常见分布的均值方差，n阶矩

定义：$E(X)=\sum_{i=1}^{+\infty}x_i P(X=x_i)=\int_{-\infty}^{+\infty}xf(x)dx$

$D(X)=Var(X)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\int_{-\infty}^{+\infty}(x-\mu)^2f(x)dx$

且 $D(X)=E(X^2)-(E(X))^2$

	密度 $f(x)$	$E(X)$	$D(X)$
二项分布 $X\sim B(n,p)$	$f(x)=p^x\cdot(1-p)^{n-x}$	np	$np(1-p)$
泊松分布 $X\sim P(\lambda)$	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ
几何分布 $X\sim Ge(p)$	$P(X=k)=p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
均匀分布 $X\sim U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0, & 其它\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 $X\sim Exp(\lambda)$	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & 其它\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布 $X\sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma分布 $X\sim\Gamma(n,\theta)$	$f(x)=\frac{\theta^n}{\Gamma(n)}\cdot x^{n-1}e^{-\theta x}$	$\frac{n}{\theta}$	$\frac{n}{\theta^2}$

正态分布：$X\sim N(\mu,\underset{\sigma}{\overset{\|}{\sigma^2}})$

$N(\mu_1,\sigma_1^2)\pm N(\mu_2,\sigma_2^2)=N(\mu_1\pm\mu_2,\sigma_1^2+\sigma_2^2)$

Gamma分布：

$\Gamma(\frac{n}{2},\frac{1}{2})=\chi^2(n)$

$\Gamma(1,\theta)=Exp(\frac{1}{\theta})$

n个 $Exp(\lambda)$ 和为 $\Gamma(n,\lambda)$

$X\sim\Gamma(n,\lambda)$，$aX\sim\Gamma(n,\frac{\lambda}{a})$

n个 $P(\lambda)$ 和为 $P(n\lambda)$

特殊函数

Gamma函数：

$\Gamma(z)=\int_0^{+\infty}t^{z-1}e^{-t}dt$

$\Gamma(z+1)=z\Gamma(z)$ $\Gamma(n+1)=n!\ (n\in N)$

$\Gamma(\frac{1}{2})=\sqrt{\pi}$

Beta函数：

$B(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}dx$

$B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$

概率论前置

分布复合：ξ有密度函数 $p(x)$，$\eta=g(\xi)$

$g(\xi)$严格单调，且 $g^{-1}(y)$ 有连续导函数，则η有密度函数 $p(g^{-1}(y))|(g^{-1}(y))'|$

大数定律：$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{a.s.}\mu$

中心极限定理：$\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sqrt{n}\sigma}\to N(0,1)$　$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$

正态分布n阶矩：$E((X-\mu)^k)=\begin{cases}\sigma^k\cdot(k-1)!!, & k偶\\ 0, & k奇\end{cases}$

估计

MLE：$L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i,\theta)$

求 θ 使 L 取 max，最大似然，

矩估计：$\begin{cases} v_1 = EX = g_1(\theta_1,\cdots,\theta_m) \\ \vdots \\ v_m = EX^m = g_m(\theta_1,\cdots,\theta_m). \end{cases}$

$\tilde{v}_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，令 $v_k=\tilde{v}_k$ 可得

无偏估计：$E_\theta\varphi(x_1,\cdots,x_n)=g(\theta)$．$\varphi$ 为 $g(\theta)$ 无偏．

MVUE．最小方差无偏估计：

$M_\theta(\varphi)=E_\theta[\varphi(x_1,\cdots,x_n)-g(\theta)]^2$ 最小

充分统计量 $L(\vec{x},\theta)=h(\vec{x})\cdot q(\varphi(\vec{x}),\theta)$

称 $\varphi(x)$ 充分．

完全统计量 $E_\theta[u(\varphi(\vec{x}))]=0 \Rightarrow P_\theta(u(\varphi(\vec{x}))=0)=1$

指数分布族：两点．二项．指数．正态．泊松．

$f(x;\theta)=s(\theta)h(x)e^{\sum_{j=1}^{k}c_j(\theta)T_j(x)}$

若参数空间 Θ 有内点，则 $(\sum_{i=1}^{n}T_1(x_i),\cdots,\sum_{i=1}^{n}T_k(x_i))$ 完全

BLS定理 $\varphi(x_1,\cdots,x_n)$ 完全充分统计 θ，且 $\psi(\varphi(x_1,\cdots,x_n))$ 为 $g(\theta)$ 无偏．则为 $g(\theta)$ 的(一致)MVUE（如 $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=S^2$ 为 σ^2 的）

Fisher信息量 $I_\theta=E_\theta\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^2=-E_\theta\left(\frac{\partial^2\ln f(x,\theta)}{\partial\theta^2}\right)$

Cramer-Rao不等式 $Var(\varphi)\geq\frac{(g'(\theta))^2}{nI(\theta)}$．$\forall\theta\in\Theta$

取到下界则为MVUE

置信区间：$P_\theta\{\varphi_1(\vec{x})\leq g(\theta)\leq\varphi_2(\vec{x})\}\geq\gamma$

统计学分布

χ^2分布：$X_i\sim N(0,1)$ $\sum_{i=1}^{n}X_i^2\sim\chi^2(n)$

均值为 n．方差为 $2n$

$\chi^2(m)+\chi^2(n)=\chi^2(m+n)$

t分布：$X\sim N(0,1)$．$Y\sim\chi^2(n)$

$Z=\frac{X}{\sqrt{Y/n}}\sim t(n)$．$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$

F分布：$X\sim\chi^2(m)$ $Y\sim\chi^2(n)$

$F=\frac{X/m}{Y/n}\sim F(m,n)$

正态分布枢轴量（构造置信区间）

目标	条件	枢轴量
μ	σ^2已知	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$
	σ^2未知	$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$
$\mu_1-\mu_2$	σ_1^2．σ_2^2已知	$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\sim N(0,1)$
	$\sigma_1^2=\sigma_2^2$未知	$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S\sqrt{1/n_1+1/n_2}}\sim t(n_1+n_2-2)$ S 整体
σ^2	μ已知	$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$
	μ未知	$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1．μ_2未知	$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

假设检验

定义：$H_0:\Theta_0 \leftrightarrow H_1:\Theta-\Theta_0$ 备择假设

第I类错误：以真为假

第II类错误：以假为真

$\varphi(x)$：X时以$\varphi(x)$概率拒绝H_0

功效函数 $\beta_\varphi(\theta)=E_\theta\varphi(x)=P_\theta((\vec{x})\in \underbrace{W_0}_{\text{否定域}})$

$\vartheta\in\Theta_0$ $\beta_\varphi(\vartheta)=P(\text{I})$

$\vartheta\notin\Theta$ $1-\beta_\varphi(\vartheta)=P(\text{II})$

UMP(一致最优检验)：φ为水平α. 若$\sup\limits_{\theta\in\Theta_0}\beta_\varphi(\theta)\le\alpha$

若$\forall\varphi$. $\beta_\varphi(\vartheta)\ge\beta_\psi(\vartheta)$. ψ为UMP

若加以无偏 对$\forall\vartheta\in\Theta_1$也成立.

对全部无偏中找最优. 为UMPU

N-P引理(构造UMP) 且无偏

对$\Theta=\{\theta_1,\vartheta_2\}$. 给定$0<\alpha<1$. 设$W_0=\{x|\lambda(x)>\lambda_0\}$. $\lambda(x)=\frac{L(x;\vartheta_2)}{L(x;\vartheta_1)}$, $\varphi_0(x)=\begin{cases}1, & \lambda(x)>\lambda_0\\ 0, & \text{其它}\end{cases}$

λ_0有$\int_{W_0}L(x,\vartheta_1)dx=\beta_{\varphi_0}(\vartheta_1)=\alpha$ (λ_0存在). 解x范围

单参数假设检验

$$f(x;\vartheta)=s(\vartheta)h(x)e^{Q(\vartheta)V(x)}$$

$s(\vartheta)>0$. $h(x)>0$ $Q(\vartheta)$关于ϑ严格增.

记$t(x)=\sum_{i=1}^n V(x_i)$ 充分统计量

① $H_0:\vartheta\le\vartheta_1 \leftrightarrow H_1:\vartheta>\vartheta_1$

$\varphi_0(x)=\begin{cases}1, & \sum_{i=1}^n V(x_i)>c\\ 0, & \text{其它}\end{cases}$ UMP

$P_{\vartheta_1}(\sum_{i=1}^n V(x_i)>c)=\alpha$

② $H_0:\vartheta\in(\vartheta_1,\vartheta_2)\leftrightarrow H_1:\vartheta\in(\vartheta_1,\vartheta_2)$

$\varphi_0(x)=\begin{cases}1, & c_1<t(x)<c_2\\ 0, & \text{其它}\end{cases}$ UMP

$\beta_{\varphi_0}(\vartheta_1)=\beta_{\varphi_0}(\vartheta_2)=\alpha$

同时满足. 解c_1, c_2.

③ $H_0:\vartheta\in[\vartheta_1,\vartheta_2]\leftrightarrow H_1:\vartheta\notin[\vartheta_1,\vartheta_2]$

$\varphi_0(x)=\begin{cases}1, & t(x)<c_1 \text{或} t_1>c_2\\ 0, & \text{其它}\end{cases}$ UMPU.

不存在UMP

④ $H_0:\vartheta=\vartheta_0\leftrightarrow H_1:\vartheta\ne\vartheta_0$

φ_0同③. 满足$\beta_{\varphi_0}(\vartheta_0)=\alpha$. 且$E_{\vartheta_0}(\varphi_0(x)\sum_{i=1}^n V(x_i))$

$=\alpha E_{\vartheta_0}(\sum_{i=1}^n V(x_i))$ UMPU

特别. 若$t(x)$关于r_0对称. 则可化为

$P_{\vartheta_0}\{t(x)<r_0-c \text{或} t(x)>r_0+c\}=\alpha$

广义似然比检验法

对样本$x_1\cdots x_n$. 似然$L(\vec{x};\theta)$

记$\lambda(x)=\frac{\sup\{L(x;\theta)|\theta\in\Theta\}}{\sup\{L(x;\theta)|\theta\in\Theta_0\}}$ 为广义似然比

记MLE 代入化简

给定α. 找λ_0使$\sup\{\beta_{\varphi_0}(\theta)|\theta\in\Theta_0\}=\alpha$

$\varphi_0=\begin{cases}1, & \lambda(x)>\lambda_0\\ 0, & \text{其它}\end{cases}$

正态分布检验

① $X \sim N(\mu, \sigma^2)$，μ, σ^2 未知.

(i)

$H_0: \mu = \mu_0 \leftrightarrow H_1: \mu \neq \mu_0$　$\lambda(x) = \left[\dfrac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{\frac{n}{2}}$

$= \left(1+\dfrac{T_0^2}{n-1}\right)^{\frac{n}{2}}$

有 $T_0 = \dfrac{\sum_{i=1}^{n}\sqrt{n(n-1)}\,(x_i-\mu_0)^2}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \dfrac{\bar{x}-\mu_0}{\sqrt{s^2/n}} \sim t(n-1)$

故否定域为 $\{|T_0| > c\}$.

(ii)

$H_0: \mu \leq \mu_0 \leftrightarrow H_1: \mu > \mu_0$. 同上. $\{T_0 > c\}$.

(iii)

$H_0: \sigma^2 = \sigma_0^2 \leftrightarrow H_1: \sigma^2 \neq \sigma_0^2$　$\lambda(x) = \begin{cases} 1. & \sigma^2 = \sigma_0^2 \\ \left(\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}\right)^{-\frac{n}{2}} e^{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma_0^2}} & \end{cases}$

令 $T_1 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2} \sim \chi^2(n-1)$

故 $W_0 \{T_1 > c_2$ 或 $T_1 < c_1\}$

(iV)

$H_0: \sigma^2 \leq \sigma_0^2 \leftrightarrow H_1: \sigma^2 > \sigma_0^2$，同上 $\{T_1 > c_2\}$

② $X \sim N(\mu_1, \sigma_1^2)$　$Y \sim N(\mu_2, \sigma_2^2)$

$X_1 \cdots X_{n_1}$　　$Y_1 \cdots Y_{n_2}$

(i) μ_1, μ_2 已知

$H_0: \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1: \sigma_1^2 \neq \sigma_2^2$　　$\leq \leftrightarrow >$ 时同理

$\lambda(x) = \dfrac{\delta^{\frac{n_1+n_2}{2}}}{\delta_1^{\frac{n_1}{2}}\ \delta_2^{\frac{n_2}{2}}}$

$\delta = \dfrac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2\right)$ ☆

$\delta_1 = \dfrac{1}{n_1}\sum_{i=1}^{n}(x_i-\bar{x})^2$　$\delta_2 = \dfrac{1}{n_2}\sum_{i=1}^{n}(y_i-\bar{y})^2$

故 $F_0 = \dfrac{\sum_{i=1}^{n}(x_i-\mu_1)^2/n_1}{\sum_{i=1}^{n}(y_i-\mu_2)^2/n_2} \sim F(n_1, n_2)$

$W_0 = \{F_0 > c_2$ 或 $F_0 < c_1\}$

(ii) μ_1, μ_2 未知. 考虑 (i) 中　$\leq \leftrightarrow >$ 时同理

需调整 F 为 $F_1 = \dfrac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2/(n_1-1)}{\sum_{i=1}^{n_2}(y_i-\bar{y})/(n_2-1)} \sim F(n_1-1, n_2-1)$

有 $W_0 = \{F_1 > c_2$ 或 $F_1 < c_1\}$.

一般取 $\int_0^{c_1} f(u)\,du = \int_{c_2}^{+\infty} f(u)\,du = \dfrac{\alpha}{2}$

f 为 F 的密度函数

(iii) $\sigma_1^2 = \sigma_2^2$ 已知数值

$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \neq \mu_2$　$\leq \leftrightarrow >$ 时同理

化简 $\lambda(x)$ 与 $\bar{x}-\bar{y}$ 有关 $\dfrac{\bar{x}-\bar{y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$

$W_0 = \{|\bar{x}-\bar{y}| > c\}$

(iV) $\sigma_1^2 = \sigma_2^2$. 未知数值.

$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \neq \mu_2$　$\leq \leftrightarrow >$ 时同理

$\lambda(x) = \left(1 + \dfrac{n_1 n_2}{n_1+n_2} \cdot \dfrac{(\bar{x}-\bar{y})^2}{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2}\right)^{\frac{n_1+n_2}{2}}$

$= \left(1 + \dfrac{T_2^2}{n_1+n_2-2}\right)^{\frac{n_1+n_2}{2}}$

$T_2 = \dfrac{\bar{x}-\bar{y}}{\sqrt{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2}} \cdot \sqrt{\dfrac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}} \sim t(n_1+n_2-2)$

有 $W_0 = \{|T| > c\}$. $2\int_c^{+\infty} t(u)\,du = c$

t 为 $t_{n_1+n_2-2}$ 密度函数