概率论与数理统计 第11次作业

Name: 宋昊原 **Student ID: 2022010755**

December 3, 2023

1 合格率区间估计

根据中心极限定理，近似地有

$$\frac{p^* - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

其中

$$p^* = 0.6$$

$$n = 100$$

以p^*估计分母中的p，可以得到$1-\alpha$置信的区间估计：

$$(p^* - \sqrt{\frac{p^*(1-p^*)}{n}}\Phi^{-1}(1-\frac{\alpha}{2}), p^* + \sqrt{\frac{p^*(1-p^*)}{n}}\Phi^{-1}(1-\frac{\alpha}{2}))$$

代入

$$\alpha = 0.05$$

解得0.95置信区间为

$$(0.504, 0.696)$$

2 幂函数分布的估计

2.1

似然函数

$$L(\theta) = \prod_{i=1}^{n} \theta X_i^{\theta-1} = \theta^n (\prod_{i=1}^{n} X_i)^{\theta-1}$$

对数似然函数

$$l(\theta) = \ln L(\theta) = n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln X_i$$

故

$$l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\ln X_i$$

故

$$\theta^* = -\frac{n}{\sum\limits_{i=1}^{n}\ln X_i}$$

这是最大值点.
代入数据有

$$\theta^* = 0.7983$$

下面求Fisher信息量

$$I(\theta) = E((\frac{d}{d\theta}(\ln\theta + (\theta-1)\ln x))^2)$$

$$= \int_0^1 (\frac{1}{\theta} + \ln x)^2 \theta x^{\theta-1} dx$$

$$= \int_0^1 \frac{x^{\theta-1}}{\theta} dx + 2\int_0^1 x^{\theta-1} \ln x dx + \int_0^1 x^{\theta-1} (\ln x)^2 dx$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}$$

标准差的估计值

$$\sigma_n = \sqrt{\frac{1}{nI(\theta)}} \approx \sqrt{\frac{1}{nI(\theta^*)}} = \frac{\theta^*}{\sqrt{n}} = 0.3992$$

3 正态总体的估计

3.1

对数似然函数

$$l(\sigma) = -n \ln(\sqrt{2\pi}\sigma) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}$$

令

$$l'(\sigma) = -\frac{n}{\sigma} + \frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{\sigma^3} = 0$$

解得

$$\sigma^* = \sqrt{\frac{\sum\limits_{i=1}^{n} (X_i - \mu)^2}{n}}$$

3.2

计算Fisher信息量

$$I(\sigma) = E((\frac{\partial \ln f}{\partial \sigma})^2) = E((-\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3})^2)$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma^3} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx - 2\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sqrt{2\pi}\sigma^5} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx + \int_{-\infty}^{+\infty} \frac{(x-\mu)^4}{\sqrt{2\pi}\sigma^7} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx$$

$$= \frac{1}{\sigma^2} - \frac{2Var(N(0,1))}{\sigma^2} + \frac{Kurt(N(0,1))}{\sigma^2} = \frac{2}{\sigma^2}$$

故σ^*标准误差的估计

$$\sqrt{\frac{1}{nI(\sigma^*)}} = \frac{\sigma}{\sqrt{2n}}$$

由于近似地有

$$\frac{\sigma^* - \sigma}{\frac{\sigma}{\sqrt{2n}}} \sim N(0,1)$$

可以给出其$1-\alpha$置信区间

$$(\sigma^* - \frac{\sigma^*}{\sqrt{2n}} \Phi^{-1}(1-\frac{\alpha}{2}), \sigma^* + \frac{\sigma^*}{\sqrt{2n}} \Phi^{-1}(1-\frac{\alpha}{2}))$$

其中

$$\Phi^{-1}$$

为标准正态累积分布函数的反函数.

4 方差不同样本的差值估计

近似地

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

以S_1^2, S_2^2估计σ_1^2, σ_2^2，有

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim N(0,1)$$

于是$1-\alpha$置信区间估计为

$$(\bar{X}-\bar{Y}-\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\Phi^{-1}(1-\frac{\alpha}{2}), \bar{X}-\bar{Y}+\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\Phi^{-1}(1-\frac{\alpha}{2}))$$

代入数据计算得

$$(-3.14,\ -0.90)$$

5 后验约会

θ的先验分布为

$$f_\Theta(\theta)=1,\ \theta\in(0,1)$$

条件分布

$$f_{X|\Theta}(x|\theta)=\frac{1}{\theta},\ x\in(0,\theta)$$

于是X的边际分布

$$f_X(x)=\int_x^1\frac{1}{\theta}d\theta=-\ln x$$

于是后验分布

$$f_{\Theta|X}(\theta|x)=\frac{1\times\frac{1}{\theta}}{-\ln x}=-\frac{1}{\theta\ln x},\ \theta\in(x,1)$$

6 硬币的后验估计

后验分布密度函数为

$$f(\theta)=\frac{\theta^x(1-\theta)^{n-x}}{\int_0^1\theta^x(1-\theta)^{n-x}d\theta}=\frac{\theta^x(1-\theta)^{n-x}}{B(x+1,n-x+1)}$$

求对数密度函数的极值，令

$$\frac{d\ln f(\theta)}{d\theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta}=0$$

解得

$$\hat{\theta}=\frac{x}{n}$$

在此题中为

$$\hat{\theta}=\frac{13}{20}=0.65$$

事实上此后验分布密度函数与似然函数成正比，因此与极大似然估计得到的结论相同.

7 正态分布的后验估计

首先求出其后验分布.
先验分布

$$f_M(\mu)=\frac{1}{\sqrt{2\pi}\sigma_0}\exp(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})$$

条件分布

$$f_{X_1,...,X_n|M}(x_1,...,x_n|\mu)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x_i-\mu)^2}{2\sigma^2})=\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2})$$

边际分布

$$f_{X_1,...,X_n}(x_1,...,x_n)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})d\mu=I(\mathbf{x})$$

与μ无关.
故后验分布为

$$g(\mu)=\frac{\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})}{I(\mathbf{x})}$$

7.1 最大后验估计

$$\frac{\partial\ln g(\mu)}{\partial\mu}=-\frac{\sum\limits_{i=1}^{n}(\mu-x_i)}{\sigma^2}-\frac{(\mu-\mu_0)}{\sigma_0^2}$$

令上述导数为0，则

$$\hat{\mu}_1=\frac{n\sigma_0^2\bar{X}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}$$

7.2 后验均值估计

$$\hat{\mu}_2=\int_{-\infty}^{+\infty}g(\mu)d\mu=\int_{-\infty}^{+\infty}\frac{\frac{\mu}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})}{I(\mathbf{x})}d\mu$$

$$=\int_{-\infty}^{+\infty}\frac{\frac{\mu}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-(\frac{n}{2\sigma^2}+\frac{1}{2\sigma_0^2})\mu^2+(\frac{n\bar{X}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2})\mu-C_1(\mathbf{X}))}{I(\mathbf{X})}d\mu$$

其中$C_1(\mathbf{X})$与μ无关.

令

$$\hat{\mu}_1=\frac{n\sigma_0^2\bar{X}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}$$

$$\gamma=\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}$$

则

$$\hat{\mu}_2=\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^nI(\mathbf{X})}\int_{-\infty}^{+\infty}\mu\exp(-\frac{\gamma}{2}(\mu-\hat{\mu}_1)^2+C_2(\mathbf{X}))d\mu$$

其中$C_2(\mathbf{X})$为配方的剩余项，与μ无关.

此时将$d\mu$调整至与指数函数内的部分相同，得

$$= \frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n I(\mathbf{X})}\frac{1}{\gamma}\int_{-\infty}^{+\infty}\exp(-\gamma(\mu-\hat{\mu}_1)^2+C_2(\mathbf{X}))d(-\gamma(\mu-\hat{\mu}_1)^2+C_2(\mathbf{X}))$$

$$+\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n I(\mathbf{X})}\int_{-\infty}^{+\infty}\hat{\mu}_1\exp(-\frac{\gamma}{2}(\mu-\hat{\mu}_1)^2+C_2(\mathbf{X}))d\mu$$

$$= 0+\frac{\hat{\mu}_1 I(\mathbf{X})}{I(\mathbf{X})}=\hat{\mu}_1$$

故后验均值估计和最大后验估计得到同一结果:

$$\hat{\mu}_1=\frac{n\sigma_0^2\bar{X}+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}$$

8 几何分布的贝叶斯估计

8.1 后验分布

先验分布

$$f_\Theta(\theta)=1,\ \theta\in(0,1)$$

用$\mathbf{X}$表示样本向量，则条件分布

$$f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)=\theta^n(1-\theta)^{\sum\limits_{i=1}^n x_i}=\theta^3(1-\theta)^{10}$$

边际分布

$$f_{\mathbf{X}}(\mathbf{x})=\int_0^1\theta^3(1-\theta)^{10}d\theta=\int_0^1(1-\theta)^3\theta^{10}d\theta$$

$$=\int_0^1(\theta^{10}-3\theta^{11}+3\theta^{12}-\theta^{13})d\theta=\frac{1}{11}-\frac{3}{12}+\frac{3}{13}-\frac{1}{14}=\frac{1}{4004}$$

故后验分布

$$g(\theta)=4004\theta^3(1-\theta)^{10},\ \theta\in(0,1)$$

8.2 后验均值分布

$$\hat{\theta}=\int_0^1 4004\theta^4(1-\theta)^{10}d\theta=\int_0^1 4004(1-\theta)^4\theta^{10}d\theta$$

$$=4004(\frac{1}{11}-\frac{4}{12}+\frac{6}{13}-\frac{4}{14}+\frac{1}{15})=\frac{4}{15}$$

9 Bayes区间估计

9.1

根据第7题，后验分布为

$$g(\mu)=\frac{\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^n(x_i-\mu)^2}{2\sigma^2}-\frac{(\mu-\mu_0)^2}{2\sigma_0^2})}{I(\mathbf{X})}$$

$$=\frac{\frac{1}{\sqrt{2\pi}\sigma_0(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\gamma}{2}(\mu-\hat{\mu}_1)^2+C_2(\mathbf{X}))}{I(\mathbf{X})}$$

这符合正态分布的形式，于是后验分布为

$$\mu \sim N(\hat{\mu}_1, \frac{1}{\gamma})$$

其中

$$\hat{\mu}_1 = \frac{n\sigma_0^2\bar{X} + \sigma^2\mu_0}{n\sigma_0^2 + \sigma^2}$$

$$\tau^2 := \frac{1}{\gamma} = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

于是

$$a = \hat{\mu}_1 - \tau\Phi^{-1}(1 - \frac{\alpha}{2}) = \frac{n\sigma_0^2\bar{X} + \sigma^2\mu_0}{n\sigma_0^2 + \sigma^2} - \sqrt{\frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}}\Phi^{-1}(1 - \frac{\alpha}{2})$$

$$b = \hat{\mu}_1 + \tau\Phi^{-1}(1 - \frac{\alpha}{2}) = \frac{n\sigma_0^2\bar{X} + \sigma^2\mu_0}{n\sigma_0^2 + \sigma^2} + \sqrt{\frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}}\Phi^{-1}(1 - \frac{\alpha}{2})$$

其中Φ^{-1}为标准正态累积分布函数的反函数.

9.2

$\sigma_0 \to \infty$时，有

$$a = \bar{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})$$

$$b = \bar{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})$$

这与经典方法得到的结论一致.

事实上，$\sigma_0 \to \infty$说明先验分布的方差为无穷大，即先验分布实际上类似于一个“$\mathbb{R}$上的均匀分布”，这相当于我们对μ一无所知的情况下得到的结论.

9.3

条件分布

$$f_{\mathbf{X}|M}(\mathbf{x}|\mu) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2})$$

边际分布

$$f_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{+\infty} f(\mu)\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2})d\mu$$

故后验分布

$$g(\mu) = \frac{f(\mu)\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2})}{\int_{-\infty}^{+\infty} f(\mu)\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp(-\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2})d\mu}$$

$f(\mu)$是“常数”，故

$$g(\mu|\mathbf{X}) = \frac{\exp(-\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2})}{\int_{-\infty}^{+\infty}\exp(-\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2})d\mu} = \frac{\exp(-\frac{n\mu^2 - 2n\bar{X}\mu + C(\mathbf{X})}{2\sigma^2})}{\int_{-\infty}^{+\infty}\exp(-\frac{n\mu^2 - 2n\bar{X}\mu + C(\mathbf{X})}{2\sigma^2})d\mu}$$

其中$C(\mathbf{X})$与μ无关，这符合正态分布形态，故后验分布为

$$\mu \sim N(\bar{X}, \frac{\sigma^2}{n})$$

最大后验区间和等尾可信区间均为

$$(\bar{X} - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}), \bar{X} + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2}))$$

这与第二问结论相同，因为认为先验分布是$f(\mu) \propto 1$实际上等价于对μ一无所知.

10 计算机实验：自助法Bootstrap（续）

10.1 θ的区间估计

在第10次作业计算机实验中的模拟程序中，求出$\hat{\theta}^*$，并求其0.025和0.975分位数，得到区间估计为

$$(123.000526,\ 184.629933)$$

10.2 区间估计的其他方式

可以使用课上提到的极大似然区间估计的方法，借助Fisher信息量$I(\theta)$来估计$Var(\theta)$.