数理统计

$$\text{try}: \quad P\left(x_l \mid y_l\right)=\frac{\color{red}{P(y_l \mid x_l)}P(x_l)}{P(y_l)}$$

$$\mathbb{Drowm\ Genius} \quad \mathcal{Drowm\ Genius} \quad \mathscr{Drowm\ Genius}$$

参考

1. 韦来生《数理统计》
2. 茆诗松《概率论与数理统计》
3. Casella《统计推推断》
4. 本人上个学期的概率论笔记《keep random》
5. slides
    1. stat Chapter1.pdf
    2. stat Chapter 2_New.pdf
    3. stat Chapter3.pdf
    4. stat Chapter 4.pdf
    5. stat Chapter 5-final.pdf

统计量分布的推导

前置概率论知识

- 随机变量商的密度函数

$$Z=\frac{X}{Y}, f_z(z)=\int_{-\infty}^{+\infty} f(yz,y)|y|\mathrm{d}y$$

- Gamma 分布

$$f_{\Gamma(n,\lambda)}(x)=\frac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x}$$

$$f_{\Gamma(1/2,1/2)}(x)=\frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-\frac{1}{2}x} \quad \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

$$E(x)=\frac{n}{\lambda} \qquad Var(x)=\frac{n}{\lambda^2}$$

- Definition: $\Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx$.
- $\Gamma(z+1)=z\Gamma(z)$, $\Gamma(1)=1$, $\Gamma(n)=(n-1)!$ for n be a positive integer.
- $\Gamma(1/2)=\sqrt{\pi}$, $\Gamma(3/2)=\frac{1}{2}\sqrt{\pi}$, $\Gamma(-1/2)=-2\sqrt{\pi}$.
- $\Gamma(\frac{1}{2}+n)=\frac{(2n-1)!!}{2^n}\sqrt{\pi}$, $\Gamma(\frac{1}{2}-n)=\frac{(-1)^n 2^n}{(2n-1)!!}\sqrt{\pi}$, $\lim_{n\to\infty}\frac{\Gamma(n+\alpha)}{\Gamma(n)\cdot n^\alpha}=1$.
- For any two positive real numbers x_1 and x_2, and for any $t\in[0,1]$, we have

$$\Gamma(tx_1+(1-t)x_2)\le\Gamma^t(x_1)\Gamma^{1-t}(x_2).$$

- 随机变量、随机向量函数的分布
- 多元正态分布
  多元统计分析（二）：多元正态分布 - 知乎 (zhihu.com)
  主要内容：多元正态分布的表示，多元正态分布线性变换，多元正态导出χ^2分布
- 特征函数、数量特征
  概率论中常见分布的数学期望、方差及其特征函数推导——连续性随机变量 - 知乎 (zhihu.com)
- 其他
  数理统计第二讲（经验分布函数，格里文科定理，统计量） - 知乎

次序统计量

联合分布

$$f(x_{(1)}\dots x_{(n)})=n!f(X_1\dots X_n)\mathbb{1}_{X_1<\dots X_n}\underset{i.i.d}{=}n!(f(X))^n\mathbb{1}_{X_1<\dots X_n}$$

边缘分布

$$1-F(x_{(1)})\overset{i.i.d}{=}(1-F(x))^n\rightarrow f(x_{(1)})=n(1-F(x))^{n-1}f(x)$$

$$F(x_{(n)})\overset{i.i.d}{=}(F(x))^n\rightarrow f(x_{(n)})=n(F(x))^{n-1}f(x)$$

$$F(x_{(m)}:至少有m个样本小于x)=\sum_{i=m}^{n}C_n^i(F(x))^i(1-F(x))^{n-i}$$

$$\rightarrow f(x_{(m)})=\frac{n!}{(n-m)!(m-1)!}F(x)^{m-1}(1-F(x))^{n-m}f(x)$$

$$F(x_{(i)}=x,x_{(j)}=y,(i<j):至少有i个样本小于x,j个样本大于x)$$

$$=\sum_{k=i}^{n}C_n^k(F(x))^k(1-F(x))^{n-k}\sum_{m=j}^{n}C_n^m(F(y))^{n-m}(1-F(y))^m$$

$$f_{ij}(x,y)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}F(x)^{i-1}(F(y)F(x))^{j-i-1}(1-F(y))^{n-j}f(x)f(y)$$

🆓极差分布

正态总体相关分布

正态样本统计量的分布($\bar{X}\ S^2$)

⭐对于服从正态分布的简单随机样本$X_1\cdots X_n,X_i\sim N(a,\sigma^2)$的样本均值$\bar{X}$和样本方差$S^2$，我们有以下结论：

1. $\bar{X}\sim N(a,\sigma^2/n)\quad i.e.\ \sqrt{n}\frac{(\bar{X}-a)}{\sigma}\sim N(0,1)\ \ :Z\ statistic$
2. $(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$
3. $\bar{X}$与S^2不相关
   证明：后两个命题证明考虑正交变换$Y=AX\ \ Y$与X的分布一样，仅仅做了旋转变换

$$\boldsymbol{A}=\begin{pmatrix}\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

Let $\boldsymbol{A} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$, which satisfies $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.

Denote $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, then we have:

$$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{n}\bar{X} \ and\ Y_1^2 + ... + Y_n^2 = X_1^2 + ... + X_n^2$$

as $\mathrm{tr}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$ such that

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2.\end{aligned}$$

- From Theorem 2, we have $Y_1, ..., Y_n$ are independent, and $Y_i \sim N(\mu_i, \sigma^2)$.
- And

$$\mu_i = a\sum_{k=1}^{n} a_{ik} = a\sqrt{n}\sum_{k=1}^{n}\frac{1}{\sqrt{n}}\cdot a_{ik} = a\sqrt{n}\boldsymbol{A}(i,:)\boldsymbol{A}(1,:)^{\mathrm{T}} = 0$$

  for $i = 2, ..., n$ as $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- $Y_2, ..., Y_n \sim N(0, \sigma^2)$ such that $Y_2/\sigma, ..., Y_n/\sigma \sim N(0, 1)$.

统计学三大分布($\chi^2 \to t\ .\ F$)

- χ^2**分布**
  - 定义：记$X_1 \cdots X_n \overset{i.i.d}{\sim} N(0,1)$，则$\sum_{i=1}^{n} X_i^2 \sim \chi_n^2$:自由度为n的卡方分布
  - 来源：若$X \sim N(0,1)$, 则$X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$. 由于Gamma分布具有独立可加性，所以$\sum_{i=1}^{n} X_i^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})\ i.e.\ \chi_n^2$
    - $\Gamma(\frac{1}{2}, \frac{1}{2})$的证明：由换元法求$pdf$可知$X^2$的分布函数: $f_{\chi^2}(x) = 2\phi(\sqrt{x})|\frac{1}{2\sqrt{x}}| = \frac{e^{-\frac{x}{2}}}{\sqrt{2\pi x}} \sim \Gamma(\frac{1}{2}, \frac{1}{2})$, since : $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
    - 可加性的证明：卷积+归纳法
    - 可加性的来源：$\Gamma(k, \theta)$分布可以视为k个$i.i.d\ \sim Exp(\theta)$的和
      - 证明：特征函数
      - 总结：指数分布$Exp(\lambda)$的独立和:伽马分布$\Gamma(n, \lambda)$→$\Gamma(1/2, 1/2)$独立和:卡方分布→卡方分布独立和极限:正态分布
    - Gamma分布可加性可以推广到非整数, 但是$\Gamma(k_1, \theta_1), \Gamma(k_2, \theta_2)$要满足等比例条件：[Gamma分布的可加性到底应该如何证明？ -]

知乎 (zhihu.com)

- 推论(从**指数分布构造卡方分布**)：如果$X_1 \cdots X_n$独立，$X_i \sim \text{Exp}(\lambda_i)$，则$\sum_{i=1}^n 2\lambda_i X_i \sim \chi^2_{2n}$
  - $2\lambda_i X_i$的pdf是$\frac{1}{2}e^{-\frac{y}{2}} \sim \chi^2_2 = \Gamma\left(1, \frac{1}{2}\right)$
  - χ^2可加性

- **t分布**
  - 定义：记$Y \sim \chi^2_n, X \sim N(0,1), X, Y$相互独立，$T = \frac{X}{\sqrt{Y/n}} \sim t_n$自由度为n的$t$分布
- **F分布**
  - 定义：记$Y \sim \chi^2_n, X \sim \chi^2_m; X, Y$相互独立$: F = \frac{X/m}{Y/n} \sim F_{m,n}$:自由度为m和n的$F$分布
- **性质**
  - $t \sim t_n \to t^2 \sim F_{(1,n)}$
  - 两个独立正态随机变量的比$\frac{X}{Y} \sim t_1 \sim \text{Standard Cauchy}$
  - $Z \sim F_{(m,n)} \to 1/Z \sim F_{(n,m)}$
  - $F_{(m,n)}(\alpha) = F_{(n,m)}(1-\alpha)$

$$
\begin{aligned}
\alpha &= P(Z < F_{(m,n)}(\alpha)) \\
&= P(1/Z > 1/F_{(m,n)}(\alpha)) \\
&= P(1/Z < F_{(n,m)}(1-\alpha))
\end{aligned}
$$

推论:

- Lemma 1: Assume $X_1, ..., X_n$ $i.i.d \sim N(a, \sigma^2)$, then

$$T = \frac{\sqrt{n}(\overline{X} - a)}{S} \sim t_{n-1}.$$

Proof: We have $\sqrt{n}(\overline{X} - a)/\sigma \sim N(0,1)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.

- Lemma 2: Assume $X_1, ..., X_m$ $i.i.d \sim N(a_1, \sigma^2)$, $Y_1, ..., Y_n$ $i.i.d \sim N(a_2, \sigma^2)$, and $X_1, ..., X_m; Y_1, ..., Y_n$ are independent, then

$$T = \frac{(\overline{X} - \overline{Y}) - (a_1 - a_2)}{S_w} \cdot \sqrt{\frac{mn}{m+n}} \sim t_{n+m-2},$$

where $(n+m-2)S_w^2 = \sum_{i=1}^m (X_i - \overline{X})^2 + \sum_{j=1}^n (Y_j - \overline{Y})^2$.
Proof: We have $\overline{X} \sim N(a_1, \sigma^2/m)$ and $\overline{Y} \sim N(a_2, \sigma^2/n)$ such that $\overline{X} - \overline{Y} \sim N(a_1 - a_2, \frac{m+n}{mn}\sigma^2)$.

- Lemma 3: Assume $X_1, ..., X_m$ $i.i.d \sim N(a_1, \sigma_1^2)$, $Y_1, ..., Y_m$ $i.i.d \sim N(a_2, \sigma_2^2)$, and $X_1, ..., X_m; Y_1, ..., Y_n$ are independent, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{m-1,n-1}.$$

Proof: We have $(m-1)S_1^2/\sigma_1^2 \sim \chi^2_{m-1}$ and $(n-1)S_2^2/\sigma_2^2 \sim \chi^2_{n-1}$.

统计量的极限分布

几类收敛(to be filled) LLN CLT

[概率论——大数定律与中心极限定理 - 知乎 (zhihu.com)]

指数分布族

指数分布族(指数可以视为一个参数空间的向量和样本空间的向量的内积:
$Q(\theta)'T(X)$
注意：指数分布-Gamma分布-χ^2分布；两点分布-二项分布-Poisson分布；正态分布都属于指数族；但是均匀分布不属于指数族!

$$f(x,\theta) = C(\theta)\exp\left\{\sum_{i=1}^{k} Q_i(\theta)\, T_i(x)\right\} h(x),$$

指数分布族的自然形式(本质就是把上个式子换个元：$\theta_{natural} = Q_i(\theta)$)

$$f(x,\theta) = C^*(\theta)\exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\} h(x),$$

由此可以引出自然参数空间(通过$H\ddot{o}lder$不等式可以证明是一个凸集)

$$\Theta^* = \left\{(\theta_1, \ldots, \theta_k) : \int_{\mathscr{X}} \exp\left\{\sum_{i=1}^{k} \theta_i T_i(x)\right\} h(x)\, dx < \infty\right\}.$$

赫尔德不等式：设$p>1$，$\frac{1}{p}+\frac{1}{q}=1$，如果$f\in L^p(\Omega)$，$g\in L^q(\Omega)$，那么
$\int_\Omega f(x)g(x)dx \le \left(\int_\Omega |f(x)|^p dx\right)^{\frac{1}{p}}\left(\int_\Omega |g(x)|^q dx\right)^{\frac{1}{q}}$

$$\Omega = \left\{ \boldsymbol{w} = (w_1, \cdots, w_k) : \int_{\mathscr{X}} \exp\left\{\sum_{j=1}^{k} w_j T_j(x)\right\} d\mu_1(x) < \infty \right\}$$

其中 $d\mu_1(x) = h(x)\,d\mu(x)$，设 $\boldsymbol{w}' = (w'_1, \cdots, w'_k)$，$\boldsymbol{w}'' = (w''_1, \cdots, w''_k) \in \Omega$，又设 $0<\alpha<1$，往证 $\boldsymbol{w} = \alpha\boldsymbol{w}' + (1-\alpha)\boldsymbol{w}'' = (w_1, \cdots, w_k) \in \Omega$. 事实上

$$\begin{aligned}
&\int_{\mathscr{X}} \exp\left\{\sum_{j=1}^{k} w_j T_j(x)\right\} d\mu_1(x) \\
&= \int_{\mathscr{X}} \left[\exp\left\{\sum_{j=1}^{k} w'_j T_j(x)\right\}\right]^{\alpha} \left[\exp\left\{\sum_{j=1}^{k} w''_j T_j(x)\right\}\right]^{1-\alpha} d\mu_1(x) \\
&\le \left[\int_{\mathscr{X}} \exp\left\{\sum_{j=1}^{k} w'_j T_j(x)\right\} d\mu_1(x)\right]^{\alpha} \left[\int_{\mathscr{X}} \exp\left\{\sum_{j=1}^{k} w''_j T_j(x)\right\} d\mu_1(x)\right]^{1-\alpha}
\end{aligned}$$

最后一个不等式是采用了 Hölder 不等式，由于 $\boldsymbol{w}', \boldsymbol{w}'' \in \Omega$，所以不等式右边两个因子中的积分存在，从而得到 $\boldsymbol{w} \in \Omega$，这说明了 Ω 是凸集，证毕.

几个常见分布的指数族形式

正态(可以发现指数族的判定与样本容量无关)

$$
\begin{aligned}
f(x,\theta) &= (\sqrt{2\pi}n)^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right) \\
&= (\sqrt{2\pi}n)^{-n}\exp\left(-\frac{n\mu^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i-\frac{1}{2\sigma^2}\sum x_i^2\right)
\end{aligned}
$$

正态分布自然参数空间:

$\left\{\left(\phi_1(x)=\frac{\mu}{\sigma^2},\phi_2(x)=-\frac{1}{2\sigma^2}\right),\mu\in\mathbb{R},\sigma^2\in\mathbb{R}^+\right\}=\{(\phi_1,\phi_2):\phi_1\in\mathbb{R},\phi_2<0\}$

指数

$$
\begin{aligned}
f(x,\theta) &= \prod\left(\lambda e^{-\lambda x_i}\mathbb{1}_{(0,\infty)}(x_i)\right) \\
&= \lambda^n\exp\left(-\lambda\sum x_i\right)\prod\mathbb{1}_{(0,\infty)}(x_i)
\end{aligned}
$$

指数分布自然参数空间:$\{-\lambda\}=(-\infty,0)$

Gamma(γ,λ) γ已知

$$
\begin{aligned}
f(x,\theta) &= \prod\left(\frac{\lambda^\gamma}{\Gamma(\gamma)}x_i^{\gamma-1}e^{-\lambda x_i}\mathbb{1}_{(0,\infty)}(x_i)\right) \\
&= \frac{\lambda^{n\gamma}}{\Gamma(\gamma)^n}\exp\left(-\lambda\sum x_i\right)\prod x_i^{\gamma-1}\mathbb{1}_{(0,\infty)}(x_i)
\end{aligned}
$$

Gamma(γ,λ)分布自然参数空间:$\{-\lambda\}=(-\infty,0)$

两点分布$b(1,p)$

$$
\begin{aligned}
f(x,\theta) &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\
&= (1-p)^n\left(\frac{p}{1-p}\right)^{\sum x_i} \\
&= (1-p)^n\exp\left(\ln\left(\frac{p}{1-p}\right)\sum x_i\right)
\end{aligned}
$$

两点分布$b(1,p)$自然参数空间：$\left\{\ln\left(\frac{p}{1-p}\right),p\in(0,1)\right\}=\mathbb{R}$

泊松$\text{Poisson}(\lambda)$

$$
\begin{aligned}
F(X,\theta) &= \frac{e^{-n\lambda}}{\prod(x_i)!}\lambda^{\sum x_i} \\
&= e^{-n\lambda}\exp\left(\ln(\lambda)\sum x_i\right)\frac{1}{\prod(x_i)!}
\end{aligned}
$$

泊松$\text{Poisson}(\lambda)$分布自然参数空间：$\{\ln(\lambda),\lambda\in\mathbb{R}\}=\mathbb{R}$

指数族的性质

1. 如果一族分布属于指数族，那么ta的分布函数的支撑集与θ无关：
   $\{x : f(x) > 0\} = \{x : h(x) > 0\}$
   通过这一性质(必要条件)可以看出均匀分布不是指数族：支撑集与区间参数θ相关

统计量的几个性质

充分统计量 s.s.

Definition: If the conditional distribution $P_\theta(\boldsymbol{X} \in A | T(\boldsymbol{X}) = t)$ does not depend on θ for any Borel set A, then $T(\boldsymbol{X})$ is a sufficient statistic.

idea:统计量$T(X)$包含了关于参数θ的全部信息，即：如果两个样本X, Y满足$T(X) = T(Y)$那么依据这两个样本关于参数θ的信息是一样的
判别s. s. 的**充要条件**：因式分解定理——统计量分布函数可以因式分解到以下形式

$$f(\boldsymbol{x}, \theta) = g(t(\boldsymbol{x}), \theta) h(\boldsymbol{x}),$$

例子：对于从任意一维分布族$\mathscr{F}$中的一个总体F抽出的一组样本X, 其对应的次序统计量$T(X) = (T_{(1)}(X) \cdots T_{(n)}(X))$是s. s.
因为样本的联合分布pdf总可以写成
$f(x_1 \cdots x_n) = f(x_{(1)}) \cdots f(x_{(n)}) = g(T(X), \theta) h(X), h(X) \equiv 1$
根据因式分解定理，可以证明$T(X)$是s. s.
引理：对充分统计量$T(X)$的可逆映射的像$\phi(T)$也是s. s.

🆓最小充分统计量 minimal s.s.

此处的最小指的是映射关系:每一次映射过程造成了信息的减少/至多不变。
$\sigma_T \subset \sigma_S$

Minimal Sufficient Statistic– Assume T is a sufficient statistic of distribution family $\mathscr{F}$. If for any sufficient statistic $S(\boldsymbol{X})$, there always exist a function $q_S(\cdot)$ such that $T(\boldsymbol{X}) = q_S(S(\boldsymbol{X}))$, then we call $T(\boldsymbol{X})$ the minimal sufficient statistic of $\mathscr{F}$.

完全统计量 c.s.

idea：在span(T)中，零无偏估计量只有0本身

Definition: Assume $\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$ is a distribution family. Θ is the parameter space. Assume $T = T(X)$ is a statistic, if for any $\varphi(T(X))$ which satisfies

$$E_\theta \varphi(T(X)) = 0 \text{ for any } \theta \in \Theta,$$

then

$$P(\varphi(T(X)) = 0) = 1 \text{ for any } \theta \in \Theta.$$

We call $T(X)$ a complete statistic (CS) of θ.

定理：对于**指数分布族**，只要**变形后得到的自然参数空间有内点**，则变形后得到的$T(X)$就是c. s. s.

Theorem: Assume the density of $\boldsymbol{X} = (X_1, ..., X_n)^{\mathrm{T}}$ is

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} \theta_i T_i(\boldsymbol{x})\right\} h(\boldsymbol{x}), \quad \boldsymbol{\theta} = (\theta_1, ..., \theta_k)^{\mathrm{T}} \in \Theta^*.$$

Let $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), ..., T_k(\boldsymbol{X}))^{\mathrm{T}}$. If there are some inner points in $\Theta^* \in R^k$, then $T(\boldsymbol{X})$ is a complete statistic.

引理：对c. s. s. $T(X)$的可逆映射的像$\phi(T)$也是c. s. s.

Basu 定理

使用Basu定理的题目 - 知乎 (zhihu.com)

简述：**有界完备充分统计量(bounded c.s.s.)与辅助统计量(分布与参数无关)独立**，对于任意$\theta \in \Theta$

- Definition: We call $T(X)$ a bounded complete statistic if for any bounded or almost bounded everywhere function $\varphi(\cdot)$ which satisfies

$$E_\theta \varphi(T(X)) = 0 \text{ for any } \theta \in \Theta,$$

then $P(\varphi(T(X)) = 0) = 1$ for any $\theta \in \Theta$ is true.

- (Basu Theorem) Assume $\mathscr{F} = \{F_\theta(x), \theta \in \Theta\}$ is a distribution family, $\boldsymbol{X} = (X_1, \dots, X_n)^{\mathrm{T}}$ is a simple random sample selected from $\mathscr{F}$. $T(X)$ is a bounded complete statistic and sufficient statistic. If the distribution of random variable $V(\boldsymbol{X})$ does not depend on θ, then for any $\theta \in \Theta$, $V(\boldsymbol{X})$ and $T(\boldsymbol{X})$ are independent.

引理：自然参数形式的指数分布族之中对于Basu定理的应用

- Lemma: Assume the density of $\boldsymbol{X} = (X_1, \dots, X_n)^{\mathrm{T}}$ is

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} \theta_i T_i(\boldsymbol{x})\right\} h(\boldsymbol{x}) \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^{\mathrm{T}} \in \Theta^*.$$

Let $T(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \dots, T_k(\boldsymbol{X}))^{\mathrm{T}}$. If the distribution of random variable $V(\boldsymbol{X})$ does not depend on θ, then for any $\theta \in \Theta$, $V(\boldsymbol{X})$ and $T(\boldsymbol{X})$ are independent.

解题思路总结

- 首先给出样本联合条件分布函数
- 将其因式分解(根据因式分解定理得到s. s.)
- 证明完备(恒成立证明)

    - 进一步设定零条件期望函数，写出期望表达式
        - 通过求导等手段证其恒为0，得到c. s. s.
    - 对于指数族：直接分解到自然形式的指数族得到c. s. s.
- 证明不完备(存在性证明)
    - 构造不恒为零的零条件期望函数
- 利用Basu定理证明有界完备s. s. 与辅助统计量独立

点估计

描述点估计量的几个性质

无偏性 unbiased

$$\begin{aligned}&E(\hat{g}(X))=g(\theta)\quad for\ any\ \theta\in\Theta\quad \text{无偏}\\&\lim_{n\to\infty}E(\hat{g_n}(X))=g(\theta)\ for\ any\ \theta\in\Theta\quad \text{渐近无偏}\end{aligned}$$

有效性 efficient

对于无偏估计$\hat{g}(X),\tilde{g}(X)$

$$\begin{aligned}&D_\theta\hat{g}(X)\le D_\theta\tilde{g}(X)\quad for\ any\ \theta\in\Theta\\&D_\theta\hat{g}(X)<D_\theta\tilde{g}(X)\quad for\ some\ \theta\in\Theta\end{aligned}$$

一致性/相合性 consistent

$$\begin{aligned}&\lim_{n\to\infty}P_\theta(|\hat{g}_n(X)-g(\theta)|\ge\epsilon)=0\ for\ any\ \theta\in\Theta\quad \text{strong consistency}\\&P_\theta(\lim_{n\to\infty}\hat{g}_n(X)=g(\theta))=0\ for\ any\ \theta\in\Theta\quad \text{weak consistency}\\&\lim_{n\to\infty}E_\theta(|\hat{g}_n(X)-g(\theta)|)^r=0\ for\ any\ \theta\in\Theta\quad \text{r}^{\text{th}}\text{ mean concistenct}\end{aligned}$$

🆓一致渐进正态估计 CAN

Slutsky引理

对于$r.v.$序列$\{X_n\},\{Y_n\}$:如果$X_n\xrightarrow{\mathscr{D}}X,Y_n\xrightarrow{P}c$, 那么：

$$X_n\pm Y_n\xrightarrow{\mathscr{D}}X\pm c\quad X_n\cdot Y_n\xrightarrow{\mathscr{D}}cX\quad \frac{X_n}{Y_n}\xrightarrow{\mathscr{D}}\frac{X}{c}$$

矩法(MM)

本质就是将要估计的$g(\theta)$用总体的各阶矩(中心矩、原点矩)表示$g(\theta)=g(\alpha,\mu)$，然后对应的矩估计量就是换成对应的样本矩$g(\hat{\alpha},\hat{\mu})$

性质

1. 依据“原点样本矩是无偏的；中心样本矩是渐进无偏的”可以判断矩估计量的无偏性
2. 依据“点样本矩、中心样本矩是强相合估计”可以判断矩估计量的相合性
    1. 定理：a.s.收敛的r.v.经过连续变换后依旧a.s.收敛；注意原点矩可以视为原点矩的连续映射的像，应用此定理可以证明其强相合性
3. CAN(一致渐近正态分布)

极大似然估计(MLE)

本质是**基于已获得样本**的联合条件分布(似然函数)的值推断总体参数取值(样本取自哪一个总体)的可能性(似然性)大小：这个值越大，样本取自该总体的可能性越高。

求法

先求样本的联合分布函数
如果**关于参数可微**，那么就用微积分的方法求解(出于计算方便，我们一般以对数似然函数为目标函数)，得到的方程可以解出$\hat{\theta}(X)_{MLE}$
如果**关于参数不可微**，那就根据定义来分析(e.g.均匀分布$U(0,\theta)$的MLE是$X_{(n)}$)

性质

1. 如果总体是指数族，只要似然方程的解是自然参数空间的内点，那么MLE的解唯一
2. MLE总可以表示为s.s.$T(X)$**的函数**：根据**因式分解定理**：$L(x,\theta)=\prod f(x_i,\theta)=\prod g(T,\theta)h(x_i)$，对$\theta$的优化集中在$g(T,\theta)$中。

估计效果的评价，UMVUE

解释：U-一致(对于全体θ)；MV-最小方差；UE-无偏估计(可估的)

为什么用UMVUE?

缩小范围(仅仅考虑无偏估计)为保证存在性。本来是考虑全体统计量取最小MSE者，但是不好求

UE改进

对于任意一个无偏估计$g(X)$ 若存在一个充分统计量$T(X)$ 那么有：

$$\begin{aligned}&E_\theta(h(T))=E_\theta(E(g(X)|T))=E_\theta(g(X))=g(\theta)\\&D_\theta(h(T):=E(g(X)|T(X)))\le D_\theta(g(X))\end{aligned}$$

等号当且仅当$g(X)=T(X),a.s.$时成立

$$\begin{aligned}&D_\theta g(X)\\&=E_\theta(g(X)-g(\theta))^2\\&=E_\theta[(g(X)-h(T))+(h(T)-g(\theta))]^2\\&=E_\theta(g(X)-h(T))^2+2E_\theta(g(X)-h(T))(h(T)-g(\theta))+D_\theta h(T)^2\\&=E_\theta(g(X)-h(T))^2+D_\theta h(T)^2\end{aligned}$$

$$\begin{aligned}&E_\theta(g(X)-h(T))(h(T)-g(\theta))\\&=E_\theta\{E(g(X)-h(T))(h(T)-g(\theta))|T\}\\&=E_\theta\{(h(T)-g(\theta))E(g(X)-h(T)|T)\}\\&=0\end{aligned}$$

UMVUE 判别

零无偏准则

idea:对于任意一个无偏估计$g(X)$，我们似乎总可以插入一些“零无偏的噪声”ϵ使得估计的效果变差；但如果有一个无偏估计$g(X)^*$对于各种噪声都是免疫的：插入任何噪声都不会影响ta的估计效(用方差衡量的话)，那么ta自然就是一个UMVUE。这种“免疫性”用数学表示就是

$$\mathrm{Cov}_\theta(g(X)^*,l(X))=0,\forall l(X):E_\theta(l(X))=0,\forall\theta$$

⭐注意，零无偏估计量全体所对应的集合$\{l(X):E_\theta l(X)=0\}$是一个线性空间

⭐这是一个UMVUE判别的充要条件

充分性:

对于任一$\hat{g}(X)$总可以表示成$g(X)^*+l(X)$的形式：

$$\begin{aligned}D_\theta(\hat{g}(X)) &= D_\theta(g(X)^* + l(X)) \\ &= D_\theta(g(X)^*) + D_\theta(l(X)) + 2Cov_\theta(g(X)^*, l(X)) \\ &= D_\theta(g(X)^*) + D_\theta(l(X)) \\ &\geq D_\theta(g(X)^*)\end{aligned}$$

必要性:
反证，如果一个UMVUE:$g(X)$与一个零无偏估计量$l(X)$条件协方差不为0, 那么可以构造一个条件方差更小的无偏估计：$g(X)^* = g(X) - al(X)$

$$\begin{aligned}D_\theta(g(X)^*) &= D_\theta(g(X) - al(X)) \\ &= D_\theta(g(X)) + a^2 D_\theta(l(X)) - 2aCov_\theta(g(X), l(X)) \\ &\leq D_\theta(g(X)) \iff a^2 D_\theta(l(X)) - 2aCov_\theta(g(X), l(X)) \leq 0\end{aligned}$$

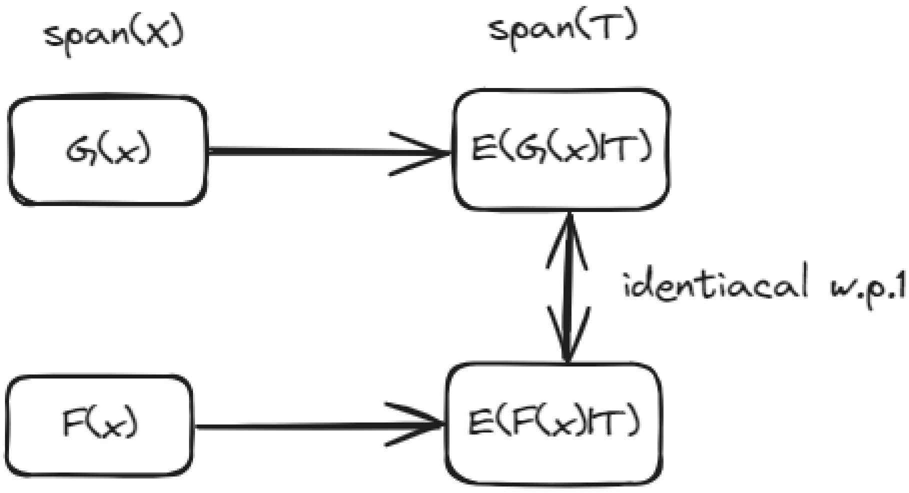
由此，我们总可以找到一个满足条件的a得到一个更有效的无偏估计，与UMVUE条件矛盾。所以零无偏条件成立

如果我们现在知道一个充分统计量$T(X)$，那么我们可以把视角放在$\text{span}(T)$上面，因为我们总可以借助$T(X)$优化无偏估计$g(X)$。此时零无偏条件变为：

$$\text{Cov}_\theta(h(T)^*, \delta(T)) = 0, \forall \delta(T) : E_\theta(\delta(T)) = 0, \forall \theta$$

c.s.s.方法，L-S定理

如果我们已知一个c. s. s. $T(X)$那么我们可以证明从ta**映射出的无偏估计量是唯一的UMVUE** (e. g. 我们得到了一个无偏估计$g(X)$, 那么$h(T) = E(g(X)|T)$就是唯一的UMVUE
⭐这是一个基于c. c. s. 的UMVUE判别的充分条件
最小方差：对于任一个无偏估计$g(X)$, 总可以通过s. s. $T(X)$优化
$h(T) = E(g(X)|T)$
唯一性：从c. s. 的性质出发：对于$\text{span}(T)$中的两个无偏估计$\tilde{g}(X), \hat{g}(X)$，ta们的差是零无偏估计量，由于$T(X)$是c. s. $\text{span}(T)$中零无偏估计量只有0 所以这两个无批判估计在概率层面上相等



证 先证唯一性. 设 $\hat{g}_1(T(\boldsymbol{X}))$ 为 $g(\theta)$ 的任一无偏估计, 令 $\delta(T(\boldsymbol{X})) = \hat{g}(T(\boldsymbol{X})) - \hat{g}_1(T(\boldsymbol{X}))$, 则 $E_\theta\delta(T(\boldsymbol{X})) = E_\theta\hat{g}(T(\boldsymbol{X})) - E_\theta\hat{g}_1(T(\boldsymbol{X})) = 0,\ \theta \in \Theta$. 由 $T(\boldsymbol{X})$ 为完全统计量, 可知 $\delta(T(\boldsymbol{X}))=0$, a.s. P_θ 成立, 即 $\hat{g}(T(\boldsymbol{X}))=\hat{g}_1(T(\boldsymbol{X}))$, a.s. P_θ 成立, 故唯一性成立.

再证一致最小方差性. 设 $\varphi(\boldsymbol{X})$ 为 $g(\theta)$ 的任一无偏估计. 令 $h(T(\boldsymbol{X})) = E(\varphi(\boldsymbol{X})|T)$, 由 $T(\boldsymbol{X})$ 为充分统计量, 故知 $h(T(\boldsymbol{X}))$ 与 θ 无关, 是统计量. 由引理 3.4.1 可知

$$E_\theta(h(T(\boldsymbol{X}))) = g(\theta), \qquad 一切\ \theta \in \Theta,$$
$$D_\theta(h(T(\boldsymbol{X}))) \leqslant D_\theta(\varphi(\boldsymbol{X})), \qquad 一切\ \theta \in \Theta.$$

由唯一性得 $\hat{g}(T(\boldsymbol{X})) = h(T(\boldsymbol{X}))$, a.e. P_θ 成立, 由上式可知

$$D_\theta(\hat{g}(T(\boldsymbol{X}))) = D_\theta(h(T(\boldsymbol{X}))) \leqslant D_\theta(\varphi(\boldsymbol{X})), \quad 一切\ \theta \in \Theta,$$

所以 $\hat{g}(T(\boldsymbol{X}))$ 为 $g(\theta)$ 的 UMVUE, 且唯一.

估计效果的边界，C-R不等式

$$D_\theta[\hat{g}(X)] \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

C-R不等式满足的一系列必要条件(**C-R分布族**)：

- The parameter space is an open set in R;
- The support set does not depend on θ;
- For any $x \in \boldsymbol{X}$ and $\theta \in \Theta$, $\frac{\partial f(x,\theta)}{\partial \theta}$ exists;
- $\frac{\partial}{\partial \theta} \int f(x, \theta)\, dx = \int \frac{\partial}{\partial \theta} f(x, \theta)\, dx$;
- The fisher information $I(\theta) = E_\theta \left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right]^2$ exists and $0 < I(\theta) = E_\theta \left[\frac{\partial \log f(X,\theta)}{\partial \theta}\right]^2 < \infty$.

C-R不等式**完整内容**:

Theorem: Assume $\mathscr{F} = \{f(x, \theta), \theta \in \Theta\}$ be a C-R distribution family. $g(\theta)$ is defined on Θ. Assume $\boldsymbol{X} = (X_1, ..., X_n)$ is a simple sample from $f(x, \theta)$, $\hat{g}(\boldsymbol{X})$ is an unbiased estimator of $g(\theta)$, and satisfies

$$\int \cdots \int \hat{g}(\boldsymbol{x}) f(\boldsymbol{x}, \theta)\, d\boldsymbol{x}$$

can be taken derivative under the integration, then

$$D_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{(g'(\theta))^2}{nI(\theta)} \quad \textit{for any } \theta \in \Theta.$$

e. g. $g(\theta) = \theta \to D_\theta(\hat{\theta}(X)) \geq \frac{1}{nI(\theta)}$
这个式子表明给定样本，对于总体参数的估计存天然的下限，这一下限与总体分布相关(Fisher信息量)

🆓证明：归根到底还是Cauchy不等式

> 书上的证明很有趣，但是有点长，理解欣赏就好

区间估计

描述置信区间的性质

1. 置信度：我们用区间估计水平的一致下限：置信系数来表示
   def :区间估计$[\hat{\theta}_1, \hat{\theta}_2]$的置信系数 $= \inf\limits_{\theta \in \Theta} P_\theta(\hat{\theta}_1 \le \theta \le \hat{\theta}_2)$
2. 置信水平α：外生给定的；只要一个区间估计的置信度不小于$1-\alpha$，则可以说ta具有$1-\alpha$的置信水平的
3. 置信限：当区间估计是单侧的，形如$(-\infty, \hat{\theta}_1], [\hat{\theta}_2, \infty)$，此时也可以给定一个置信水平$\alpha$来定义一个具有$1-\alpha$的置信水平的置信限
4. 精度：也就是区间的大小(在一维空间上就是长度)

枢轴变量法构造置信区间

✍️ **Pivotal Variable Method**

所谓枢轴变量就是一个r. v. $\phi(T(X), \theta)$，其中$T(X)$是对θ的一个点估计；$\phi(T(X), \theta)$的分布与θ无关，从而我们可以依据这一分布来给出一个与θ无关的，关于$\phi(\cdot)$具有$1-\alpha$水平的置信区间；然后再进行变量代换就可以得到关于θ的$1-\alpha$水平的置信区间。

正态总体下的估计

	一个正态总体	
待估参数	参数信息	枢轴变量及其分布
μ	σ^2已知	$\sqrt{n}\frac{\bar{x}-\mu}{\sigma^2} \sim N(0,1)$
	σ^2未知	$\sqrt{n}\frac{\bar{x}-\mu}{S^2} \sim t_{n-1}$
σ^2	μ已知	$\frac{n\sum(x_i-\mu)^2}{\sigma^2} \sim \chi^2_n$
	μ未知	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

		两个正态总体$X \sim N(a, \sigma_1^2), Y \sim N(b, \sigma_2^2)$
待估参数	样本/参数信息	枢轴变量及其分布
b-a	样本容量一样	设$Z_i = X_i + Y_i \sim N(b-a, \sigma_1^2 + \sigma_2^2)$
	σ_1^2, σ_2^2已知	$\bar{Y} - \bar{X} \sim N\left(b-a, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$
	$\sigma_1^2 = \sigma_2^2 = \sigma^2$未知	$\frac{\bar{Y}-\bar{X}-(b-a)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2}$
	其他情况	用大样本CLT近似, $\mathrm{Z-stat}: \frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{\frac{S_x^2}{m}+\frac{S_y^2}{n}}}$

非正态总体下的估计

分布类型	枢轴变量及其分布	CI(α)
指数分布 Exp(λ)	$2\lambda\sum x_i \sim \chi^2_{2n}$	$\left[\chi^2_{2n}(1-\frac{\alpha}{2}) \le 2\lambda\sum x_i \le \chi^2_{2n}(\frac{\alpha}{2})\right]$
均匀分布 $U(0,\theta)$	$z=\frac{\theta}{X_{(n)}} \xrightarrow{pdf} nz^{-(n+1)}\mathbb{1}_{(1,\infty)}(z)$	$\left[1 \le \frac{\theta}{X_{(n)}} \le \frac{1}{\sqrt[n]{a}}\right]$

假设检验(这里仅涉及参数检验)

注意一个点：我们的假设是对于参数空间的，而检验是在样本空间上做的

什么是假设检验？假设检验速览

1. 参数空间上的假设命题$H_0 \to H_1$
   一般地，H_0是我们人为设定的，也就是我们所感兴趣的假设；从数学上说，H_0可以表示为命题$\theta \in \Theta_0 \subset \Theta$，对应地，$H_1$可以表示为命题$\theta \in \Theta_1 = \Theta\ /\Theta_0$ ，也就是取补集。根据Θ_0是否是单点集，可以把ta分为简单假设和复合假设。
2. 基于一份样本的检验：
   现在，我们通过简单那随机抽样得到了一份样本$X_1,\cdots X_n$，我们可以先假设H_0成立，然后从样本中提取一些信息，即计算一个统计量$T(X,\theta_0),\theta_0 \in \Theta_0$，根据边界值以及$T$的分布所决定的：样本空间$\mathscr{S}$上的拒绝域$D$，来决定“根据这一样本，是否要拒绝原假设”。
   现在，我们终于可以定义检验函数$\phi(T)$来描述一个假设检验：

$$\phi(T)=\begin{cases}1 \iff T\in D\\ r \iff T\in\partial D:\text{这一条仅限离散型用到}\\ 0 \iff T\notin D\end{cases}$$

   当然，对于用什么统计量是很有说法的。有一种情况我们单独拿出来讨论：当我们基于MLE，我们要用的统计量就变成了似然比λ。此时似然比的表达式一般都很复杂，不会有很好的分布性质。但我们可以从中提取一个分布良好的统计量T作为原像，并找到一个双射$g(\cdot)$，使得$\lambda=g(T)$。此时$P(\lambda=c)=P(g(T)=c)=P(T=g^{-1}(c))$，从而我们可以把分析对象改为$T$，得到可以求解的$\phi(T)$。
3. 如何评价一个检验
   由于样本的随机性，我们不能保证基于样本判定假设正误的绝对可靠性，对此我们用该检验**犯错误**的概率衡量：

	$\phi(T)=0$	$\phi(T)=1$
$\theta\in\Theta_0$	√	$\mathrm{I}-\mathrm{error}$
$\theta\in\Theta_1$	$\mathrm{II}-\mathrm{error}$	√

我们对此引入**功效函数(势函数)**：给定总体参数条件下的拒绝概率 $P(\phi=1|\theta)\overset{def}{=}\beta_\phi(\theta)$，ta与两类错误的发生概率的关系是：

$$\beta_\phi(\theta)=\mathbb{1}_{\theta\in\Theta_0}P(\mathrm{I}-\mathrm{error})+\mathbb{1}_{\theta\in\Theta_1}(1-P(\mathrm{II}-\mathrm{error}))$$

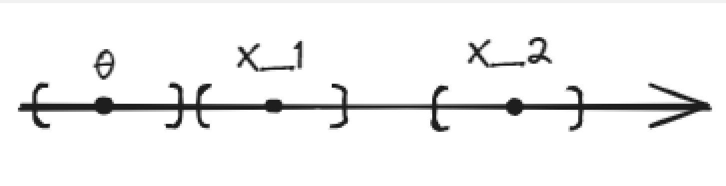
作为一个优秀的检验，我们希望$\phi(\cdot)$所带来的两类错误的概率都越小越好，但是这很难做到：一般地，降低$\mathrm{I}-\mathrm{error}$的方法是缩小$D$，但是这一操作往往会导致$\mathrm{II}-\mathrm{error}$的发生概率升高。
面对这一权衡，我们依据$\mathrm{Neyman}-\mathrm{Pearson}$原则:在约定$\mathrm{I}-\mathrm{error}$发生概率上限(显著性水平)的条件下最小化$\mathrm{II}-\mathrm{error}$的发生概率，此时对应最优检验称为$\mathrm{UMPT}$。这么做的动机是：我们认为$\mathrm{I}-\mathrm{error}$的成本更高，我们更不希望这一类错误的发生

正态总体假设检验:基于统计量与原假设的"距离"判断、

✍ 对于复合原假设，我们需要额外证明$\forall\theta_0\in\Theta_0,\beta_\phi(\theta_0)\le\alpha$，也就是证明功效函数的单调性

检验 & 置信区间

置信区间是基于样本给出的参数真实值的区间估计，不止于判决原假设的正误，置信区间还给出了一定置信水平下，真实参数的取值范围与原假设的差距大小：

> 下图"X_1"是第一份样本得出的$(1-\alpha)$置信区间，下图"X_1"是第二份样本得出的$(1-\alpha)$置信区间，θ是简单原假设的参数值。可以看出第二份样本"X_2"下得出的θ估计显著异于原假设，而"X_1"的差距就没那么显著
>
> 
>

检验 & p-值

基于样本和原假设，我们可以算出一个检验统计量$T(X,\theta_0)$，此时可以计算出在原假设成立条件下，抽样到比已有样本更“偏离原假设”的样本的概率：

$$P(|T(X_{new},\theta_0)|>|T(X,\theta_0)|)\overset{def}{=}\mathrm{p-value}$$

> 如果这个值很小，说明在原假设条件下，抽到已有样本的概率很低，从而说明原假设不太可信，可以拒绝
>
> 
>

检验结论总结(书上表格)

表 5.2.1　单个正态总体均值的假设检验

	H_0	H_1	检验统计量及其分布	否定域
σ^2 已知	$\mu=\mu_0$	$\mu\neq\mu_0$	$U=\sqrt{n}(\bar{X}-\mu_0)/\sigma$ $U\vert\mu=\mu_0\sim N(0,1)$	$\vert U\vert>u_{\alpha/2}$
	$\mu\leqslant\mu_0$	$\mu>\mu_0$		$U>u_\alpha$
	$\mu\geqslant\mu_0$	$\mu<\mu_0$		$U<-u_\alpha$
σ^2 未知	$\mu=\mu_0$	$\mu\neq\mu_0$	$T=\sqrt{n}(\bar{X}-\mu_0)/S$ $T\vert\mu=\mu_0\sim t_{n-1}$	$\vert T\vert>t_{n-1}(\alpha/2)$
	$\mu\leqslant\mu_0$	$\mu>\mu_0$		$T>t_{n-1}(\alpha)$
	$\mu\geqslant\mu_0$	$\mu<\mu_0$		$T<-t_{n-1}(\alpha)$

表 5.2.2　单个正态总体方差的假设检验

	H_0	H_1	检验统计量及其分布	否定域
μ 已知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi_\mu^2=nS_\mu^2/\sigma_0^2$ $\chi_\mu^2\vert\sigma_0^2\sim\chi_n^2$	$nS_\mu^2/\sigma_0^2<\chi_n^2(1-\alpha/2)$ 或 $nS_\mu^2/\sigma_0^2>\chi_n^2(\alpha/2)$
	$\sigma^2\leqslant\sigma_0^2$	$\sigma^2>\sigma_0^2$		$nS_\mu^2/\sigma_0^2>\chi_n^2(\alpha)$
	$\sigma^2\geqslant\sigma_0^2$	$\sigma^2<\sigma_0^2$		$nS_\mu^2/\sigma_0^2<\chi_n^2(1-\alpha)$
μ 未知	$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$ $\chi^2\vert\sigma_0^2\sim\chi_{n-1}^2$	$(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha/2)$ 或 $(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha/2)$
	$\sigma^2\leqslant\sigma_0^2$	$\sigma^2>\sigma_0^2$		$(n-1)S^2/\sigma_0^2>\chi_{n-1}^2(\alpha)$
	$\sigma^2\geqslant\sigma_0^2$	$\sigma^2<\sigma_0^2$		$(n-1)S^2/\sigma_0^2<\chi_{n-1}^2(1-\alpha)$

表 5.2.3　两个正态总体均值差的假设检验

	H_0	H_1	检验统计量及其分布	否定域
σ_1^2 σ_2^2 已知	$\mu_2-\mu_1=\mu_0$	$\mu_2-\mu_1\neq\mu_0$	$U=\dfrac{\bar{Y}-\bar{X}-\mu_0}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}$ $U\vert\mu_0\sim N(0,1)$	$\vert U\vert>u_{\alpha/2}$
	$\mu_2-\mu_1\leqslant\mu_0$	$\mu_2-\mu_1>\mu_0$		$U>u_\alpha$
	$\mu_2-\mu_1\geqslant\mu_0$	$\mu_2-\mu_1<\mu_0$		$U<-u_\alpha$
σ_1^2 σ_2^2 未知	$\mu_2-\mu_1=\mu_0$	$\mu_2-\mu_1\neq\mu_0$	$T_w=\dfrac{\bar{Y}-\bar{X}-\mu_0}{S_w}\sqrt{\dfrac{mn}{m+n}}$ $T_w\vert\mu_0\sim t_{n+m-2}$ $S_w^2=\dfrac{(m-1)S_1^2+(n-1)S_2^2}{n+m-2}$	$\vert T_w\vert>t_{n+m-2}\left(\dfrac{\alpha}{2}\right)$
	$\mu_2-\mu_1\leqslant\mu_0$	$\mu_2-\mu_1>\mu_0$		$T_w>t_{n+m-2}(\alpha)$
	$\mu_2-\mu_1\geqslant\mu_0$	$\mu_2-\mu_1<\mu_0$		$T_w<-t_{n+m-2}(\alpha)$

表 5.2.4　两个正态总体方差比的假设检验

	H_0	H_1	检验统计量及其分布	否定域
μ_1 μ_2 已知	$\sigma_2^2=\sigma_1^2$	$\sigma_2^2\neq\sigma_1^2$	$F_*=S_{2*}^2/S_{1*}^2$ $F_*\vert_{\sigma_2^2=\sigma_1^2}\sim F_{n,m}$ $S_{1*}^2=\dfrac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2$ $S_{2*}^2=\dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\mu_2)^2$	$F_*<F_{n,m}(1-\alpha/2)$ 或 $F_*>F_{n,m}(\alpha/2)$
	$\sigma_2^2\leqslant\sigma_1^2$	$\sigma_2^2>\sigma_1^2$		$F_*>F_{n,m}(\alpha)$
	$\sigma_2^2\geqslant\sigma_1^2$	$\sigma_2^2<\sigma_1^2$		$F_*<F_{n,m}(1-\alpha)$
μ_1 μ_2 未知	$\sigma_2^2=\sigma_1^2$	$\sigma_2^2\neq\sigma_1^2$	$F=S_2^2/S_1^2$ $F\vert_{\sigma_2^2=\sigma_1^2}\sim F_{n-1,m-1}$ $S_1^2=\dfrac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2$ $S_2^2=\dfrac{1}{n-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2$	$F<F_{n-1,m-1}(1-\alpha/2)$ 或 $F>F_{n-1,m-1}(\alpha/2)$
	$\sigma_2^2\leqslant\sigma_1^2$	$\sigma_2^2>\sigma_1^2$		$F>F_{n-1,m-1}(\alpha)$
	$\sigma_2^2\geqslant\sigma_1^2$	$\sigma_2^2<\sigma_1^2$		$F<F_{n-1,m-1}(1-\alpha)$

似然比检验: literally

如果我们用似然程度去检验原假设的话，那么就是接下来要记录的似然比检验。如果原假设条件下的“极大似然值”和无约束条件下的“极大似然值”(也就是正常求解MLE的过程所得到的结果)**很接近**，那么我们就倾向于接受原假设；反之则拒绝原假设。

$$L_{\Theta_0}(x) = \sup_{\theta \in \Theta_0} f(x, \theta)$$
$$L_{\Theta}(x) = \sup_{\theta \in \Theta} f(x, \theta)$$

$$\lambda(x) = \frac{L_{\Theta}(x)}{L_{\Theta_0}(x)} \geq 1 \text{ (似然比)}$$

$$\phi(x) = \begin{cases} 1, \lambda(x) > c \\ r, \lambda(x) = c \text{ (检验函数)} \\ 0, \lambda(x) < c \end{cases}$$

由于$\lambda(x)$未必有很好的分布，我们通常从$\lambda(x)$的表达式中提取一个**具有良好分布的充分统计量**：T 使得$\lambda(x) = g(T)$，其中$g(\cdot)$是**单调函数**，从而可以将$\phi(x)$改写为$\phi(T)$的形式，更好得到边界值。

> **书上有很多例子，自己动手算吧，没啥能记的**

这里再记一个关于$\lambda(x)$极限分布的Wilks定理

Theorem

Assume the dimensions of Θ and Θ_0 are k and s respectively. If $k - s = t > 0$ and the density function of the sample satisfies the CR-conditions in Page 109, then for the testing problem $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_1$, we have

$$2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi_t^2$$

under H_0 and as $n \to \infty$.

Dimension: Suppose the sample are generated from $N(\mu, \sigma^2)$, and we want to test $H_0 : \mu = 0 \leftrightarrow H_1 : \mu \neq 0$. Then the dimension of $\Theta_0 = \{\theta = (0, \sigma^2), \sigma^2 > 0\}$ is 1; the dimension of $\Theta = \{\theta = (\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0\}$ is 2.

UMPT & UMPUT

> 直到现在，我们还没有像在点估计那里一样，对假设检验进行优化、比较。现在我们给出评价一个假设检验的具体内容。

由于我们以犯第二类错误的概率评价，我们可以由此定义水平为α的UMPT : ϕ

$$
\begin{aligned}
&\forall \theta_0 \in \Theta_0 : \beta_\phi(\theta_0) \le \alpha : \phi \in \Phi(\alpha) \\
&\forall \theta_1 \in \Theta_1, \phi_1 \in \Phi(\alpha) : \beta_\phi(\theta_1) \ge \beta_{\phi_1}(\theta_1)
\end{aligned}
$$

也就是ϕ不犯第二类错误的概率在所有水平为α的检验中最低

N-P 引理

那么如何得到一个UMPT？只有在一定条件下$\mathrm{N-P\ Lemma}$才能给出一个基于(原始定义下的)似然比检验的UMPT

> 书上的证明很有趣，但是有点长，理解欣赏就好

(NP Lemma) Suppose the density function of sample $\mathbf{X}$ is $f(\mathbf{x}, \theta)$ and there are only two possible values for θ. Consider the following testing problem:

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1, \tag{3}$$

then for any given $0 < \alpha < 1$, we have

Existence & UMPT: There must exist testing function $\varphi(\mathbf{x})$, and fixed numbers $c, r > 0$ satisfies: (i).

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \text{if } f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) > c; \\ r, & \text{if } f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) = c; \\ 0, & \text{if } f(\mathbf{x}, \theta_1)/f(\mathbf{x}, \theta_0) < c. \end{cases} \tag{4}$$

(ii) $E_{\theta_0}[\varphi(\mathbf{X})] = \alpha$. This $\varphi(\mathbf{X})$ is the UMPT of (3).

基于N-P引理求一般假设问题的UMPT

基于$\mathrm{N-P\ Lemma}$我们可以逐步推出其他假设形式的UMPT：举一个例子

$$
\begin{aligned}
&\{H_0 : \mu = 0, H_1 : \mu = \mu_1 > 0\}\text{可以通过引理求出UMPT} : \phi \\
\to &\{H_0 : \mu = 0, H_1 : \mu > 0\}\text{的UMPT是}\phi, \text{如果UMPT和}\mu\text{无关} \\
\to &\{H_0 : \mu \le 0, H_1 : \mu > 0\}\text{的UMPT是}\phi, \text{如果}\beta(x)\text{关于}\mu\text{单增} \\
&\text{也就是}\phi\text{在}H_0\text{上具有}\alpha\text{的检验水平}
\end{aligned}
$$

一般情况下就是：

$$
\begin{aligned}
&\{H_0:\mu\in\Theta_0,H_1:\mu\in\Theta_1\}\text{的UMPT是}\phi,\text{如果}\beta(x)\text{关于}\mu\text{单调}\\
&\text{也就是}\phi\text{在}H_0\text{上具有}\alpha\text{的检验水平}\\
\leftarrow&\{H_0:\mu=\partial\Theta_0,H_1:\mu\in\Theta_1\}\text{的UMPT是}\phi,\text{如果UMPT和}\mu\text{无关}\\
\leftarrow&\{H_0:\mu=\partial\Theta_0,H_1:\mu=\mu_1\in\Theta_1\}\text{可以通过引理求出UMPT}:\phi
\end{aligned}
$$

这一套流程的必要条件是：

1. 原假设只能有一个边界点，具体说就是参数空间是$\mathbb{R}$子集，且原假设是单边的
2. $\beta(x)$在H_0上关于参数单调
3. $\phi(x)$与表达式参数无关
   而对于指数族$\mathscr{F}=\{f(x,\theta)=c(\theta)\exp(Q(\theta)T(x))h(x)\}$，功效函数的单调性与$Q(\theta)$的单调性有关（详见定理5.4.2/3）

UMPUT

至于UMPUT，无非是进一步限制了考虑的检验函数的范围，从而更有可能求出一个最优的检验。毕竟UMPT不是总能求出来的。

Definition

Suppose φ is a test of testing problem $H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1$. If the power function of φ—$\beta_\varphi(\theta)$ satisfies that: for any $\theta_0\in\Theta_0$, we have $\beta_\varphi(\theta_0)\le\alpha$, and for any $\theta_1\in\Theta_1$, we have $\beta_\varphi(\theta_1)\ge\alpha$, then we call φ as an unbiased test with significance level α.

Definition

Denote $\mathcal{U}_\alpha$ as a set contains all the unbiased tests with significance level α for testing problem $H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1$. Suppose $\varphi\in\mathcal{U}_\alpha$, and for any $\varphi_1\in\mathcal{U}_\alpha$, we have

$$\beta_\varphi(\theta)\ge\beta_{\varphi_1}(\theta)\quad\theta\in\Theta_1,$$

then we call φ as the uniformly most powerful unbiased test.

所谓无偏检验满足以下条件的检验函数$\phi(x)$：

$$\forall\theta_0\in\Theta_0,\theta_1\in\Theta_1,\beta_\phi(\theta_0)\le\alpha\le\beta_\phi(\theta_1)$$

即犯第一类错误的概率不大于**不犯**第二类错误的概率

- If φ is an UMPT, then it must be an UMPUT. Let $\varphi^* \equiv \alpha$, then φ^* is a test with significance level α such that $\beta_\varphi(\theta_1) \geq \beta_{\varphi^*}(\theta_1) \equiv \alpha$, therefore

$$\beta_\varphi(\theta_1) \geq \alpha \geq \beta_\varphi(\theta_0).$$

- The UMPUT of the following testing problems exist.

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0,$$

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \leftrightarrow H_1 : \theta < \theta_1 \quad \text{or} \quad \theta > \theta_2.$$

> 此处的$\phi^* \equiv \alpha$指的是，无论样本如何，均有α概率随机拒绝原假设(类似于掷骰子)；最后两类检验问题的UMPT不存在

指数族分布专题

[六大常用分布的矩估计和最大似然估计推导过程_常见分布的似然估计量-CSDN博客]

指数族分布	样本联合pdf
$N(\mu, \sigma^2)$	$f(x,\theta) = (\sqrt{2\pi}n)^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$
$\text{Exp}(\lambda)$	$f(x,\theta) = \prod\left(\lambda e^{-\lambda x_i}\mathbb{1}_{(0,\infty)}(x_i)\right)$
$\Gamma(\gamma,\lambda), \gamma$已知	$f(x,\theta) = \frac{\lambda^{n\gamma}}{\Gamma(\gamma)^n}\exp\left(-\lambda\sum x_i\right)\prod x_i^{\gamma-1}\mathbb{1}_{(0,\infty)}(x_i)$
$b(1,p)$	$f(x,\theta) = p^{\sum x_i}(1-p)^{n-\sum x_i}$
$\text{Poisson}(\lambda)$	$f(x,\theta) = \frac{e^{-n\lambda}}{\prod(x_i)!}\lambda^{\sum x_i}$

指数族分布	自然参数形式pdf	自然参数空间
$N(\mu, \sigma^2)$	$(\sqrt{2\pi}n)^{-n}\exp\left(-\frac{n\mu^2}{2\sigma^2}\right)\exp\left(\frac{\mu}{\sigma^2}\sum x_i - \frac{1}{2\sigma^2}\sum x_i^2\right)$	$\mathbb{R}\times\mathbb{R}^-$
$\text{Exp}(\lambda)$	$\lambda^n \exp\left(-\lambda\sum x_i\right)\prod\mathbb{1}_{(0,\infty)}(x_i)$	$\mathbb{R}^-$
$\Gamma(\gamma,\lambda), \gamma$已知	$\frac{\lambda^{n\gamma}}{\Gamma(\gamma)^n}\exp\left(-\lambda\sum x_i\right)\prod x_i^{\gamma-1}\mathbb{1}_{(0,\infty)}(x_i)$	$\mathbb{R}^-$
$b(1,p)$	$(1-p)^n \exp\left(\ln\left(\frac{p}{1-p}\right)\sum x_i\right)$	$\mathbb{R}$

指数族分布	自然参数形式pdf	自然参数空间
Poisson(λ)	$e^{-n\lambda}\exp\left(\ln(\lambda)\sum x_i\right)\frac{1}{\prod(x_i)!}$	$\mathbb{R}$

指数族分布	$N(\mu,\sigma^2)$	Exp(λ) 均值$\frac{1}{\lambda}$	$b(1,p)$	Poisson(λ)
c. s. s.	$\left(\sum x_i, \sum x_i^2\right)$	$\sum x_i$	$\sum x_i$	$\sum x_i$
MLE	$\left(\bar{X}, \frac{\sum(x_i-\bar{x})^2}{n}=\frac{(n-1)S^2}{n}\right)$	$\bar{X}$	$\bar{X}$	$\bar{X}$
UMVUE	$\left(\bar{X}, S^2=\frac{\sum x_i^2}{n-1}\right)$	$\bar{X}$	$\bar{X}$	$\bar{X}$

❗均匀分布专题

😩少有的不属于指数族的常见分布

对于均匀分布，我们经常通过线性变换把区间参数消去

$X\sim U(a,b)\to\frac{X-a}{b-a}\sim U(0,1)$

$X\sim U(0，\theta)\to\frac{X}{\theta}\sim U(0,1)$

$X\sim U\left(-\frac{\theta}{2},\frac{\theta}{2}\right)\to\frac{X}{\theta}\sim U\left(-\frac{1}{2},\frac{1}{2}\right)$

$U(0,1)$ 的次序统计量

对于$U(0,1)$，我们还有进一步的结论：围绕Beta分布

[数理统计4：均匀分布的参数估计，次序统计量的分布，Beta分布 - 江景景景页 - 博客园]()

[如何通俗理解 beta 分布？ - 知乎]()

次序统计量：

$f_{(1)}(x)=n(1-x)^{n-1}=\frac{\Gamma(n+1)}{\Gamma(1)\Gamma(n)}x^{1-1}(1-x)^{n-1}$

$\sim Be(1,n=\mathbf{1+(n-1)})$

$f_{(n)}(x)=n(x)^{n-1}=\frac{\Gamma(n+1)}{\Gamma(1)\Gamma(n)}(x)^{n-1}(1-x)^{1-1}$

$\sim Be(n=\mathbf{1+(n-1)},1)$

$f_{(m)}(x)=\frac{n!}{(m-1)!(n-m)!}(1-x)^{n-m}x^{m-1}=\frac{\Gamma(n+1)}{\Gamma(m)\Gamma(n-m+1)}(1-x)^{n-m+1-1}x^{m-1}$

$\sim Be(m=\mathbf{1+(m-1)B,1+(n-m)})$

极差：

$f_{(n)}-f_{(1)}\overset{pdf}{=}n(n-1)r^{n-2}(1-r)=\frac{\Gamma(n+1)}{\Gamma(n-1)\Gamma(2)}r^{n-1-1}(1-r)^{2-1}\sim Be(n-1,2)$

极差与$f_{(n-1)}$同分布，等效于容量为$(n-1)$的样本的最大值的分布(to be justified, just an inspiration)

对于$Be(a,b)$：

$$E[X]=\frac{a}{a+b}=\mu$$
$$Var[X]=\frac{ab}{(a+b)^2(a+b+1)}=\frac{\mu(1-\mu)}{\alpha+\beta+1}=\frac{\mu(1-\mu)}{\varphi+1}$$

$U(0,\theta)$ 的估计

s.s. & c.s.s. $X_{(n)}$

其中充分统计量的证明可以直接用因式分解(次序统计量这是普适的充分统计量)
而完全统计量的证明则需要按照定义引入零无偏估计量
$\phi(X_{(n)}):E_\theta(\phi(X_{(n)}))=0$然后证明$\phi=0,\ a.s.$（书上例题2.8.2）

MLE $X_{(n)}$

这个是因为样本的条件联合分布函数$f=\theta^{-n}\mathbb{1}_{X_{(n)}\le\theta}(x)$，关于$\theta$递增，所以应该是分母尽可能的小；同时可以发现$\theta<X_{(n)}\to f(x)=0$，所以$\theta$的下界就是$X_{(n)}$

UE & UMVUE $\frac{1+n}{n}X_{(n)}$

这个就算一下X_n的期望，然后对其代数变换一下就好了；至于UMVUE，是从上面c. s. s. 引出来的(用L-S定理)

均匀分布$U(0,\theta)$	结果
s. s.	$X_{(n)}$
c. s. s.	$X_{(n)}$
MLE	$X_{(n)}$
UMVUE	$\frac{1+n}{n}X_{(n)}$