数据科学与工程数学基础

作业提交规范及第 6 次作业

教师：黄定江
助教：刘文辉、徐艺玮

2023 年 12 月 28 日

作业提交规范

1. 作业提交形式：建议使用 Word 或 LaTeX 编写所得到的电子文档。若使用 Word 编写，将其另存为 PDF 形式，然后提交 PDF 文档。若使用 LaTeX 编写，将其编译成 PDF 形式，然后提交 Tex 和 PDF 两个文档。若使用手写拍照的方式，请务必规整好各题解答的图片，并整合在一个 PDF 文档中，只发图片格式的作业概不批改！
2. 作业命名规范：提交的电子文档必须命名为："**学号_姓名**"。命名示例：52200000000_刘某某。
3. 作业提交途径：点击打开每次作业的传送门地址：**第 6 次作业提交传送门**，无需注册和登录，按要求输入个人学号和姓名，然后上传作业文档即可。注意：传送门将会在截至时间点到达后自动关闭。
4. 作业更改说明：如果需要修改已经提交的作业，只要在截至日期前，再次上传更改后的作业（切记保持同名），即可覆盖已有作业。
5. 作业评分说明：正常提交作业的按照实际评分记录；逾期补交作业的根据逾期情况在实际评分基础上酌情扣分；未交作业的当次作业记为 0 分；若发现存在抄袭的作业时，相似的两份作业均会被记为 0 分。

第 6 次作业

提交截至时间：**2023/12/19 下周二 12:00（中午）**

理论部分

习题 1. *(互信息) 假设* $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \cdots \rightarrow X_n$ 是一个马尔科夫链，即

$$p\left(x_1, x_2, \ldots, x_n\right)=p\left(x_1\right) p\left(x_2 \mid x_1\right) \cdots p\left(x_n \mid x_{n-1}\right)$$

试化简 $I\left(X_1 ; X_2, \ldots, X_n\right)$

解.

$$\begin{aligned}
I\left(X_1 ; X_2, \ldots, X_n\right) & =H\left(X_1\right)-H\left(X_1 \mid X_2, \ldots, X_n\right) \\
& =H\left(X_1\right)-\left[H\left(X_1, X_2, \ldots, X_n\right)-H\left(X_2, \ldots, X_n\right)\right] \\
& =H\left(X_1\right)-\left[\sum_{i=1}^n H\left(X_i \mid X_{i-1}, \ldots, X_1\right)-\sum_{i=2}^n H\left(X_i \mid X_{i-1}, \ldots, X_2\right)\right] \\
& =H\left(X_1\right)-\left[\left(H\left(X_1\right)+\sum_{i=2}^n H\left(X_i \mid X_{i-1}\right)\right)-\left(H\left(X_2\right)+\sum_{i=3}^n H\left(X_i \mid X_{i-1}\right)\right)\right] \\
& =H\left(X_2\right)-H\left(X_2 \mid X_1\right) \\
& =I\left(X_2 ; X_1\right) \\
& =I\left(X_1 ; X_2\right)
\end{aligned}$$

习题 2. *(通过 KL 散度理解 MLE) 假设* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 来自密度为 $p(\mathbf{x})$ 的分布 P，试说明如果采用具有密度函数 $q_\theta(\mathbf{x})$ 的分布族 Q_θ 来计算 *MLE*，那么 *MLE* 将试图找到在 *KL* 散度意义上最接近真实分布 P 的分布 Q_θ 。

即证明

$$arg \max_\theta \prod_{i=1}^n q_\theta\left(\mathbf{x}_i\right) \Longleftrightarrow \arg \min_\theta D_{\mathrm{kl}}\left(P \| Q_\theta\right)$$

解.

$$\begin{aligned}
& \arg \max_\theta \prod_{i=1}^n q_\theta\left(\mathbf{x}_i\right) \Longleftrightarrow \arg \min_\theta -\frac{1}{n} \sum_{i=1}^n \log q_\theta\left(\mathbf{x}_i\right) \\
& \xrightarrow{P} \arg \min_\theta -E_P \log q_\theta(\mathbf{x}) \Longleftrightarrow \arg \min_\theta -\int p(\mathbf{x}) \log q_\theta(\mathbf{x}) d \mathbf{x} \\
& \Longleftrightarrow \arg \min_\theta H\left(P, Q_\theta\right) \Longleftrightarrow \arg \min_\theta \left(H\left(P, Q_\theta\right)-H(P)\right) \\
& \Longleftrightarrow \arg \min_\theta \left\{-\int p(\mathbf{x}) \log q_\theta(\mathbf{x}) d \mathbf{x}+\int p(\mathbf{x}) \log p(\mathbf{x}) d \mathbf{x}\right\} \\
& \Longleftrightarrow \arg \min_\theta \left\{-\int p(\mathbf{x}) \log \frac{q_\theta(\mathbf{x})}{p(\mathbf{x})}\right\} \Longleftrightarrow \arg \min_\theta D_{\mathrm{kl}}\left(P \| Q_\theta\right)
\end{aligned}$$

其实，从优化模型参数角度来说，最小化负对数似然，交叉熵 *(多分类问题中)*，*KL* 散度这 *3* 种方式是一样的。

习题 3. 设某种电子器件的寿命（以 h 计）T 服从双参数的指数分布，其概率密度为

$$f(t)=\begin{cases}\frac{1}{\theta}e^{-(t-c)/\theta} & t\geq c\\ 0 & \text{其他}\end{cases}$$

其中 $c,\theta(c,\theta>0)$ 为未知参数．自一批这种器件中随机地取 n 件进行寿命试验．设它们的失效时间依次为 $x_1\leqslant x_2\leqslant\cdots\leqslant x_n$。

(1) 求 θ 与 c 的最大似然估计值.

(2) 求 θ 与 c 的矩估计量

解. (1) 易知似然函数为

$$L(\theta,c)=\frac{1}{\theta^n}\exp\left\{\frac{nc-\sum_{i=1}^n x_i}{\theta}\right\}.$$

所以

$$\ln L(\theta,c)=-n\ln\theta+\frac{nc-\sum_{i=1}^n x_i}{\theta}.$$

对 θ 求偏导，并令导数为 0，可得 $\frac{n}{\theta}+\frac{nc-\sum_{i=1}^n x_i}{\theta^2}=0$. 可得 $\theta=\frac{\sum_{i=1}^n x_i}{n}-c$. 对该函数求二阶导并将 θ 代入, 可得

$$\frac{n}{\theta^2}+\frac{2nc-2\sum_{i=1}^n x_i}{\theta^3}=-\frac{n}{\theta^2}<0,$$

这说明求得的 θ 确实是极大值点。因此, θ 的最大似然估计值为 $\theta=\frac{\sum_{i=1}^n x_i}{n}-c$.

另外，在上式中, 通过简单地观测可以发现 $\ln L(\theta,c)$ 的数值会随着 c 的增加而增加, 故为使得上式最大, 应当使 c 最大. 故 c 的最大似然估计值为 $c=x_{(1)}$.

(2) 该分布的期望和二阶矩分别为

$$E(X)=\int_c^{+\infty}\frac{x}{\theta}\exp\left\{-\frac{x-c}{\theta}\right\}\mathrm{d}x=\theta+c.$$

$$E\left(X^2\right)=\int_c^{+\infty}\frac{x^2}{\theta}\exp\left\{-\frac{x-c}{\theta}\right\}\mathrm{d}x=c^2+2c\theta+2\theta^2.$$

该分布的方差 $\operatorname{Var}(X)=E\left(X^2\right)-E(X)^2=\theta^2$. 通过联立方程组，可求得 θ 的矩估计为 $\theta=\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})}{n}}$，$c$ 的矩估计为 $c=\bar{x}-\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})}{n}}$.

习题 4. 设总体 X 的概率密度为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}x^{(1-\theta)/\theta} & 0<x<1\\ 0 & \text{其他}\end{cases}\quad 0<\theta<+\infty$$

$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本。

(1) 验证 θ 的最大似然估计量是 $\hat{\theta}=\frac{-1}{n}\sum_{i=1}^n\ln X_i$

(2) 证明 $\hat{\theta}$ 是 θ 的无偏估计量。

解. *(1)* 易知似然函数为

$$L(\theta)=\frac{1}{\theta^n}\left(\prod_{i=1}^n x_i\right)^{\frac{1-\theta}{\theta}}.$$

所以 $\ln L(\theta)=-n\ln\theta+\frac{1-\theta}{\theta}\ln\left(\prod_{i=1}^n x_i\right)$. 对 θ 求导，并令导数为 0，可得

$$-\frac{n}{\theta}-\frac{1}{\theta^2}\ln\left(\prod_{i=1}^n x_i\right)=0.$$

可得 $\hat{\theta}=-\frac{1}{n}\Sigma_{i=1}^n\ln X_i$. 对该函数求二阶导并将 $\hat{\theta}$ 代入, 可得 $-\frac{n^3}{\left(\sum_{i=1}^n\ln X_i\right)^2}<0$, 这说明求得的 θ 确实是极大值点。故原命题得证。

(2) 首先求得

$$E(\ln x)=\int_0^1\ln x\frac{1}{\theta}x^{\frac{1-\theta}{\theta}}\mathrm{d}x=-\theta.$$

所以 $E(\hat{\theta})=-\frac{-n\theta}{n}=\theta$. 故原命题得证。

习题 5. 假设总体 $X\sim N\left(\mu,\sigma^2\right)$ (σ^2 已知), $X_1,X_2,\ldots,X_n$ 为来自总体 X 的样本, 由过去的经验和知识, 我们可以确定 μ 的取值比较集中在 μ_0 附近, 离 μ_0 越远, μ 取值的可能性越小, 于是我们假定 μ 的先验分布为正态分布

$$\pi(\mu)=\frac{1}{\sqrt{2\pi\sigma_\mu^2}}\exp\left[-\frac{1}{2\sigma_\mu^2}\left(\mu-\mu_0\right)^2\right]\quad(\mu_0,\sigma_\mu\text{已知})$$

求 μ 的后验概率分布。

解. 样本分布密度为

$$q(\mathbf{x}\mid\mu)=\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n\left(x_i-\mu\right)^2\right]$$

于是后验密度函数为

$$h(\mu\mid\mathbf{x})=\frac{q(\mathbf{x}\mid\mu)\cdot\pi(\mu)}{f_{\mathbf{x}}(\mathbf{x})}=\frac{q(\mathbf{x}\mid\mu)\cdot\pi(\mu)}{\int_{-\infty}^{+\infty}q(\mathbf{x}\mid\mu)\cdot\pi(\mu)d\mu}$$

$$\propto\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n\left(x_i-\mu\right)^2\right]\cdot\exp\left[-\frac{1}{2\sigma_\mu^2}\left(\mu-\mu_0\right)^2\right]$$

化简得

$$h(\mu\mid\mathbf{x})\propto\exp\left[-\frac{(\mu-t)^2}{2\eta^2}\right]$$

其中 $t=\frac{\frac{n}{\sigma^2}\bar{x}+\frac{1}{\sigma_\mu^2}\mu_0}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}$， $\eta^2=\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}$, 于是

$$\mu\mid\mathbf{x}\sim N\left(\frac{\frac{n}{\sigma^2}\bar{x}+\frac{1}{\sigma_\mu^2}\mu_0}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}},\frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_\mu^2}}\right).$$

习题 6. 假设总体 $X \sim P(\lambda), X_1, X_2, \ldots, X_n$ 为来自总体 X 的样本, 假定 λ 的先验分布为伽玛分布 $\Gamma(\alpha, \beta)$, 求 λ 的后验期望估计（平方损失下的贝叶斯估计）。

解. 因为 λ 的先验密度函数 $\pi(\lambda)$ 为伽玛分布 $\Gamma(\alpha, \beta)$, 即

$$\pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}$$

样本分布密度函数为:

$$q(\mathbf{x} \mid \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \ldots, x_n!} e^{-n\lambda} \propto \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}$$

所以

$$h(\lambda \mid \mathbf{x}) \propto \lambda^{\alpha+\sum_{i=1}^{n} x_i - 1} e^{-(\beta+n)\lambda}$$

即

$$\lambda \mid \mathbf{x} \sim \Gamma\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n\right)$$

故 λ 的后验期望估计为

$$\hat{\lambda} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n} = \frac{n}{\beta + n} \bar{x} + \frac{\beta}{\beta + n} \frac{\alpha}{\beta}$$

它是样本均值 $\bar{x}$ 和先验分布 $\Gamma(\alpha, \beta)$ 的均值 $\frac{\alpha}{\beta}$ 的加权平均。