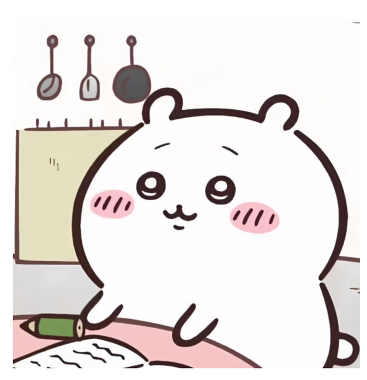
数理统计 第三章 区间估计



区间估计的基本概念

区间估计	对两个统计量$T_1(X),T_2(X)$,若$T_1(X)\leqslant T_2(X)$,则称$(T_1(X),T_2(X))$为区间估计
置信系数	若一个参数的区间估计$(\hat\theta_L(X),\hat\theta_U(X))$，称$\inf\limits_{\theta\in\Theta}P_\theta\left\{\hat\theta_L(X)\leqslant\theta\leqslant\hat\theta_U(X)\right\}$为此估计区间的置信系数
置信区间	$(\hat\theta_L(X),\hat\theta_U(X))$为$\theta$的一个区间估计，对于给定一个$\alpha\in(0,1)$,如果$P_\theta\left\{\hat\theta_L(X)\leqslant\theta\leqslant\hat\theta_U(X)\right\}\geqslant 1-\alpha$对$\forall\theta\in\Theta$成立 则称$(\hat\theta_L(X),\hat\theta_U(X))$为$\theta$的一个置信水平为$1-\alpha$的置信区间(confident interval) 特别地，如果$P_\theta\left\{\hat\theta_L(X)\leqslant\theta\leqslant\hat\theta_U(X)\right\}=1-\alpha$对$\forall\theta\in\Theta$成立，则称$(\hat\theta_L(X),\hat\theta_U(X))$为$\theta$的一个置信水平为$1-\alpha$的同等置信区间
置信限	望目指标：有目标值　区间 望大指标：越大越好　置信下限$\hat\theta_L:P\left\{\theta\geqslant\hat\theta_L(X)\right\}\geqslant 1-\alpha,\forall\theta\in\Theta$ 望小指标：越小越好　置信上限$\hat\theta_U:P\left\{\theta\leqslant\hat\theta_U(X)\right\}\geqslant 1-\alpha,\forall\theta\in\Theta$
置信域	$\theta\in\mathbb{R}^K$,若$P\{\theta\in S(X)\}\geqslant 1-\alpha,\forall\theta\in\Theta$,称区域$S(X)$为$\theta$的置信水平为$1-\alpha$的置信域
分位数	$X\sim F$称$\xi_p=\inf\{X:F(x)\geqslant p\}$，若$F(x)$连续且严增，$\xi_p=F^{-1}(p),P\{X\leqslant\xi_p\}=p$
	U_α称为$N(0,1)$的(上侧)α分位数，由分布对称,$U_\alpha=-U_{1-\alpha}$

用枢轴量发求置信区间的基本步骤

1.找一个参数θ的好的点估计$T(X)$(一般MLE)(充分统计量)

2.构造$S(T,\theta)$，要求其不含参数，分布$T(x)\sim f_T(t,\theta)$完全已知，称$S(T,\theta)$为枢轴量(感觉就是通过性质往，χ^2,t,F或者充分统计量的已知PDF上凑一凑)

3.给定α，令$P_\theta\{a\leqslant S(T,\theta)<b\}=1-\alpha$，求出$a,b$(取法不唯一，一般$a=F_{1-\frac{\alpha}{2}},b=F_{\frac{\alpha}{2}}$),原则上
- a,b取法对称(尾部概率对称)
  $\Downarrow$
  $P\{S(T,\theta)\leqslant c)\}=P\{S(T,\theta)\leqslant d)\}=\dfrac{\alpha}{2}$
- 尽可能使$b-a$短
  $\Downarrow$
  条件利用充分

4.由$a\leqslant S(T,\theta)<b$反解$\hat\theta_L(X)\leqslant\theta\leqslant\hat\theta_U(X)$，则$P\{a\leqslant S(T,\theta)<b\}=P\left\{\hat\theta_L\leqslant\theta\leqslant\hat\theta_U\right\}=1-\alpha$

条件一定要充分利用

单样本正态总体的置信区间,$X_1,\dots,X_n\overset{\text{IID}}{\sim}N(\mu,\sigma^2)$

求μ的$C.I$	σ^2已知	$\left(\overline{X}-\dfrac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}},\overline{X}+\dfrac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}}\right)$	$P\left\{\left\lvert\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right\rvert\leqslant u_{\frac{\alpha}{2}}\right\}=1-\alpha$ $\sim N(0,1)$
	σ^2未知	$\left(\overline{X}-\dfrac{S_n}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1),\overline{X}+\dfrac{S_n}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1)\right)$	$P\left\{\left\lvert\dfrac{\sqrt{n}(\overline{X}-\mu)}{S_n}\right\rvert\leqslant t_{\frac{\alpha}{2}}(n-1)\right\}=1-\alpha$ $\sim t(n-1)$
求σ^2的$C.I$	μ已知	$\left(\dfrac{\sum(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\dfrac{\sum(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$	$P\left\{\chi^2_{1-\frac{\alpha}{2}}(n)\leqslant\dfrac{\sum(X_i-\mu)^2}{\sigma^2}\leqslant\chi^2_{\frac{\alpha}{2}}(n)\right\}=1-\alpha$ $\sim\chi^2(n)$
	μ未知	$\left(\dfrac{(n-1)S_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\dfrac{(n-1)S_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$	$P\left\{\chi^2_{1-\frac{\alpha}{2}}(n-1)\leqslant\dfrac{(n-1)S_n^2}{\sigma^2}\leqslant\chi^2_{\frac{\alpha}{2}}(n-1)\right\}=1-\alpha$ $\sim\chi^2(n)$

两样本正态总体的置信区间

1.对μ_1,μ_2的估计($B-F$问题)

$X_1,\dots,X_n\overset{\text{IID}}{\sim}N(\mu_1,\sigma_1^2),\overline{X}=\dfrac{1}{m}\sum X_i,S_{1m}^2=\dfrac{1}{m-1}\sum(X_i-\overline{X})^2$

$Y_1,\dots,Y_n\overset{\text{IID}}{\sim}N(\mu_2,\sigma_2^2),\overline{Y}=\dfrac{1}{m}\sum Y_i,S_{1m}^2=\dfrac{1}{m-1}\sum(Y_i-\overline{Y})^2$

σ_1,σ_2均已知	$\overline{X}-\overline{Y}\mp u_{\frac{\alpha}{2}}\left(\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}\right)$
$\sigma_1^2=\sigma_2^2=\sigma^2$未知	$\overline{X}-\overline{Y}\mp\dfrac{\sqrt{m+n}}{\sqrt{mn(m+n-2)}}t_{\frac{\alpha}{2}}(m+n-2)\sqrt{(m-1)S_{1m}^2+(n-1)S_{2n}^2}$
$\dfrac{\sigma_2^2}{\sigma_1^2}=\theta$已知	$\overline{X}-\overline{Y}\mp\dfrac{\sqrt{m\theta+n}}{\sqrt{mn(m+n-2)}}t_{\frac{\alpha}{2}}(m+n-2)\sqrt{(m-1)S_{1m}^2+(n-1)S_{2n}^2/\theta}$
$m=n$	$\overline{z}\mp\dfrac{\sqrt{\sum(z_i-\overline{z})}}{\sqrt{n(n-1)}}t_{\frac{\alpha}{2}}(n-1)$
m,n均充分大	$\overline{X}-\overline{Y}\mp u_{\frac{\alpha}{2}}\sqrt{S_{1m}^2/m+S_{2n}^2/n}$

2.对$\dfrac{\sigma_1^2}{\sigma_2^2}$的估计

μ_1,μ_2均未知时	$\left[\dfrac{S_{1m}^2/S_{2n}^2}{F_{\frac{\alpha}{2}}(m-1,n-1)},\dfrac{S_{1m}^2/S_{2n}^2}{F_{1-\frac{\alpha}{2}}(m-1,n-1)}\right]$
μ_1,μ_2均已知时	$\left[\dfrac{S_{1m}^{*2}/S_{2n}^{*2}}{F_{\frac{\alpha}{2}}(m,n)},\dfrac{S_{1m}^{*2}/S_{2n}^{*2}}{F_{1-\frac{\alpha}{2}}(m,n)}\right]$

一些特殊情况下的置信区间

1.无法用常用分位数表示

例：$X_1,\dots,X_n\overset{\text{IID}}{\sim}U(0,\theta)$,求$\theta$的$C.I$

θ的$MLE:\hat\theta=X_{(n)}$，$X_{(n)}\sim f_n(t)=\dfrac{1}{\theta^n}nt^{n-1},t\in(0,\theta)$

令$S=\dfrac{X_\theta}{\theta},f_s(s)=ns^{n-1},s\in(0,1)$，令$P\left\{a\leqslant\dfrac{X_{(n)}}{\theta}\leqslant b\right\}=1-\alpha$，$X_{(n)}$分布无法用常用分位数来表示

但$\left\{a\leqslant\dfrac{X_{(n)}}{\theta}\leqslant b\right\}=1-\alpha\iff\left\{\dfrac{X_{(n)}}{b}\leqslant\theta\leqslant\dfrac{X_{(n)}}{a}\right\}=1-\alpha,C.I$的区间长度为$\left(\dfrac{1}{a}-\dfrac{1}{b}\right)X_{(n)}$

求CI的过程中令$\displaystyle\int_a^b ns^{n-1}\mathrm{d}s$,需使得$\dfrac{1}{a}-\dfrac{1}{b}$取 min，解得$b=1,a=\sqrt[n]{\alpha},C.I$为$\left(X_{(n)},\dfrac{X_{(n)}}{\sqrt[n]{\alpha}}\right)$

2.置信区间只有单侧

例：$X_1,\dots,X_n\overset{\text{IID}}{\sim}\mathrm{E}(\lambda)$,求平均寿命$\dfrac{1}{\lambda}$的$C.I$

对产品寿命进行估计，估计区间只有下界

$T=\sum X_i\Rightarrow\dfrac{1}{\lambda},\mathrm{E}(\lambda)=\Gamma(1,\lambda),T=\sum X_i\sim\Gamma(n,\lambda),2\lambda T\sim\Gamma\left(n,\dfrac{1}{2}\right)=\chi^2(2n)$，要求的形式：$P\left\{\dfrac{1}{\lambda}\geqslant\hat\theta_L(x)\right\}=1-\alpha$(只求下限)

因此令$P\{2\lambda T\leqslant c\}=1-\alpha,2\lambda T\sim\chi^2(2n)\Rightarrow c=\chi^2_\alpha(2n)\Rightarrow P\left\{a\lambda T\leqslant\chi^2_\alpha(2n)\right\}=1-\alpha\iff P\left\{\dfrac{1}{\lambda}\geqslant\dfrac{2T}{\chi^2_\alpha(2n)}\right\}=1-\alpha$

则$\dfrac{1}{\lambda}$的置信度为$1-\alpha_n$的置信下限：$\dfrac{2T}{\chi^2_\alpha(2n)}$即$C.I$为$\left(\dfrac{2T}{\chi^2_\alpha(2n)},+\infty\right)$

3.无法直接计算C,I，只能在大样本情况下使用渐进分布

例：$X_1,\dots,X_n\overset{\text{IID}}{\sim}b(1,p)$,求$p$的$C.I$

$T=\sum X_i\Rightarrow p,T=\sum X_i\sim B(n,p)$，无法构造枢轴量，

但知道极限分布$\dfrac{T-np}{\sqrt{npq}}\xrightarrow{\text{L}}N(0,1)$，则$P\left\{-u_{\frac{\alpha}{2}}\leqslant\dfrac{T-np}{\sqrt{npq}}\leqslant u_{\frac{\alpha}{2}}\right\}=1-\alpha$

由$-u_{\frac{\alpha}{2}}\leqslant\dfrac{T-np}{\sqrt{npq}}\leqslant u_{\frac{\alpha}{2}}$,解得$p$的$C.I$为$\dfrac{n}{n+u_{\frac{\alpha}{2}}^2}\left(\overline{X}+\dfrac{u_{\frac{\alpha}{2}}^2}{2n}\mp\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}+\dfrac{u_{\frac{\alpha}{2}}^2}{4n^2}}\right)$

例：$X_1,\dots,X_n\overset{\text{IID}}{\sim}\mathrm{p}(\lambda)$,求$\lambda$的$C.I$

$T=\sum X_i\sim P(n\lambda)$,类似地,$\dfrac{T-n\lambda}{\sqrt{n\lambda}}\xrightarrow{\text{L}}N(0,1)$，$\left\lvert\dfrac{T-n\lambda}{\sqrt{n\lambda}}\right\rvert\leqslant u_{\frac{\alpha}{2}}\Rightarrow\lambda$的$C.I$为$\overline{X}+\dfrac{u_{\frac{\alpha}{2}}^2}{2n}\mp u_{\frac{\alpha}{2}}\sqrt{\dfrac{u_{\frac{\alpha}{2}}^2}{4n^2}+\dfrac{\overline{X}}{n}}$

信念分布

设有一样本$X\sim N(\theta,1)$，则$\overline{X}\sim N\left(\theta,\dfrac{1}{n}\right)\Rightarrow\sqrt{n}(\overline{X}-\theta)\sim N(0,1)$

$P\left\{\sqrt{n}(\overline{X}-\theta)\right\}=\Phi(t)\Rightarrow P\left\{\theta>\overline{X}-\dfrac{t}{\sqrt{n}}\right\}=\Phi(t)\Rightarrow P\left\{\theta<\overline{X}-\dfrac{t}{\sqrt{n}}\right\}=1-\Phi(t)$

$P\{\theta<u\}=1-\Phi\left(\sqrt{n}\left(\overline{X}-\mu\right)\right)$　如果将θ视作随机变量：$F(\theta)=1-\Phi\left(\sqrt{n}\left(\overline{X}-\mu\right)\right)$

但注意到θ是常数，不能存在分布，我们称此处的$F(\theta)$为信念分布(fiducial distribution)