqrt{2n\theta}}{\sqrt{\pi}}\int_{0}^{+\infty}xe^{-\frac{x^2}{2}}dx=\sqrt{\frac{2n\theta}{\pi}}$$

然后$bE(T)=\sqrt{\theta}$即可

3. $X_2,\ldots,X_{10}$ 为取自正态总体 $X \sim N\left(0,(\sqrt{2})^2\right)$ 的样本, 记 $T=\sum_{i=1}^{10}\left(X_i-\bar{X}\right)^2$, 计算方差 DT

思路：求方差不一定就是$E(X^2)-E^2(X)$，可以转化成已知函数求

χ^2**均值为**n**，方差为**$2n$

$T=\sum_{i=1}^{10}\left(X_i-\bar{X}\right)^2=2\sum_{i=1}^{10}\left(\frac{X_i-\bar{X}}{\sqrt{2}}\right)^2=2\times\frac{(10-1)S^2}{(\sqrt{2})^2} \longrightarrow \frac{T}{2} \sim \chi^2(9)$

于是 $D\left(\frac{T}{2}\right)=2\times 9=18 \Longrightarrow \frac{1}{4}DT=18 \Longrightarrow DT=72$

4. 假设总体 $X \sim U(0,\theta), X_1,X_2,\cdots,X_n$ 为一组样本, 求:

(1) 未知参数 θ 的极大似然估计量 $\hat{\theta}_1$;

(2) 若 $\hat{\theta}_2=k\hat{\theta}_1$ 是 θ 的无偏估计量, k 应取何值?

(1) **设**$x_1,x_2,\cdots x_n$**为样品**$X_1,X_2,\cdots X_n$**的观测值**(这句话必须要写)

X_i 的概率密度函数为 $f(x_i)=\begin{cases}\frac{1}{\theta}, 0 \leqslant x_i \leqslant \theta \\ 0, \text{其它}\end{cases}$

参数 θ 的似然函数为 $L(\theta)=\left(\frac{1}{\theta}\right)^n$, 为 θ 的减函数, 由于 $0 \leqslant x_i \leqslant \theta$, 在 $\theta=\max\{x_i\}$ 处, 似然函数取得最大值. 故 θ 的最大似然估计量为 $\hat{\theta}_1=\max\limits_{1\leq i\leq n}\{X_i\}$.

(补：极大似然估计值为$\hat{\theta}_1=\max\limits_{1\leq i\leq n}x_i$)

(2) 令 $Y=\max\{X_1,X_2,\cdots,X_n\}$, 则 $F_Y(y)=(F_X(y))^n$

$$f_Y(y)=nf_X(y)(F_X(y))^{n-1}=\begin{cases}\frac{ny^{n-1}}{\theta^n} & ,0 \leqslant y \leqslant \theta \\ 0, & \text{其它}\end{cases}$$

$E\left(\hat{\theta}_1\right)=EY=\int_{-\infty}^{\infty}yf_Y(y)dy=\frac{n}{n+1}\theta$, 若 $E\left(\hat{\theta}_2\right)=kE\left(\hat{\theta}_1\right)=\theta \Rightarrow k=\frac{1+n}{n}$.

5. 设 A,B,C 为三个随机事件, A 与 C 相互独立, B 与 C 相互独立, 则 $A\cup B$ 与 C 相互独立的充要条件是 (AB 与 C 相互独立)

拆解分析题:

$$\begin{aligned}P[(A\cup B)C]&=P(AC\cup BC)=P(AC)+P(BC)-P(ABC)\\&=P(A)P(C)+P(B)P(C)-P(ABC)\end{aligned}$$

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

所以若$P[(A\cup B)C]=P(A\cup B)P(C)$，则$P(ABC)=P(AB)P(C)$，所以满足$AB$与$C$相互独立

6. 设随机事件A与B相 互独立, A与 C相互独立, $BC=\varnothing$, $P(A)=P(B)=\frac{1}{2}$, $P(AC \mid AB\cup C)=\frac{1}{4}$, 求 $P(C)$.

解:

由条件概率公式知

$$P(AC \mid AB \cup C)=\frac{P(AC \cap (AB \cup C))}{P(AB \cup C)}$$

由事件的运算性质分配律可知 $AC \cap (AB \cup C)=ABC \cup AC=AC$

由 A 与 C 相互独立可 知 $P(AC)=P(A)P(C)$.

由 $BC=\varnothing$ **可知** AB **与** C **互不相容**, 所以 $P(AB \cup C)=P(AB)+P(C)$, 由 A 与 B 相互独立可知 $P(AB)=P(A)P(B)$, 即

$$P(AC \mid AB \cup C)=\frac{P(A)P(C)}{P(A)P(B)+P(C)}=\frac{0.5P(C)}{0.5 \times 0.5+P(C)}=\frac{1}{4}$$

7. 不断抛掷两颗骰子, 设 $A=\{$ 两颗骰子点数之和为 5$\}$, $B=\{$ 两颗骰子点数之和为 7$\}$, 求 A 在 B 之前发生的概率
   解: 设 $C=\{A$ 在 B 之前发生 $\}$
   考虑第一次抛掷的三种结果:
   $D_1=\{$ 第一次抛扶出现 $A\}$
   $D_2=\{$ 第一次抛郑出现 $B\}$
   $D_3=\{$ 第一次抛郑 A, B 均末出现 $\}$
   则 D_1, D_2, D_3 构成样本空间的有限划分，故使用全概率公式有:
   $$\begin{aligned}P(C)&=P(C \mid D_1)P(D_1)+P(C \mid D_2)P(D_2)+P(C \mid D_3)P(D_3)\\&=1 \times P(A)+0 \times P(B)+P(C){\color{red}P(\bar{A}\bar{B})}\\&=1 \times \frac{1}{9}+0 \times \frac{1}{6}+P(C)\left(1-\frac{1}{9}-\frac{1}{6}\right)\\&=\frac{1}{9}+\frac{13}{18}P(C) \Rightarrow \frac{5}{18}P(C)=\frac{1}{9} \Rightarrow P(C)=\frac{2}{5}\end{aligned}$$

书本

1. 设将5个球随意地放入3个盒子中,求每个盒子内至少有一个球的概率.

   分类5 0;4 1;3 2 事件数=$3+C_5^1A_3^3+C_5^2A_3^3$
2. 两台电子仪器的寿命分别为 X_1, X_2, 且 $X_1 \sim N(40,36), X_2 \sim N(45,9)$, 若要在 45 小时 的期间内使用这种仪器, 问选用哪一台仪器较好? 若在 52 小时内使用呢?

   要计算的是$P\{0<X<45\}$！ 不要直接想的太简单 选X_1
3. 二维随机变量 (X, Y) 的联合分布函数是

   $$F(x, y)=A\left(B+\arctan \frac{x}{2}\right)\left(C+\arctan \frac{y}{3}\right), \quad(x, y) \in \mathbf{R}^2,$$

   试求: (1) 系数 A、B、C; (2) 边缘分布函数.

   (1) $\lim\limits_{x \to -\infty} F(x,y)=0, \lim\limits_{y \to -\infty} F(x,y)=0, \lim\limits_{\substack{x \to +\infty\\ y \to +\infty}} F(x,y)=1$

   $A(B-\pi/2)(C+\arctan \frac{y}{3})=0$ 推出B、C，最后用x，y趋于无穷等于1求A

   (2)$F_X(x)=\lim\limits_{y \to \infty} F(x,y) \qquad F_Y(x)=\lim\limits_{x \to \infty} F(x,y)$
4. 条件概率$f_{X|Y}(x|y)$要用y的形式表示x的范围，然后补上y的范围，如果是矩形区域则直接写两个范围
5. 已知随机变量 X、Y 相互独立, X 服从参数为 λ 的指数分布, Y 服从区间 $(0, h)(h>0)$ 上的均匀分布, 写出 $X+Y$ 的概率密度.

   $G=\{(x, z) \mid x>0,0<z-x<h\}=\{(x, z) \mid {\color{red}x>0, x<z<x+h}\}$,

   画xoz图

   $$0<z<h, f_Z(z)=\int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) d x=\int_0^z \lambda e^{-\lambda x} \frac{1}{h} d x=\frac{1}{h}\left(1-e^{-\lambda z}\right);$$
   $$z \geqslant h, f_Z(z)=\int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) d x=\int_{z-h}^z \lambda e^{-\lambda x} \frac{1}{h} d x=\frac{1}{h} e^{-\lambda z}\left(e^{\lambda h}-1\right);$$

6. 设随机变量 X 与 Y 相互独立, $X \sim N(0,1), Y \sim N(0,1)$, 求 $Z=X/Y$ 的概率密度.
$G=\{(y,z) \mid yz \in R, y \in R\}$,

$$f_Z(z)=\int_{-\infty}^{+\infty}|y| f(yz, y) dy=2 \int_0^{+\infty} y \frac{1}{2\pi} e^{-\frac{y^2 z^2+y^2}{2}} dy=\frac{1}{\pi\left(z^2+1\right)}$$

7. 已知随机变量 $X \sim U(-\pi, \pi)$, 试求 $Y=\cos X$ 和 $Y^2=\cos^2 X$ 的数学期望.

$$f(x)=\begin{cases}\frac{1}{2\pi}, -\pi<x<\pi \\ 0, \text{ 其它}\end{cases}, E(Y)=\int_{-\infty}^{+\infty} \cos x f(x) dx=\int_{-\pi}^{\pi} \cos x \frac{1}{2\pi} dx=0,$$

$$E\left(Y^2\right)=\int_{-\infty}^{+\infty} \cos^2 x f(x) dx=\int_{-\pi}^{\pi} \cos^2 x \frac{1}{2\pi} dx=\frac{1}{2}$$

8. 除非μ已知，否则标准差、方差就是样本标准差、样本方差