

中期考试复习

G Z T*

2025 年 12 月 15 日

目录

A 高等数理统计	1
A.1 统计量	1
A.2 点估计	15
A.3 假设检验	30
A.4 区间估计	47
A.5 统计决策与 Bayes 分析	53
B 风险管理	61
B.1 风险度量	61
B.2 Copula	64
C 随机过程	74
D 机器学习基本方法	84

A 高等数理统计

A.1 统计量

Definition A.1.1 (Gamma 分布族). 在 $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$ 上用密度函数

$$p(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

表示的概率分布称为 *Gamma* 分布, 记为 $\Gamma(\alpha, \lambda)$. 它的期望与方差分别为

$$\mathbb{E}X = \frac{\alpha}{\lambda}, \mathbb{V}ar X = \frac{\alpha}{\lambda^2}$$

*School of Statistics, RUC, Beijing, 100872, China (gzt@ruc.edu.cn).

其特征函数 $f(t) = (1 - \frac{it}{\lambda})^{-\alpha}$, 于是可以看出当 $X_i \sim \Gamma(\alpha_i, \lambda), i = 1, 2, \dots, n$ 且 X_i 相互独立, 则其和为

$$X_1 + \dots + X_n \sim \Gamma(\alpha_1 + \dots + \alpha_n, \lambda).$$

Gamma 分布有两个重要子族: 指数分布族与 χ^2 分布族。

(1) 指数分布族: *Gamma* 分布中令 $\alpha = 1$ 即得指数分布, 记为 $Exp(\lambda) := \Gamma(1, \lambda)$, 其 pdf 为

$$p(x; \lambda) = \lambda e^{-\lambda x}, x > 0$$

(2) χ^2 分布族: *Gamma* 分布中令 $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ 即得 $\chi^2(n) =: \Gamma(\frac{n}{2}, \frac{1}{2})$, 其 pdf 为

$$p(x; n) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, x > 0.$$

Definition A.1.2 (Beta 分布族). 记 $D = (0, 1)$, 定义在 (D, \mathcal{B}_D) 上用密度函数

$$p(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

表示的概率分布称为 *Beta* 分布, 记为 $\beta(a, b)$, 其中 a, b 是两个正参数。它的期望与方差分别为

$$\mathbb{E}X = \frac{a}{a+b}, \mathbb{V}arX = \frac{ab}{(a+b)^2(a+b+1)}.$$

有几类特殊的分布与 *Beta* 分布有关。

(1) 均匀分布: $X \sim U(0, 1) = \beta(1, 1), X_{(k)} \sim \beta(k, n-k+1)$, 更一般地, $X \sim U(\theta_1, \theta_2), \frac{X_{(1)}-\theta_1}{\theta_2-\theta_1} \sim \beta(1, n), \frac{X_{(n)}-\theta_1}{\theta_2-\theta_1} \sim \beta(n, 1)$

(2) 若 $X_1 \sim \Gamma(\alpha_1, \lambda), X_2 \sim \Gamma(\alpha_2, \lambda)$ 则 $\frac{X_1}{X_1+X_2} \sim \beta(\alpha_1, \alpha_2)$ 且与 $X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$ 独立。

Definition A.1.3 (Fisher Z 分布). 定义在 $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$ 上用密度函数

$$p(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{x^{a-1}}{(1+x)^{a+b}}$$

表示的概率分布称为 *Fisher Z* 分布, 记为 $Z(a, b)$. 其中 a, b 为正参数。它的期望与方差为

$$\mathbb{E}X = \frac{a}{b-1}, \mathbb{V}arX = \frac{a(a+b-1)}{(b-1)^2(b-2)}.$$

我们有 *Z* 分布与 *Beta* 分布之间的关系:

(1) 若 $X \sim \beta(a, b)$, 则 $Y = X/(1-X) \sim Z(a, b)$.

(2) 若 $X \sim Z(a, b)$, 则 $Y = \frac{X}{1+X} \sim \beta(a, b)$.

Z 分布与 *F* 分布之间也存在关系。若 $X \sim Z(n_1/2, n_2/2)$, 则 $Y = (n_2/n_1)X$ 的密度函数为

$$p(y; n_1, n_2) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{y^{n_1/2-1}}{\left(1 + \frac{n_1}{n_2}y\right)^{\frac{n_1+n_2}{2}}}, y > 0.$$

Remark 1. 这里给出一些关于分布的关系。

若 $X \sim Exp(\lambda) = \Gamma(1, \lambda)$

(1) $X_{(1)} \sim Exp(n\lambda), \sum X_i \sim \Gamma(n, \lambda), kX \sim \Gamma(1, \frac{\lambda}{k}), 2\lambda X \sim \Gamma(1, \frac{1}{2}) = \chi^2(2), 2\lambda \sum X_i \sim \Gamma(n, \frac{1}{2}) = \chi^2(2n).$

(2) $e^{-\lambda X} \sim U(0, 1)$

(3) 若 $X, Y \sim Exp(\lambda), i.i.d.$, 则 $\frac{X}{X+Y} \sim U(0, 1).$

如果任意 CDF $F(x; \theta)$ 连续且严格递增

$$\begin{aligned} F(X; \theta) &\sim U(0, 1) \\ -\ln F(X; \theta) &\sim Exp(1) \\ -2 \ln F(X; \theta) &\sim \Gamma\left(1, \frac{1}{2}\right) = \chi^2(2) \\ -2 \sum_{i=1}^n \ln F(X_i; \theta) &\sim \Gamma\left(n, \frac{1}{2}\right) = \chi^2(2n) \end{aligned}$$

PROBLEM A.1.1. X_1, \dots, X_n 来自于总体 $N(\mu, \sigma^2), \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2.$

(1) 证明 $t = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1);$

(2) 证明 $Z = \frac{X_1 - \bar{X}}{S}$ 分布与参数 (μ, σ^2) 无关 (所以是辅助统计量), 并求其分布;

(3) 求常数 c 使得 $t_c = c(X_{n+1} - \bar{X})/S$ 服从 t 分布。

SOLUTION. (1)

$$\begin{aligned} U &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \\ t &= \frac{\sqrt{n}(\bar{X} - \mu)}{S} = \frac{U}{\sqrt{V/(n-1)}} \sim t(n-1). \end{aligned}$$

(2) 设 $Y_i = \frac{X_i - \mu}{\sigma} \sim N(0, 1)$, 于是

$$\begin{aligned} \bar{X} &= \mu + \sigma \bar{Y}, X_1 - \bar{X} = \sigma(Y_1 - \bar{Y}), S^2 = \sigma^2 \frac{1}{n-1} \sum (Y_i - \bar{Y})^2 \\ Z &= \frac{X_1 - \bar{X}}{S} = \frac{Y_1 - \bar{Y}}{\sqrt{\frac{1}{n-1} \sum (Y_i - \bar{Y})^2}} \end{aligned}$$

只依赖于标准正态分布 Y_i , 与参数无关。

又 $Y_1 - \bar{Y} \sim N(0, \frac{n-1}{n})$, $\sum (Y_i - \bar{Y})^2 \sim \chi^2(n-1)$, 则

$$Z = \frac{Y_1 - \bar{Y}}{S_Y} = \frac{\sqrt{\frac{n-1}{n}} U}{\sqrt{(n-1)S_Y^2/(n-1)}} = \sqrt{\frac{n-1}{n}} t_{n-1},$$

其中 t_{n-1} 是自由度为 $n-1$ 的 t 分布, 其密度函数为:

$$f(t) = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi}\Gamma(\frac{n-1}{2})} \left(1 + \frac{t^2}{n-1}\right)^{-n/2},$$

则 $Z = \sqrt{\frac{n-1}{n}}t$ 的密度函数为 (代入 $t = g(z) = \sqrt{\frac{n}{n-1}}z$, 以及 $g'(z) = \sqrt{\frac{n}{n-1}}$):

$$f_Z(z) = \frac{\sqrt{n}}{n-1} \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi}\Gamma(\frac{n-1}{2})} \left(1 + \frac{nz^2}{(n-1)^2}\right)^{-n/2} \dots$$

(3)

$$\begin{cases} X_{n+1} - \bar{X} \sim N\left(0, \frac{n+1}{n}\sigma^2\right) \Rightarrow \frac{X_{n+1} - \bar{X}}{\sqrt{\frac{n+1}{n}\sigma}} \sim N(0, 1) \\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \end{cases} \Rightarrow \sqrt{\frac{n}{n+1}} \frac{X_{n+1} - \bar{X}}{S} \sim t(n-1).$$

次序统计量

回顾一下次序统计量及其相关知识。

Theorem A.1.4. 若 X 的 PDF 为 $p(x)$, CDF 为 $F(x)$, 则 $X_{(k)}$ 的密度函数为:

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1-F(x)]^{n-k} p(x).$$

于是有

$$\begin{aligned} p_n(x) &= np(x)[F(x)]^{n-1} & F_n(x) &= [F(x)]^n, \\ p_1(x) &= np(x)[1-F(x)]^{n-1} & F_1(x) &= 1 - [1-F(x)]^n. \end{aligned}$$

Theorem A.1.5. $(X_{(i)}, X_{(j)}), i < j$ 的联合分布的密度函数为

$$p_{ij}(y, z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(y)]^{i-1} [F(z) - F(y)]^{j-i-1} [1 - F(z)]^{n-j} p(y)p(z), y \leq z,$$

且 $(X_{(1)}, X_{(n)})$ 的联合分布为

$$p_{1,n}(y, z) = n(n-1)[F(z) - F(y)]^{n-2} p(y)p(z), y \leq z.$$

n 个次序统计量的联合分布的密度函数为

$$f(y_1, \dots, y_n) = n! \prod_{i=1}^n f(y_i), \quad y_1 < y_2 < \dots < y_n.$$

Theorem A.1.6. 对于 $X_1, \dots, X_n \sim U(0, 1)$ i.i.d, 我们有

$$p(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & others \end{cases} \quad F(x) = \begin{cases} 0, & 0 < x \leq 1, \\ 1 & \end{cases}$$

则 $X_{(k)}$ 的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k} \sim Be(k, n-k)$$

恰为 Beta 分布 $Be(k, n - k + 1)$ 的密度函数。则

$$\mathbb{E}[X_{(k)}] = \frac{k}{n+1}, \mathbb{V}ar(X_{(k)}) = \frac{k(n+1-k)}{(n+1)^2(n+2)}.$$

更一般地，对于 $U(a, b)$, $\mathbb{E}[X_{(k)}] = a + \frac{k(b-a)}{n+1}$.

PROBLEM A.1.2. 随机变量 $X \sim U(0, 1), X_{(1)}, \dots, X_{(n)}$ 是该分布上的 n 个次序统计量，试证明: $\left(\frac{X_{(1)}}{X_{(r)}}, \dots, \frac{X_{(r-1)}}{X_{(r)}}\right)$ 与 $(X_{(r+1)}, \dots, X_{(n)})$ 相互独立。

SOLUTION. 令 $U_i = \frac{X_{(i)}}{X_{(r)}}, i = 1, \dots, r-1, V = X_{(r)}, W_j = X_{(r+j)}, j = 1, \dots, n-r$. 我们现在想把次序统计量 $(X_{(1)}, \dots, X_{(n)})$ 映射到新的随机变量 $(U_1, \dots, U_{r-1}, V, W_1, \dots, W_{n-r})$ 上。则上面规则的逆变换为 1-1 的

$$\begin{cases} X_{(1)} &= U_1 V \\ \vdots & \\ X_{(r-1)} &= U_{r-1} V \\ X_{(r)} &= V \\ X_{(r+1)} &= W_1 \\ \vdots & \\ X_{(n)} &= W_{n-r} \end{cases}$$

于是我们有如下 Jacobi 矩阵 $(n \times n)$

$$\begin{pmatrix} V & 0 & \cdots & 0 & U_1 & 0 & \cdots & 0 \\ 0 & V & \cdots & 0 & U_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V & U_{r-1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

且 $|\det(J)| = V^{r-1}$. 于是 $(U_1, \dots, U_{r-1}, V, W_1, \dots, W_{n-r})$ 的联合密度函数为

$$\begin{aligned} f_{U,V,W}(u_1, \dots, u_{r-1}, v, w_1, \dots, w_{n-r}) &= f_{X_{(1)}, \dots, X_{(n)}}(x_1(u, v, v), \dots, x_n(u, v, w)) \cdot v^{r-1} \\ &= n!v^{r-1}\mathbb{I}_{\{0 < u_1 < \dots < u_{r-1} < 1\}}\mathbb{I}_{\{0 < v < w_1 < \dots < w_{n-r} < 1\}} \end{aligned}$$

题目中要求 (U, W) 的边际密度，只需要对 v 积分掉，注意积分区间

$$\begin{aligned} f_{U,W} &= n! \frac{w_1^r}{r} \mathbb{I}_{\{0 < u_1 < \dots < u_{r-1} < 1\}} \mathbb{I}_{\{0 < v < w_1 < \dots < w_{n-r} < 1\}} \\ &= (r-1)! \mathbb{I}_{\{0 < u_1 < \dots < u_{r-1} < 1\}} \cdot \frac{n!}{r!} w_1^r \mathbb{I}_{\{0 < v < w_1 < \dots < w_{n-r} < 1\}} \\ &= f_U(u_1, \dots, u_{r-1}) \cdot f_W(w_1, \dots, w_{n-r}). \end{aligned}$$

联合密度函数能完全分离，于是 U, W 相互独立（联合密度的拆分要注意密度函数的归一性，比如 U 实际上就是 r 个次序统计量的联合密度函数）。

分布的近似

Theorem A.1.7 (Slutsky). 设 $\{Z_n\}, \{U_n\}$ 是两个随机变量序列，若

$$Z_n \xrightarrow{L} Z, U_n \xrightarrow{P} c = \text{constant}$$

则有

(i)

$$Z_n + U_n \xrightarrow{L} Z + c$$

(ii)

$$U_n Z_n \xrightarrow{L} cZ$$

(iii)

$$Z_n / U_n \xrightarrow{L} Z/c (c \neq 0)$$

PROBLEM A.1.3. 设 $\{X_n, n \geq 1\}$ 是独立同分布且具有有限方差的随机变量序列且均值为 μ . 试证明 $Z_n = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ 依分布收敛于一个标准正态随机变量.

SOLUTION.

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{S} =: T_n \cdot \frac{\sigma}{S},$$

由中心极限定理可知

$$T_n \xrightarrow{d} N(0, 1).$$

又 $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ 是样本方差，由弱大数定律

$$S_n^2 \xrightarrow{P} \sigma^2, \frac{\sigma}{S_n} \xrightarrow{P} 1.$$

由于 Slutsky 定理即证明了结论。

Theorem A.1.8 (Delta 方法). 设 $\{a_n\}$ 为一列趋于 ∞ 的数列， b 为常数，并且对随机变量序列 $\{Z_n\}$ 有

$$a_n(Z_n - b) \xrightarrow{L} Z$$

又设 $g(\cdot)$ 为可微函数，且 g' 在 b 处连续，则有

$$a_n[g(Z_n) - g(b)] \xrightarrow{L} g'(b)Z$$

PROBLEM A.1.4. 设 X_1, \dots, X_n 是来自两点分布 $b(1, \theta)$ 的一个样本, 其中 $0 < \theta < 1$, 现要求样本均值 \bar{X} 的函数

$$g_1(\bar{X}) = \frac{1}{\bar{X}}, \quad g_2(\bar{X}) = \sqrt{\bar{X}(1 - \bar{X})}, \quad g_3(\bar{X}) = \bar{X}(1 - \bar{X}).$$

的渐进分布。

SOLUTION. 由题意得 $\mathbb{E}\bar{X} = \theta, \text{Var}(\bar{X}) = \frac{1}{n}\theta(1 - \theta)$, 由中心极限定理可知,

$$\sqrt{\frac{n}{\theta(1 - \theta)}}(\bar{X} - \theta) \xrightarrow{L} N(0, 1)$$

令 $a_n = \sqrt{\frac{n}{\theta(1 - \theta)}}$, 显然是趋于 ∞ 的数列, 则由 Delta 方法, 取 $g_1(x) = \frac{1}{x}, g_2(x) = \sqrt{x(1 - x)}, g_3(x) = x(1 - x), Z \stackrel{d}{=} N(0, 1)$. 由

$$a_n[g(Z_n) - g(b)] \xrightarrow{L} g'(b)Z$$

得

$$\begin{aligned} \sqrt{\frac{n}{\theta(1 - \theta)}}(g_1(\bar{X}) - \frac{1}{\theta}) &\xrightarrow{L} -\frac{1}{\theta^2}Z, \\ \sqrt{\frac{n}{\theta(1 - \theta)}}(g_2(\bar{X}) - \sqrt{\theta(1 - \theta)}) &\xrightarrow{L} \frac{1 - 2\theta}{2\sqrt{\theta(1 - \theta)}}Z, \\ \sqrt{\frac{n}{\theta(1 - \theta)}}(g_3(\bar{X}) - \theta(1 - \theta)) &\xrightarrow{L} (1 - 2\theta)Z. \end{aligned}$$

即,

$$\begin{aligned} g_1(\bar{X}) &\xrightarrow{L} N\left(\frac{1}{\theta}, \frac{1 - \theta}{n\theta^3}\right), \\ g_2(\bar{X}) &\xrightarrow{L} N\left(\sqrt{\theta(1 - \theta)}, \frac{(1 - 2\theta)^2}{4n}\right), \theta \neq 1/2, \\ g_3(\bar{X}) &\xrightarrow{L} N\left(\theta(1 - \theta), \frac{\theta(1 - \theta)(1 - 2\theta)^2}{n}\right), \theta \neq 1/2. \end{aligned}$$

NOTE OF PROBLEM A.1.4. 在讨论 $g_2(\bar{X})$ 当 $\theta = 1/2$ 时, 可以考虑二阶 Delta 方法.

Theorem A.1.9 (矩的近似). 设 x_1, \dots, x_n 是来自总体 X 的一个样本, 该总体 X 的四阶中心矩 $\mu_4 = \mathbb{E}(X - \mathbb{E}X)^4$ 有限, 记 $\mathbb{E}X = \mu, \text{Var}X = \sigma^2$, 若函数 $h(x)$ 的四阶导数存在且有界, 则有

$$\begin{aligned} \mathbb{E}[h(\bar{x})] &= h(\mu) + \frac{1}{n}h''(\mu)\sigma^2 + o\left(\frac{1}{n^2}\right) \\ \text{Var}[h(\bar{x})] &= \frac{1}{n}[h'(\mu)]^2\sigma^2 + \frac{1}{n^2}\left[h'(\mu)h''(\mu)\mu_3 + \frac{1}{2}[h''(\mu)]^2\sigma^4 + h'(\mu)h'''(\mu)\sigma^4\right] + o\left(\frac{1}{n^3}\right) \end{aligned}$$

PROBLEM A.1.5. 在 Bernoulli 分布 $b(1, p)$ 中, 其方差随着期望 $EX = p$ 变化而变化。能否找到一个函数 $h(\bar{x})$ 使得其方差近似不依赖于 p ? 即

$$\text{Var}(h(\bar{x})) = c + O\left(\frac{1}{n^2}\right),$$

其中 c 不依赖于 p (可以依赖于 n), 这里的 \bar{x} 表示容量为 n 的样本均值。

SOLUTION. 由定理 1.9, 首先 $h(\bar{x})$ 的方差要近似到 $\frac{1}{n}$, 即取

$$\text{Var}[h(\bar{x})] = \frac{1}{n} [h'(p)]^2 \sigma^2 + O\left(\frac{1}{n^2}\right),$$

其中 $\sigma^2 = \text{Var}(X) = p(1 - p)$. 此时要使方差为常数, 等价于使

$$[h'(p)]^2 \sigma^2 = e \quad \text{or} \quad h'(p) = \sqrt{\frac{e}{p(1 - p)}}.$$

解上面的微分方程, 可得

$$h(p) = \sqrt{e} \int \sqrt{\frac{1}{p(1 - p)}} dp + d,$$

其中 d 为任意常数。利用变换 $p = \sin^2 \theta$, 将上述不定积分转化为

$$h(p) = \sqrt{e} 2 \arcsin \sqrt{p} + d.$$

e 与 d 的取值是无关紧要的, 则不妨取 $e = \frac{1}{4}, d = 0$, 则

$$h(p) = \arcsin \sqrt{p}.$$

则按此启示, 令 $h(\bar{x}) = \arcsin \sqrt{\bar{x}}$ 即为要求的方差稳定变换, 其方差近似为

$$\text{Var}[h(\bar{x})] \simeq \frac{1}{4n}.$$

NOTE OF PROBLEM A.1.5. 本题来自《高等数理统计》例 1.15 (方差稳定变换) .

- 在指数分布 $\text{Exp}(\frac{1}{\theta})$ 场合, 方差稳定变换是 $h(\bar{x}) = \ln \bar{x}$.

PROBLEM A.1.6. 设 \bar{x} 是来自 Poisson 分布的一个样本均值。

- (1) 寻找 $h(\bar{x}) = e^{-\bar{x}}$ 的期望与方差的近似值;
- (2) 寻求一个 $h(\bar{x})$, 使其方差近似为常数.

SOLUTION.

(1) Poisson 分布的均值与方差均为 λ . 由定理 1.9 可知

$$\begin{aligned}\mathbb{E}[h(\bar{x})] &= h(\mu) + \frac{1}{2n}h''(\mu)\sigma^2 + O(n^{-2}) \\ &= e^{-\lambda} + \frac{1}{2n}e^{-\lambda}\lambda + O(n^{-2}),\end{aligned}$$

方差为

$$\begin{aligned}Var[h(\bar{x})] &= \frac{1}{n}[h'(\mu)]^2\sigma^2 + \frac{1}{n^2}\left[h'(\mu)h''(\mu)\mu_3 + \frac{1}{2}[h''(\mu)]^2\sigma^4 + h'(\mu)h'''(\mu)\sigma^4\right] + O(n^{-3}), \\ &= \left(\frac{\lambda}{n} + \frac{3\lambda^2 - 2\lambda}{2n^2}\right)e^{-2\lambda} + O(n^{-3}),\end{aligned}$$

从而得到近似的期望与方差.

(2) 由定理 1.9 要使方差近似到 n^{-1} , 与 λ 无关, 即令 $h'(\mu)^2\sigma^2 = c$, 解微分方程为 $h(\lambda) = 2\sqrt{c\lambda} + d$, 其中 c, d 为任意常数, 不妨取 $c = 1/4, d = 0$ 可得 $h(\bar{x}) = \sqrt{\bar{x}}$ 是所求的方差稳定变换, 其方差近似为 $Var[h(\bar{X})] \simeq \frac{1}{4n}$.

充分统计量

在 $T = t$ 条件下, 样本 X 的条件分布不依赖于 θ .

$$P_\theta(X = x|T = t) = P(X = x|T = t).$$

Theorem A.1.10 (因子分解定理). 设 $\{\mathcal{X}, \mathcal{B}, \{p_\theta, \theta \in \Theta\}$ 为可控结构, μ 为控制测度, 记 $p_\theta(x) = dP_\theta/d\mu$, 又设 $T = T(x)$ 是统计量, 则 T 为充分统计量的充要条件为存在

- (i) 非负可测函数 $h(x)$,
- (ii) 可测函数 $g_\theta(t)$ 使得对任意的 $\theta \in \Theta$ 都有

$$p_\theta(x) = g_\theta(T(X))h(x), a.s.\mu.$$

Definition A.1.11 (完备性). 设 $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ 是一个参数结构, 其中 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, 加入对于可测函数 $\phi(x)$, 有

$$\int \phi(x)dP_\theta = 0, \forall \theta \in \Theta$$

总能推出 $\phi(x) = 0 a.s. P_\theta$, 则称此结构是完备的或分布族 \mathcal{P} 是完备的。

Definition A.1.12 (完备统计量). 设 T 是统计结构 $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ 上的统计量, 其诱导的统计结构为 $(\mathcal{T}, \mathcal{B}, \mathcal{P}^T)$, 假如后者是完备结构则称 T 为完备统计量。

Remark 2. 假如原结构 $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ 完备, 则诱导的统计结构是完备的, 但也存在诱导结构是完备的但原结构不完备, 比如正态分布 $\{N(0, \sigma^2) : \sigma \in \mathbb{R}^+\}$ 不完备, 但是统计量 $T = \sum X_i^2$ 是完备统计, 这是因为 $T = \sum X_i^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2\sigma^2})$ 是完备分布族。

Definition A.1.13 (指数结构). 设 $\{\mathcal{X}, \mathcal{B}, \{p_\theta, \theta \in \Theta\}$ 为可控参数结构, 假如其密度函数可以表示为如下形式

$$p_\theta(x) = c(\theta) \exp \left\{ \sum_{j=1}^k c_j(\theta) T_j(x) \right\} h(x) \quad (\text{A.1})$$

且其支撑 $\{x : p_\theta(x) > 0\}$ 与 θ 无关则称此结构为指数结构。

Remark 3. 若在指数型分布(A.1)中重新设置参数 $w = (w_1, \dots, w_k)$, 其中

$$w_j = c_j(\theta), j = 1, \dots, k$$

假如你能从此函数方程组中唯一地解出 $\theta = \theta(w_1, \dots, w_k)$, 那么再令 $c^*(w) = c(\theta(w))$, 代回(A.1), 即得密度函数另一种形式

$$p_\theta(x) = c^*(w) \exp \left\{ \sum_{j=1}^k w_j T_j(x) \right\} h(x)$$

被称为标准形式。

Theorem A.1.14. 设 X_1, \dots, X_n 是来自指数型分布族标准形式的一个样本则有

(i) 统计量

$$(T_1(X), \dots, T_k(X)) = \left(\sum_{i=1}^n T_1(x_i), \dots, \sum_{i=1}^n T_k(x_i) \right)$$

是指指数型分布族的充分统计量

(ii) 充分统计量的期望和协方差分别为

$$\begin{aligned} \mathbb{E}_w(T_j(X)) &= -\frac{\partial \ln c(w)}{\partial w_j}, j = 1, \dots, k \\ Cov_w(T_i(X), T_j(X)) &= -\frac{\partial^2 \ln c(w)}{\partial w_i \partial w_j}, i, j = 1, \dots, k \end{aligned}$$

其中 $c(w) = [c^*(w)]^n$.

Theorem A.1.15. 设 (X_1, \dots, X_n) 是来自指数型分布的一个样本, 其联合密度函数为

$$p_w(x) = c(w) \exp \left\{ \sum_{j=1}^k w_j T_j(x) \right\} h(x)$$

其中 $(T_1(X), \dots, T_k(X)) = (\sum_{i=1}^n T_1(x_i), \dots, \sum_{i=1}^n T_k(x_i))$, 假如 Ω 有内点, 则 $T_1(x), \dots, T_k(x)$ 是完备统计量。

PROBLEM A.1.7. 设 X_1, \dots, X_n 是来自 Weibull 分布的一个样本，其分布函数为：

$$F(x) = \begin{cases} 1 - \exp\left\{-\left(\frac{x}{\eta}\right)^m\right\}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

其中 $m > 0$ 为形状参数， $\eta > 0$ 为尺度参数，求 $Y = \min(X_1, \dots, X_n)$ 服从的分布。

SOLUTION. 由题意得， $Y := X_{(1)}$ 即次序统计量的最小值。对于 $y > 0$ ($y \leq 0$ 时显然分布函数为 0)，

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(y \geq X_{(1)}) = 1 - P(y \leq X_{(1)}) \\ &= 1 - P(y < X_1, y < X_2, \dots, y < X_n) \\ &= 1 - \prod_{i=1}^n P(y < X_i) \\ &= 1 - (1 - F(y))^n \\ &= 1 - \exp\left\{-n \cdot \left(\frac{y}{\eta}\right)^m\right\} \\ &= 1 - \exp\left\{-\left(\frac{y}{n^{-1/m} \cdot \eta}\right)^m\right\}. \end{aligned}$$

则 $Y = \min(X_1, \dots, X_n)$ 服从形状参数为 m ，尺度参数为 $n^{-1/m} \cdot \eta$ 的 Weibull 分布。

NOTE OF PROBLEM A.1.7. 本题来自《高等数理统计》习题 1.31.

PROBLEM A.1.8. 设二维随机变量 $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ 服从二元正态分布，其均值向量为零向量，协方差矩阵为

$$\begin{pmatrix} \sigma^2 + r^2 & \sigma^2 - r^2 \\ \sigma^2 - r^2 & \sigma^2 + r^2 \end{pmatrix}, \sigma, r > 0$$

求其参数 (σ^2, r^2) 的充分统计量。

SOLUTION. 其二元密度函数为

$$\begin{aligned} f(\mathbf{x}) &= f(x_1, x_2) \propto \exp\left\{\frac{1}{2}((\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu))\right\} \\ \ln f(x_1, x_2) &\propto \left((x_1, x_2) \frac{1}{4} \begin{pmatrix} \sigma^{-2} + r^{-2} & \sigma^{-2} - r^{-2} \\ \sigma^{-2} - r^{-2} & \sigma^{-2} + r^{-2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) \\ &\propto \sigma^{-2}(x_1 + x_2)^2 + r^{-2}(x_1 - x_2)^2. \end{aligned}$$

由因子分解定理可知 $T = ((X_1 + X_2)^2, (X_1 - X_2)^2)$ 是充分统计量。

PROBLEM A.1.9. 设 X_1, \dots, X_n 时来自如下密度函数的一个样本，分别求未知参数 θ 得充分统计量：

1. (幂分布) $p_\theta(x) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$;
2. (Pareto 分布) $p_\theta(x) = \frac{\theta a^\theta}{x^{\theta+1}}$, $x > a$, $\theta > 0$, $a > 0$ 已知;
3. (Laplace 分布) $p_\theta(x) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, $x \in \mathbb{R}$, $\theta > 0$.

SOLUTION.

1. 由于 $p_\theta(\mathbf{x}) = \theta^n (\prod_{i=1}^n x_i)^{\theta-1} \mathbf{I}(0 < x_i < 1)$, 则由因子分解定理可知 $T(\mathbf{x}) = \prod_{i=1}^n x_i$ 即为所求的充分统计量.
2. 因子分解定理同上, $T(\mathbf{x}) = \prod_{i=1}^n x_i$.
3. 因子分解定理同上, $T(\mathbf{x}) = \sum_{i=1}^n |x_i|$.

□

PROBLEM A.1.10. 设 $X_1, \dots, X_r \sim M(n; p_1, \dots, p_r)$, 设 $(X_{1j}, \dots, X_{rj}, j = 1, \dots, m)$ 是来自多项分布的一个多维样本, 证明 $(\sum_{j=1}^m X_{1j}, \dots, \sum_{j=1}^m X_{rj})$ 是 (p_1, \dots, p_r) 的充分统计量.

SOLUTION. 记 $\mathbf{X}_j = (X_{1j}, \dots, X_{rj})$, $j = 1, \dots, m$. 由 $P(\mathbf{X}_j = \mathbf{x}_j) \propto p_1^{x_{1j}} \cdots p_r^{x_{rj}}$, 得

$$P(\mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_m = \mathbf{x}_m) \propto p_1^{\sum_{j=1}^m x_{1j}} \cdots p_r^{\sum_{j=1}^m x_{rj}},$$

由因子分解定理可知, $(\sum_{j=1}^m X_{1j}, \dots, \sum_{j=1}^m X_{rj})$ 是 (p_1, \dots, p_r) 的充分统计量.

PROBLEM A.1.11. 假设在 μ 的条件下, X, Y 相互独立服从均值为 μ 的泊松分布, μ 是一个随机变量, 其密度函数为 $\frac{\lambda^\nu \mu^{\nu-1} e^{-\lambda\mu}}{\Gamma(\nu)}$, $\mu > 0, \nu, \lambda > 0$. 证明 X, Y 的联合矩母函数为

$$\mathbb{E}(e^{sX+tY}) = \lambda\nu \{ \lambda - (e^s - 1) - (e^t - 1) \}^{-\nu}$$

并且求出 (X, Y) 的均值和协方差矩阵, 如果 $\lambda = \frac{\nu}{\xi}$ 并且 $\nu \rightarrow \infty$ 会出现什么情况?

SOLUTION. 由题意可得

$$\begin{aligned} p(X = x | \mu) &= \frac{\mu^x}{x!} e^{-\mu}, x = 0, 1, 2, \dots \\ p(Y = y | \mu) &= \frac{\mu^y}{y!} e^{-\mu}, y = 0, 1, 2, \dots \end{aligned}$$

所以

$$\mathbb{E}(e^{sX} | \mu) = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} e^{sx} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(e^s \mu)^x}{x!} = e^{-\mu} e^{e^s \mu} = e^{-\mu(1-e^s)}$$

同理可得,

$$E(e^{tY}|\mu) = e^{-\mu(1-e^t)}$$

因为在 μ 的条件下 X, Y 相互独立, 根据矩母函数的性质可得

$$E(e^{sX+tY}|\mu) = E(e^{sX}|\mu)E(e^{tY}|\mu) = e^{-\mu(1-e^s+1-e^t)}$$

根据重期望公式可得,

$$E[e^{sX+tY}] = E[E(e^{sX+tY}|\mu)]$$

所以

$$\begin{aligned} E(e^{sX+tY}) &= \int_0^{+\infty} E(e^{sX+tY}|\mu) \lambda^\nu \mu^{\nu-1} e^{-\lambda\mu} / \Gamma(\nu) d\mu \\ &= \int_0^{+\infty} e^{-\mu(2-e^s-e^t)} \lambda^\nu \mu^{\nu-1} e^{-\lambda\mu} / \Gamma(\nu) d\mu \\ &= \lambda^\nu (\lambda + 2 - e^s - e^t)^{-\nu} / \Gamma(\nu) \int_0^{+\infty} e^{-\mu(\lambda+2-e^s-e^t)} (\mu(\lambda + 2 - e^s - e^t))^{\nu-1} d\mu (\lambda + 2 - e^s - e^t) \\ &= (\lambda^\nu (\lambda + 2 - e^s - e^t)^{-\nu} / \Gamma(\nu)) \Gamma(\nu) \\ &= \lambda^\nu \{ \lambda - (e^s - 1) - (e^t - 1) \}^{-\nu} \end{aligned}$$

证毕.

令 $t = 0$, 则 $E(e^{sX}) = E(e^{sX+0Y}) = \lambda^\nu \{ \lambda - (e^s - 1) \}^{-\nu}$

记 $g(s) = E(e^{sX})$

根据矩母函数的性质

$$\mathbb{E}[X^n] = g^{(n)}(0)$$

所以,

$$\begin{aligned} g'(s) &= \lambda^\nu \nu \{ \lambda - (e^s - 1) \}^{-\nu-1} e^s \\ g''(s) &= \lambda^\nu \nu [(\nu+1)(\lambda - (e^s - 1))^{-\nu-2} e^{2s} + (\lambda - (e^s - 1))^{-\nu-1} e^s] \\ g'(0) &= \nu \lambda^{-1} \\ g''(0) &= \nu(\nu+1) \lambda^{-2} + \nu \lambda^{-1} \end{aligned}$$

所以,

$$\begin{aligned} \mathbb{E}[X] &= g'(0) = \nu \lambda^{-1} \\ \mathbb{E}[X^2] &= g''(0) = \nu(\nu+1) \lambda^{-2} + \nu \lambda^{-1} \\ \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}^2[X] = \nu \lambda^{-2} + \nu \lambda^{-1} \end{aligned}$$

同理可得,

$$\text{Var}(Y) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \nu \lambda^{-2} + \nu \lambda^{-1}$$

$$\mathbb{E}[X|\mu] = \sum_{x=0}^{\infty} \frac{x \mu^x}{x!} e^{-\mu} = \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} e^{-\mu} = \mu$$

同理，可得

$$\mathbb{E}[Y|\mu] = \mu$$

因为在 μ 的条件下， X, Y 相互独立，所以 $\mathbb{E}(XY|\mu) = \mathbb{E}(X|\mu)\mathbb{E}(Y|\mu) = \mu^2$ 又因为

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[\mathbb{E}(XY|\mu)] \\ &= \int_0^{+\infty} \mu^2 \lambda^\nu \mu^{\nu-1} e^{-\lambda\mu} \Gamma(\nu) d\mu \\ &= (\lambda^{-2}/\Gamma(\nu)) \int_0^{+\infty} (\lambda\mu)^{\nu+1} e^{-\lambda\mu} d\lambda \mu \\ &= (\lambda^{-2}/\Gamma(\nu)) \Gamma(\nu + 2) \\ &= \lambda^{-2} \nu(\nu + 1)\end{aligned}$$

$$\begin{aligned}Cov(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \lambda^{-2} \nu(\nu + 1) - \lambda^{-2} \nu^2 \\ &= \nu \lambda^{-2}\end{aligned}$$

所以

$$\mathbb{E}[(X, Y)] = (\nu \lambda^{-1}, \nu \lambda^{-1})$$

$$\Sigma = \begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(Y, X) & Var(Y) \end{bmatrix} = \begin{bmatrix} \nu \lambda^{-2} + \nu \lambda^{-1} & \nu \lambda^{-2} \\ \nu \lambda^{-2} & \nu \lambda^{-2} + \nu \lambda^{-1} \end{bmatrix}$$

如果 $\lambda = \nu/\xi$ 并且 $\nu \rightarrow \infty$, 就有

$$\begin{aligned}\lim_{\nu \rightarrow \infty} \nu \lambda^{-1} &= \xi \\ \lim_{\nu \rightarrow \infty} \nu \lambda^{-2} &= 0 \\ \lim_{\nu \rightarrow \infty} (\nu \lambda^{-2} + \nu \lambda^{-1}) &= \xi\end{aligned}$$

此时，

$$\mathbb{E}[(X, Y)] = (\xi, \xi)$$

$$\Sigma = \begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(Y, X) & Var(Y) \end{bmatrix} = \begin{bmatrix} \xi & 0 \\ 0 & \xi \end{bmatrix}$$

A.2 点估计

PROBLEM A.2.1. $X_1, \dots, X_n \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, θ 是未知参数。证明: $\frac{X_{(1)} + X_{(n)}}{2}$ 是 θ 的强相合估计并且在 MSE 下相合。

SOLUTION. 记 $U_i = X_i - \theta + \frac{1}{2} \sim U(0, 1)$, 则

$$X_{(i)} = \theta - \frac{1}{2} + U_{(i)}, T_n = \frac{X_{(1)} + X_{(n)}}{2} = \theta - \frac{1}{2} + \frac{U_{(2)} + U_{(n)}}{2},$$

由于 $P(\lim U_{(1)} = 0) = 1, P(\lim U_{(n)} = 1) = 1$,

$$T_n - \theta = -\frac{1}{2} + \frac{U_{(2)} + U_{(n)}}{2} \xrightarrow{a.s.} 0.$$

于是强相合估计成立。

均方误差求 $MSE(T_n) = \mathbb{E}[(T_n - \theta)^2] = \text{Var}(T_n) + |\mathbb{E}T_n - \theta|^2$.

由于 $\mathbb{E}U_{(1)} = \frac{1}{n+1}, \mathbb{E}U_{(n)} = \frac{n}{n+1}$, 易得 T_n 是无偏估计量。则需要计算 $\text{Var}(T_n)$.

$$\begin{aligned} \text{Var}(T_n) &= \frac{1}{4}\text{Var}(U_{(1)} + U_{(n)}) \\ &= \frac{1}{4}[\text{Var}U_{(1)} + \text{Var}U_{(n)} + 2\text{Cov}(U_{(1)}, U_{(n)})] \\ &= \frac{1}{2(n+1)(n+2)} \rightarrow 0 \end{aligned}$$

MSE 下相合。

PROBLEM A.2.2. 设 $X_1, \dots, X_n \sim U(0, \theta)$, 证明 $X_{(n)}$ 是 θ 的渐进无偏相合估计。

SOLUTION. 易得 $X_{(n)}$ 的 pdf 为

$$p(t; \theta) = nt^{n-1}\theta^{-n}, 0 < t < \theta,$$

其期望 $\mathbb{E}X_{(n)} = \frac{n\theta}{n+1}$, 则渐进无偏性成立。

另外, 对于任意的 $\epsilon > 0$,

$$\begin{aligned} p_\theta(|X_{(n)} - \theta| \geq \epsilon) &= P_\theta(X_{(n)} \leq \theta - \epsilon) \\ &= \int_0^{\theta-\epsilon} \frac{nt^{n-1}}{\theta^n} dt = \left(\frac{\theta-\epsilon}{\theta}\right)^n \rightarrow 0 \end{aligned}$$

因此相合性成立。

PROBLEM A.2.3. 设随机变量 $X \sim p(x; \theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$. X_1, \dots, X_n 为独立同分布的样本, 令 $T(X)$ 为 θ 的充分统计量,

求:(1) $T(X)$;(2) $\mathbb{E}[T(X)]$;(3) $\text{Var}[T(X)]$;(4) $I(\theta)$;(5) $\frac{1}{\theta}$ 的 UMVUE.

SOLUTION. 其实这是 Beta 分布的特殊形式 $X \sim \beta(\theta, 1)$, $\mathbb{E}X = \frac{\theta}{\theta+1}$, $\text{Var}X = \frac{\theta}{(\theta+1)^2(\theta+2)}$.

(1) 样本联合密度为

$$\begin{aligned} p(\mathbf{x}; \theta) &= \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1} = \theta^n \left(\exp \left\{ \sum_{i=1}^n \ln x_i \right\} \right)^{\theta-1} \\ &= g \left(\prod_{i=1}^n x_i; \theta \right) \cdot h(x) \end{aligned}$$

最后一个等式是因为因子分解定理, 于是 $T(X) = \prod X_i$ 或者 $\sum \ln X_i$ 或者 $-\sum \ln X_i$. 最后一个更为常用因为如果 $X \sim \beta(\theta, 1) \Rightarrow -\log X \sim \text{Exp}(\theta)$: $f_Y(y) = \theta e^{-\theta y}$. 于是 $-\sum \log X_i \sim \Gamma(n, \theta)$: $f = \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y}$.

(2)(3) 取 $T(X) = -\sum \log X_i \sim \Gamma(n, \theta)$, 则 $\mathbb{E}[T(X)] = \frac{n}{\theta}$, $\text{Var}[T(X)] = \frac{n}{\theta^2}$.

(4) 求其联合密度函数的对数似然

$$\begin{aligned} l(\theta) &= n \log \theta + \sum (\theta - 1) \log x_i \\ l''(\theta) &= -\frac{n}{\theta^2} \\ I(\theta) &= -\mathbb{E}[l''(\theta)] = \frac{n}{\theta^2}. \end{aligned}$$

(5) 求 $\frac{1}{\theta}$ 的 UMVUE. 由于 $\mathbb{E}[-\log X_i] = 1/\theta$, 且 $T = -\sum \log X_i$ 是充分完备统计量, 考虑估计量

$$\hat{g} = \frac{T}{n}, \mathbb{E}[\hat{g}] = \frac{1}{\theta}$$

是无偏估计且是充分完备统计量的函数, 于是 $\hat{g} = T/n$ 是 $1/\theta$ 的 UMVUE.

PROBLEM A.2.4. $X_1, \dots, X_n \sim \{p(x; \mu, \sigma) = \frac{1}{2\sigma} \exp\{-|x - \mu|/\sigma\}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$.

(1) 求 μ, σ 的矩估计;

(2) 求 μ, σ 的 MLE.

SOLUTION.

(1) 由题意得 X 的期望为

$$\begin{aligned} \mathbb{E}X &= \int \frac{x}{2\sigma} \exp\{-|x - \mu|/\sigma\} dx \\ &= \int \frac{x - \mu + \mu}{2\sigma} \exp\{-|x - \mu|/\sigma\} dx \\ &= \mu. \end{aligned}$$

X 的二阶矩为

$$\begin{aligned}
 \mathbb{E}(X^2) &= \int \frac{x^2}{2\sigma} \exp\{-|x - \mu|/\sigma\} dx \\
 &= \int \frac{(x - \mu + \mu)^2}{2\sigma} \exp\{-|x - \mu|/\sigma\} dx \\
 &= \int \frac{(x - \mu)^2}{2\sigma} \exp\{-|x - \mu|/\sigma\} dx + \mu^2 \\
 &= \int_0^\infty \frac{t^2}{2\sigma} \exp\{-t/\sigma\} dt + \mu^2 \\
 &= \sigma^2 \int_0^\infty u^2 \exp(-u) du + \mu^2 \\
 &= 2\sigma^2 + \mu^2.
 \end{aligned}$$

于是 $\hat{\mu}_1 = \bar{X}$, $\hat{\sigma}_1 = \sqrt{\frac{1}{2} \left[\frac{1}{n} \sum X_i^2 - (\frac{1}{n} \sum X_i)^2 \right]}$.

(2) 对数似然函数为

$$l = \ln L = \ln \prod p(x_i) \propto -n \ln \sigma - \frac{1}{\sigma} \sum |x_i - \mu|$$

则

$$\begin{aligned}
 \frac{\partial l}{\partial \mu} &= -\frac{1}{\sigma} \frac{\partial}{\partial \mu} |x_i - \mu| = \frac{1}{\sigma} \sum sgn(x_i - \mu), \\
 \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum |x_i - \mu|,
 \end{aligned}$$

则极大似然估计

$$\hat{\mu}_2 = X_m = m_{0.5}, \quad \hat{\sigma}_2 = \frac{1}{n} \sum |X_i - \hat{\mu}_2|.$$

PROBLEM A.2.5. 对 Poisson 分布 $P(\theta)$:

- (1) 求 $I(\theta)$;
- (2) 求 $I(\frac{1}{\theta})$;
- (3) 找一个函数 $g(\cdot)$, s.t. $g(\theta)$ 的 Fisher 信息与 θ 无关.

SOLUTION. 易知 Poisson 分布族是 Cramer-Rao 正则族, 其 Fisher 信息存在。

(1)

$$\begin{aligned}
 p(x; \theta) &= \frac{\theta^x}{x!} e^{-\theta x} \\
 I(\theta) &= -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \theta^2}\right] \\
 &= \mathbb{E}(X/\theta^2) = \frac{1}{\theta}.
 \end{aligned}$$

(2) 令 $\lambda = \frac{1}{\theta}$, 则 $p_\lambda(x) = \frac{1}{\lambda^x x!} e^{-\frac{1}{\lambda}}$, 于是

$$I\left(\frac{1}{\theta}\right) = I(\lambda) = \frac{1}{\lambda^3} = \theta^3.$$

(3) 令 $\lambda = \theta^k$, $p_\lambda(x) = \frac{\lambda^x}{x!} e^{-\lambda^{\frac{1}{k}}}$.

$$\begin{aligned} I(\lambda) &= -\mathbb{E}\left[\frac{\partial^2 \ln p_\lambda}{\partial \lambda^2}\right] \\ &= -\mathbb{E}\left[\frac{X}{\lambda^2 k} - \frac{1}{k}\left(\frac{1}{k} - 1\right)\lambda^{\frac{1}{k}-2}\right] \\ &= \frac{\lambda^{\frac{1}{k}-2}}{k^2}. \end{aligned}$$

即

$$I(\theta^k) = \frac{\theta^{1-2k}}{k^2},$$

则只需令 $1-2k=0, k=1/2$ 即可使得 $I(\theta^{1/2})=4$ 与 θ 无关。即取 $g(\theta)=\theta^{1/2}$.

PROBLEM A.2.6. 设 X_1, \dots, X_n 是来自于 $N(\mu, 1)$ 的一个样本, 令 $g(\mu) = P(X_1 \leq 0)$, 试求 $g(\mu)$ 的 UMVUE.

SOLUTION. 已知正态分布 $N(\mu, \sigma^2)$ 在 σ^2 已知时, $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ 是 μ 的充分统计量, 而正态分布族是指数型分布族, 于是 T 是充分完备统计量。

则寻找 $g(\mu)$ 的 UMVUE 可以转化为寻找 $g(\mu)$ 的一个无偏估计, 再将之对充分完备统计量求条件期望即可。

取 $\phi(X) = \mathbf{I}(X_1 \leq 0)$, 则 $\mathbb{E}\phi(X) = \mathbb{E}\mathbf{I}(X_1 \leq 0) = P(X_1 \leq 0) = g(\mu)$. 于是 $\phi(X)$ 是 $g(\mu)$ 的无偏估计。又因为 T 是充分完备统计量, UMVUE 即为 $\mathbb{E}[\phi(X)|T]$.

$$\begin{aligned} \mathbb{E}[\phi(X)|T=t] &= \mathbb{E}[\mathbf{I}(X_1 \leq 0)|\bar{X}=t] \\ &= P(X_1 \leq 0|\bar{X}=t) \\ &= P(X_1 - \bar{X} \leq -t). \end{aligned} \tag{A.2}$$

接下来我们考虑 $X_1 - \bar{X}$ 的分布。由正态分布的性质可知, 正态随机变量的线性组合仍然服从正态分布, 则只需求 $\mathbb{E}[X_1 - \bar{X}], \text{Var}(X_1 - \bar{X})$ 即可。

$$\mathbb{E}[X_1 - \bar{X}] = 0,$$

$$\begin{aligned} \text{Var}(X_1 - \bar{X}) &= \text{Var}(X_1) + \text{Var}(\bar{X}) - 2\text{Cov}(X_1, \frac{X_1 + \dots + X_n}{n}) \\ &= 1 + \frac{1}{n} - 2 \cdot \frac{1}{n} \\ &= \frac{n-1}{n}. \end{aligned}$$

于是 $X_1 - \bar{X} \sim N(0, \frac{n-1}{n})$, 则(A.2)) 写为

$$\begin{aligned} P(X_1 - \bar{X} \leq -t) &= P\left(\frac{X_1 - \bar{X}}{\sqrt{\frac{n-1}{n}}} \leq -\sqrt{\frac{n-1}{n}}t\right) \\ &= \Phi\left(-\sqrt{\frac{n-1}{n}}t\right), \end{aligned}$$

其中 $\Phi(\cdot)$ 是标准正态分布的分布函数。于是, $g(\mu)$ 的 UMVUE 即为 $\Phi\left(-\sqrt{\frac{n-1}{n}}\bar{X}\right)$.

PROBLEM A.2.7. 设 X_1, \dots, X_n 是来自 $b(1, p)$ 的一个样本, 求 $p(1-p)$ 的 UMVUE.

SOLUTION. 首先证明 $T(X) = \sum_{i=1}^n X_i \sim b(n, p)$ 是 p 的充分完备统计量, 其中充分统计量可根据 $P(X_1 = x_1, \dots, X_n = x_n | T = t) = \binom{n}{t}^{-1}$ 与 p 无关从而得证; 也可求联合分布函数从而由因子分解定理直接得出:

$$P(X_1 = x_1, \dots, X_n = x_n) = p^{\sum x_i} (1-p)^{n-\sum x_i}.$$

又二项分布族是完备分布族则统计量 T 诱导的分布族仍完备, T 为完备统计量。

此外, 也可以由二项分布族是指数分布族, 其标准形式 (参数写为 θ) 为

$$\begin{aligned} p_\theta(\mathbf{x}) &= \binom{n}{\sum x} \theta^{\sum x} (1-\theta)^{n-\sum x} = \binom{n}{\sum x} (1-\theta)^n \exp\left\{x \ln \frac{\theta}{(1-\theta)}\right\} \\ &= (1+e^w)^{-n} \exp\left\{w \sum x\right\} \binom{n}{\sum x}, \end{aligned}$$

其中 $w = \ln \frac{\theta}{1-\theta} \Rightarrow \theta = \frac{e^w}{1+e^w}, w \in \Omega = \mathbb{R}$ 有内点, 所以 $T(X) = \sum_{i=1}^n X_i$ 是充分完备统计量。

下面找 $g(p) = p(1-p)$ 的无偏估计。

法一: 注意到

$$\begin{aligned} \mathbb{E}[T(X)] &= np \Rightarrow \mathbb{E}\left[\frac{T}{n}\right] = p, \\ \mathbb{E}[T(T-1)] &= \sum_{t=0}^n t(t-1) \binom{n}{t} p^t (1-p)^{n-t} = n(n-1)p^2 \Rightarrow \mathbb{E}\left[\frac{T(T-1)}{n(n-1)}\right] = p^2 \end{aligned}$$

由 Lehmann-Scheffé 定理, 如果找到 $h(T)$ 使得 $\mathbb{E}[h(T)] = g(p) = p(1-p)$, 则 $h(T)$ 就是 $g(p)$ 的 UMVUE.

用以上两个无偏估计:

$$\begin{aligned} \mathbb{E}\left[\frac{T}{n} - \frac{T(T-1)}{n(n-1)}\right] &= p - p^2 = g(p), \\ \Rightarrow h(T) &= \frac{T(n-T)}{n(n-1)}. \end{aligned}$$

法二: 令 $\varphi(X) = \mathbb{I}_{\{X_1=1, X_2=0\}}$, 则

$$\mathbb{E}\varphi(X) = P(X_1 = 1, X_2 = 0) = p(1-p),$$

令

$$\begin{aligned} h(t) &= \mathbb{E}[\varphi(X)|T=t] = \mathbb{E}\left[\mathbb{I}_{\{X_1=1, X_2=0\}}|\sum X_i=t\right] \\ &= \frac{P(X_1=1, X_2=0, \sum_{i=1}^n X_i=t)}{P(\sum_{i=1}^n X_i=t)} = \frac{\binom{n-2}{t-1}}{\binom{n}{t}} = \frac{t(n-t)}{n(n-1)}. \end{aligned}$$

则 $h(T) = \frac{T(n-T)}{n(n-1)}$ 是 $g(p)$ 的 UMVUE.

PROBLEM A.2.8. 设 $T(x)$ 是 $g(\theta)$ 的无偏估计, $\text{Var}(T(X)) < \infty$.

证明: $T(X)$ 为 $g(\theta)$ 的 UMVUE 的充要条件是对于任一 0 的无偏估计 $\phi(x)$, 如果 $\text{Var}(\varphi(X)) < \infty$, 则 $\text{Cov}(\varphi(X), T(X)) = 0$.

SOLUTION. $\Rightarrow:$

若 $T(X)$ 是 $g(\theta)$ 的 UMVUE, 则对于其余 $g(\theta)$ 的无偏估计 $T'(X)$, 总有 $\text{Var}(T(X)) \leq \text{Var}(T'(X))$.

构造 $T'(X) := T(X) + a\varphi(X)$, $\mathbb{E}[T(X) + a\varphi(X)] = g(\theta)$, 则 $T'(X)$ 是 $g(\theta)$ 的无偏估计。于是

$$\begin{aligned} \text{Var}(T(X) + a\varphi(X)) &\geq \text{Var}(T(X)), \\ \text{Var}(T(X)) + a^2 \text{Var}(\varphi(X)) + 2a\text{Cov}(T(X), \varphi(X)) &\geq \text{Var}(T(X)), \\ a^2 \text{Var}(\varphi(X)) + 2a\text{Cov}(T(X), \varphi(X)) &\geq 0. \end{aligned}$$

最后一个不等式左端是一个关于 a 的一元二次方程, 要使上述不等式成立, 该一元二次方程根的判别式应小于等于 0, 即

$$\Delta = 4[\text{Cov}(T(X), \varphi(X))]^2 \leq 0,$$

则 $\text{Cov}(T(X), \varphi(X)) = 0$, 充分性得证。

$\Leftarrow:$

若 $T(X)$ 是满足 $\text{Cov}(T(X), \varphi(X)) = 0$ 的 $g(\theta)$ 的无偏估计量。现设 $g(\theta)$ 的任一无偏估计量 $T^*(X)$, 计算 $\text{Var}(T^*)$.

$$\begin{aligned} \text{Var}(T^*) &= \text{Var}(T^* - T + T) \\ &= \text{Var}(T) + \text{Var}(T^* - T) + 2\text{Cov}(T, T^* - T) \\ &= \text{Var}(T) + \text{Var}(T^* - T) \\ &\geq \text{Var}(T) \end{aligned}$$

倒数第二个等式是因为 $T^* - T$ 是 0 的无偏估计量。必要性得证。

PROBLEM A.2.9. 设 X_1, \dots, X_n 是来自 $N(\mu, \sigma^2)$ 的一个样本, 记 $\theta = (\mu, \sigma^2)$, 固定 c , 令 $q_c(\theta)$ 表示 c 下方总体的比例, 求 $q_c(\theta)$ 的 MLE 和 UMVUE.

SOLUTION. (1) 先求参数 $\theta = (\mu, \sigma^2)$ 的 MLE, 再利用 MLE 的不变性进行计算。
正态分布的对数似然函数为

$$\begin{aligned} l(\mu, \sigma^2; \mathbf{X}) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \\ \frac{\partial l}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = 0, \quad \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (X_i - \mu)^2 = 0 \\ \Rightarrow \hat{\mu} &= \bar{X}, \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}. \end{aligned}$$

而 $q_c(\theta) = \mathbb{P}(X_1 \leq c) = \Phi\left(\frac{c-\mu}{\sigma}\right)$, 故由 MLE 的不变性可知

$$\hat{q}_c(\theta) = \Phi\left(\frac{c-\hat{\mu}}{\hat{\sigma}}\right).$$

(2) $N(\mu, \sigma^2)$ 为指数分布族且其充分完备统计量为 $T = (\bar{X}, S^2)$, 这里的 $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$. 由于 $\mathbb{I}_{\{X_1 \leq c\}}$ 为 $q_c(\theta)$ 的无偏估计, 则 $q_c(\theta)$ 的 UMVUE 可以是 $\mathbb{E}[\mathbb{I}_{\{X_1 \leq c\}}|T]$.

构造辅助统计量 (其分布与参数无关的统计量) $Z = \frac{X_1 - \bar{X}}{S}$, 根据 Basu 定理 (充分统计量与辅助统计量互相独立), Z, T 互相独立, 于是

$$\begin{aligned} \mathbb{E}[I_{\{X_1 \leq c\}}|T] &= \mathbb{P}(X_1 \leq c|T) = \mathbb{P}\left(Z \leq \frac{c-\bar{X}}{S} \middle| T\right) \\ &= \mathbb{P}\left(Z \leq \frac{c-\bar{X}}{S}\right) = \mathbb{P}(kZ \leq kV), \end{aligned}$$

其中 $V = \frac{c-\bar{X}}{S}, k = \frac{\sqrt{n}}{n-1}$.

可以计算 Z 的密度函数为 (关于 Z 的分布见 Problem A.1):

$$f_Z(z) = \frac{\sqrt{n}}{n-1} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \left(1 - \frac{nz^2}{(n-1)^2}\right)^{\frac{n}{2}-2}, \quad -\frac{n-1}{\sqrt{n}} \leq Z \leq \frac{n-1}{\sqrt{n}},$$

令 $u = kz$, 则

$$f(u) = \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} (1-u^2)^{\frac{n}{2}-2}, \quad -1 \leq u \leq 1,$$

且当 $kv \leq -1$ 时, $\mathbb{P}(u \leq kv) = 0$; $kv \geq 1$ 时 $\mathbb{P}(u \leq kv) = 1$. 当 $kv \in (-1, 1)$ 时,

$$\mathbb{P}(u \leq kv) = \int_{-1}^{kv} f(u) du = \int_0^{\frac{1}{2}+\frac{1}{2}kv} \frac{\Gamma(n-1)}{\Gamma(\frac{n-2}{2})\Gamma(\frac{n-2}{2})} V^{n/2-2} (1-V)^{n/2-2} dV = G(1/2+1/2kv),$$

其中第二个等式是令 $u = 2v - 1, G(\cdot)$ 表示 Beta 分布在 $Be\left(\frac{n-2}{2}, \frac{n-2}{2}\right)$ 的分布函数。

PROBLEM A.2.10. 令随机变量 $X \sim U(0, \theta), \theta > 0$ 是未知参数。定义动作空间范围 $[0, \infty)$, 损失函数 $L(\theta, d) = (\theta-d)^2$, 其中 d 是动作选择。考虑决策规则 $d_\mu(x) = \mu x, \mu \geq 0$.

问：当 μ 为何值时 d_μ 是无偏的？并证明 $\mu = \frac{3}{2}$ 是 d_μ 满足可容许的必要条件。

SOLUTION. 由于 $X \sim U(0, \theta)$, $\mathbb{E}[X] = \theta/2$, $Var(X) = \frac{1}{12}\theta^2$, 要使 d_μ 是 θ 无偏的，则

$$\mathbb{E}[d_\mu(X)] = \mu\mathbb{E}[X] = \mu\frac{\theta}{2} = \theta$$

于是 $\mu = 2$.

要证明 $\mu = 3/2$, d_μ 是可容许的，只需要计算期望损失最小

$$\begin{aligned}\mathbb{E}[L(\theta, d)] &= \mathbb{E}[\theta - \mu X]^2 \\ &= \mathbb{E}[\theta^2 - 2\theta\mu X + \mu^2 X^2] \\ &= \theta^2 - 2\theta\mu\mathbb{E}[X] + \mu^2\mathbb{E}[X^2] \\ &= \theta^2\left(\frac{1}{3}\mu^2 - \mu + 1\right).\end{aligned}$$

要使期望损失最小，则 $\mu = \frac{3}{2}$, 必要性得证。

PROBLEM A.2.11. 设 X_1, \dots, X_n 是 i.i.d. 随机变量， $0 < \theta < 1$, $P(X_1 = -1) = \frac{1-\theta}{2}$, $P(X_1 = 1) = \frac{\theta}{2}$, $P(X_i = 0) = \frac{1}{2}$

- (1) 求 θ 的 MLE $\hat{\theta}_1$, 并说明是否无偏;
- (2) 求 θ 的矩估计 $\hat{\theta}_2$;
- (3) 计算 θ 的无偏估计的方差的 C-R 下界.

SOLUTION.

(1) 由题意知，对数似然函数为

$$l(\theta) = \left[\sum_{i=1}^n I(x_i = -1) \right] \ln(1 - \theta) + \left[\sum_{i=1}^n I(x_i = 1) \right] \ln \theta - n \ln 2,$$

求其极大值点可得

$$\hat{\theta}_1 = \frac{\sum I(x_i = 1)}{\sum I(x_i = 1 \text{ or } -1)}$$

不妨设 $n_1 = \sum I(x_i = -1)$, $n_2 = \sum I(x_i = 0)$, $n_3 = \sum I(x_i = 1)$, $n = n_1 + n_2 + n_3$. 则极大似然估计可以写为

$$\hat{\theta}_1 = \begin{cases} \frac{n_3}{n_1+n_3}, & n_2 \neq n \\ c(\in [0, 1]), & n_2 = n \end{cases}$$

考虑是否无偏，由于

$$\mathbb{E}(\hat{\theta}_1) = \mathbb{E}\left(\frac{n_3}{n_1+n_3}\right) = \mathbb{E}\left[\mathbb{E}\left(\frac{n_3}{n_1+n_3} | n_1+n_3\right)\right]$$

又 $P(X_1 = 1 | X_1 = -1 \text{ or } X_1 = 1) = \theta$. 可知当 $n_1 + n_3 = m$ 的条件下, n_3 服从二项分布 $b(m, \theta)$.

所以 $\mathbb{E}(\frac{n_3}{n_1+n_3} | n_1 + n_3 = m) = \theta$. 则

$$\mathbb{E}(\hat{\theta}_1) = \mathbb{E}(\theta) = \theta.$$

另一种思考:

$$\begin{aligned}\mathbb{E}\hat{\theta}_1 &= \mathbb{E}[\mathbb{E}(\hat{\theta}_1 | n_2)] \\ &= \sum_{k=0}^{n-1} \mathbb{E}(\hat{\theta}_1 | n_2 = k) P(n_2 = k) + \mathbb{E}(\hat{\theta}_1 | n_2 = n) P(n_2 = n) \\ &= \sum_{k=0}^{n-1} \binom{n}{k} \left(\frac{1}{2}\right)^n \frac{P(X_1 = 1)}{1 - P(X_1 = 0)} + \left(\frac{1}{2}\right)^n c \\ &= \theta(1 - \frac{1}{2^n}) + c \frac{1}{2^n} \rightarrow \theta.\end{aligned}$$

则 $MLE\hat{\theta}_1$ 不是 θ 的无偏估计而是渐进无偏估计。

(2) 由 $\mathbb{E}X = \theta - \frac{1}{2}$, 带入样本矩可得

$$\hat{\theta}_2 = \frac{1}{2} + \frac{\sum_{i=1}^n X_i}{n}$$

(3)

$$\begin{aligned}\frac{\partial^2 l}{\partial \theta^2} &= -\frac{\sum I(x_i = -1)}{(1-\theta)^2} - \frac{\sum I(x_i = 1)}{\theta^2}, \\ i(\theta) &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial \theta^2}\right) = \frac{n}{2\theta(1-\theta)}.\end{aligned}$$

则其 C-R 下界为 $i^{-1}(\theta) = \frac{2\theta(1-\theta)}{n}$.

NOTE OF PROBLEM A.2.11. 本题来自《高等数理统计》习题 2.35.

PROBLEM A.2.12. 设 X_1, \dots, X_n i.i.d. X_1 取值有 4 种可能, 概率分别为 $p_1 = 1 - \theta, p_2 = \theta - \theta^2, p_3 = \theta^2 - \theta^3, p_4 = \theta^3$. 以 N_j 记 X_1, \dots, X_n 中出现各种可能结果的次数, $\sum N_j = n$.

- (1) 确定 a_1, a_2, a_3, a_4 使得 $T = \sum_j a_j N_j$ 为 θ 的无偏估计;
- (2) 比较 $Var(T)$ 与 θ 的无偏估计方差的 C-R 下界进行比较。

SOLUTION.

(1) 由多项分布的性质可以得知 $N_j \sim b(n, p_j)$.

$$\begin{aligned}\mathbb{E}T &= n \sum a_j p_j = \theta \\ &= n[a_1(1-\theta) + a_2(\theta-\theta^2) + a_3(\theta^2-\theta^3) + a_4\theta^3] = \theta\end{aligned}$$

比较系数得 $a_1 = 0, a_2 = a_3 = a_4 = \frac{1}{n}$.

(2)

$$\begin{aligned}Var(T) &= \frac{1}{n^2} \left[\sum_{j=2}^3 Var(N_j) \right] \\ &= \frac{1}{n^2} [n - N_1] = \frac{\theta(1-\theta)}{n}\end{aligned}$$

下面求 θ 无偏估计方差的 C-R 下界。由于 $N_1, N_2, N_3, N_4 \sim Multinomial$, 则

$$\begin{aligned}p(N_1 = n_1, N_2 = n_2, N_3 = n_3, N_4 = n_4) &\propto (1-\theta)^{n_1} (\theta-\theta^2)^{n_2} (\theta^2-\theta^3)^{n_3} \theta^{3n_4} \\ l = \ln p &= \ln(1-\theta) + (n_2 + 2n_3 + 3n_4) \ln \theta \\ \frac{\partial \ln p}{\partial \theta} &= \frac{-(n_1 + n_2 + n_3)}{1-\theta} + \frac{n_2 + 2n_3 + 3n_4}{\theta} \\ \frac{\partial^2 \ln p}{\partial \theta^2} &= \frac{n_1 + n_2 + n_3}{(1-\theta)^2} - \frac{n_2 + 2n_3 + 3n_4}{\theta^2}\end{aligned}$$

于是

$$\begin{aligned}i(\theta) &= \mathbb{E}\left(-\frac{\partial^2 \ln p}{\partial \theta^2}\right) \\ &= \frac{n(\theta^2 + \theta + 1)}{\theta(1-\theta)}.\end{aligned}$$

则 C-R 下界为 $\frac{\theta(1-\theta)}{n(\theta^2+\theta+1)} < \frac{\theta(1-\theta)}{n} = Var(T)$. (“ $<$ ” 是因为 $\theta^2 + \theta + 1 > 1$.)

NOTE OF PROBLEM A.2.12. 本题来自《高等数理统计》习题 2.25.

PROBLEM A.2.13.

- (1) 设 $X \sim p(x; \theta) = \frac{f(x)}{h(\theta)}, \theta \leq x \leq b$. θ 是未知参数, $h(\theta) > 0, f(x) > 0$, 证明若 $T(X)$ 是 $g(\theta)$ 的无偏估计, 则 $-T(X) = \frac{g(X)h'(X)+g(X)h'(X)}{f(X)}$.
- (2) 设 X_1, \dots, X_n 是来来自于 (1) 中分布的一个样本, 则 $g(\theta)$ 的唯一的 UMVUE 为 $T(X) = g(X_{(1)}) - \frac{g'(X_{(1)})}{n} \frac{h(X_{(1)})}{f(X_{(1)})}$.

SOLUTION.

(1) 由于 $\mathbb{E}T(X) = g(\theta)$, 得

$$\begin{aligned}\int_{\theta}^b \frac{T(x)f(x)}{h(\theta)} dx &= g(\theta) \\ \int_{\theta}^b T(x)f(x)dx &= g(\theta)h(\theta) \\ g'(\theta)h(\theta) + g(\theta)h'(\theta) &= -T(\theta)f(\theta)\end{aligned}$$

将 θ 换为 X 即得要证明的结论。

(2) 首先我们证明上述分布是完备分布族。这只需要找到 θ 的任意一个无偏估计量 $\varphi(x)$, 由于 $\mathbb{E}\varphi(X) = \int_{\theta}^b \varphi(x) \frac{f(x)}{h(\theta)} dx = 0 \Rightarrow \varphi(x) \equiv 0$ 所以该分布是完备分布族。接下来求 θ 的充分统计量。

$$p(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i)/h^n(\theta) I(\theta \leq x_{(1)}) I(x_{(n)} \leq b)$$

由因子分解定理可知 $X_{(1)}$ 是该分布的充分统计量, 故是完备充分统计量。则我们只需要证明题目中给出的式子是无偏的即可。 $X_{(1)}$ 的密度函数 $p_1(t) = \frac{nf(t)}{h(\theta)}[1 - F(t)]^{n-1}, \theta \leq t \leq b$.

$$\begin{aligned}\mathbb{E}g(\theta) &= \int_{\theta}^b g(\theta)p_1(t)dt \\ &= \int_{\theta}^b -g(\theta)d[1 - F(t)]^n \\ &= g(\theta) + \int_{\theta}^b \frac{g'(t)h(\theta)[1 - F(t)]}{nf(t)} p_1(t)dt\end{aligned}$$

又因为 $F(t) = \int_{\theta}^t \frac{f(x)}{h(\theta)} dx, h(\theta)[1 - F(t)] = \int_t^b f(x)dx$, 可得 $h(\theta)[1 - F(x)] = h(x)$, 代入前面的公式可得

$$\begin{aligned}\int_{\theta}^b g(x)p_1(x)dx &= g(\theta) + \int_{\theta}^b \frac{g'(x)}{n} \frac{h(x)}{f(x)} p_1(x)dx \\ \int_{\theta}^b [g(x) - \frac{g'(x)}{n} \frac{h(x)}{f(x)}]p_1(x)dx &= g(\theta) \\ \mathbb{E}[g(X_{(1)}) - \frac{g'(X_{(1)})}{n} \frac{h(X_{(1)})}{f(X_{(1)})}] &= g(\theta).\end{aligned}$$

就证明了题中的结论。

NOTE OF PROBLEM A.2.13. 本题来自《高等数理统计》习题 2.16.

PROBLEM A.2.14. 设 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} P(\lambda), \bar{X}, S^2$ 分别为样本均值和样本方差。

- (1) 证明 \bar{X} 是 λ 的 UMVUE;
- (2) 证明 $\mathbb{E}(S^2|\bar{X}) = \bar{X}$, 并利用这一等式证明 $Var(S^2) > Var(\bar{X})$.

SOLUTION.

- (1) 由于 $P(X_1 = x_1, \dots, X_n = x_n) = \frac{\lambda^{\sum x_i}}{\prod(x_i!)} e^{-n\lambda}$ 为指数分布族, 且 $\sum X_i$ 是充分统计量, 则 \bar{X} 是充分完备统计量, 又 $\mathbb{E}\bar{X} = \lambda$, 故 \bar{X} 是 λ 的 UMVUE.
- (2) 由重期望公式得 $\mathbb{E}[\mathbb{E}(S^2|\bar{X})] = \mathbb{E}S^2 = \lambda$, 故 $\mathbb{E}(S^2|\bar{X})$ 既是 λ 的无偏估计又是充分完备统计量 \bar{X} 的函数, 则 $\mathbb{E}(S^2|\bar{X})$ 是 λ 的 UMVUE, 又因为 UMVUE 的唯一性可知 $\mathbb{E}(S^2|\bar{X}) = \bar{X}$.

由重方差公式可知

$$\begin{aligned} Var(S^2) &= \mathbb{E}[Var(S^2|\bar{X})] + Var[\mathbb{E}(S^2|\bar{X})] \\ &= \mathbb{E}[Var(S^2|\bar{X})] + Var(\bar{X}) \\ &> Var(\bar{X}). \end{aligned}$$

□

Definition A.2.1 (Fisher 信息量). 设统计结构 $\{\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\}\}$ 可控, 有随机向量

$$S_\theta(X) = \left(\frac{\partial \ln p_\theta(x)}{\partial \theta_1}, \dots, \frac{\partial \ln p_\theta(x)}{\partial \theta_k} \right)^\top$$

满足

- (1) $S_\theta(X)$ 对一切 $\theta \in \Theta$ 有定义;
- (2) $\mathbb{E}_\theta(S_\theta(X)) = 0, \forall \theta \in \Theta$;
- (3) $\mathbb{E}_\theta \|S_\theta(X)\|^2 < \infty$,

则把 $S_\theta(X)$ 的协方差阵

$$I(\theta) = \mathbb{V}ar_\theta(S(X)) = \mathbb{E}_\theta[S_\theta(X)S_\theta^\top(X)]$$

称为 Fisher 信息矩阵。

如果进一步假设 $\ln p_\theta(x)$ 对 θ 的二阶导存在且积分与微分可以交换次序 (在多数场合成立, 比如指数族), 则 Fisher 信息可如下计算

$$I_{ij} = -\mathbb{E}_\theta \left[\frac{\partial^2 \ln p_\theta(x)}{\partial \theta_i \partial \theta_j} \right], 1 \leq i, j \leq k.$$

由独立变量组成的向量的 Fisher 信息不论其组成部分是否同分布都等于各组成部分 Fisher 信息之和。

Definition A.2.2 (Cramer-Rao 正则族). 分布族称为 $C\text{-}R$ 正则族如果

- (1) $\Theta \in \mathbb{R}^k$ 是开矩形;
- (2) $\partial \ln p / \partial \theta_i, i = 1, \dots, k$ 对所有 θ 都存在;
- (3) 支撑 $\{x : p_\theta(x) > 0\}$ 与 θ 无关;
- (4) 对 $p_\theta(x)$, 积分微分可交换;
- (5) 对一切 $1 \leq i, j \leq k, \forall \theta, \mathbb{E}_\theta \left| \frac{\partial \ln p}{\partial \theta_i} \frac{\partial \ln p}{\partial \theta_j} \right| < \infty$.

指数族为 $C\text{-}R$ 正则族。

对于 $C\text{-}R$ 正则族, 充分统计量诱导的统计结构的 Fisher 信息量与原信息结构的 Fisher 信息量相等 (实际上是充要条件: $I^T(\theta) = I(\theta) \Leftrightarrow T(X)$ 是充分统计量)。

Theorem A.2.3 (C-R 不等式). 设 $\{p_\theta(x), \theta \in \Theta \subset \mathbb{R}^k\}$ 为 $C\text{-}R$ 正则族, 设 $g(\theta) = (g_1(\theta), \dots, g_s(\theta))^\top, s \leq k$, 且 $\partial g_i(\theta) / \partial \theta_j$ 对一切 $i = 1, \dots, s, j = 1, \dots, k$ 都存在。假设 $T(X)$ 是 $g(\theta)$ 的模平方可积的无偏估计, 记 $\Delta = \mathbb{E}(TS^\top) = \frac{d}{d\theta}g(\theta)$. 则有

$$\mathbb{V}ar_\theta(T(X)) \geq \Delta I^{-1} \Delta^\top,$$

其中 $\Delta I^{-1} \Delta^\top$ 称为 $g(\theta)$ 无偏估计协方差的下界或者 $C\text{-}R$ 下界。

$s = k = 1$ 时退化为

$$\mathbb{V}ar_\theta(T(X)) \geq \left(\frac{\partial g(\theta)}{\partial \theta} \right)^2 / I(\theta).$$

Remark 4. 对信息不等式我们要牢记其 $C\text{-}R$ 正则族的条件。如果不满足条件时有这样的无偏估计存在, 其方差比直接算得的 $C\text{-}R$ 下界还要小。比如 $p_\theta(x) = e^{-(x-\theta)}$ 其支撑与 θ 有关显然不是 $C\text{-}R$ 正则族, 直接计算其 $C\text{-}R$ 下界为

$$1/I = 1/nI_1 = 1/n\mathbb{E} \left(\frac{\partial \ln p}{\partial \theta} \right)^2 = 1/n.$$

而知其充分统计量为 $X_{(1)}$, 直接根据其充分统计量构造无偏估计得到 UMVUE 为 $X_{(1)} - \frac{1}{n}$ 的方差为 $\frac{1}{n^2}$.

Remark 5. 如果某个无偏估计的方差达到了 $C\text{-}R$ 下界, 那么它显然就是 UMVUE 了。但如果达不到也有可能是 UMVUE。

Definition A.2.4 (有效无偏估计). 通常把达到 $C\text{-}R$ 下界的无偏估计称为有效无偏估计, 而把无偏估计的方差与其 $C\text{-}R$ 下界之比的倒数称为该估计的效。设 $\{p_\theta(x), \theta \in \Theta\}$ 是 Cramer-Rao 正则族, $g(\theta)$ 为可估参数, $T(X)$ 是 $g(\theta)$ 的一个无偏估计, 则称

$$\frac{[g'(\theta)]^2 I^{-1}(\theta)}{\mathbb{V}ar_\theta(T(X))}$$

为估计的效。如果效等于 1 则称 $T(X)$ 为 $g(\theta)$ 的有效无偏估计。

Theorem A.2.5. 设 $X = (X_1, \dots, X_n)$ 的分布为指数型分布

$$p(\mathbf{x}; \theta) = \exp\{\theta T(\mathbf{x}) + c(\theta) + d(\mathbf{x})\},$$

则 $g(\theta)$ 存在有效无偏估计的充要条件是存在常数 a, b 使得 $g(\theta) = a\mathbb{E}_\theta[T(X)] + b$.

PROBLEM A.2.15. 设 $X_1, \dots, X_n \sim b(k, p)$, $\hat{p} = \bar{X}/k$, 证明: \hat{p} 是 p 的有效估计。

SOLUTION. 首先证明无偏性

$$\mathbb{E}[\bar{X}] = kp \Rightarrow \mathbb{E}[\hat{p}] = \frac{\mathbb{E}[\bar{X}]}{k} = p.$$

求单个观测 $X_i \sim B(k, p)$ 的 Fisher 信息量

$$\begin{aligned} P(X_i = x) &= \binom{k}{x} p^x (1-p)^{k-x}, x = 1, \dots, k, \\ l(p) &= \log \binom{k}{x} + x \log p + (k-x) \log(1-p), \\ \Rightarrow l''(p) &= -\frac{x}{p^2} - \frac{k-x}{(1-p)^2}. \\ I_1(p) &= -\mathbb{E}[l''(p)] = \frac{k}{p(1-p)}. \end{aligned}$$

则样本量为 n 时总的 Fisher 信息量为

$$I_n(p) = nI_1(p) = \frac{nk}{p(1-p)}.$$

对任意的无偏估计量 \tilde{p} 有

$$\mathbb{V}ar(\tilde{p}) \geq \frac{1}{I_n(p)} = \frac{p(1-p)}{nk}.$$

下面求 $\hat{p} = \bar{X}/k$ 的方差

$$\begin{aligned} \mathbb{V}ar(X_i) &= kp(1-p), \mathbb{V}ar(\bar{X}) = \frac{kp(1-p)}{n}, \\ \Rightarrow \mathbb{V}ar(\hat{p}) &= \frac{p(1-p)}{nk}, \end{aligned}$$

达到 C-R 下界, 于是是有效估计。

另一种方法:

$$\begin{aligned} P(\mathbf{X} = \mathbf{x}) &= \prod_{i=1}^n \binom{k}{x_i} p^{\sum x_i} (1-p)^{k-\sum x_i} = \exp \left\{ n\bar{x} \log \frac{p}{1-p} + k \log(1-p) + \sum \log \binom{k}{x_i} \right\}, \\ \mathbb{E}_p[T(X)] &= \mathbb{E}_p[\bar{X}] = kp, g(p) = p. \end{aligned}$$

则不妨令 $a = \frac{1}{k}, b = 0$ 可知 p 的有效无偏估计存在且为

$$\hat{p} = aT(X) + b = \frac{\bar{X}}{k}$$

得证。

Definition A.2.6 (矩估计). 设 X_1, \dots, X_n 是来自总体的一个样本, 记 r 阶原点矩核和 r 阶样本原点矩如下

$$\mu_r = \mathbb{E}X_i^r, m_r = \frac{1}{n} \sum X_i^r.$$

如果某参数 θ 可以表示为总体前 k 阶矩的函数

$$\theta = g(\mu_1, \dots, \mu_k)$$

则我们可以用 $\hat{\theta}(X) = g(m_1, \dots, m_k)$ 估计 θ , 称为 θ 的矩估计。矩的估计是不唯一的。

PROBLEM A.2.16. 设 $(X_1, Y_1), \dots, (X_n, Y_n)$ 独立同分布, X, Y 独立且 $X \sim \text{Exp}(\lambda), Y \sim \text{Exp}(\mu)$.

(1) 求 λ, μ 的 MLE;

(2) 令 $Z_i = \min(X_i, Y_i), \Delta_i = \begin{cases} 1, & Z_i = X_i \\ 0, & Z_i = Y_i \end{cases}$. 在只知道 Z_i, Δ_i 的情形下求 λ, μ 的 MLE.

SOLUTION. (1) 先求其联合分布的对数似然函数

$$l(\lambda, \mu) = n \ln \lambda - \lambda \sum_{i=1}^n X_i + n \ln \mu - \mu \sum_{i=1}^n Y_i,$$

分别对 λ, μ 求导数并令为 0 即得

$$\hat{\lambda} = \frac{1}{\bar{X}}, \hat{\mu} = \frac{1}{\bar{Y}}.$$

(2) 相当于 X_i, Y_i 是潜在的生存空间, 我们只能观测到较早失效的时间和失效原因 (X 或 Y). 我们知道 $Z \sim \text{Exp}(\lambda + \mu)$, 下面推导 Z_i, Δ_i 的联合概率分布

$$\begin{aligned} f_{Z,\Delta}(z, 1) &= f_X(z)P(Y > z) = \lambda e^{-\lambda z} e^{-\mu z} = \lambda e^{-(\lambda+\mu)z}, z > 0, \\ f_{Z,\Delta}(z, 0) &= f_Y(z)P(X > z) = \mu e^{-\mu z} e^{-\lambda z} = \mu e^{-(\lambda+\mu)z}, z > 0. \end{aligned}$$

则其似然函数为

$$L(\lambda, \mu) = \prod_{i=1}^n [\lambda e^{-(\lambda+\mu)Z_i}]^{\Delta_i} [\mu e^{-(\lambda+\mu)Z_i}]^{1-\Delta_i}$$

记 $n_1 = \sum_{i=1}^n \Delta_i$ 表示来自 X 失效的个数, $n_0 = n - n_1, S = \sum_{i=1}^n Z_i$, 则对数似然为

$$l(\lambda, \mu) = n_1 \ln \lambda + n_0 \ln \mu - (\lambda + \mu)S.$$

于是 MLE 为

$$\hat{\lambda} = \frac{n_1}{S}, \hat{\mu} = \frac{n_0}{S}.$$

注意到 $\hat{\lambda} + \hat{\mu} = n/S$ 正好是 Z 的指数分布参数 $\lambda + \mu$ 的 MLE.

A.3 假设检验

Definition A.3.1. 设 $\phi(x)$ 为定义在 \mathcal{X} 上的可测函数，满足条件 $0 \leq \phi(x) \leq 1$ ，则 $\phi(x)$ 称为检验函数，简称检验。在 $\phi(x)$ 仅取两个值 0, 1 时为非随机化检验否则为随机化检验，其势函数为 $g(\theta) = \mathbb{E}_\theta[\phi(x)]$ 。

非随机化检验的拒绝域 $W = \{x; \phi(x) = 1\}$ 。在随机化检验中，假设样本观测值为 x ，则我们以概率 $\phi(x)$ 拒绝原假设；以概率 $1 - \phi(x)$ 接受原假设。在 $\theta \in \Theta_0$ 时， $g(\theta)$ 是检验犯第一类错误（拒真）的概率，而在 $\theta \in \Theta_1$ 时， $1 - g(\theta)$ 是检验犯第二类错误（取伪）的概率。

Example A.3.2. 电话交换台单位时间内接到的呼唤次数服从 Poisson 分布 $P(\lambda)$, $\lambda > 0$ 为单位时间内接到的平均呼唤次数。为了考察该交换台在单位时间内接到的呼唤次数是否不超过 1 可以建立以下两个假设：

$$H_0 : \lambda \leq 1 \quad vs \quad H_1 : \lambda > 1$$

设 $\mathbf{x} = (x_1, \dots, x_n)$ 是该台 n 次记录，即 \mathbf{x} 是来自 $P(\lambda)$ 的样本 $\mathbf{X} = (X_1, \dots, X_n)$ 的一个观测值。取检验统计量 $T = \sum X_i$ ，它是 λ 的充分完备统计量。一般来说，在原假设 H_0 成立时检验统计量的值较小，而在备择假设成立时检验统计量的值较大。于是存在一个临界值 c ，在 $T \geq c$ 时认为数据样本与原假设矛盾，从而拒绝原假设；在 $T < c$ 时不拒绝原假设。该检验的拒绝域可用检验统计量表示如下：

$$W = \{\mathbf{x} : \sum_{i=1}^n x_i \geq c\}.$$

检验犯第一类错误的概率为

$$\alpha(\lambda) = \sum_{k=c}^{\infty} \frac{(n\lambda)^k}{k!} \exp\{-n\lambda\}, \lambda \leq 1$$

而检验犯第二类错误的概率为

$$\beta(\lambda) = \sum_{k=0}^{c-1} \frac{(n\lambda)^k}{k!} \exp\{-n\lambda\} = 1 - \sum_{k=c}^{\infty} \frac{(n\lambda)^k}{k!} \exp\{-n\lambda\}, \lambda > 1.$$

于是该例题中检验的势函数为

$$g(\lambda) = P_\theta(\mathbf{X} \in W) = \sum_{k=c}^{\infty} \frac{(n\lambda)^k}{k!} \exp\{-n\lambda\}$$

是 λ 的严格单调增函数。

由以上式子可以得到当样本容量 n 固定时，不可能使得犯这两类错误的概率都减少，Neyman 和 Pearson 的假设检验的基本思想就是使第一类错误的概率限制在某一个范围内然后寻找使犯第二类错误的概率尽可能小的检验。在这种思想指导下，寻找一个好的检验法就是对选定的一个较小的数 $\alpha \in (0, 1)$ ，在满足

$$g(\theta) = P_\theta(\mathbf{X} \in W) \leq \alpha, \theta \in \Theta_0$$

的检验中寻找这样的检验，使得在 $\theta \in \Theta_1$ 时 $g(\theta)$ 尽可能大。

由于本题检验的势函数 $g(\lambda)$ 关于 λ 单增，给定 α 后，要使犯第一类错误的概率不超过 α ，只需要

$$g(1) = \sum_{k=c}^{\infty} \frac{n^k}{k!} \exp\{-n\} \leq \alpha.$$

为此我们可以取 $n = 10$ ，然后取定不同的 c 值进行研究。若取 $\alpha = 0.05$ ，由于

$$c = 16 : g(1) = 0.049 < 0.05,$$

$$c = 15 : g(1) = 0.083 > 0.05$$

根据 $N-P$ 假设检验理论的基本思想，取 $c = 16$ 。

但上述问题中水平 $\alpha = 0.05$ 并没有被足量地使用，因为我们可以计算犯第一类错误的概率为 $g(1) = 0.049$ ，于是我们为了构造出水平 α 被足量地使用的检验，考虑随机化检验：

$$\phi(\mathbf{x}) = \begin{cases} 1, & T \geq c = 16 \quad (\text{i.e. } \mathbf{x} \in W) \\ r, & T = 15 \\ 0, & T \leq 14 \end{cases}$$

该检验下犯第一类错误的概率为

$$\begin{aligned} \alpha(\lambda) &= \mathbb{E}_{\lambda} \phi(\mathbf{X}) = P(\mathbf{x} \in W) + rP(T = 15) \\ &= \sum_{k=16}^{\infty} \frac{(10\lambda)^k}{k!} \exp\{-10\lambda\} + r \cdot \frac{(10\lambda)^{15}}{15!} \exp\{-10\lambda\} \\ &\leq \sum_{k=16}^{\infty} \frac{(10\lambda)^k}{k!} \exp\{-10\lambda\} + r \cdot \frac{(10)^{15}}{15!} \exp\{-10\lambda\} \\ &= 0.048740 + r \cdot 0.034718 = 0.05 \Rightarrow r = 0.036. \end{aligned}$$

这个随机化检验的实施过程如下：若样本观测值 x_1, \dots, x_{10} 满足 $T = \sum x_i \geq 16$ 则拒绝原假设；若 $\sum x_i \leq 14$ 则不拒绝原假设；若 $\sum x_i = 15$ 则先做一个成功概率为 $r = 0.036$ 的伯努利试验，若成功则拒绝原假设，反之则不拒绝原假设。

Theorem A.3.3. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是来自分布族 $(P_{\theta} : \theta \in \Theta)$ 的样本， $T(\mathbf{X})$ 为 θ 的充分统计量。则对于任何一个检验函数 $\phi(x)$ ，存在另一个只依赖于 $T(x)$ 的检验函数与 $\phi(x)$ 相互等价。

以上定理告诉我们当 θ 的充分统计量存在时，关于 θ 的任何检验问题其优良检验只需要在充分统计量构成的检验函数中去寻找即可。

Neyman-Pearson 基本引理

如果一个假设只含一个元素则称为简单假设，否则称为复合假设。现在考虑简单原假设 $H_0 : \theta = \theta_0$ 对简单备择假设 $\theta = \theta_1$ 的检验问题。

Definition A.3.4. 在检验问题 (Θ_0, Θ_1) 中，设 $\phi(x)$ 是水平为 α 的检验，如果对于任何一个水平为 α 的检验 $\phi_1(x)$ 都有

$$\mathbb{E}_{\theta_1}[\phi(X)] \geq \mathbb{E}_{\theta_1}[\phi_1(X)]$$

则称 $\phi(x)$ 是水平为 α 的最优势检验 MPT.

N-P 引理证明了再简单原假设对简单备择假设的检验问题中, MPT 一定存在并且可以具体构造出 MPT 检验函数。

Theorem A.3.5 (N-P 基本引理). 设 P_{θ_0} 和 P_{θ_1} 是可测空间 $(\mathcal{X}, \mathcal{B})$ 上的两个不同的概率测度, 关于某个 σ 有限的测度 μ , 有

$$p(x; \theta_0) = \frac{dP_{\theta_0}}{d\mu}, \quad p(x; \theta_1) = \frac{dP_{\theta_1}}{d\mu}$$

则在简单检验问题中,

(1) 对给定的 $\alpha \in (0, 1)$ 存在一个检验函数 $\phi(x)$ 及常数 $k \geq 0$ 使得

$$\mathbb{E}_{\theta_0} \phi(X) = \alpha, \tag{A.3}$$

$$\phi(x) = \begin{cases} 1, & p(x; \theta_1) > k \cdot p(x; \theta_0) \\ 0, & p(x; \theta_1) < k \cdot p(x; \theta_0) \end{cases} \tag{A.4}$$

(2) 由以上两个式子确定的检验函数 $\phi(x)$ 是水平为 α 的 MPT. 反之, 如果 $\phi(x)$ 是水平为 α 的 MPT, 则一定存在常数 $k \geq 0$, 使得 $\phi(x)$ 满足(A.4).

Remark 6. 满足(A.4)式的检验函数 $\phi(x)$ 通常称为似然比检验函数。在集合 $\{x : p(x; \theta_0) > 0 \text{ or } p(x; \theta_1) > 0\}$ 上, 定义似然比检验函数为

$$\lambda(x) = \frac{p(x; \theta_1)}{p(x; \theta_0)}.$$

在 $\lambda(x)$ 比较大时 $p(x; \theta_1)$ 比较大, 原假设为真时观察到样本点 x 的可能性比备择假设为真时观察到样本点 x 的可能性小, 由此很自然地在 $\lambda(x)$ 比较大时拒绝原假设。N-P 基本引理告诉我们在简单的假设检验中, MPT 是似然比检验, 反之似然比检验也是 MPT.

Remark 7. 在似然比 $\lambda(x)$ 具有连续分布时, MPT 检验函数可取为非随机化的形式

$$\phi(x) = \begin{cases} 1, & \lambda(x) \geq k \\ 0, & \lambda(x) < k \end{cases}$$

其中 k 由(A.3)式, 即 $\mathbb{E}_{\theta_0} \phi(X) = P_{\theta_0}\{\lambda(X) \geq k\} = \alpha$ 确定。在 $\lambda(x)$ 为离散随机变量时, 至多在集合 $\{x : \lambda(x) = k\}$ 上实施随机化, MPT 检验函数可取

$$\phi(x) = \begin{cases} 1, & \lambda(x) > k \\ r, & \lambda(x) = k \\ 0, & \lambda(x) < k \end{cases}$$

其中 k, r 由下面的式子确认

$$P_{\theta_0}(\lambda(X) \geq k) \geq \alpha > P_{\theta_0}(\lambda(X) > k) := \alpha_1,$$

$$r = \frac{\alpha - \alpha_1}{P_{\theta_0}(\lambda(X) = k)}.$$

Definition A.3.6 (一致最优势检验). 设 $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ 为一参数统计结构, 考虑检验问题 (Θ_0, Θ_1) . 设 $\phi(x)$ 是水平为 α 的检验, 如果对于任意一个水平为 α 的检验 $\phi_1(x)$, 都有

$$\mathbb{E}_\theta[\phi(X)] \geq \mathbb{E}_\theta[\phi_1(X)], \forall \theta \in \Theta_1$$

则称检验 $\phi(x)$ 是水平为 α 的一致最优势检验 *UMPT*.

Definition A.3.7. 设 $\{p(x; \theta) : \theta \in \Theta\}$ 是含有实参数 θ 的概率密度族, 其中 Θ 是实直线的一个区间。如果存在实值统计量 $T(X)$, 使得对任意的 $\theta_1 < \theta_2$, 都有

- (1) 概率分布 P_{θ_1} 与 P_{θ_2} 不同;
- (2) 似然比 $\lambda(x) = \frac{p(x; \theta_2)}{p(x; \theta_1)}$ 是 $T(X)$ 的非降函数 (或非增函数), 则称概率密度族关于 $T(X)$ 具有非降 (或非增) 单调似然比 *MLE*.

Remark 8. 单参数指数分布族

$$p(x; \theta) = c(\theta) \exp\{Q(\theta)T(x)\}h(x),$$

其中 $c(\theta) > 0$. 假设 $Q(\theta)$ 是 θ 的严增 (或严减) 函数, 则在 $\theta_1 < \theta_2$ 时其似然比

$$\lambda(x) = \frac{c_1(\theta)}{c_2(\theta)} \exp\{[Q(\theta_2) - Q(\theta_1)]T(x)\},$$

是 $T(x)$ 的严增 (或严减) 函数。故单参数指数分布族关于 $T(x)$ 有 *MLR*。这里的 $T(x)$ 是充分统计量。

二项分布族、负二项分布族、*Poisson* 分布族、正态分布族 (均值已知方差未知或者均值未知方差已知的情况) 和指数分布族都是单参数指数组型分布族, 它们关于充分统计量均有 *MLR*。

Theorem A.3.8 (定理 3.8: 单边假设检验的 *MLR*). 设单参数概率密度族 $\{p(x; \theta) : \theta \in \Theta \subset \mathbb{R}\}$ 关于实值统计量 $T(x)$ 具有非降 *MLR*, 则关于单边假设检验问题

$$H_0 : \theta \leq \theta_0 \text{ v.s. } H_1 : \theta > \theta_0,$$

(1) 存在水平为 α 的 *UMPT* 检验函数

$$\phi(T(x)) = \begin{cases} 1, & T(x) \geq c \\ r, & T(x) = c \\ 0, & T(x) < c \end{cases}$$

其中常数 r, c 由下式确定

$$\mathbb{E}_{\theta_0}\phi(T(X)) = \alpha.$$

(2) 这个检验的势函数 $g(\theta) = \mathbb{E}_\theta\phi(T(X))$ 是非降的, 且在集合 $\{\theta : g(\theta) < 1\}$ 上是严格增加的。

(3) 在一切使得 $\mathbb{E}_{\theta_0}\phi(X) = \alpha$ 的检验函数中, 由以上两式确定的检验函数 $\phi(T(X))$ 使得对任意的 $\theta < \theta_0$, $\mathbb{E}_\theta\phi(X)$ 都达到最小。

由 *N-P* 基本引理知, 满足以上两式的检验函数 $\phi(T(x))$ 是该简单假设检验问题 $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta = \theta_1 > \theta_0$ 的水平为 α 的 *MPT*, 由于 θ_1 的任意性, $\phi(T(x))$ 也是检验问题 $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta > \theta_0$ 的水平为 α 的 *UMPT*.

PROBLEM A.3.1. [例 3.5] 若有 N 件产品，其中含有 m 件不合格品。今从中不放回地随机抽取 n 件进行检验，设 n 件中含有 x 件不合格品。考察如下的单边假设检验问题： $H_0 : 0 \leq m \leq m_0$ vs $H_1 : m_0 < m \leq N$ ，求此检验水平为 α 的 UMPT。

SOLUTION.

$$p(x; m) = \frac{\binom{m}{x} \cdot \binom{N-m}{n-x}}{\binom{N}{n}},$$

显然此概率密度族不是单参数指数型分布族。但由于

$$\begin{aligned} \frac{p(x; m+1)}{p(x; m)} &= \frac{\binom{m+1}{x} \cdot \binom{N-m-1}{n-x}}{\binom{m}{x} \binom{N-m}{n-x}} \\ &= \frac{(m+1)(N-m-n+x)}{(N-m)(m+1-x)} \end{aligned}$$

是关于 x 的严格增函数，所以此概率密度族关于 $T(x) = x$ 具有非降 MLR。由定理 3.8 可知题中的检验问题存在水平为 α 的 UMPT，检验函数为

$$\phi(x) = \begin{cases} 1, & x > c \\ r, & x = c \\ 0, & x < c \end{cases}$$

其中 c 由

$$\begin{aligned} P_{m_0}\{x > c\} &\leq \alpha \leq P_{m_0}\{x \geq c\}, \\ \sum_{x \geq c+1} \frac{\binom{m_0}{x} \binom{N-m_0}{n-x}}{\binom{N}{n}} &\leq \alpha \leq \sum_{x \geq c} \frac{\binom{m_0}{x} \binom{N-m_0}{n-x}}{\binom{N}{n}} \end{aligned}$$

确定。而

$$r = \frac{\alpha - P_{m_0}\{x > c\}}{P_{m_0}\{x = c\}}.$$

PROBLEM A.3.2. [例 3.6] 设 $X = (X_1, \dots, X_n) \sim \{N(0, \sigma^2) : \sigma^2 > 0\}$ ，求 $H_0 : \sigma^2 \leq \sigma_0^2$ v.s. $H_1 : \sigma^2 > \sigma_0^2$ 的水平为 α 的 UMPT

SOLUTION. 样本联合密度为

$$p(\mathbf{x}; \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ \frac{-\sum x_i^2}{2\sigma^2} \right\},$$

这是一个单参数指数型分布族，其中 $Q(\sigma^2) = -\frac{1}{2\sigma^2}$ 是 σ^2 严格增函数，则其 UMPT 存在且仅依赖于其充分统计量 $T(X) = \sum X_i^2$ ，拒绝域为 $W = \{x : T(x) \geq c\}$ ，其中 c 由

$\mathbb{E}_{\sigma_0^2} \phi(T(X)) = \alpha$ 确定。
在 $\sigma^2 = \sigma_0^2$ 时, $T(X) \sim \sigma_0^2 \chi^2(n)$, 故 $c = \sigma_0^2 \chi_{1-\alpha}^2(n)$.

PROBLEM A.3.3. [习题 3.8, 4.4] 设 $X = (X_1, \dots, X_n)$ 来自密度函数为

$$p(x; \theta) = \exp\{-(x - \theta)\}, x \geq \theta$$

的样本

- (1) 考虑如下检验问题: $H_0: \theta = 0$ v.s. $H_1: \theta > 0$, 试构造水平为 α 的 UMPT;
- (2) 试基于 $X_{(1)}$ 构造置信水平为 $1 - \alpha$ 的置信区间。

SOLUTION. (1) 样本联合密度为

$$p(\mathbf{x}; \theta) = \exp\left\{-\left(\sum_{i=1}^n x_i - n\theta\right)\right\}, x_{(1)} \geq \theta,$$

对于 $\theta_1 < \theta_2$, 求其似然比

$$\lambda(x) = \frac{p(\mathbf{x}; \theta_2)}{p(\mathbf{x}; \theta_1)} = \exp\{n(\theta_2 - \theta_1)\}, x \geq \theta_2 > \theta_1,$$

知 $\lambda(x)$ 是 $T(x) = x_{(1)}$ 的非降函数, 即该概率密度组关于 $T(x) = x_{(1)}$ 具有 MLR。由定理 3.8 知存在水平为 α 的 UMPT 检验函数

$$\phi(x) = \begin{cases} 1, & x_{(1)} \geq c \\ 0, & \text{others} \end{cases}$$

其中 c 由 $\mathbb{E}_0 \phi(x_{(1)}) = \alpha$ 确定。以下求 c , 求得 $x_{(1)}$ 的概率密度函数为

$$p_\theta(t) = n \exp\{n(\theta - t)\}, t \geq \theta,$$

于是 $\mathbb{E}_0 \phi(x_{(1)}) = \int_c^\infty n \exp\{-nt\} dt = \alpha$, 解得 $c = -\ln \alpha/n$.

(2) 由因子分解定理可知 $X_{(1)}$ 是 θ 的充分统计量, 取枢轴量 $X_{(1)} - \theta \sim Exp(n)$. 所以 $2n(X_{(1)} - \theta) \sim \Gamma(1, \frac{1}{2}) = \chi^2(2)$. 则可以由下式构造置信区间

$$P\{\chi_{\alpha_1}^2(2) \leq 2n(X_{(1)} - \theta) \leq \chi_{1-\alpha_2}^2(2)\} = 1 - \alpha,$$

其中 $\alpha_1 + \alpha_2 = \alpha$. 考虑 $\theta \leq X_{(1)}$, 可取 $\alpha_1 = 0, \alpha_2 = \alpha$.

Theorem A.3.9 (定理 3.15). 设样本 X 服从多参数指数型分布

$$p(x; \theta, \mathbf{r}) = c(\theta, \mathbf{r}) \cdot \exp\left\{\theta \cdot u(x) + \sum_{i=1}^k r_i \cdot t_i(x)\right\} \cdot h(x)$$

(1) $H_0 : \theta \leq \theta_0$ v.s. $H_1 : \theta > \theta_0$ 水平为 α 的 UMPUT 为

$$\phi(u, t) = \begin{cases} 1, & u > c(t) \\ r(t), & u = c(t) \\ 0, & u < c(t) \end{cases}$$

其中 $r(t) \in [0, 1]$ 和 $c(t)$ 由下式确定

$$\mathbb{E}_{\theta_0}[\phi(U, T)|T = t] = \alpha$$

考虑反向检验要将检验中取值 0 和 1 的条件取反号。

(2) $H_0 : \theta_1 \leq \theta \leq \theta_2$ v.s. $H_1 : \theta < \theta_1$ or $\theta > \theta_2$ 水平为 α 的 UMPUT 为

$$\phi(u, t) = \begin{cases} 1, & u < c_1(t) \text{ or } u > c_2(t) \\ r_i(t), & u = c_i(t), i = 1, 2 \\ 0, & c_1(t) < u < c_2(t) \end{cases}$$

其中 $r_i(t) \in [0, 1]$ 和 $c_i(t)$ 由下式确定

$$\mathbb{E}_{\theta_1}[\phi(U, T)|T = t] = \mathbb{E}_{\theta_2}[\phi(U, T)|T = t] = \alpha$$

考虑反向检验要将检验中取值 0 和 1 的条件取反号。

(3) $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta \neq \theta_0$ 水平为 α 的 UMPUT 为

$$\phi(u, t) = \begin{cases} 1, & u < c_1(t) \text{ or } u > c_2(t) \\ r_i(t), & u = c_i(t), i = 1, 2 \\ 0, & c_1(t) < u < c_2(t) \end{cases}$$

其中 $r_i(t) \in [0, 1]$ 和 $c_i(t)$ 由下式确定

$$\mathbb{E}_{\theta_0}[\phi(U, T)|T = t] = \alpha$$

$$\mathbb{E}_{\theta_0}[U\phi(U, T)|T = t] = \alpha\mathbb{E}_{\theta_0}[U|T = t]$$

考虑反向检验要将检验中取值 0 和 1 的条件取反号。

Remark 9. 利用这个检验可以进行两个 Poisson 总体或者两个二项总体的比较。

设相互独立的两个总体 $x \sim P(\lambda), y \sim P(\mu)$, 要比较两个总体的差异就是要比较参数 λ, μ 的差异。为了应用以上定理, 要把他们联合分布表示为

$$p(x, y; \lambda, \mu) = e^{-(\lambda+\mu)} \exp\{\theta \cdot u + r \cdot y\} \cdot (x!y!)^{-1},$$

其中 $\theta = \ln(\mu/\lambda), r = \ln \lambda, u = y, t = x + y$. 等价于关于 θ 的检验问题。按照以上定理只要在直线 $x + y = t$ 的整数点上寻找最优势的条件检验:

$$P(y|x + y = t) = \binom{t}{y} p^y (1-p)^{t-y}, y = 0, 1, \dots, t$$

这是二项分布 $b(t, p)$, $p = \frac{\mu}{\lambda+\mu} = \frac{e^\theta}{1+e^\theta}$, 于是就转变为关于二项分布参数 p 的检验问题。上述方法也适用于两个样本 $X_1, \dots, X_m, Y_1, \dots, Y_n$ 是否来自同一个总体。这是因为 $\sum X_i \sim P(m\lambda)$, $\sum Y_i \sim P(n\mu)$, 检验这两个样本是否来自同一个 Poisson 总体等价于检验 $m\lambda$ 和 $n\mu$ 的比值是否为 $m:n$.

设相互独立的两个总体 $x \sim b(m, p_1), y \sim b(n, p_2)$, 其联合分布为

$$P(x, y; p_1, p_2) = (1 - p_1)^m (1 - p_2)^n \exp\{\theta \cdot u + r \cdot y\} \cdot \binom{m}{x} \binom{n}{y}$$

其中 $\theta = \ln \frac{p_2(1-p_1)}{p_1(1-p_2)}$, $r = \ln \frac{p_1}{1-p_2}$, $u = y$, $t = x + y$, 则比较两个总体的差异即参数 p_i 就化为 θ 检验问题。只要在 $x + y = t$ 的整数点上寻找最优势的条件检验, 在给定 $x + y = t$ 后, y 的条件分布为

$$P(y|x + y = t) = c_i(\theta) \binom{m}{t-y} \binom{n}{y} e^{\theta y}, y = 0, 1, \dots, t$$

其中 $c_i(\theta) = \left[\sum_{y=0}^t \binom{m}{t-y} \binom{n}{y} e^{\theta y} \right]^{-1}$, 在 $\theta = 0$ ($p_1 = p_2$) 时这个条件分布就是超几何分布

$$P(y|x + y = t) = \binom{m}{t-y} \binom{n}{y} \left[\binom{m+n}{t} \right]^{-1}, y = 0, 1, \dots, t$$

用它可以确定 $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$ 的检验问题的临界值.

PROBLEM A.3.4. 现有来自于两个独立总体的 Bernoulli 随机变量, 第一个总体的成功概率为 p_1 , 有 m 个独立观测 X_1, \dots, X_m , 第二个总体的成功概率为 p_2 , 有 n 个独立观测 Y_1, \dots, Y_n . 构造一个 UMP 检验, 检验两个总体成功概率是否相等。

SOLUTION. 上述两个总体的联合分布可以表示成指数族分布的形式:

$$\begin{aligned} P(x, y; p_1, p_2) &= \binom{m}{x} p_1^x (1 - p_1)^{m-x} \cdot \binom{n}{y} p_2^y (1 - p_2)^{n-y} \\ &= (1 - p_1)^m (1 - p_2)^n \cdot \exp\{\theta \cdot u + r \cdot t\} \cdot \binom{m}{x} \binom{n}{y}, \end{aligned}$$

其中 $\theta = \ln \frac{p_2(1-p_1)}{p_1(1-p_2)}$, $r = \ln \frac{p_1}{1-p_1}$, $u = y$, $t = x + y$. 则比较 p_1, p_2 异同的检验就变为了关于 θ 的检验问题。即:

$$H_0 : p_1 = p_2 \text{ vs } H_1 : p_1 \neq p_2 \Leftrightarrow H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0$$

由定理 3.15, 只要在直线 $x + y = t$ 的整数点上寻找最优势的条件检验, 在给定 $x + y = t$

后， $y = u$ 的条件分布为

$$\begin{aligned}
 P(u|t, \theta) &= P_\theta(y = u|x + y = t) = \frac{P(y = u, x = t - u)}{P(x + y = t)} \\
 &= \frac{\binom{m}{t-u} p_1^{t-u} (1-p_1)^{m-t+u} \binom{n}{u} p_2^u (1-p_2)^{n-u}}{\sum_{y=y^*}^{t^*} \binom{m}{t-y} p_1^{t-y} (1-p_1)^{m-t+y} \binom{n}{y} p_2^y (1-p_2)^{n-y}} \\
 &= \frac{\binom{m}{t-u} p_1^{t-u} (1-p_1)^{m-t+u} \binom{n}{u} p_2^u (1-p_2)^{n-u}}{(1-p_1)^m (1-p_2)^n \sum_{y=y^*}^{t^*} \binom{m}{t-y} \binom{n}{y} \left(\frac{p_1}{1-p_1}\right)^{t-y} \left(\frac{p_2}{1-p_2}\right)^y} \\
 &= \frac{\binom{m}{t-u} p_1^{-u} (1-p_1)^u \binom{n}{u} p_2^u (1-p_2)^{-u}}{\sum_{y=y^*}^{t^*} \binom{m}{t-y} \binom{n}{y} \left(\frac{p_1}{1-p_1}\right)^{-y} \left(\frac{p_2}{1-p_2}\right)^y} \\
 &= \frac{\binom{m}{t-u} \binom{n}{u} \exp(u \ln \frac{p_2(1-p_1)}{p_1(1-p_2)})}{\sum_{y=y^*}^{t^*} \binom{m}{t-y} \binom{n}{y} \exp(y \ln \frac{p_2(1-p_1)}{p_1(1-p_2)})} \\
 &= \frac{\binom{m}{t-u} \binom{n}{u} \exp(u\theta)}{\sum_{y=y^*}^{t^*} \binom{m}{t-y} \binom{n}{y} \exp(y\theta)} \\
 &= c_t(\theta) \cdot \binom{m}{t-u} \binom{n}{u} e^{\theta u}
 \end{aligned}$$

是指数分布族，其中 $t^* = \min(n, t)$, $y^* = \max(0, t - m)$. 由定理 3.15 双边假设检验问题水平为 α 的 UMPUT 为

$$\phi(u, t) = \begin{cases} 1, & u < c_1(t) \text{ or } u > c_2(t) \\ r_i(t), & u = c_i(t), i = 1, 2 \\ 0, & c_1(t) < u < c_2(t) \end{cases}$$

其中 $c_i(t), r_i(t)$ 由下面两式确定

$$\begin{aligned}
 \mathbb{E}_{\theta_0}[\phi(U, T)|T = t] &= \alpha \\
 \mathbb{E}_{\theta_0}[U\phi(U, T)|T = t] &= \alpha \mathbb{E}_{\theta_0}[U|T = t]
 \end{aligned}$$

又当 H_0 成立时， $p(u|t; \theta) = \frac{\binom{m}{t-u} \binom{n}{u}}{\binom{m+n}{t}}$ 是超几何分布，则 $\mathbb{E}_{\theta_0}[U|T = t] = t \cdot \frac{n}{m+n}$.

PROBLEM A.3.5. 假设 (X_1, X_2, \dots, X_n) 是来自于具有未知形状参数 θ 和未知尺度参数 γ 的伽马分布的随机样本。已知常数 $\theta_0 > 0$ 和 $\gamma_0 > 0$ 。

- (1) 对于假设检验 $H_0 : \theta \leq \theta_0$ 对比 $H_1 : \theta > \theta_0$ 和 $H_0 : \theta = \theta_0$ 对比 $H_1 : \theta \neq \theta_0$ ，证明存在 UMPU 检验，其拒绝域基于 $V = \prod_{i=1}^n \left(\frac{X_i}{\bar{X}}\right)$ ，其中 \bar{X} 是样本均值。
- (2) 对于假设检验 $H_0 : \gamma \leq \gamma_0$ 对比 $H_1 : \gamma > \gamma_0$ ，证明 UMPU 检验在 $\sum_{i=1}^n X_i > C(\prod_{i=1}^n X_i)$ 时拒绝 H_0 ，其中 C 是某个函数。

SOLUTION.

1. 设 $Y = \log(\prod_{i=1}^n X_i)$ 和 $U = n\bar{X}$ 。随机样本 (X_1, \dots, X_n) 的联合密度可以写成

$$f(x_1, \dots, x_n; \theta, \gamma) = \left[\frac{1}{\Gamma(\theta)\gamma^\theta} \right]^n e^{\theta Y - U/\gamma - Y},$$

它属于指数族分布。根据 Shao (2003) 中的定理 6.4, UMPU 检验是 Y 和 U 的函数。根据 Basu 定理, $V_1 = Y - n \log\left(\frac{U}{n}\right) = \log\left(\prod_{i=1}^n \frac{X_i}{\bar{X}}\right)$ 满足 Shao (2003) 中引理 6.7 的条件。因此, UMPU 检验的拒绝域可以通过使用 V_1 来确定。由于 $V = e^{V_1}$, UMPU 检验的拒绝域也可以通过使用 V 来确定。

2. 设 $U = \log(\sum_{i=1}^n X_i)$ 和 $Y = n\bar{X}$ 。随机样本 (X_1, \dots, X_n) 的联合密度可以写成

$$f(x_1, \dots, x_n; \theta, \gamma) = \left[\frac{1}{\Gamma(\theta)\gamma^\theta} \right]^n e^{-Y/\gamma + (\theta-1)U},$$

根据 Shao (2003) 中的定理 6.4, 对于假设检验 $H_0 : \gamma \leq \gamma_0$ 对比 $H_1 : \gamma > \gamma_0$, UMP 检验为

$$T^* = \begin{cases} 1 & \text{当 } Y > C_1(U), \\ 0 & \text{当 } Y \leq C_1(U), \end{cases}$$

其中 C_1 是一个函数, 满足 $E(T^*|U) = \alpha$ 当 $\gamma = \gamma_0$ 。结果通过设定 $C(x) = C_1(\log x)$ 得到。

Theorem A.3.10 (定理 3.16: 正态总体参数的检验问题). 设样本服从多参数指数型分布, 密度函数如下

$$p(x; \theta, \mathbf{r}) = c(\theta, \mathbf{r}) \cdot \exp \left\{ \theta \cdot u(x) + \sum_{i=1}^k r_i \cdot t_i(x) \right\} \cdot h(x) \quad (\text{A.5})$$

其中 θ 为被检验的参数, $\mathbf{r} = (r_1, \dots, r_k)$ 为多余参数 (不需要被检验)。若参数空间 Ω 有内点, 则 $(U(X), T(X)) = (U(X), T_1(X), \dots, T_k(X))$ 是参数 (θ, \mathbf{r}) 的充分完备统计量, (1) 如果存在一个统计量 $V = V(U, T)$, 在给定 $T = t$ 后, V 是 U 的单调增函数, 且当 $\theta = \theta_0$ 时 V 与 T 相互独立。则 $H_0 : \theta \leq \theta_0$ v.s. $H_1 : \theta > \theta_0$ 的单边假设检验问题的水平 α 的 UMPUT 为

$$\phi(v) = \begin{cases} 1, & v > c \\ r, & v = c \\ 0, & v < c \end{cases}$$

其中 $r \in [0, 1]$ 和 c 不依赖于 t 且由下式确定

$$\mathbb{E}_{\theta_0}[\phi(V)] = \alpha.$$

如果考虑反向检验 $H_0 : \theta \geq \theta_0$ v.s. $H_1 : \theta < \theta_0$, 则上式检验函数中的不等式要变号。

(2) 如果存在一个统计量 $V = V(U, T)$, 在给定 $T = t$ 后 V 是 U 的单调函数, 且当

$\theta = \theta_1$ 或 θ_2 时, V 与 T 相互独立。则 $H_0 : \theta_1 \leq \theta \leq \theta_2$ v.s. $H_1 : \theta < \theta_1$ or $\theta > \theta_2$ 的双边假设检验问题的水平 α 的 UMPUT 为

$$\phi(v) = \begin{cases} 1, & v < c_1 \text{ or } v > c_2 \\ r, & v = c_i, i = 1, 2 \\ 0, & c_1 < v < c_2 \end{cases}$$

其中 $r_i \in [0, 1]$ 和 c_i 不依赖于 t 且由下式确定

$$\mathbb{E}_{\theta_1}[\phi(V)] = \mathbb{E}_{\theta_2}[\phi(V)] = \alpha.$$

同样反向检验要交换检验函数中取值 1 和 0 的条件。

(3) 如果存在一个统计量 $V = a(T) \cdot U + b(T)$ ($a(T) \neq 0$), 即给定 $T = t$ 后 V 是 U 的线性函数, 且当 $\theta = \theta_0$ 时 V 与 T 相互独立。则 $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta \neq \theta_0$ 的双边假设检验问题的水平 α 的 UMPUT 为

$$\phi(v) = \begin{cases} 1, & v < c_1 \text{ or } v > c_2 \\ r, & v = c_i, i = 1, 2 \\ 0, & c_1 < v < c_2 \end{cases}$$

其中 $r_i \in [0, 1]$ 和 c_i 不依赖于 t 且由下式确定

$$\mathbb{E}_{\theta_0}[\phi(V)] = \alpha, \mathbb{E}_{\theta_0}[V \cdot \phi(V)] = \alpha \mathbb{E}_{\theta_0}[V].$$

同样反向检验要交换检验函数中取值 1 和 0 的条件。

PROBLEM A.3.6. [例 3.12] 设样本 $X = (X_1, \dots, X_m) \sim N(\mu_1, \sigma^2)$ 与样本 $Y = (Y_1, \dots, Y_n) \sim N(\mu_2, \sigma^2)$ 相互独立。由于这两个正态总体的方差相等, 于是比较两个正态总体的差异的检验问题就是比较这两个正态总体的均值的检验问题; $H_0 : \mu_1 = \mu_2$ v.s. $H_1 : \mu_1 \neq \mu_2$.

SOLUTION. 样本的联合密度为

$$\begin{aligned} p(x, y; \mu_1, \mu_2, \sigma^2) &= (2\pi\sigma^2)^{-\frac{m+n}{2}} \exp \left\{ -\frac{\sum_{i=1}^m (x_i - \mu_1)^2 + \sum_{i=1}^n (y_i - \mu_2)^2}{2\sigma^2} \right\} \\ &= (2\pi\sigma^2)^{-\frac{m+n}{2}} \exp \left\{ -\frac{m\mu_1^2 + n\mu_2^2}{2\sigma^2} \right\} \cdot \exp \{ \theta \cdot u + r_1 \cdot t_1 + r_2 \cdot t_2 \} \end{aligned}$$

其中

$$\begin{aligned} \theta &= \frac{mn(\mu_2 - \mu_1)}{(m+n)\sigma^2}, r_1 = \frac{m\mu_1 + n\mu_2}{(m+n)\sigma^2}, r_2 = -\frac{1}{2\sigma^2} \\ u &= \bar{y} - \bar{x}, t_1 = m\bar{x} + n\bar{y}, t_2 = \sum_{i=1}^m x_i^2 + \sum_{i=1}^n y_i^2 \end{aligned}$$

这时所考虑的检验问题变为 $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. 根据定理 A.3.10 的 (3), 首先寻找统计量 V 使固定 T_1, T_2 后 V 是 U 的线性函数。在 $\theta = 0$ 时 V 与 (T_1, T_2) 相互独立。为此, 令

$$V = \frac{U}{\sqrt{T_2 - \frac{T_1^2}{m+n}}},$$

则对固定的 T_1, T_2, V 是 U 的线性函数, 在 $\theta = 0$ 时 (此时 $\mu_1 = \mu_2 = \mu$, 且合样本 $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ 相互独立且同分布), 这时 σ^2, μ 的充分完备统计量为 T_1, T_2, V 的分布与参数 μ, σ^2 . 由 Basu 引理, V 与 (T_1, T_2) 互相独立, 且可以证明 $\theta = 0$ 时, V 的分布关于原点对称. 根据定理 A.3.10(3), 该检验问题的 UMPUT 的拒绝域为 $\{x, y : |v| \geq c\}$. 进一步令

$$\begin{aligned} W &= \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{V}{\sqrt{1 - \frac{mn}{m+n} V^2}} \\ &= \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{Y} - \bar{X}}{\sqrt{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2}} \end{aligned}$$

则 $|W|$ 是 $|V|$ 的单调增函数且在 $\theta = 0$ 时, $W \sim t(m+n-2)$, 所以该检验问题的 UMPUT 拒绝域为 $\{x, y : |w| \geq t_{1-\alpha/2}(m+n-2)\}$. 类似地对于单边检验问题 $H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$ 和 $H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$ 的 UMPUT 的拒绝域分别为 $\{x, y : w \leq t_{1-\alpha}(m+n-2)\}$, $\{x, y : w \geq t_{1-\alpha}(m+n-2)\}$

PROBLEM A.3.7. 设样本 X_1, \dots, X_m 和样本 Y_1, \dots, Y_n 相互独立, 分别来自于正态总体 $N(\mu_1 \sigma_1^2), N(\mu_2, \sigma_2^2)$.

(1) 试证明: 在 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时,

$$V = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2} \sim \beta\left(\frac{n-1}{2}, \frac{m-1}{2}\right)$$

(2) 试证明: 检验问题 $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$ 的 UMPUT 的拒绝域为 $\{x, y : v \leq c_1 \text{ or } v \geq c_2\}$, 其中 c_1, c_2 由下面两式确定

$$\begin{aligned} \int_{c_1}^{c_2} Be(x \mid \frac{n-1}{2}, \frac{m-1}{2}) dx &= 1 - \alpha, \\ \int_{c_1}^{c_2} \left(x - \frac{n-1}{m+n-2}\right) Be(x \mid \frac{n-1}{2}, \frac{m-1}{2}) dx &= 0. \end{aligned}$$

SOLUTION.

$$(1) \sigma_1^2 = \sigma_2^2 = \sigma^2$$

$$\frac{\sum_{i=1}^m (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(m-1)} = \Gamma\left(\frac{m-1}{2}, \frac{1}{2}\right)$$

$$\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{(n-1)} = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

下述 $X_1 \sim \Gamma(d_1, \lambda), X_2 \sim \Gamma(d_2, \lambda)$ 且 X_1, X_2 相互独立时

$$\begin{aligned} & \text{若 } Y_1 = X_1 + X_2 \text{ 独立 且 } Y_2 \sim \beta(d_1, d_2) \\ & Y_2 = \frac{X_1}{X_1 + X_2} \end{aligned}$$

$$\begin{aligned} p_{X_1}(x|d_1, \lambda) &= \frac{\lambda^{d_1}}{\Gamma(d_1)} x^{d_1-1} e^{-\lambda x} \quad p_{X_2}(x|d_2, \lambda) = \frac{\lambda^{d_2}}{\Gamma(d_2)} x^{d_2-1} e^{-\lambda x} \\ \begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = \frac{X_1}{X_1 + X_2} \end{cases} \Rightarrow \begin{cases} X_1 = Y_1 Y_2 \\ X_2 = Y_1 - Y_1 Y_2 \end{cases} \quad J = \begin{vmatrix} Y_2 & Y_1 \\ 1 - Y_2 & Y_1 \end{vmatrix} = -Y_1 \\ p_{Y_1, Y_2}(y_1, y_2) &= \frac{\lambda^{d_1}}{\Gamma(d_1)} y_1^{d_1-1} e^{-\lambda y_1} \cdot y_1 \cdot \frac{\lambda^{d_2}}{\Gamma(d_2)} y_2^{d_2-1} e^{-\lambda(y_1+y_2)} \\ &= \frac{\lambda^{d_1+d_2}}{\Gamma(d_1)\Gamma(d_2)} y_1^{d_1+d_2-1} e^{-\lambda y_1} \cdot y_2^{d_2-1} (1-y_2)^{d_2-1} \\ &= \Gamma(d_1+d_2) \cdot \beta(d_1, d_2) \end{aligned}$$

$$\text{故 } V = \frac{\sum(Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2 + \sum(Y_i - \bar{Y})^2} \sim \beta\left(\frac{n-1}{2}, \frac{m-1}{2}\right)$$

(2) 对于 $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$ 基于对称分布的假设检验

UMPUT 拒绝域中 C_1, C_2 由下式确认

$$\begin{aligned} \mathbb{E}_{\theta_0} \phi(V) &= \alpha \quad \text{其中 } \phi(V) = \begin{cases} 1 & V \leq C_1 \text{ or } V \geq C_2 \\ 0 & \text{others} \end{cases} \quad \text{为何无 } V \text{ (边缘?)} \\ \mathbb{E}_{\theta_0} (V \phi(V)) &= \alpha \mathbb{E}_{\theta_0} V \end{aligned}$$

$$\begin{aligned} \textcircled{1} \quad \mathbb{E}_{\theta_0} \phi(V) &= P(V \leq C_1 \text{ or } V \geq C_2) = \int_0^{C_1} \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx + \int_{C_2}^1 \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx = \alpha \\ \text{即 } \int_{C_1}^{C_2} \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx &= 1 - \alpha \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \mathbb{E}_{\theta_0} (V \phi(V)) &= \int_0^{C_1} x \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx + \int_{C_2}^1 x \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx \\ &= \int_0^{C_1} x \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx - \int_{C_2}^{C_1} x \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx \\ &= \frac{n-1}{m+n-2} - \int_{C_1}^{C_2} x \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx = \alpha \cdot \frac{n-1}{m+n-2} = \left(1 - \int_{C_1}^{C_2} x \text{Be}(x) dx\right) \frac{n-1}{m+n-2} \\ \Rightarrow \int_{C_1}^{C_2} (x - \frac{n-1}{m+n-2}) \text{Be}(x | \frac{n-1}{2}, \frac{m-1}{2}) dx &= 0 \quad (\text{Be}(x) 表示 pdf, 为与 \beta 分布区分) \end{aligned}$$

$$F = \frac{V}{1-V} \cdot \frac{m-1}{n-1}$$

$$\Rightarrow V \sim \beta\left(\frac{n-1}{2}, \frac{m-1}{2}\right)$$

$$\text{则 } W = \frac{V}{FV} \sim \mathcal{Z}\left(\frac{n-1}{2}, \frac{m-1}{2}\right)$$

$$F = \frac{m-1}{n-1} V \sim F(n, m)$$

$$EF = \frac{m}{m-2}$$

PROBLEM A.3.8. 假设 X_1, \dots, X_n 是来自于 $\Gamma(g, b)$ 分布的一组样本, $\Gamma(g, b)$ 分布密度函数为

$$\frac{1}{\Gamma(g)b^g} x^{g-1} e^{-x/b}, 0 < x, 0 < b, g.$$

给出下列情形的 UMP 检验的拒绝域:

(1) $H: b \leq b_0$ vs $b > b_0$, 给定 g ;

(2) $H: g \leq g_0$ vs $g > g_0$, 给定 b .

SOLUTION.

1. 当 g 给定时, $p(x; b)$ 时单参数概率密度族且似然比为

$$\lambda(x) = \left(\frac{b_0}{b_1}\right)^g \exp\left\{\left(\frac{1}{b_0} - \frac{1}{b_1}\right) \sum_{i=1}^n x_i\right\}, b_1 > b_0$$

因此 $p(x; b)$ 关于实值统计量 $T(X) = \sum_{i=1}^n X_i$ 具有非降 MLR。由定理 3.8 可知, 关于检验问题的 UMPT 形如

$$\phi(T(X)) = \begin{cases} 1, & T(x) \geq c \\ 0, & T(x) < c \end{cases}$$

且满足 $\mathbb{E}_{b_0} \phi(T(X)) = \alpha$. 又 $T(X) \sim \Gamma(ng, b)$, 则取 $c = \Gamma_{1-\alpha}(ng, b_0)$, 得到检验问题的 UMPT.

2. 当 b 给定时, $p(x; g)$ 是单参数概率密度族, 似然比为

$$\lambda(x) = \frac{\Gamma(g_0)}{\Gamma(g_1)} b^{g_0-g_1} (\prod_{i=1}^n x_i)^{g_0-g_1}, g_1 > g_0$$

因此概率密度函数关于统计量 $T(X) = \prod X_i$ 具有非降 MLR. 由定理 3.8 可知检验问题的 UMPT 形如

$$\phi(T(X)) = \begin{cases} 1, & T(x) \geq c \\ 0, & T(x) < c \end{cases}$$

且满足 $\mathbb{E}_{g_0} \phi(T(X)) = \alpha$.

由于 $P(\prod X_i > c) = P(\frac{1}{n} \sum \log(X_i) > d)$ 可以通过蒙特卡罗方法得到 d , 使得 $P(\frac{1}{n} \sum \log(X_i) \geq d) = \alpha$. 从而检验函数就可以确定。

PROBLEM A.3.9. 设 $(X_1, Y_1), \dots, (X_n, Y_n)$ 是来自二维正态分布族的样本。试构造一套判断规则:

当 $T \leq c$ 时,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix} \right],$$

当 $T > c$ 时,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left[\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix} \right],$$

且同时控制犯两类错误的概率。

PROBLEM A.3.10. 设 $X_i \sim N(\mu, r^2 \sigma_i^2)$, 其中 σ_i^2 为常数, μ 未知。考虑如下检验问题: $H_0 : r = 1$ v.s. $H_1 : r \neq 1$. 求此检验水平为 α 的 UMPUT.

似然比检验

N-P 基本引理告诉我们，在简单原假设对简单备择假设的检验问题中，最优势检验由似然比检验给出。事实上，似然比检验方法也可用在复合假设的检验问题中构造检验。

设 $X_1, \dots, X_n \sim p(x; \theta)$ 在简单检验 $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta = \theta_1$ 中

$$\lambda(x) = \frac{p(x; \theta_1)}{p(x; \theta_0)},$$

现在考虑复合检验 $H_0 : \theta \in \Theta_0$ v.s. $H_1 : \theta \in \Theta_1$,

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_1} p(x; \theta)}{\sup_{\theta \in \Theta_0} p(x; \theta)} = \frac{p(x; \hat{\theta}_1)}{p(x; \hat{\theta}_0)}.$$

其中 $\hat{\theta}_i$ 是 H_i 成立时 θ 的 MLE. $\lambda(x)$ 比较大时备择假设成立观察到样本点 x 的可能性比较大，因此拒绝原假设，故取拒绝域 $\{x : \lambda(x) \geq c\}$.

以上似然比检验的写法是对分布的检验比较常用，对于参数假设检验时也可以使用

$$\lambda(x) = \frac{\sup_{\theta \in \Theta} p(x; \theta)}{\sup_{\theta \in \Theta_0} p(x; \theta)} = \frac{p(x; \hat{\theta})}{p(x; \hat{\theta}_0)}.$$

PROBLEM A.3.11. [例 3.14] 设有样本 $X = (X_1, \dots, X_n)$. 考虑检验问题原假设和备择假设分别是 H_0 : 样本来自正态分布 $N(\mu, \sigma^2)$ v.s H_1 : 样本来自双参数指数分布 $Exp(\mu, \sigma)$. 二者的密度函数分别为

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

$$p(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{x - \mu}{\sigma}\right\}, x \geq \mu$$

原假设成立时 μ, σ 的 MLE 分别为

$$\hat{\mu}_0 = \bar{X}, \hat{\sigma}_0 = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2},$$

备择假设成立时 μ, σ 的 MLE 分别为

$$\hat{\mu}_1 = X_{(1)}, \hat{\sigma}_1 = \bar{X} - X_{(1)}.$$

所以该检验问题的似然比为

$$\lambda(x) = \frac{\prod p_1(x_i; \hat{\mu}_1, \hat{\sigma}_1)}{\prod p_1(x_i; \hat{\mu}_0, \hat{\sigma}_0)} = \left(\sqrt{2\pi e^{-1}}\right)^n D^n,$$

其中

$$D = \frac{\sqrt{n \sum (x_i - \bar{x})^2}}{\sum (x_i - x_{(1)})},$$

由于 $\lambda(x)$ 关于 D 严格增加，拒绝域取为 $\{x : D \geq c\}$. 由于

$$D = \frac{\sqrt{n \sum (x_i - \bar{x})^2}}{\sum (x_i - x_{(1)})} = \frac{\sqrt{n \sum (u_i - \bar{u})^2}}{\sum (u_i - u_{(1)})}$$

其中 $u_i = \frac{x_i - \mu}{\sigma}$, 所以不论原假设和备择假设哪个为真, D 的分布与 μ, σ 均无关。原假设为真时, 检验犯第一类错误的概率为

$$P\{D \geq c | u_i \sim N(0, 1)\}$$

利用随机模拟法对于不同的样本容量 n , 球的原假设为真时 D 的分位数值, 从而得到检验的临界值 c . 备择假设为真时, 检验犯第二类错误的概率为

$$P\{D \geq c | u_i \sim Exp(0, 1)\}$$

利用随机模拟法可以求得犯第二类错误的概率.

Theorem A.3.11 (似然比统计量的极限分布:简单原假设的检验问题). $X = (X_1, \dots, X_n) \sim \{p(x; \theta) : \theta \in \Theta\}$, 参数空间 $\Theta \in \mathbb{R}^k$ 且含有内点, θ_0 是 Θ 中一个内点, 假设该密度函数族满足下列四个条件:

- (1) $\int_x \frac{\partial p(x; \theta)}{\partial \theta_i} d\mu(x) = \int_x \frac{\partial^2 p(x; \theta)}{\partial \theta_i \partial \theta_j} d\mu(x) = 0, i, j = 1, \dots, k.$
- (2) $I(\theta) > (I_{ij}(\theta))_{k \times k} > 0, \forall \theta \in \Theta$
- (3)
- (4)

除上述四个条件外, 再假设 θ 的 MLE 为 $\hat{\theta}$, 考虑如下双边检验问题

$$H_0 : \theta = \theta_0 \text{ v.s. } H_1 : \theta \neq \theta_0,$$

该检验问题的似然比统计量为

$$\lambda(X) = \frac{\prod p(X_i; \hat{\theta})}{\prod p(x_i; \theta_0)}$$

那么在原假设 H_0 成立时, $2 \ln \lambda(X)$ 随 n 增大依分布收敛于 $\chi^2(k)$.

PROBLEM A.3.12. [习题 3.16] 设样本 $X = (X_1, \dots, X_n) \sim p(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left\{-\frac{|x-\mu|}{\sigma}\right\}, x \geq \mu$

- (1) 试求检验问题 $H_0 : \sigma = 1$ v.s. $H_1 : \sigma \neq 1$ 的 UMPUT
- (2) 试求检验问题 $H_0 : \mu = 0$ v.s. $H_1 : \mu \neq 0$ 的 UMPUT

SOLUTION. 利用似然比检验求 UMPUT, 样本联合密度函数为

$$p(\mathbf{x}; \mu, \sigma) = \frac{1}{\sigma^n} \exp\left\{-\frac{\sum x_i - n\mu}{\sigma}\right\} \mathbb{I}_{\{x_{(1)} \geq \mu\}}$$

(1) 计算当 H_0, H_1 分别成立时的 MLE

$$\begin{aligned} \hat{\mu}_0 &= X_{(1)}, \hat{\sigma}_0 = 1; \\ \hat{\mu}_1 &= X_{(1)}, \hat{\sigma}_1 = \bar{X} - X_{(1)}, \end{aligned}$$

因此似然比为

$$\lambda(x) = \frac{p(\mathbf{x}; \hat{\mu}_1, \hat{\sigma}_1)}{p(\mathbf{x}; \hat{\mu}_0, \hat{\sigma}_0)} = (\bar{X} - X_{(1)})^{-n} e^{-n} \exp\{n(\bar{X} - X_{(1)})\},$$

不妨记 $T = n(\bar{X} - X_{(1)})$, $LR = 2 \ln \lambda(X) = 2T - 2n \ln T + 2n \ln n - n \sim \chi^2(2)$ 即可进行似然比检验。

当 H_0 成立时, $T = n(\bar{X} - X_{(1)}) \sim \Gamma(n-1, 1)$, 于是 $V = 2T \sim \Gamma(n-1, \frac{1}{2}) = \chi^2(2n-2)$. 令 $g(T) = LR$, 可知 $g(T)$ 关于 T 是先递减后增加的凸函数, 因此假设检验的拒绝域为 $W = \{LR > c\} = \{T < c_1 \text{ or } T > c_2\} = \{V < 2c_1 \text{ or } V > 2c_2\}$, 此时 UMPUT 为

$$\phi(t) = \begin{cases} 1, & t \leq c_1 \text{ or } t \geq c_2 \\ 0, & c_1 < t < c_2 \end{cases}$$

实际可取 $c_1 = \frac{1}{2}\chi_{\alpha/2}^2(2n-2)$, $c_2 = \frac{1}{2}\chi_{1-\alpha/2}^2(2n-2)$.

(2) 计算当 H_0, H_1 分别成立时的 MLE

$$\begin{aligned} \hat{\mu}_0 &= 0, \hat{\sigma}_0 = \bar{X}; \\ \hat{\mu}_1 &= X_{(1)}, \hat{\sigma}_1 = \bar{X} - X_{(1)}, \end{aligned}$$

因此似然比为

$$\lambda(X) = \left(\frac{\bar{X}}{\bar{X} - X_{(1)}} \right)^n \mathbb{I}_{\{X_{(1)} \geq 0\}}$$

可以凑出 F 分布: 因为 $X_{(1)} \sim Exp(\frac{n}{\lambda}, \mu)$, $X_{(1)} - \mu \sim Exp(\frac{n}{\lambda})$, 于是 $\frac{2n}{\lambda}(X_{(1)} - \mu) \sim \Gamma(1, \frac{1}{2}) = \chi^2(2)$. 又因为 $n(\bar{X} - X_{(1)}) \sim \Gamma(n-1, \lambda)$, $\frac{2n}{\lambda}(\bar{X} - X_{(1)}) \sim \Gamma(n-1, \frac{1}{2}) = \chi^2(2n-2)$.

于是

$$\frac{\frac{2n}{\lambda}(X_{(1)} - \mu)/2}{\frac{2n}{\lambda}(\bar{X} - X_{(1)})/(2n-2)} \sim F(2, 2n-2).$$

PROBLEM A.3.13. 设 $X_1, \dots, X_n \sim f(x; \theta_1) = \frac{1}{\theta_1} e^{-\frac{x}{\theta_1}}$, $x > 0$, $Y_1, \dots, Y_n \sim f(y; \theta_2) = \frac{1}{\theta_2} e^{-\frac{y}{\theta_2}}$, $y > 0$. 两个样本相互独立, 对检验问题 $H_0 : \theta_1 = \theta_2$ v.s. $H_1 : \theta_1 \neq \theta_2$, 试给出广义似然比检验法, 并求检验统计量在 H_0 下的分布。

SOLUTION. (1) 在全参数空间下, 指数分布的 MLE 时样本均值

$$\hat{\theta}_1 = \bar{X}, \hat{\theta}_2 = \bar{Y},$$

全参数似然函数最大值

$$\begin{aligned} \sup_{\Omega} L(\theta_1, \theta_2) &= \prod_{i=1}^m \frac{1}{\bar{X}} e^{-X_i/\bar{X}} \cdot \prod_{j=1}^n \frac{1}{\bar{Y}} e^{-Y_j/\bar{Y}} \\ &= \bar{X}^{-m} e^{-m} \cdot \bar{Y}^{-n} e^{-n} \end{aligned}$$

(2) 在 H_0 空间 $\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta > 0\}$ 下, 总体合并, 联合样本容量为 $m+n$, MLE 为

$$\hat{\theta}_0 = \frac{m\bar{X} + n\bar{Y}}{m+n},$$

此时似然函数最大值为

$$\sup_{\Omega_0} L = \hat{\theta}_0^{-(m+n)} e^{-(m+n)}.$$

(3) 广义似然比统计量为

A.4 区间估计

基于连续随机变量构造置信区间

一般来说, 欲构造 θ 的置信水平为 α 的置信区间, 我们首先考虑 θ 的 MLE 或者充分统计量, 据此得到统计量 $T(X)$ 寻找枢轴量然后构造 θ 的置信区间, 或大样本时构造 θ 的近似置信区间。

在统计量 $T(X)$ 是随机变量时, 寻找枢轴量不是一件难事。设 $T(X)$ 的分布函数为 $G(t, \theta) = P_\theta(T(X) \leq t)$, 那么 $G(T(X), \theta)$ 服从区间 $[0, 1]$ 上的均匀分布。这是 $G(T(X), \theta)$ 可被取为枢轴量。

PROBLEM A.4.1. 设 $X = (X_1, \dots, X_n)$ 是来自总体密度 $p(x; \theta) = \frac{\theta}{x^2}, 0 < \theta \leq x < \infty$ 的总体的样本, 试求 θ 的置信水平为 $1 - \alpha$ 的置信区间。

SOLUTION. 对连续随机变量构造置信区间, 我们考虑寻找充分统计量或 MLE。最小的次序统计量 $X_{(1)}$ 是 θ 的 MLE 和充分统计量, 所以我们基于 $X_{(1)}$ 构造 θ 的区间估计。

$T = X_{(1)}$ 的密度函数和分布函数分别为

$$\begin{aligned} g(t, \theta) &= \frac{n\theta^n}{t^{n+1}}, & 0 < \theta \leq t < \infty \\ G(t, \theta) &= 1 - \frac{\theta^n}{t^n}, & 0 < \theta \leq t < \infty \end{aligned}$$

则枢轴量取为 $G(X_{(1)}, \theta) = 1 - \frac{\theta^n}{X_{(1)}^n} \sim U(0, 1)$. 选取两个常数 $c, d (0 < c < d < 1)$, 使得 $d - c = 1 - \alpha$, 则有

$$P_\theta \left\{ c \leq 1 - \frac{\theta^n}{X_{(1)}^n} \leq d \right\} = 1 - \alpha.$$

从而得到 θ 的置信区间 $[\sqrt[n]{1-d}X_{(1)}, \sqrt[n]{1-c}X_{(1)}]$. 接下来考虑如何选取 c, d 使得置信区间的平均长度最短。

令 $l = (1-c)^{\frac{1}{n}} - (1-d)^{\frac{1}{n}}$, 由约束条件 $1 - \alpha = d - c, d = 1 - \alpha + c, l = (1-c)^{\frac{1}{n}} - (\alpha - c)^{\frac{1}{n}}$, $l'(c) = -\frac{1}{n}(1-c)^{-\frac{1}{n}-1} + \frac{1}{n}(\alpha - c)^{-\frac{1}{n}-1}$.

考虑函数 $f(x) = \frac{1}{n}x^{\frac{1}{n}-1}, n > 1$ 时 $f(x)$ 显然是减函数, 又 $\alpha - c < 1 - c$, 则 $l(c)$ 是增函数, 则 $c = 0, l(c)$ 取最小值, 此时 $d = 1 - \alpha$. 则平均长度最短的置信区间为 $[\alpha^{\frac{1}{n}}X_{(1)}, X_{(1)}]$.

法 2. $F(X) = 1 - \frac{\theta}{X} \sim U(0, 1)$, 于是 $\frac{\theta}{X} \sim U(0, 1)$. 考察 $Y = -\ln \frac{X}{\theta}$ 的分布, 易得 $Y \sim EXP(1)$, $Y_{(1)} = -\ln \frac{\theta}{X_{(1)}} \sim EXP(n)$.

我们选取 c, d 使得 $P(c \leq Y_{(1)} \leq d) = \int_c^d n \exp(-ny) dy = 1 - \alpha$. 由于 $n \exp(-ny)$ 在 $y > 0$ 上单调递减, 要使区间最短取 $c = 0$ 从而 $d = -\frac{\ln \alpha}{n}$.

于是 $P(0 \leq Y_{(1)} \leq -\frac{\ln \alpha}{n}) = P(0 \leq -\ln(\frac{\theta}{X_{(1)}}) \leq -\frac{\ln \alpha}{n}) = P(\alpha^{\frac{1}{n}} X_{(1)} \leq \theta \leq x_{(1)}) = 1 - \alpha$.

变式: 求 $UMAU$

首先说明 $T = X_{(1)}$ 是充分统计量, 其密度函数为

$$g(t, \theta) = \frac{n\theta^n}{t^{n+1}}, 0 < \theta \leq t < \infty$$

对于 $\theta_1 \geq \theta_2$ 求其似然比函数

$$\lambda(t) = \frac{g(t, \theta_1)}{g(t, \theta_2)} = \frac{\theta_1^n}{\theta_2^n} I(t \geq \theta_1)$$

是单增函数, 所以对检验问题 $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, 其 $UMPUT$ 为

$$\phi(t) = \begin{cases} 1, & t \leq c_1 \text{ or } t \geq c_2 \\ 0, & \text{others} \end{cases}$$

且

$$g(\theta) = \mathbb{E}_\theta \phi(x) = \begin{cases} 1, & c_2 \leq \theta \\ \frac{\theta^n}{c_2^n}, & c_1 \leq \theta \leq c_2 \\ 1 + \frac{\theta^n}{c_2^n} - \frac{\theta^n}{c_1^n}, & \theta \leq c_1 \leq c_2 \end{cases}$$

从 $g(\theta)$ 的图像上看, 为了使得 $g(\theta)$ 在 θ_0 处最小, 必有 $c_1 = \theta_0$, 又因为 $g(\theta_0) = \alpha, c_2 = \frac{\theta_0}{\alpha^{1/n}}$. 则接受域为 $\theta_0 \leq X_{(1)} \leq \frac{\theta_0}{\alpha^{1/n}}$ 即 $\alpha^{1/n} X_{(1)} \leq \theta_0 \leq X_{(1)}$. 由 $UMAU$ 对应 $UMPUT$ 可知 $UMAU$ 为 $[\alpha^{1/n} X_{(1)}, X_{(1)}]$.

在离散时稍显复杂

PROBLEM A.4.2. 设样本 $X = (X_1, \dots, X_n) \sim b(n, p)$, 求 p 的置信区间。

SOLUTION. (1) 在大样本情形下给出一个近似置信区间

p 的 MLE 为 $\hat{p} = \bar{X}$, 由中心极限定理可知和, 在 $n \rightarrow \infty$ 时,

$$\sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \xrightarrow{L} N(0, 1)$$

所以在样本容量 n 充分大时

$$P \left\{ -\lambda \leq \sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \leq \lambda \right\} \approx 1 - \alpha,$$

其中 $\lambda = U_{1-\alpha/2}$, 从而反解出参数 p 的置信区间

在实际运用中常常把枢轴量分母中的 $\sqrt{p(1-p)}$ 中的 p 也用 $\hat{p} \xrightarrow{P} p$ 近似, 一般要求 $n > 30$.

(2) 因为 p 的 MLE 为 \bar{X} , 充分统计量为 $T = \sum X_i$, 所以我们基于 T 构造 p 的置信区间。 T 的分布函数为

$$G(t, p) = P_p(T \leq t) = \sum_{i=1}^{[t]} \binom{n}{t} p^i (1-p)^{n-i}$$

不难验证下列等式成立

$$\sum_{i=1}^k \binom{n}{t} p^i (1-p)^{n-i} = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \int_p^1 u^k (1-u)^{n-k-1} du, k = 0, 1, \dots, n-1$$

由此以及恒等式 $\sum_{i=1}^n \binom{n}{t} p^i (1-p)^{n-i} = 1$ 可知, T 的分布函数 $G(t, p)$ 是 p 的连续严格减函数。

由样本 X_1, \dots, X_n 算得的 $T = \sum X_i$ 必是一个正整数, 不妨令其为 k , 在 $k > 0$ 时, p 的水平为 $1 - \alpha$ 的置信下限是下列方程的解

$$\sum_{i=1}^{k-1} \binom{n}{t} p^i (1-p)^{n-i} = 1 - \alpha,$$

即

$$\frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k)} \int_0^p u^k (1-u)^{n-k-1} du = \alpha$$

$$Be(p|k, n-k+1) = \alpha,$$

其中 Be 是 p 为随机变量的 Beta 分布的分布函数, 又 $F = \frac{B}{1-B} \cdot \frac{n}{m} \sim F(2m, 2n)$, 所以可等价变换为

$$F\left(\frac{p}{1-p} \cdot \frac{n-k+1}{k} \middle| 2k, 2(n-k+1)\right) = \alpha,$$

由此求得 p 的置信水平为 $1 - \alpha$ 的置信下限为下列方程的解

$$\frac{p}{1-p} \cdot \frac{n-k+1}{k} = F_\alpha(2k, 2(n-k+1)) = \frac{1}{F_{1-\alpha}(2(n-k+1), 2k)},$$

于是

$$\hat{p}_L = \frac{k}{k + (n-k+1) \cdot F_{1-\alpha}(2(n-k+1), 2k)}.$$

同理可求得置信上限:

$$\sum_{i=1}^k \binom{n}{t} p^i (1-p)^{n-i} = \alpha, k < n$$

则...

$$\hat{p}_U = \frac{k}{k + (n-k+1) \cdot F_{1-\alpha}(2(n-k+1), 2k)},$$

$$\hat{p}_U = \frac{(k+1) \cdot F_{1-\alpha}(2(k+1), 2(n-k))}{(n-k) \cdot F_{1-\alpha}(2(k+1), 2(n-k)) + k+1}.$$

其中 $\alpha_1 + \alpha_2 = \alpha$, 实用时常取 $\alpha_1 = \alpha_2 = \alpha/2$.

$k = 0$ 时

$$\hat{p}_L = 0, \hat{p}_U = \frac{F_{1-\alpha}(2, 2n)}{nF_{1-\alpha}(2, 2n) + 1},$$

$k = 1$ 时

$$\hat{p}_L = \frac{n}{n + F_{1-\alpha}(2, 2n)}, \hat{p}_U = 1,$$

PROBLEM A.4.3. [习题 4.6] 设总体 $X \sim U\{1, 2, \dots, N\}$, 其中 N 为未知参数。设 $X = (X_1, \dots, X_n)$ 为来自 X 的样本, 试基于 $X_{(n)}$ 构造关于 N 的置信水平为 $1 - \alpha$ 的置信上限。

SOLUTION. $X = (X_1, \dots, X_n)$ 的联合分布为

$$P(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{N^n} \mathbb{I}_{\{x_{(n)} \leq N\}}.$$

则 $X_{(n)}$ 是 N 的 MLE 和充分统计量。

求 $X_{(n)}$ 的分布函数为

$$P(X_{(n)} \leq k) = \prod_{i=1}^n P(X_i \leq k) = \frac{k^n}{N^n}, k = 0, 1, \dots, N$$

则对于任意的 $y \geq 0$,

$$G(y; N) = P(X_{(n)} \leq y) = \frac{[y]^n}{N^n}$$

是关于 N 的严格递减函数。则由定理 4.3 可知, N 的置信水平为 $1 - \alpha$ 的置信上下限为

$$\hat{N}_L = \sup_N \{N : G(X_{(n)-}; N) \geq 1 - \alpha_1\}, \hat{N}_U = \inf_N \{N : G(X_{(n)}; N) \leq \alpha_2\}$$

其中 $\alpha_1 + \alpha_2 = \alpha$.

解得

$$\hat{N}_L = \left\lceil \frac{X_{(n)} - 1}{\sqrt[n]{1 - \alpha_1}} \right\rceil, \hat{N}_U = \left\lfloor \frac{X_{(n)}}{\sqrt[n]{\alpha_2}} \right\rfloor.$$

PROBLEM A.4.4. 设参数空间 Θ 是直线上含有内点的一个区间。证明: 如果 $[\hat{\theta}_L(x), \hat{\theta}_U(x)]$ 是参数 θ 的置信水平为 $1 - \alpha$ 的 UMAU 置信区间, 则对 θ 的任一置信水平为 $1 - \alpha$ 的无偏置信区间 $[\hat{\theta}_L^*(x), \hat{\theta}_U^*(x)]$ 都有

$$\mathbb{E}_\theta [\hat{\theta}_U(x) - \hat{\theta}_L(x)] \leq \mathbb{E}_\theta [\hat{\theta}_U^*(x) - \hat{\theta}_L^*(x)]$$

SOLUTION. 由定理条件可知

$$\begin{aligned}
 \mathbb{E}_\theta[\hat{\theta}_U(x) - \hat{\theta}_L(x)] &= \iint_{\hat{\theta}_L(x)}^{\hat{\theta}_U(x)} d\theta' dP_\theta(x) \\
 &= \iint_{\{x: \hat{\theta}_L(x) \leq \theta' \leq \hat{\theta}_U(x)\}} dP_\theta(x) d\theta' \\
 &= \int_{\theta' \neq \theta} P_\theta\{\hat{\theta}_L(x) \leq \theta' \leq \hat{\theta}_U(x)\} d\theta' \\
 &= \int_{\theta' \neq \theta} P_\theta\{\hat{\theta}_L^*(x) \leq \theta' \leq \hat{\theta}_U^*(x)\} d\theta' \\
 &\leq \int_{\theta' \neq \theta} P_\theta\{\hat{\theta}_L^*(x) \leq \theta' \leq \hat{\theta}_U^*(x)\} d\theta' \\
 &= \mathbb{E}_\theta[\hat{\theta}_U^*(x) - \hat{\theta}_L^*(x)]
 \end{aligned}$$

证毕. □

PROBLEM A.4.5. 人们普遍认为，长寿命灯泡的平均寿命不超过普通灯泡的两倍。在实验中，记录普通灯泡的寿命 X 和长寿命灯泡的寿命 Y ，分别为 $X = 1, Y = 5$ ，设灯泡寿命分别是均值为 λ 和 μ 的指数随机变量，构造：

- (1) $H_0 : \lambda > \frac{1}{2}\mu$, 显著性水平 $\alpha = 0.05$ 的假设检验;
- (2) $\frac{\lambda}{\mu}$ 的 95% 置信区间.

SOLUTION.

(1) 由 Γ 分布的性质可知

$$\begin{aligned}
 \Gamma(1, \lambda) &= EXP(\lambda), \\
 \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) &= \chi^2(n)
 \end{aligned}$$

则

$$\frac{2X}{\lambda} \sim \Gamma\left(1, \frac{1}{2}\right) \sim \chi^2(2), \quad \frac{2Y}{\mu} \sim \Gamma\left(1, \frac{1}{2}\right) \sim \chi^2(2)$$

由于 X, Y 相互独立与 F 分布的性质可知

$$\frac{\frac{2X}{\lambda}/2}{\frac{2Y}{\mu}/2} = \frac{\mu X}{\lambda Y} \sim F(2, 2)$$

从而

$$\frac{\lambda}{\mu} = \frac{X}{Y} \frac{1}{F(2, 2)}$$

于是原假设检验问题等价为 $H_0 : \frac{\lambda}{\mu} > \frac{1}{2}$ vs $H_1 : \frac{\lambda}{\mu} \leq \frac{1}{2}$. 检验函数记为

$$\phi\left(\frac{X}{Y}\right) = \begin{cases} 1, & \frac{X}{Y} \leq c \\ 0, & \frac{X}{Y} > c \end{cases}$$

其中 c 由 $\mathbb{E}_{\frac{\lambda}{\mu}=\frac{1}{2}}\phi\left(\frac{X}{Y}\right) = \alpha$ 确定。解得 $c = \frac{1}{2F_{1-\alpha}(2,2)}$.

(2) 由上问得选取枢轴量 $G(X, Y, \lambda, \mu) = \frac{\mu X}{\lambda Y} \sim F(2, 2)$, 对给定 $\alpha = 0.05$, 选取两个常数 $c, d(c < d)$ 使得

$$P(c \leq G \leq d) \geq 1 - \alpha,$$

不失一般性不妨选取 $c = F_{\frac{\alpha}{2}}(2, 2), d = F_{1-\frac{\alpha}{2}}(2, 2)$, 最终的置信区间就可以写为 $\left[\frac{1}{d} \frac{X}{Y}, \frac{1}{c} \frac{X}{Y}\right]$.

PROBLEM A.4.6. 设 $X_1, \dots, X_m \sim N(\mu_1, \sigma_1^2), Y_1, \dots, Y_n \sim N(\mu_2, \sigma_2^2)$ 是两个独立随机样本。

- (1) 试对 $\frac{\sigma_2^2}{\sigma_1^2}$ 构造一个水平为 $1 - \alpha$ 的置信区间;
- (2) 若 σ_1^2, σ_2^2 均已知, 试对 $\mu_2 - \mu_1$ 构造一个水平为 $1 - \alpha$ 的置信区间.

SOLUTION. (1) 记样本方差

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, S_2^2 = \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2.$$

由抽样分布理论

$$\frac{(m-1)S_1^2}{\sigma_1^2} := U_1 \sim \chi^2(m-1), \frac{(n-1)S_2^2}{\sigma_2^2} := U_2 \sim \chi^2(n-1),$$

且两者独立。下面构造枢轴量, 考虑

$$F = \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} = \frac{\frac{(n-1)S_2^2/\sigma_2^2}{n-1}}{\frac{(m-1)S_1^2/\sigma_1^2}{m-1}} = \frac{U_2/(n-1)}{U_1/(m-1)} \sim F(n-1, m-1)$$

于是我们有

$$\begin{aligned} P\left(F_{\alpha/2}(n-1, m-1) \leq \frac{\sigma_1^2 S_2^2}{\sigma_2^2 S_1^2} \leq F_{1-\alpha/2}(n-1, m-1)\right) &= 1 - \alpha \\ \Rightarrow P\left(\frac{S_2^2/S_1^2}{F_{1-\alpha/2}(n-1, m-1)} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{S_2^2/S_1^2}{F_{\alpha/2}(n-1, m-1)}\right) &= 1 - \alpha. \end{aligned}$$

则得到置信区间。同时注意 $F_{\alpha/2}(n-1, m-1) = 1/F_{1-\alpha/2}(m-1, n-1)$.

(2) $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n})$ 且独立。于是有

$$\begin{aligned}\bar{Y} - \bar{X} &\sim N\left(\mu_2 - \mu_1, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right) \\ Z &= \frac{(\bar{Y} - \bar{X}) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1) \\ \Rightarrow P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) &= 1 - \alpha \\ \Rightarrow P\left((\bar{Y} - \bar{X}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \leq \mu_2 - \mu_1 \leq (\bar{Y} - \bar{X}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) &= 1 - \alpha\end{aligned}$$

置信区间为

$$\left[(\bar{Y} - \bar{X}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, (\bar{Y} - \bar{X}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right]$$

A.5 统计决策与 Bayes 分析

PROBLEM A.5.1. 计算两个高斯分布之间的 Kullback-Leibler(KL) 散度, 即 $KL(\mathbb{P}_{f,\Sigma_1} \mathbb{P}_{g,\Sigma_2})$, 其中 $f, g \in \mathbb{R}^n$ 是均值向量, Σ_1, Σ_2 是 n 阶对称正定矩阵 (协方差矩阵)。我们考虑的是简化形式, 球形高斯分布分布, 其中 $\Sigma_1 = \Sigma_2 = \sigma^2 I_n$. 我们证明 $KL(\mathbb{P}_f, \mathbb{P}_g) = \|f - g\|^2 / (2\sigma^2)$.

SOLUTION. KL 散度定义为

$$KL(P, Q) = \mathbb{E}_P \left[\log \frac{p(x)}{q(x)} \right].$$

对于高斯分布 $\mathcal{N}(f, \sigma^2 I_n)$ 和 $\mathcal{N}(g, \sigma^2 I_n)$, 它们的概率密度函数分别为:

$$\begin{aligned}p(x) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|x - f\|^2\right) \\ q(x) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|x - g\|^2\right),\end{aligned}$$

又 KL 散度公式可以写作 $KL(P_f, P_g) = \int p(x) \log \frac{p(x)}{q(x)} dx$, 带入 $p(x), q(x)$ 表达式

$$\begin{aligned}KL(P_f, P_g) &= \int \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|x - f\|^2\right) \log \frac{\exp\left(-\frac{1}{2\sigma^2}\|x - f\|^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\|x - g\|^2\right)} dx \\ &= \frac{1}{2\sigma^2} \int \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\|x - f\|^2\right) (\|x - g\|^2 - \|x - f\|^2) dx \\ &= \mathbb{E}_P[\|x - g\|^2] - \mathbb{E}_P[\|x - f\|^2].\end{aligned}$$

我们知道 $\mathbb{E}_P[x] = f$, $\mathbb{E}_P[\|x - f\|^2] = n\sigma^2$. 计算 $\mathbb{E}_P[\|x - g\|^2]$ 时, 使用二次型展开:

$$\begin{aligned}\mathbb{E}_P[\|x - g\|^2] &= \mathbb{E}_P[\|(x - f) + (f - g)\|^2] \\ &= \mathbb{E}_P[\|x - f\|^2] + 2\mathbb{E}_P[(x - f)^\top(f - g)] + \|f - g\|^2\end{aligned}$$

由于 $\mathbb{E}_P[(x - f)] = 0$, 所以中间项为 0, 最终得到:

$$\mathbb{E}_P[\|x - g\|^2] = n\sigma^2 + \|f - g\|^2.$$

第一项是 $n\sigma^2$, 因为这是关于 $\mathcal{N}(f, \sigma^2 I_n)$ 的自变量。

将两部分代入 KL 散度公式, 得到:

$$KL(P_f, P_g) = \frac{1}{2\sigma^2} ((n\sigma^2 + \|f - g\|^2) - n\sigma^2) = \frac{\|f - g\|^2}{2\sigma^2}.$$

NOTE OF PROBLEM A.5.1. 本题来自于 Wiki 练习 3.6.1.

PROBLEM A.5.2.

1. 假设 Y_1, \dots, Y_n 独立同分布并来源于 $f(y; \theta)$, θ 是一个尺度参数。假设有一个 1-1 变换将 θ 映射为 $\phi = \phi(\theta)$, 样本中关于 ϕ 的 Fisher 信息量 $i(\phi)$ 和关于 θ 的 Fisher 信息量 $i(\theta)$ 之间的关系是什么?
2. 假设 X 是二项分布 $b(k, \theta)$ 的一个观测值, θ 的 Jeffrey 先验是什么?

SOLUTION.

1. 由 Fisher 信息量的定义可知

$$i(\theta) = E\left[\frac{\partial \ln f(y; \theta)}{\partial \theta}\right]^2$$

则经过参数变换后的 Fisher 信息量为

$$\begin{aligned}i(\phi) &= E\left[\frac{\partial \ln f(y; \theta)}{\partial \phi}\right]^2 \\ &= E\left[\frac{\partial \ln f(y; \theta)}{\partial \theta}\right]^2 / \left[\frac{\partial \phi}{\partial \theta}\right]^2 \\ &= E\left[\frac{\partial \ln f(y; \theta)}{\partial \theta}\right]^2 \times \frac{1}{\phi'^2} \\ &= i(\theta) \times \frac{1}{\phi'^2}\end{aligned}$$

因此我们得到了 $i(\phi)$ 和 $i(\theta)$ 之间的关系: $i(\phi) = i(\theta) \times \frac{1}{\phi'^2}$.

2. 首先

$$f(x|\theta) = \binom{k}{x} \theta^x (1-\theta)^{k-x}$$

然后我们对 θ 求 Fisher 信息量

$$\begin{aligned} i(\theta) &= E\left[\frac{\partial \ln f(y; \theta)}{\partial \theta}\right]^2 \\ &= -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}\right] \\ &= E\left[\frac{x}{\theta^2} + \frac{k-x}{(1-\theta)^2}\right] \\ &= \frac{k\theta}{\theta^2} + \frac{k-k\theta}{(1-\theta)^2} \\ &= \frac{k}{\theta(1-\theta)} \end{aligned}$$

由 Jeffrey prior 的定义知

$$p(\theta) \propto [\theta(1-\theta)]^{-1/2}$$

所以 θ 的 Jeffrey prior 为

$$p(\theta) = \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)} \theta^{-1/2} (1-\theta)^{-1/2}, \theta \in (0, 1)$$

即 $\theta \sim Beta(1/2, 1/2)$.

□

PROBLEM A.5.3. 设 X_1, \dots, X_n 是来自正态总体 $N(0, \sigma^2)$ 的一个样本。问 $S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是不是 σ^2 的容许估计？

SOLUTION. 对于估计使用平方损失函数下的风险

$$R(\sigma^2, S_{n-1}^2) = \mathbb{E}(S_{n-1}^2 - \sigma^2)^2 = \frac{2\sigma^4}{n-1}$$

我们考虑形如 $\delta(x) = cS_{n-1}^2$ 的估计，在同样的二次损失函数下计算其风险

$$\begin{aligned} R(\sigma^2, \delta(x)) &= \mathbb{E}(cS_{n-1}^2 - \sigma^2)^2 \\ &= c^2 \mathbb{E}(S_{n-1}^2 - \sigma^2)^2 + \sigma^4(1-c)^2 \\ &= \sigma^4 \left[\frac{2c^2}{n-1} + (1-c)^2 \right], \end{aligned}$$

令 $g(c) = \frac{2c^2}{n-1} + (1-c)^2$, 这是关于 c 的二次函数，且在 $c_0 = \frac{n-1}{n+1}$ 处达到最小，所以令

$$S_{n+1}^2 = \frac{1}{n+1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

则有

$$R(\sigma^2, S_{n-1}^2) > R(\sigma^2, S_{n+1}^2) = \frac{2\sigma^4}{n+1}.$$

这就证明了 S_{n-1}^2 是非容许的。实际上 S_{n+1}^2 也是非容许的，比如取 $\hat{\sigma}^2 = \frac{1}{n+2} \sum_{i=1}^n x_i^2$ ，它的风险更小。

PROBLEM A.5.4. 假设 X_1, \dots, X_n 是独立同分布的随机变量，给定 μ 的情况下，服从正态分布 $N(\mu, \sigma_0^2)$ ， σ_0^2 已知。同时假设 μ 的先验分布是正态分布，均值 ξ_0 和方差 ν_0 已知。

给定 μ ，令 X_{n+1} 是来自相同分布的单个未来观测值，独立于 X_1, \dots, X_n 。证明：给定 (X_1, \dots, X_n) ， X_{n+1} 服从正态分布，均值为

$$\left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1} \left\{ \frac{\bar{X}}{\sigma_0^2/n} + \frac{\xi_0}{\nu_0} \right\}$$

方差为

$$\sigma_0^2 + \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1}.$$

SOLUTION. 对于 μ 的先验分布， $\mu \sim N(\xi_0, \nu_0)$ ，

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\nu_0}} \exp \left\{ -\frac{(\mu - \xi_0)^2}{2\nu_0} \right\}$$

令 $x = (x_1, \dots, x_n)$ 表示 $X = (X_1, \dots, X_n)$ 的观测值。

$$\begin{aligned} f(x; \mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{(x_i - \mu)^2}{2\sigma_0^2} \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2} \right\} \end{aligned}$$

由 $\pi(\mu | x) \propto \pi(\mu)f(x; \mu)$ ：

$$\pi(\mu | x) \propto \exp \left\{ -\frac{(\mu - \xi_0)^2}{2\nu_0} - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2} \right\}$$

其中：

$$\begin{aligned} \frac{(\mu - \xi_0)^2}{\nu_0} + \frac{\sum_{i=0}^n (x_i - \mu)^2}{\sigma_0^2} &= \frac{\mu^2 - 2\mu\xi_0^2 + \xi_0^2}{\nu_0} + \frac{n\mu^2 - 2n\bar{x}\mu + \sum_{i=1}^n x_i^2}{\sigma_0^2} \\ &= \mu^2 \left(\frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right) - 2\mu \left(\frac{\bar{X}}{\sigma_0^2/n} + \frac{\xi_0}{\nu_0} \right) + C_1 \\ &= \frac{1}{\nu_1} (\mu - \xi_1)^2 + C_2 \end{aligned}$$

其中 $\nu_1 = \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1}$ ， $\xi_1 = \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1} \left\{ \frac{\bar{x}}{\sigma_0^2/n} + \frac{\xi_0}{\nu_0} \right\}$ ， C_1 和 C_2 与 μ 无关。

因此

$$\pi(\mu | x) \propto \exp \left\{ -\frac{1}{2\nu_1} (\mu - \xi_1)^2 \right\}$$

所以给定 $X = (X_1, \dots, X_n)$, 得到 μ 的后验分布为正态分布, 其均值 $\xi_1 = \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1} \left\{ \frac{\bar{X}}{\sigma_0^2/n} + \frac{\xi_0}{\nu_0} \right\}$,

$$\text{方差 } \nu_1 = \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1}$$

对于 X_{n+1} 的期望, 由重期望公式:

$$\begin{aligned} \mathbb{E}(X_{n+1}) &= \mathbb{E}(\mathbb{E}(X_{n+1} | \mu)) \\ &= \mathbb{E}(\mu) \\ &= \xi_1 \\ &= \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1} \left\{ \frac{\bar{X}}{\sigma_0^2/n} + \frac{\xi_0}{\nu_0} \right\} \end{aligned}$$

对于 X_{n+1} 的方差, 由条件方差公式:

$$\begin{aligned} Var(X_{n+1}) &= Var(\mathbb{E}(X_{n+1} | \mu)) + \mathbb{E}(Var(X_{n+1} | \mu)) \\ &= Var(\mu) + \mathbb{E}(\sigma_0^2) \\ &= \nu_1 + \sigma_0^2 \\ &= \sigma_0^2 + \left\{ \frac{1}{\sigma_0^2/n} + \frac{1}{\nu_0} \right\}^{-1} \end{aligned}$$

PROBLEM A.5.5. 非参数贝叶斯中 Dirichlet 过程先验下, 如何计算后验分布。

SOLUTION. 它们使用的是 *Dirichlet* 过程先验 (*Dirichlet Process Priors, DPPs*)。作为关键结果的式子可以表示成对于集合 A ,

$$(Y_{n+1} \in A | y^n, \alpha, H) = \frac{\alpha H(A)}{\alpha + n} + \frac{n}{\alpha + n} \hat{F}_n(A)$$

其中 Y_{n+1} 是未来值, $y^n = (y_1, \dots, y_n)^T$ 是样本数据, H 是基准测度, α 是集中参数, \hat{F}_n 是经验分布函数。这样的的预测分布有两个明显的弊端: (1) 它有一个非常小的项; (2) 如果 $S = \mathbb{R} \setminus \{y_1, \dots, y_n\}$, 那么 $P(Y_{n+1} \in S | y^n) = \alpha H/(n + \alpha)$, 所以当 H 是有限测度时, 只要选择足够大的 n 就可以使 $P(Y_{n+1} \in S | y^n) < \epsilon$, ϵ 为任意给定, 预测分布收敛到 0。预测分布也有一个不太明显的限制, 就是由于 *Dirichlet* 过程先验把大量概率单位分给了离散分布, 预测分布也把大量概率单位分给了离散分布。

Dirichlet 过程先验的这种用法是 *Dirichlet* 分布作为多项式分布共轭先验的一般化。如果假定 Y 的取值为 $1, \dots, J$, 由多项式分布和 *Dirichlet* 先验得到的预测分布为

$$P(Y_{n+1} = j | y^n, \alpha) = \frac{\alpha_j + n_j}{\sum_{j=1}^J \alpha_j + n} = \frac{\alpha_j}{\sum_{j=1}^J \alpha_j + n} + \frac{n}{\sum_{j=1}^J \alpha_j + n} \frac{n_j}{n}$$

其中 n_j 是 $N_j = \#\{k|Y_i = j\}$ 的结果, $\alpha = (\alpha_1, \dots, \alpha_J)$ 是给定 $\theta = (\theta_1, \dots, \theta_J), \theta_j \geq 0$ 和 $\sum_{j=1}^J \theta_j = 1$ 的 $Dirichlet(\alpha_1, \dots, \alpha_J)$ 密度中的超参数。多项式分布为

$$P(n_1, \dots, n_J | \theta) = \binom{n}{n_1, \dots, n_J} \prod_{j=1}^J \theta_j^{n_j}$$

显然, 如果 H 是一个概率且 $\alpha_j = \alpha H(A_j)$, 其中 A_j 是 n 个区间 $y_j \in A_j$ 的不相关穷举集, 那么上述两个预测分布在形式上相同。

PROBLEM A.5.6. 分层 Poisson 模型 $y_i | \theta_i \stackrel{i.i.d.}{\sim} P(\theta_i)$, $\theta_i | \beta \stackrel{i.i.d.}{\sim} EXP(\beta)$, $\beta \sim U(0, +\infty)$.

- (1) 求 $\mathbb{E}(\theta_j | y, \beta)$;
- (2) 令 $\gamma = \frac{1}{1+\beta}$, 证明其后验分布 $\gamma | y \sim \beta(n\bar{y} - 1, n + 1)$;
- (3) 求 $\mathbb{E}(\theta_j | y)$;
- (4) 在频率学派中 $\hat{\theta}_j = y_j$, 在什么情况下 $\mathbb{E}(\theta_j | y) > \hat{\theta}_j$.

SOLUTION.

$$Y_j | \theta_j \sim P(\theta_j)$$

$$L(\theta_1, \dots, \theta_n | \vec{y}) = \prod_{j=1}^n \frac{\theta_j^{y_j} e^{-\theta_j}}{y_j!}$$

由于 $\theta_j | \beta \sim \text{Exp}(\beta)$ $f(\theta_j | \beta) = \beta e^{-\beta \theta_j}$

$\theta_1, \dots, \theta_n$ 相互独立 $f(\theta_1, \dots, \theta_n | \beta) = \prod_{j=1}^n f(\theta_j | \beta)$

后验分布 \propto 先验 \times 似然

$$f(\theta_1, \dots, \theta_n | \vec{y}) \propto f(\theta_1, \dots, \theta_n | \beta) L(\theta_1, \dots, \theta_n | \vec{y})$$

$$\propto \prod_{j=1}^n \frac{\theta_j^{y_j} e^{-\theta_j}}{y_j!} \prod_{j=1}^n \beta e^{-\beta \theta_j}$$

$\theta_j | y_j$ 后验分布 $\sim T(y_j + 1, 1 + \beta)$

$$\text{且 } E(\theta_j | y_j, \beta) = \frac{y_j + 1}{1 + \beta}$$

由于 β 的先验分布是均匀分布

$$\text{后验分布 } f(\beta | \vec{y}) \propto \beta^n e^{-\beta \bar{y}} \sim T(n+1, n\bar{y})$$

$$\text{且 } 1 = \frac{1}{1+\beta}, \text{ 且 } f(1 | \vec{y}) = \beta(n\bar{y} - 1, n+1)$$

$$\text{后验 } f(\theta_j | y_j) = \int_{\theta_j} f(\theta_j | \vec{y}) d\theta_j$$

$$= \iint_{\theta_j} f(\theta_j | \vec{y}) f(\theta_j | \beta) d\theta_j d\beta$$

$$= \int_{y_j}^1 \text{且}(\theta_j | \vec{y}, 1) f(1 | \vec{y}) d\theta_j$$

$$= \int_{y_j}^1 \frac{T(y_j + 1)}{\Gamma(n+1)\Gamma(n\bar{y})} (1-\beta)^n \beta^{n\bar{y}-2} d\beta$$

$$= (y_j + 1) \frac{\Gamma(n\bar{y} - n)}{\Gamma(n\bar{y} - 1)\Gamma(n)} \int_{y_j}^1 (1-\beta)^n \beta^{n\bar{y}-1} d\beta$$

$$= (y_j + 1) \frac{\Gamma(n\bar{y} - n)}{\Gamma(n\bar{y} - 1)\Gamma(n)} \cdot \frac{\Gamma(n\bar{y})}{\Gamma(n+1)} (1-\beta)^{n\bar{y}}$$

$$= (y_j + 1) \frac{n\bar{y}}{n\bar{y} + n}$$

且 $E(\theta_j | y_j) > \hat{\theta}_j = y_j$

$$\text{且 } y_j < \frac{n\bar{y} - 1}{n + 1}$$

$$\text{注 II } f(\theta_j | \vec{y}, \beta) = f(\theta_j | y_j, \beta)$$

$$= \frac{f(\theta_j, y_j, \beta)}{f(y_j, \beta)}$$

$$= \frac{f(\beta) f(\theta_j | \beta) f(y_j | \theta_j, \beta)}{f(\beta) f(y_j | \beta)}$$

$$= \frac{f(\theta_j | \beta) f(y_j | \theta_j, \beta)}{f(y_j | \beta)}$$

$$\text{且 } f(y_j | \beta) = \int_0^\infty f(y_j, \theta_j | \beta) d\theta_j$$

$$= \int_0^\infty f(\theta_j | \beta) f(y_j | \theta_j, \beta) d\theta_j$$

$$= \int_0^\infty \beta e^{-\beta \theta_j} \frac{\theta_j^{y_j}}{y_j!} e^{-\theta_j} d\theta_j$$

$$= \frac{\beta}{y_j!} \int_0^\infty e^{-\theta_j(1+\beta)} \theta_j^{y_j} d\theta_j$$

$$= \frac{\beta}{y_j!} \frac{T(y_j + 1)}{(1+\beta)^{y_j + 1}} = \frac{\beta}{(1+\beta)^{y_j + 1}}$$

从 $f(\theta_j | y_j, \beta) \sim T(y_j + 1, 1 + \beta)$

$$(f(x, y) = \frac{T(x+y)}{\Gamma(x+y)} = \int_0^\infty t^{x+y-1} (1-t)^{x+y-2} dt)$$

$$(T(x+y) = xT(x))$$

$$\text{且 } 1 = \frac{1}{1+\beta}$$

$$f(\vec{y} | \beta) = \prod_j f(y_j | \beta) = \frac{\beta^n}{(1+\beta)^{n\bar{y} + n}}$$

$$\text{且 } \beta = \frac{1}{1-\bar{y}} = \frac{1}{1-x}$$

$$f(\beta | y) = \frac{f(\beta, y)}{f(y)} = \frac{f(\beta) f(y | \beta)}{f(y)}$$

$$\propto (1-\bar{y})^n / \bar{y}^{n\bar{y} + 1} \sim \beta^{(n\bar{y} - 1, n+1)}$$

PROBLEM A.5.7. 假设 Y_1, \dots, Y_n 相互独立分别服从均值为 $\psi\pi_j$ 的 Poisson 分布，其中 $0 < \pi_j < 1$, $\sum \pi_j = 1$. 请找到一个基于 Y_1, \dots, Y_n 的边际似然并证明相比于使用全似然，使用边际似然并没有损失 ψ 的信息。

PROBLEM A.5.8. 随机变量 $X \sim p(x|\theta) = \theta e^{-\theta x}$, 对 X 作 n 次观察, 得到 x_1, \dots, x_n . 如今规定设备可靠性 $g(t_0, \theta) = e^{-t_0\theta}$. 其中 t_0 是已知正实数。若设 θ 的先验分布为 Gamma 分布 $\Gamma(\alpha, \frac{1}{\tau})$, 试在平方损失函数下求 $g(t_0, \theta)$ 的 Bayes 估计。

SOLUTION. 计算 θ 的后验分布，利用分布的核：

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto p(\mathbf{x}|\theta)\pi(\theta) \propto \pi(\theta) \prod_{i=1}^n p(x_i|\theta) \\ &\propto \theta^{\alpha-1} e^{\theta/\tau} \theta^n e^{-\theta \sum x_i} \\ &\propto \theta^{n+\alpha-1} e^{-\theta(1/\tau + \sum x_i)} \sim \Gamma(a, b),\end{aligned}$$

其中 $a = n + \alpha, b = \frac{1}{\tau} + \sum_{i=1}^n x_i$.

现求 $g(t_0, \theta) = e^{-t_0\theta}$ 在平方损失下的 Bayes 估计，即后验均值 $\hat{g}_{Bayes} = \mathbb{E}[e^{-t_0\theta}|\mathbf{x}]$.

$$\begin{aligned}\mathbb{E}[e^{-t_0\theta}|\mathbf{x}] &= \int_0^\infty e^{t_0\theta} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{b\theta} d\theta \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty \theta^{a-1} e^{-(b+t_0)\theta} d\theta,\end{aligned}$$

积分项是 $\frac{1}{\Gamma(a)} = \frac{\Gamma(a)}{(b+t_0)^a}$, 因此

$$\mathbb{E}[e^{-t_0\theta}|\mathbf{x}] = \frac{b^a}{\Gamma(a)} \cdot \frac{\Gamma(a)}{(b+t_0)^a} = \left(\frac{b}{b+t_0}\right)^a,$$

其中 $a = n + \alpha, b = \frac{1}{\tau} + \sum_{i=1}^n x_i$.

PROBLEM A.5.9. [习题 5.34] 设随机变量 $X \sim$ 几何分布，即 $P(X = k) = \theta(1 - \theta)^k, k = 0, 1, 2, \dots$ 其中参数 θ 的先验分布为均匀分布 $U(0, 1)$. 若对 X 作三次观察，观察值为 2, 3, 5，在平方损失函数下求 θ 的 Bayes 估计。

先求后验分布。利用分布的核

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto p(\mathbf{x}|\theta)\pi(\theta) \\ &\propto \pi(\theta) \prod_{i=1}^n p(x_i|\theta) \\ &\propto \mathbb{I}_{\{0 < \theta < 1\}} \prod_{i=1}^n [\theta(1 - \theta)^{x_i}] \\ &\propto \left[\theta^n (1 - \theta)^{\sum_{i=1}^n x_i}\right] \mathbb{I}_{\{0 < \theta < 1\}} \\ &\propto \left[\theta^{n+1-1} (1 - \theta)^{(\sum_{i=1}^n x_i+1)-1}\right] \mathbb{I}_{\{0 < \theta < 1\}} \\ &\sim \beta\left(n+1, \sum_{i=1}^n x_i + 1\right)\end{aligned}$$

则后验分布是 Beta 分布，取 $n = 3, \mathbf{x} = (2, 3, 5)$, 后验分布为 $\beta(4, 11)$, 则其 Bayes 估计为 $\mathbb{E}[\theta] = \frac{4}{11+4} = \frac{4}{15}$.

PROBLEM A.5.10. [习题 5.49] 设 $X \sim b(n, p), 0 < p < 1$. 证明: $\delta = x/n$ 在损失函数 $L(p, \delta) = (\delta - p)^2/p(1 - p)$ 下是 p 的最小最大估计。

SOLUTION. 取 p 的先验分布为均匀分布 $U(0, 1)$ 则由贝叶斯公式

$$\begin{aligned}\pi(p|x) &\propto p(x|p)\pi(p) \\ &\propto \binom{n}{x} p^x (1-p)^{n-x} \mathbb{I}_{\{0 \leq p \leq 1\}} \sim \beta(x+1, n+1-x),\end{aligned}$$

最小最大估计即 *Bayes* 估计，因此之需求 *Bayes* 估计。

在损失函数 $L(p, \delta) = (\delta - p)^2/p(1-p)$ 下对于决策 δ 的后验风险函数为

$$\begin{aligned}R_\pi(\delta|x) &= \mathbb{E}_{p|x}[L(p, \delta)] \\ &= \int_0^1 L(p, \delta) Be(p|x+1, n+1-x) dp \\ &= A \int_0^1 (\delta - p)^2 Be(p|x, n-x) dp.\end{aligned}$$

A 对于极值没有影响，后半部分就相当于后验分布 Beta 分布在平方损失下的后验分布, $\delta^\pi = \mathbb{E}p = \frac{x}{n}$.

B 风险管理

B.1 风险度量

Definition B.1.1 (风险价值). 给定置信水平 $\alpha \in (0, 1)$. 在置信水平 α 下损失 L 的投资组合的 VaR 由最小数字 l 给出，使得损失 L 超过 l 的概率不大于 $1 - \alpha$. 定义为

$$VaR_\alpha = VaR_\alpha(L) = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.$$

Example B.1.2 (正态和 t 损失分布的 VaR). (1) 假设损失分布 $F_L \sim N(\mu, \sigma^2)$, 置信水平 $\alpha \in (0, 1)$, 于是有

$$VaR_\alpha = \mu + \sigma \Phi^{-1}(\alpha),$$

证明过程比较简单，我们只需要证明 $F_L(VaR) = \alpha$.

$$P(L \leq VaR_\alpha) = P\left(\frac{L - \mu}{\sigma} \leq \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)).$$

(2) 假设损失 L 满足 $(L - \mu)/\sigma$ 服从一个自由度为 ν 的标准 t 分布，即 $L \sim t(\nu, \mu, \sigma^2)$, $\nu > 2$ 时损失分布的一阶矩二阶矩为 $\mathbb{E}[L] = \mu$, $\text{Var}(L) = \frac{\nu\sigma^2}{\nu-2}$, 因此注意 σ 不是分布函数的标准差。我们得到

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha).$$

Definition B.1.3. 偏矩时对分布左部或右部风险的度量，我们考虑损失分布右尾的内在风险和被用于度量风险的上偏矩。给定指数 $k \geq 0$ 和参考点 q , 上偏矩 $UPM(k, q)$ 定义如下

$$UPM(k, q) = \int_q^\infty (l - q)^k dF_L(l) \in [0, \infty].$$

k 取值越高我们的风险度量越保守，这是因为给参考点 q 的偏差更大的权重。

Definition B.1.4 (预期损失). 对于损失 L 有 $\mathbb{E}[|L|] < \infty$, 在置信水平 $\alpha \in (0, 1)$ 下分布函数 F_L 的预期损失定义为

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 q_u(F_L) du,$$

其中 $q_u(F_L) = F_L^\leftarrow(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$.

根据定义, ES 与 VaR 相关

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(L) du \geq \text{VaR}_\alpha,$$

对于有连续密度函数且可积的损失 L , 可以得到一个更直观的表达式, 根据该表达式, 预期损失可以解释为超过 VaR 时所发生的期望损失。

$$ES_\alpha = \frac{\mathbb{E}[L \mathbb{I}_{\{L \geq q_\alpha(L)\}}]}{1-\alpha} = \mathbb{E}[L | L \geq \text{VaR}_\alpha].$$

只需要证明 $\mathbb{E}[L \mathbb{I}_{\{L \geq q_\alpha(L)\}}] = \int_\alpha^1 F_L^\leftarrow(u) du$:

$$\mathbb{E}[L \mathbb{I}_{\{L \geq q_\alpha(L)\}}] = \mathbb{E}[F_L^\leftarrow(U); F_L^\leftarrow(U) \geq F_L^\leftarrow(\alpha)] = \mathbb{E}[F_L^\leftarrow(U); U \geq \alpha].$$

其中 $U \sim U[0, 1]$, 初等概率论的知识告诉我们随机变量 $F_L^\leftarrow(U)$ 的分布函数为 F_L .

Example B.1.5 (高斯损失分布的预期损失). 假设损失分布 F_L 服从均值为 μ 方差为 σ^2 的正态分布, 固定置信水平 $\alpha \in (0, 1)$, 有

$$ES_\alpha = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

证明过程如下

$$\begin{aligned} ES_\alpha &= \mu + \sigma \mathbb{E} \left[\frac{L - \mu}{\sigma} \middle| \frac{L - \mu}{\sigma} \geq q_\alpha \left(\frac{L - \mu}{\sigma} \right) \right], \\ ES_\alpha(\tilde{L}) &= \frac{1}{1-\alpha} \int_{\Phi^{-1}(\alpha)}^\infty l \phi(l) dl = \frac{1}{1-\alpha} [-\phi(l)]_{\Phi^{-1}(\alpha)}^\infty = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}. \end{aligned}$$

Example B.1.6 (学生 t 损失分布的预期损失). 假设损失 L 满足 $\tilde{L} = (L - \mu)/\sigma \sim t(\nu)$. 通过上一个例题的论证, 我们可以得到 $ES_\alpha = \mu + \sigma ES_\alpha(\tilde{L})$. 标准 t 分布的预期损失可以直接积分得出

$$ES_\alpha(\tilde{L}) = \frac{g_\nu(t^{-1}(\alpha))}{1-\alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu - 1} \right)$$

其中 t_ν 是标准 t 分布的 CDF, g_ν 是 pdf.

Theorem B.1.7. 对于一致性风险度量 $\rho : L \mapsto \rho(L)$ 有以下公理

1. 单调性: $L_1 \leq L_2 \Rightarrow \rho(L_1) \leq \rho(L_2)$
2. 平移不变性: $L \in \mathcal{M}, l \in \mathbb{R}, \rho(L + l) = \rho(L) + l$.

3. 次可加性: $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$.

4. 齐次性: 对于 $\lambda > 0$, $\rho(\lambda L) = \lambda \rho(L)$.

齐次性和次可加性的限制被放松, 得到更大一类的凸风险度量, 取而代之的是凸性特征: $\lambda \in [0, 1]$, $\rho(\lambda L_1 + (1 - \lambda)L_2) \leq \lambda \rho(L_1) + (1 - \lambda)\rho(L_2)$.

Example B.1.8 (可违约债券的非次可加性 VaR). 考虑两个零息债券组成的组合, 其期限为一年, 且两个债券违约时相互独立的。假定两个债券违约概率完全相同都为 $p = 0.9\%$. 债券的现价和面值均为 100, 债券支付的利息为 5%。如果没有违约的话一年之后将收到 105 元如果违约则收不到任何钱。把一个单位债券的损失记作 $L_i, i = 1, 2$ (负的损失相当于收益), 我们有

$$\begin{aligned} P(L_i = -5) &= 1 - p = 0.991 \quad \text{No default} \\ P(L_i = 100) &= p = 0.009 \quad \text{Default} \end{aligned}$$

设 $\alpha = 0.99$. 我们有 $P(L_i < -5) = 0$ 和 $P(L_i \leq -5) = 0.991 > \alpha$, 因此 $VaR_\alpha(L_i) = -5$.

现在考虑一个由公司债券组成的组合, 相应损失为 $L = L_1 + L_2$, 由独立性我们可以得到

$$\begin{aligned} P(L = -10) &= (1 - p)^2 = 0.982081 && \text{No default} \\ P(L = 95) &= 2p(1 - p) = 0.017838 && 1 \text{ Default} \\ P(L = 200) &= p^2 = 0.000081 && 2 \text{ Default} \end{aligned}$$

由于 $P(L \leq -10) = 0.982 < 0.99, P(L \leq 95) > 0.99$ 于是 $VaR_\alpha(L) = 9 > -10 = VaR_\alpha(L_1) + VaR_\alpha(L_2)$. 因此 VaR 是非次可加性的。

Example B.1.9 (预期损失的一致性). 另一方面, 预期损失 ES 是一致性的风险度量。从分位数的相应属性可以立即得到平移不变性, 单调性和齐次性。例如, 它认为

$$ES_\alpha(\lambda L) = \frac{1}{1 - \alpha} \int_\alpha^1 q_u(\lambda L) du = \frac{1}{1 - \alpha} \int_\alpha^1 \lambda q_u(L) du = \lambda ES_\alpha(L),$$

类似的论点适用于平移不变性和单调性。我们假设 L_1, L_2 和 $L_1 + L_2$ 具有连续分布, 于是有

$$ES_\alpha(L) = \frac{1}{1 - \alpha} \mathbb{E}[L \mathbb{I}_{\{L \geq q_\alpha(L)\}}],$$

设 $I_i := \mathbb{I}_{\{L_i \geq q_\alpha(L_i)\}}, i = 1, 2, I_{12} := I_{\{L_1 + L_2 \geq q_\alpha(L_1 + L_2)\}}$. 于是我们有

$$\begin{aligned} (1 - \alpha)(ES_\alpha(L_1) + ES_\alpha(L_2) - ES_\alpha(L_1 + L_2)) \\ &= \mathbb{E}[L_1 I_1] + \mathbb{E}[L_2 I_2] - \mathbb{E}[(L_1 + L_2) I_{12}] \\ &= \mathbb{E}[L_1(I_1 - I_{12})] + \mathbb{E}[L_2(I_2 - I_{12})] \end{aligned}$$

考虑第一项并假设 $\{L_1 \geq q_\alpha(L_1)\}$, 它遵循 $I_1 - I_{12} \geq 0$, 因此 $L_1(I_1 - I_{12}) \geq q_\alpha(L_1)(I_1 - I_{12})$. 假设另一方面 $\{L_1 < q_\alpha(L_1)\}$, 它遵循 $L_1(I_1 - I_{12}) \geq q_\alpha(L_1)(I_1 - I_{12})$. 同样的道

理也适用于 L_2 , 所以在这两种情况下我们都可以得到结论

$$\begin{aligned}
 & (1 - \alpha)(ES_\alpha(L_1) + ES_\alpha(L_2) - ES_\alpha(L_1 + L_2)) \\
 & \geq \mathbb{E}[q_\alpha(L_1)(I_1 - I_{12})] + \mathbb{E}[q_\alpha(L_2)(I_1 - I_{12})] \\
 & \geq q_\alpha(L_1)\mathbb{E}[I_1 - I_{12}] + q_\alpha(L_2)\mathbb{E}[I_1 - I_{12}] \\
 & \geq q_\alpha(L_1)((1 - \alpha) - (1 - \alpha)) + q_\alpha((1 - \alpha) - (1 - \alpha)) = 0
 \end{aligned}$$

这证明了次可加性。

PROBLEM B.1.1. 用 X 表示火灾损失随机变量, 假设对于阈值 $u = 1$, 超额损失 $Y = X - u | X > u$ 服从广义帕累托分布, 分布函数为 $G(y) = 1 - (1 + \xi \frac{y}{\sigma})^{-\frac{1}{\xi}}$, 其中形状参数 $\xi = 0.5$, 尺度参数 $\sigma = 1.5$. 如果阈值变为 $v = 1.2$, 求超额损失 $Y = X - v | X > v$ 的分布函数。如果进一步假设 $F_X(u) = 0.5$, 求 $\alpha = 95\%$ 水平下火灾损失的 VaR 和 TVaR 值。

SOLUTION. 对于 GPD 的超阈值分布当阈值从 u 提高到 $v > u$, 其超额损失仍服从 GPD 且形状参数不变, 尺度参数变为 $\sigma_v = \sigma_u + \xi(v - u)$.

B.2 Copula

Definition B.2.1 (连接函数 Copula). 一个 d 维连接函数是一个在 $[0, 1]^d$ 上的有标准均匀边际分布的分布函数。

连接函数的多元分布函数保留记号 $C(\mathbf{u}) = C(u_1, \dots, u_d)$. 因此 C 是一个 $C : [0, 1]^d \rightarrow [0, 1]$ 的映射, 即一个单位超立方体到一个单位区间的映射, 下面是必须满足的三个性质:

- (1) 若对任意的 $i, u_i = 0 \Rightarrow C(u_1, \dots, u_d) = 0$.
- (2) 若对任意的 $i \in \{1, \dots, d\}, u_i \in [0, 1] \Rightarrow C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$.
- (3) 若对所有满足 $a_i \leq b_i$ 的 $(a_1, \dots, a_d), (b_1, \dots, b_d) \in [0, 1]^d$, 那么我们有

$$\sum_{i_1=1}^2 \cdots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0$$

其中对于任意的 $j \in \{1, \dots, d\}$, 有 $u_{j1} = a_j, u_{j2} = b_j$.

Theorem B.2.2. 设 F 是一个分布函数并且用 F^\leftarrow 表示它的广义逆, 即函数 $F^\leftarrow(u) = \inf\{x : F(x) \geq u\}$.

- (1) 分位数转换: 若 $U \sim U(0, 1)$ 有一个标准均匀分布, 则 $P(F^\leftarrow(U) \leq x) = F(x)$.
- (2) 概率转换: 若 X 有分布函数 F , 其中 F 是一个连续一元分布函数, 则 $F(x) \sim U(0, 1)$.

证明. 设 $x \in \mathbb{R}, u \in (0, 1)$. 在第一部分我们有

$$F(x) \geq u \Leftrightarrow F^\leftarrow(u) \leq x,$$

从中满足

$$P(F^\leftarrow(U) \leq x) = P(U \leq F(x)) = F(x).$$

在第二部分我们可以推断

$$P(F(X) \leq u) = P(F^\leftarrow \circ F(X) \leq F^\leftarrow(u)) = P(X \leq F^\leftarrow(u)) = F \circ F^\leftarrow(u) = u.$$

□

这是随机模拟的关键。如果我们能生成一个均匀变量 U 并计算分布函数 F 的逆函数，那么我们就可以从那个分布函数中取样。

Theorem B.2.3 (Sklar 定理). 设 F 是一个联合分布函数，其中边际分布为 F_1, \dots, F_d . 那么在 $\bar{\mathbb{R}} = [-\infty, \infty]$ 上存在一个连接函数 $C : [0, 1]^d \rightarrow [0, 1]$, 使得对于所有的 x_1, \dots, x_d

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (\text{B.1})$$

如果边际分布是连续的，那么 C 是唯一的。相反，如果 C 是一个连接函数并且 F_1, \dots, F_d 是一元分布函数，那么上述定义的函数 F 是一个联合分布函数，具有边际分布 F_1, \dots, F_d .

证明. 设 \mathbf{X} 是一个有分布函数 F 并且具有连续边际 F_1, \dots, F_d 的随机向量，并且对于 $i = 1, \dots, d$, 令 $U_i = F_i(X_i)$, 那么 $U_i \sim U(0, 1)$, $F^\leftarrow(U_i) = X_i, a.s.$ 设 C 表示 (U_1, \dots, U_d) 的分布函数，由 Copula 函数的定义可知它是一个连接函数。对于任意的 x_1, \dots, x_d 在 $\bar{\mathbb{R}}$, 我们推论

$$\begin{aligned} F(x_1, \dots, x_d) &= P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= P(F_1^\leftarrow(U_1) \leq x_1, \dots, F_d^\leftarrow(U_d) \leq x_d) \\ &= P(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) \\ &= C(F_1(x_1), \dots, F_d(x_d)) \end{aligned} \quad (\text{B.2})$$

如果我们以参数 $x_i = F_i^\leftarrow(u_i)$, $0 \leq u_i \leq 1$, $i = 1, \dots, d$ 计算(B.1)则得到

$$C(u_1, \dots, u_d) = F(F_1^\leftarrow(u_1), \dots, F_d^\leftarrow(u_d))$$

这以 F 和它的边际分布形式给出了 C 的一个显式表达并因此说明了唯一性。

对于相反的陈述，假设 C 是一个 Copula, F_1, \dots, F_d 是任意的一元分布函数，通过用分布函数 C 把 \mathbf{U} 取成任意随机向量并设 $\mathbf{X} := \{F_1^\leftarrow(U_1), \dots, F_d^\leftarrow(U_d)\}$, 我们用分布函数公式(B.1)构造一个随机向量，接着可以根据从公式(B.2)开始的完全一样的等式顺序来确定满足公式(B.1)的 \mathbf{X} 的分布函数。 □

Definition B.2.4 (分布函数 F 的 Copula). 如果随机向量 \mathbf{X} 有联合分布函数 F , 且具有连续的边际分布 F_1, \dots, F_d , 则 F 的 Copula 是 $(F_1(X_1), \dots, F_d(X_d))$ 的分布函数 C .

Example B.2.5 (二元伯努利分布的连接函数). 设 (X_1, X_2) 有一个二元伯努利分布满足

$$\begin{aligned} P(X_1 = 0, X_2 = 0) &= \frac{1}{8}, & P(X_1 = 1, X_2 = 1) &= \frac{3}{8} \\ P(X_1 = 0, X_2 = 1) &= \frac{2}{8}, & P(X_1 = 1, X_2 = 0) &= \frac{2}{8} \end{aligned}$$

显然 $P(X_1 = 0) = P(X_2 = 0) = \frac{3}{8}$ 且 X_1, X_2 的边际分布 F_1, F_2 相同，由 Sklar 定理我们知道

$$P(X_1 \leq x_1, X_2 \leq x_2) = C(P(X_1 \leq x_1), P(X_2 \leq x_2))$$

对所有的 x_1, x_2 和一些连接函数 C 成立。因为 $\text{Range}F_1 = \text{Range}F_2 = \{0, \frac{3}{8}, 1\}$ (F_i 的取值范围)，显然对 C 的唯一约束是 $C(\frac{3}{8}, \frac{3}{8}) = \frac{1}{8}$. 任何满足这个约束的 Copula 就是 (X_1, X_2) 的 Copula 且有无数多个这样的连接函数。

Theorem B.2.6 (Fréchet 边界). 对于每个 Copula $C(u_1, \dots, u_d)$, 我们有边界

$$\max \left(\sum_{i=1}^d u_i + 1 - d, 0 \right) \leq C(\mathbf{u}) \leq \min(u_1, \dots, u_d)$$

证明. 第二个不等式的依据是，对于所有的 i ,

$$\bigcap_{1 \leq j \leq d} \{U_j \leq u_j\} \subset \{U_i \leq u_i\}$$

对第一个不等式，我们观察到

$$\begin{aligned} C(\mathbf{u}) &= P \left(\bigcap_{1 \leq i \leq d} \{U_i \leq u_i\} \right) = 1 - P \left(\bigcup_{1 \leq i \leq d} \{U_i > u_i\} \right) \\ &\geq 1 - \sum_{i=1}^d P(U_i > u_i) = 1 - d + \sum_{i=1}^d u_i. \end{aligned}$$

上边界和下边界分别记为 W, M .

□

PROBLEM B.2.1. 对于矩阵 $\mathbf{A}_{m \times n}$, 定义 $\|\mathbf{A}\|_p = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^p \right)^{\frac{1}{p}}$.

证明: $\|\mathbf{A}\|_p$ 是矩阵范数, 且 $\|\mathbf{A}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{A}\|_p = \max_{ij} |a_{ij}|$.

SOLUTION. (1) 矩阵范数要满足三条公理:

- 正定性: $\|\mathbf{A}\|_p \geq 0$, 且 $\|\mathbf{A}\|_p = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$.
- 齐次性: $\|t\mathbf{A}\|_p = |t| \cdot \|\mathbf{A}\|_p$.
- 三角不等式: $\|\mathbf{A} + \mathbf{B}\|_p \leq \|\mathbf{A}\|_p + \|\mathbf{B}\|_p$. (这里可以将 $\|\mathbf{A}\|_p$ 看作向量的 p 范数, 由 Minkowski 不等式得到, 其中 $p \geq 1$.)

(2) 设 $M = \max_{ij} |a_{ij}| > 0$ (若全零则显然成立). 假设矩阵中 k 个元素的绝对值等于 M , 则

$$\begin{aligned} \|\mathbf{A}\|_p &= M \left(k + \sum_{|a_{ij}| < M} \left(\frac{|a_{ij}|}{M} \right)^p \right)^{1/p} \\ \lim_{p \rightarrow \infty} \|\mathbf{A}\|_p &= \lim_{p \rightarrow \infty} M(k + o(1))^{1/p} = M = \max_{i,j} |a_{ij}| =: \|\mathbf{A}\|_\infty. \end{aligned}$$

之后也可以证明这也是一个矩阵范数。

PROBLEM B.2.2. 设 $X, Y \sim i.i.d.$ 且有 $X = \epsilon + \eta, \epsilon \sim N(0, 1), P(\eta = 0) = 0.991, P(\eta = -10) = 0.009$. 计算 $\text{VaR}(X), \text{VaR}(Y), \text{VaR}(X + Y)$, 并说明 VaR 不满足次可加性。

SOLUTION. 取 α 为

PROBLEM B.2.3.

1. 叙述一致性风险度量方法所需要满足的条件。
2. 假设损失随机变量

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma^2 \end{pmatrix} \right],$$

求 $\text{VaR}(X_1), \text{VaR}(X_2), \text{VaR}(X_1 + X_2)$.

3. 验证次可加性。

SOLUTION. 1. 一致性风险度量 (*Coherent Risk Measure*) 由 Artzner et al. (1999) 提出, 一个风险度量 $\rho : \mathcal{X} \rightarrow \mathbb{R}$ 要满足以下四条公理:

1. 单调性: 若 $X \leq Y a.s.$, 则 $\rho(X) \geq \rho(Y)$. (损失越大风险度量值越大)
 2. 平移不变性: 对任意实数 a , 有 $\rho(X + a) = \rho(X) + a$. (增加确定的损失 a , 风险值也增加 a)
 3. 次可加性: $\rho(X + Y) \leq \rho(X) + \rho(Y)$. (分散化降低风险)
 4. 正齐次性: 对任意 $\lambda \geq 0$, 有 $\rho(\lambda X) = \lambda\rho(X)$.
2. 由题目条件, 损失为正, 所以 VaR 为右尾分位数, 通常 $\text{VaR}_\alpha(X) = F_X^{-1}(\alpha)$ 表示在置信水平 α 下的最大可能损失。

对于 $X_i \sim N(\mu_i, \sigma_i^2)$:

$$\text{VaR}_\alpha(X_i) = \mu_i + \sigma_i \cdot z_\alpha, i = 1, 2,$$

其中 $z_\alpha = \Phi^{-1}(\alpha)$ 是标准正态分布的 α 分位数。

对于 $X_1 + X_2$, 由于正态分布的线性组合仍未正态分布:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2).$$

所以

$$\text{VaR}_\alpha(X_1 + X_2) = (\mu_1 + \mu_2) + \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \cdot z_\alpha}.$$

在 $\alpha > 0.5 \Rightarrow z_\alpha > 0$ 情形下, 由于 $\rho \leq 1$ 时, VaR 在正态分布下满足次可加性。注意: Artzner 等指出 VaR 一般不满足次可加性 (对于非正态分布可能违反)。

PROBLEM B.2.4. 讨论常用的一致性风险度量期望亏空 ES 的次可加性。

SOLUTION. 对于正态分布 $X_i \sim N(\mu_i, \sigma_i^2)$,

$$\begin{aligned} ES_\alpha(X_i) &= \mathbb{E}[X_i | X_i \geq VaR_\alpha(X_i)] \\ &= \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(X) du \\ &= \mu_i + \sigma_i \cdot \frac{1}{1-\alpha} \int_\alpha^1 z_u du, \end{aligned}$$

令 $z_u = \Phi^{-1}(u), u = \Phi(z), du = \phi(z)dz$. 当 $u = \alpha$ 时, $z = z_\alpha; u = 1$ 时 $z \rightarrow \infty$.

$$\int_\alpha^1 z_u du = \int_{z_\alpha}^\infty z \phi(z) dz = \int_\infty^{z_\alpha} d\phi(z) = \phi(z_\alpha).$$

因此 $ES_\alpha(X_i) = \mu_i + \sigma_i \cdot \frac{\phi(z_\alpha)}{1-\alpha} = \mu_i + k\sigma_i$, 其中 ϕ 是标准正态分布的概率密度函数。

对于两个正态分布变量验证次可加性:

$$ES_\alpha(X_1 + X_2) = \mu_1 + \mu_2 + k\sigma_P,$$

其中 $\sigma_P = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$.

PROBLEM B.2.5. 免赔额 d , 生存函数为 $\bar{F}(x)$. (1) 用 $\bar{F}(x)$ 表示保险公司的期望赔付; (2) 用 $\bar{F}(x)$ 表示未索赔期望赔付; (3) 设计方案求最优 d^* .

SOLUTION. 设免赔额为 d , 损失随机变量 $X \geq 0$ 的生存函数为 $\bar{F}(x) = \mathbb{P}(X > x)$.

(1) 保险公司的赔付为 $Y = \max(0, X - d)$. 期望赔付为:

$$\mathbb{E}[Y] = \int_d^\infty (x - d) f(x) dx.$$

利用分部积分, 可得:

$$\mathbb{E}[Y] = \int_d^\infty \bar{F}(x) dx.$$

因此:

$$\mathbb{E}[\text{赔付}] = \int_d^\infty \bar{F}(x) dx$$

(2) “未索赔赔付”指被保险人自留的损失:

$$Z = \min(X, d).$$

其期望为:

$$\mathbb{E}[Z] = \int_0^d x f(x) dx + d \cdot \bar{F}(d).$$

利用分部积分：

$$\int_0^d xf(x) dx = -d\bar{F}(d) + \int_0^d \bar{F}(x) dx.$$

所以：

$$\mathbb{E}[Z] = \int_0^d \bar{F}(x) dx.$$

因此：

$$\boxed{\mathbb{E}[\text{未索赔赔付}] = \int_0^d \bar{F}(x) dx}$$

(3) 考虑被保险人的总期望支出 (保费 + 自留损失)：

$$T(d) = P(d) + \mathbb{E}[\min(X, d)],$$

其中保费 $P(d) = (1 + \theta)\mathbb{E}[Y] = (1 + \theta) \int_d^\infty \bar{F}(x) dx$ ($\theta > 0$ 为安全附加系数)。

因此：

$$T(d) = (1 + \theta) \int_d^\infty \bar{F}(x) dx + \int_0^d \bar{F}(x) dx.$$

对 d 求导：

$$T'(d) = (1 + \theta)(-\bar{F}(d)) + \bar{F}(d) = -\theta\bar{F}(d).$$

由于 $\theta > 0$ 且 $\bar{F}(d) \geq 0$, 有 $T'(d) \leq 0$, 故 $T(d)$ 是 d 的减函数。

因此, 最小化 $T(d)$ 需要取尽可能大的 d , 即:

$$d^* \rightarrow \infty \quad (\text{不投保}).$$

在现实中, 由于风险厌恶, 最优免赔额会是有限值, 但本模型中仅考虑期望值, 故得到角点解。

$$\boxed{d^* \rightarrow \infty}$$

PROBLEM B.2.6. 设 (x) 岁个体的寿命 T_x 的分布为 $\mathbb{P}(T_x < t)$:

1. 什么是死力 μ_{x+t} ? 请用 μ_{x+t} 表示死亡概率 tq_x ;
2. 对于 Makeham 模型 $\mu_x = A + BC^x$, 请描述估计参数的过程;
3. Lee-Carter 模型呢?

SOLUTION. (1) 死力 (Force of Mortality) 定义为

$$\mu_{x+t} = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(t < T_x \leq t+h | T_x > t)}{h}.$$

已知生存函数

$$S_x(t) = \mathbb{P}(T_x > t) = \exp\left(-\int_0^t \mu_{x+s} ds\right),$$

死亡概率

$$tq_x = \mathbb{P}(T_x \leq t) = 1 - S_x(t) = 1 - \exp\left(-\int_0^t \mu_{x+s} ds\right).$$

(2) Makeham 定律:

$$\mu_x = A + BC^x, A \geq 0, B > 0, C > 1.$$

估计过程:

1. 数据: 有各年龄 x 的中心死亡率 m_x 或者死亡人数 D_x 与暴露人数 E_x , 且 $m_x \approx \mu_{x+1/2}$.
2. 线性化方法: 对两个不同年龄 $x, x+1$ 的死力差分:

$$\begin{aligned}\mu_{x+1} - \mu_x &= BC^x(C-1) \\ \ln(\mu_{x+1} - \mu_x) &= \ln[B(C-1)] - x \ln C\end{aligned}$$

对 x 回归可得 $\ln C$ 和截距从而得到 B . 但需先估计 A : 对于很高龄 x , BC^x 很小, $\mu_x \approx A$. 可先取高龄 μ_x 的平均值作为 A 的初值, 然后计算 $\mu_x - A$ 再拟合 BC^x .

3. 现代方法: 直接使用最大似然估计。假设 $D_x \sim Poisson(E_x \cdot \mu_x)$, 且 $\mu_x = A + BC^x$. 忽略掉常数的对数似然:

$$l(A, B, C) = \sum_x [D_x \ln(A + BC^x) - E_x(A + BC^x)].$$

用数值优化 (如 Newton-Raphson) 求 A, B, C 的 MLE, 注意约束 $A \geq 0, B > 0, C > 1$.

PROBLEM B.2.7. 若一个个体有三种状态, 其中状态 0 表示生存, 1 表示残疾, 2 表示死亡, 基本转移关系为: $0 \xrightarrow{\mu_{01}(t)} 1 \xrightarrow{\mu_{12}(t)} 2, 0 \xrightarrow{\mu_{02}(t)} 2$. 其中 $\mu(t)$ 表示转移强度, 计算 x 岁的人在 t 年内死亡的概率。

SOLUTION. 记

$$\begin{aligned}{}_tp_x^{00} &= P(\text{在年龄 } x+t \text{ 时状态 } 0 \mid \text{年龄 } x \text{ 时状态 } 0) \\ {}_tp_x^{01} &= P(\text{在年龄 } x+t \text{ 时状态 } 1 \mid \text{年龄 } x \text{ 时状态 } 0) \\ {}_tp_x^{02} &= P(\text{在年龄 } x+t \text{ 时状态 } 2 \mid \text{年龄 } x \text{ 时状态 } 0)\end{aligned}$$

满足 ${}_tp_x^{00} + {}_tp_x^{01} + {}_tp_x^{02} = 1$.

Kolmogorov 前向方程 (年龄相关):

$$\begin{aligned}\frac{d}{dt}{}_tp_x^{00} &= -[\mu_{01}(x+t) + \mu_{02}(x+t)] {}_tp_x^{00}, \\ \frac{d}{dt}{}_tp_x^{01} &= \mu_{01}(x+t) {}_tp_x^{00} - \mu_{12}(x+t) {}_tp_x^{01}, \\ \frac{d}{dt}{}_tp_x^{02} &= \mu_{02}(x+t) {}_tp_x^{00} + \mu_{12}(x+t) {}_tp_x^{01}.\end{aligned}$$

初始条件:

$${}_0p_x^{00} = 1, \quad {}_0p_x^{01} = 0, \quad {}_0p_x^{02} = 0.$$

求解 ${}_tp_x^{00}$

$${}_tp_x^{00} = \exp \left[- \int_0^t (\mu_{01}(x+s) + \mu_{02}(x+s)) ds \right].$$

求解 ${}_tp_x^{01}$ 方程:

$$\frac{d}{dt}{}_tp_x^{01} + \mu_{12}(x+t) {}_tp_x^{01} = \mu_{01}(x+t) {}_tp_x^{00}.$$

积分因子:

$$R(t) = e^{\int_0^t \mu_{12}(x+r) dr}.$$

解:

$${}_tp_x^{01} = e^{-\int_0^t \mu_{12}(x+r) dr} \int_0^t \mu_{01}(x+s) {}_sp_x^{00} \cdot e^{\int_0^s \mu_{12}(x+r) dr} ds.$$

代入 ${}_sp_x^{00}$ 表达式:

$${}_tp_x^{01} = \int_0^t \mu_{01}(x+s) e^{-\int_0^s [\mu_{01}(x+r) + \mu_{02}(x+r)] dr} e^{-\int_s^t \mu_{12}(x+r) dr} ds.$$

求解 ${}_tp_x^{02}$ (死亡概率) 由微分方程:

$$\frac{d}{dt}{}_tp_x^{02} = \mu_{02}(x+t) {}_tp_x^{00} + \mu_{12}(x+t) {}_tp_x^{01},$$

初始 ${}_0p_x^{02} = 0$, 得:

$${}_tp_x^{02} = \int_0^t \mu_{02}(x+s) {}_sp_x^{00} ds + \int_0^t \mu_{12}(x+s) {}_sp_x^{01} ds.$$

第一项表示直接从状态 0 到状态 2 的死亡概率, 第二项表示先转移到状态 1 再从状态 1 到状态 2 的死亡概率

x 岁的人在 t 年内死亡的概率为:

$${}_tp_x^{02} = \int_0^t \mu_{02}(x+s) {}_sp_x^{00} ds + \int_0^t \mu_{12}(x+s) {}_sp_x^{01} ds$$

其中

$${}_sp_x^{00} = \exp \left[- \int_0^s (\mu_{01}(x+r) + \mu_{02}(x+r)) dr \right],$$

$${}_s p_x^{01} = \int_0^s \mu_{01}(x+u) {}_u p_x^{00} \exp \left[- \int_u^s \mu_{12}(x+r) dr \right] du.$$

PROBLEM B.2.8. Copula 是一种多元累积分布函数，其中每个变量的边际概率分布在区间 $[0, 1]$ 均匀。Copula 用于描述/建模随机变量之间的依赖关系。

考虑一个随机向量 (X_1, \dots, X_d) ，假设其边际函数是连续的，即边际的 CDF $F_i(x) = \mathbb{P}(X_i \leq x)$ 是连续函数。通过对每个分量应用概率积分变换，随机向量

$$(U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$$

的边际分布均匀分布于区间 $[0, 1]$ 。

于是 (X_1, \dots, X_d) 的 Copula 被定义为 (U_1, \dots, U_d) 的联合累积分布函数：

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d).$$

Copula C 钟包含了关于 (X_1, \dots, X_d) 各组分之间依赖结构的所有信息，而边际累积分布函数 F_i 包含了关于 X_i 边际分布的所有信息。

反过来，也可以从多元概率分布的一般类别中生成伪随机样本，也就是说，给定从 Copula 函数中生成样本 (U_1, \dots, U_d) 的程序，所需要的样本可以构造为

$$(X_1, \dots, X_d) = (F_1^{-1}(U_1), \dots, F_d^{-1}(U_d)).$$

于是 Copula 函数可以重写为

$$C(u_1, \dots, u_d) = \mathbb{P}(X_1 \leq F_1^{-1}(u_1), \dots, X_d \leq F_d^{-1}(u_d)).$$

定义 (二元)：假设有两个随机变量 $X \sim F_X(x), Y \sim F_Y(y)$ ，则联合分布 $H(x, y)$ 可以通过一个 Copula 函数 $C(u, v)$ 表示为

$$H(x, y) = C(F_X(x), F_Y(y)),$$

其中 $u = F_X(x), v = F_Y(y)$ 是边缘分布的 CDFs.

在量化金融中，Copula 函数被应用于风险管理、投资组合管理和优化以及衍生品定价。

Name of Copula	Bivariate Copula $C_\theta(u, v)$	parameter θ
Ali-Mikhail-Haq	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta \in [-1, 1]$
Clayton	$[\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-1/\theta}$	$\theta \in [-1, 0) \cup (0, \infty)$
Gumbel	$\exp \left\{ - [(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta} \right\}$	$\theta \in [1, \infty)$
Independence	uv	

表 1: Table with the most important Archimedean copulas

PROBLEM B.2.9. 记多维函数 $W(u) = \max\{\sum_{j=1}^d u_j - d + 1, 0\}$ 和 $M(u) = \min_{1 \leq j \leq d}\{u_j\}$.

要求:

- (1) 对任意 d 维 Copula 函数, 证明 $W(u) \leq C(u) \leq M(u), u \in [0, 1]^d$.
- (2) 设随机变量 X 服从标准正态分布, 定义多元随机变量 $Z_1 = (X, X^2, \dots, X^d)$, $Z_2 = (X, X^{-1})$, 分别求 Z_1 和 Z_2 的分布对应的 Copula 函数。

SOLUTION.

(1) Copula 定义: 若 U_1, \dots, U_d 是边缘为 $U(0, 1)$ 的随机变量, 其联合分布函数为 C , 则

$$C(u_1, \dots, u_d) = \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d).$$

显然有 $\{U_1 \leq u_1, \dots, U_d \leq u_d\} \subset \{U_j \leq u_j\}$ 对每个 j 都成立, 所以 $C(\mathbf{u}) \leq \mathbb{P}(U_j \leq u_j) = u_j$ 对每个 j , 因此 $C(\mathbf{u}) \leq \min_j\{u_j\} = M(\mathbf{u})$.

至于下界, 用概率的 Fréchet 不等式:

$$\mathbb{P}(\cap_{j=1}^d A_j) \geq \sum_{j=1}^d \mathbb{P}(A_j) - (d-1).$$

取 $A_j = \{U_j \leq u_j\}$, 则 $\mathbb{P}(A_j) = u_j$, 所以:

$$C(\mathbf{u}) \geq \sum_{j=1}^d u_j - (d-1).$$

同时 $C(\mathbf{u}) \geq 0$, 所以:

$$C(\mathbf{u}) \geq \max \left\{ \sum_{j=1}^d u_j - d + 1, 0 \right\} = W(\mathbf{u}).$$

PROBLEM B.2.10. 设 X, Y 是随机变量, 联合分布函数为

$$H(x, y) = (1 + e^{-x} + e^{-y})^{-1}, \forall x, y \in \mathbb{R}.$$

试证明:

- (1) X, Y 各自服从单变量标准 logistic 分布;
- (2) X, Y 的 Copula 由下式给出:

$$C(u, v) = \frac{uv}{u+v-uv}.$$

SOLUTION.

(1)

$$F_X(x) = \lim_{y \rightarrow \infty} H(x, y) = \frac{1}{1 + e^{-x}}.$$

正是标准 *logistic* 分布的 CDF, 位置参数为 0, 尺度参数为 1. 由对称性 $F_Y(y) = \frac{1}{1+e^{-y}}$.

(2) Copula 定义为:

$$C(u, v) = H(F_X^{-1}(u), F_Y^{-1}(v)),$$

其中 $F_X^{-1}(u)$ 是分位数函数。

已知 $F_X(t) = \frac{1}{1+e^{-t}} := p \Rightarrow t = \ln\left(\frac{p}{1-p}\right)$. 即

$$F_X(u) = \ln\left(\frac{u}{1-u}\right), F_Y(v) = \ln\left(\frac{v}{1-v}\right).$$

代入可得:

$$H = \frac{1}{1 + \frac{1-u}{u} + \frac{1-v}{v}} = \frac{uv}{u+v-uv} = C(u, v).$$

C 随机过程

PROBLEM C.0.1. [离散-连续混合型贝叶斯公式] 设 X 是离散随机变量, 当 $X = x$ 时, 连续型随机变量 Y 有条件密度 $h(y|x)$, 试证

$$P(X = x|Y = y) = \frac{P(X = x)h(y|x)}{\sum_{x'} P(X = x')h(y|x')}$$

SOLUTION. 由贝叶斯公示的连续-离散混合形式

$$P(X = x|Y = y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

PROBLEM C.0.2. 设 X_1, X_2, \dots 是独立随机变量, $X_j \sim U[-j, j]$, 证明 Lindeberg 条件成立并给出 $\sum_{i=1}^n X_i$ 的极限分布。

SOLUTION.

$$\begin{aligned} \mathbb{E}X_j &= 0, \mathbb{V}arX_j = \frac{(2j)^2}{12} = \frac{j^2}{3}, \\ s_n^2 &= \sum_{j=1}^n \mathbb{V}arX_j = \sum \frac{j^2}{3} = \frac{1}{3} \cdot \frac{n(n+1)(2n+1)}{6}, \\ s_n^2 &\sim \frac{n^3}{9} (n \rightarrow \infty). \end{aligned}$$

Lindeberg 条件: 对任意的 $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}[X_j^2 \mathbb{I}_{\{|X_j| > \epsilon s_n\}}] = 0.$$

由 Lindeberg-Feller 中心极限定理

$$\frac{\sum X_j}{s_n} \xrightarrow{d} N(0, 1)$$

...

PROBLEM C.0.3. 设 $X_1, \dots, X_n \sim U(0, 1)$, 求 $\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} \right]$.

SOLUTION. 记 $S_n = \frac{1}{n} \sum X_i$, $T_n = \frac{1}{n} \sum X_i^2$, 由强大数定律

$$\begin{aligned} S_n &\xrightarrow{a.s.} \mathbb{E}[X_i] = \frac{1}{2}, T_n \xrightarrow{a.s.} \mathbb{E}[X_i^2] = \frac{1}{3}. \\ &\Rightarrow \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} = \frac{T_n}{S_n} \xrightarrow{a.s.} \frac{2}{3}. \end{aligned}$$

下面想要交换极限与期望, 要证明一致可积性或者找到一个控制函数通过控制收敛定理证明可交换:

$$0 \leq \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} \leq \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i} = 1,$$

于是由控制收敛定理

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} \right] = \frac{2}{3}.$$

PROBLEM C.0.4. (1) 写出标准布朗运动的定义和伊藤公式, 解释一下什么是 Markov 过程, 什么是鞅, 什么是停时; (2) 记 $x(t) = \int_0^t B_s ds$, $0 < s < t$, 求 $\text{Cov}(X_t, X_s)$; (3) 求 $P(u < B(t) \leq v | T_x \leq t)$, 其中 $T_x = \inf\{t > 0, B(t) = x\}$.

SOLUTION. a. 标准布朗运动 (维纳过程) $B(t)$ 是满足以下条件的随机过程:

1. $B(0) = 0$ 几乎必然
2. 具有独立增量: 对任意 $0 \leq t_1 < t_2 < \dots < t_n$, 增量 $B(t_2) - B(t_1), B(t_3) - B(t_2), \dots, B(t_n) - B(t_{n-1})$ 相互独立
3. 具有平稳正态增量: 对任意 $0 \leq s < t$, $B(t) - B(s) \sim N(0, t-s)$
4. 样本路径 $t \mapsto B(t)$ 几乎必然连续

b. 伊藤公式: 对伊藤过程 $dX_t = \mu_t dt + \sigma_t dB_t$, B_t , $f(t, x)$ 是二次连续可微函数, 则:

$$df(t, X_t) = f_t dt + f_x dX_t + \frac{1}{2} f_{xx} dX_t dX_t,$$

其中 $dX_t dX_t = \sigma_t^2 dt$. c. 随机过程 $\{X(t), t \geq 0\}$ 是 Markov 过程, 如果对未来状态的预测只依赖于当前状态, 而不依赖于过去历史。即对任意 $0 \leq s < t$ 和可测集 A :

$$P(X(t) \in A | X(u), 0 \leq u \leq s) = P(X(t) \in A | X(s))$$

d. 鞅: 随机过程 $\{M(t), t \geq 0\}$ 是关于域流 $\{\mathcal{F}_t\}$ 的鞅, 如果:

1. $M(t)$ 是 \mathcal{F}_t -可测的
2. $\mathbb{E}[|M(t)|] < \infty$ 对任意 $t \geq 0$
3. 对任意 $0 \leq s < t$, 有 $\mathbb{E}[M(t) | \mathcal{F}_s] = M(s)$ 几乎必然
- e. 停时: 随机变量 τ 是关于域流 $\{\mathcal{F}_t\}$ 的停时, 如果对任意 $t \geq 0$:

$$\{\tau \leq t\} \in \mathcal{F}_t$$

即根据到时间 t 为止的信息, 可以判断停时 τ 是否已经发生。

(2) 不失一般性, 设 $s < t$ 。

$$Cov(X_t, X_s) = \mathbb{E}[X_t X_s] - \mathbb{E}[X_t] \mathbb{E}[X_s]$$

由于布朗运动的期望为 0, $\mathbb{E}[X_t] = \mathbb{E}[X_s] = 0$, 所以:

$$Cov(X_t, X_s) = \mathbb{E}[X_t X_s] = \mathbb{E}\left[\int_0^t B(u) du \int_0^s B(v) dv\right]$$

交换积分与期望的顺序:

$$= \int_0^t \int_0^s \mathbb{E}[B(u) B(v)] dudv$$

由于 $\mathbb{E}[B(u) B(v)] = \min(u, v)$, 且 $s < t$, 有:

$$= \int_0^t \int_0^s \min(u, v) dudv = \int_0^s \left[\int_0^v u du + \int_v^s v du \right] dv$$

计算内层积分:

$$\int_0^v u du = \frac{v^2}{2}, \quad \int_v^s v du = v(s-v)$$

所以:

$$= \int_0^s \left[\frac{v^2}{2} + v(s-v) \right] dv = \int_0^s \left[vs - \frac{v^2}{2} \right] dv = \left[\frac{vs^2}{2} - \frac{v^3}{6} \right]_0^s = \frac{s^3}{2} - \frac{s^3}{6} = \frac{s^3}{3}$$

因此:

$$Cov(X_t, X_s) = \frac{1}{3} \min(s^3, t^3) = \frac{s^3}{3} \quad (s < t)$$

(3) 由布朗运动的强马尔科夫性和反射原理。设 $x > 0$ ($x < 0$ 的情况类似)。由反射原理, 对于 $y \leq x$:

$$P(B(t) \leq y, T_x \leq t) = P(B(t) \geq 2x - y)$$

因此:

$$P(B(t) \leq y | T_x \leq t) = \frac{P(B(t) \geq 2x - y)}{P(T_x \leq t)}$$

其中 $P(T_x \leq t) = 2P(B(t) \geq x) = 2\Phi\left(-\frac{x}{\sqrt{t}}\right)$, Φ 是标准正态分布函数。

对于 $u < B(t) \leq v$, 考虑两种情况:

情况 1: $v \leq x$

$$P(u < B(t) \leq v \mid T_x \leq t) = \frac{P(B(t) \geq 2x - v) - P(B(t) \geq 2x - u)}{P(T_x \leq t)}$$

情况 2: $u < x < v$

$$P(u < B(t) \leq v \mid T_x \leq t) = \frac{P(B(t) \geq 2x - v) - P(B(t) \geq 2x - u) + P(u < B(t) \leq v)}{P(T_x \leq t)}$$

由于布朗运动在击中 x 后继续演化，条件概率需要考虑路径是否经过 x 。更精确地：

$$P(u < B(t) \leq v \mid T_x \leq t) = \frac{P(u < B(t) \leq v, T_x \leq t)}{P(T_x \leq t)}$$

由反射原理：

$$P(u < B(t) \leq v, T_x \leq t) = \begin{cases} P(B(t) \geq 2x - v) - P(B(t) \geq 2x - u), & v \leq x \\ P(B(t) \geq x) + P(u < B(t) \leq v) - P(B(t) \geq 2x - u), & u < x < v \end{cases}$$

因此：

$$\boxed{P(u < B(t) \leq v \mid T_x \leq t) = \frac{P(u < B(t) \leq v, T_x \leq t)}{2\Phi(-x/\sqrt{t})}}$$

其中分子按上述情况分别计算。

PROBLEM C.0.5. 设 $\{X_n, n = 1, 2, \dots\}$ 关于它自己是鞅，令 $Z_n = X_n - X_{n-1}, X_0 = 0$.
证明： $\text{Var}(X_n) = \sum_{k=1}^n \text{Var}(Z_k)$.

SOLUTION. 由于鞅性

$$\mathbb{E}[X_n \mid X_0, \dots, X_{n-1}] = X_{n-1}$$

则

$$\mathbb{E}[Z_n \mid X_0, \dots, X_{n-1}] = \mathbb{E}[X_n - X_{n-1} \mid X_0, \dots, X_{n-1}] = X_{n-1} - X_{n-1} = 0,$$

所以 Z_n 是鞅差序列， $\mathbb{E}Z_n = 0$.

此外，对于 $j < k$,

$$\mathbb{E}[Z_j Z_k] = \mathbb{E}[\mathbb{E}[Z_j Z_k \mid X_0, \dots, X_{k-1}]] = \mathbb{E}[Z_j \mathbb{E}[Z_k \mid \mathcal{F}_{k-1}]] = 0,$$

于是 Z_n 是不相关序列。

我们有 $X_n = \sum_{k=1}^n Z_k$, 由方差公式

$$\text{Var}(X_n) = \text{Var}\left(\sum_{k=1}^n Z_k\right) = \sum_{k=1}^n \text{Var}(Z_k).$$

- PROBLEM C.0.6.** 随机过程 $\{B(t), t \geq 0\}$, $B(0) = 0$ 是一个标准布朗运动。 (1) 计算条件概率 $\mathbb{P}\{B(1) \leq 0 | B(2) \leq 0\}$;
 (2) 已知 $B(1) = 0$ 的条件下, 写出 $B(s)$ ($0 < s < 1$) 的条件密度。

SOLUTION. (1) 令 $B(1) = X, B(2) - B(1) = Y, B(2) = B(2) - B(1) + B(1) = X + Y$. 由布朗运动的独立增量性可知 X, Y 相互独立且 $X \sim N(0, 1), Y \sim N(0, 1)$, 所以 $X + Y \sim N(0, 2)$. 所求概率即为

$$\mathbb{P}(X \leq 0 | X + Y \leq 0) = \frac{\mathbb{P}(X \leq 0, X + Y \leq 0)}{\mathbb{P}(Y \leq 0)}.$$

令 $X = R \cos \theta, Y = R \sin \theta, X + Y = \sqrt{2}R \sin(\theta + \frac{\pi}{4}), \theta \sim U[0, 2\pi]$. 则

$$X \leq 0 \Rightarrow \theta \in \left[\frac{\pi}{2}, \frac{3}{2}\pi\right], X + Y \leq 0 \Rightarrow \theta \in \left[\frac{3}{4}\pi, \frac{7}{4}\pi\right],$$

$$\mathbb{P}(X \leq 0, X + Y \leq 0) = \frac{\frac{3}{2}\pi - \frac{3}{4}\pi}{2\pi} = \frac{8}{3}.$$

另一种方法:

$$\begin{aligned} \mathbb{P}(B(1) \leq 0, B(2) \leq 0) &= \mathbb{P}(B(1) \leq 0, B(1) + [B(2) - B(1)] \leq 0) \\ &= \mathbb{P}(B(1) \leq 0, B(2) - B(1) \leq -B(1)) = \int_{-\infty}^0 \mathbb{P}(B(2) - B(1) \leq -x) \phi(x) dx \\ &= \int_{-\infty}^0 \Phi(-x) d\Phi(x) = \int_0^\infty \Phi(x) \phi(-x) dx \\ &= \int_0^\infty \Phi(x) d\Phi(x) = \int_{\frac{1}{2}}^1 y dy = \frac{3}{8}. \end{aligned}$$

另外如果过程从 x 开始, $B(0) = x, B(t) \sim N(x, t)$.

又 $\mathbb{P}(X + Y \leq 0) = \frac{1}{2}$, 则结果为 $\frac{3}{4}$.

(2) 这是一个布朗桥, 对于标准布朗运动已知 $B(0) = B(1) = 0$, 则对于 $0 < s < 1$, 条件分布 $B(s) | B(1) = 0$ 是正态分布。

对于二元正态随机向量 X, Y , 均值为 (μ_X, μ_Y) , 协方差矩阵为 $\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ 给定 $Y = y, X$ 的条件分布为

$$X | Y = y \sim N \left(\mu_X + \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y), \sigma_X^2 (1 - \rho^2) \right).$$

应用到布朗桥: $B(s) \sim N(0, s), B(1) \sim N(0, 1), \text{Cov}(B(s), B(1)) = \min(s, 1) = s, \rho = \sqrt{s}$.

$$\mathbb{E}[B(s) | B(1) = 0] = 0, \text{Var}[B(s) | B(1) = 0] = s(1 - s).$$

PROBLEM C.0.7. 求 $\mathbb{E}[\exp(\sigma(B_s - B_t))], s \geq t$.

SOLUTION.

(1) 布朗运动的性质：对于 $s \geq t, B_s - B_t \sim N(0, s - t)$.

(2) 矩生成函数： $X \sim N(0, \sigma^2), \mathbb{E}[e^{tX}] = e^{\frac{1}{2}\sigma^2 t^2}$.

则 $\mathbb{E}[\exp(\sigma(B_s - B_t))] = e^{\frac{1}{2}\sigma^2(s-t)}$.

PROBLEM C.0.8. 设 $\{B(t), t \geq 0\}$ 是标准布朗运动，验证 $\{X(t) = (1-t)B\left(\frac{t}{1-t}\right), 0 \leq t \leq 1\}$ 是 Brown 桥。

SOLUTION. 首先，由于 $\{B(t)\}$ 是 Gauss 过程， $X(t)$ 是 $B(t)$ 线性变换，故 $\{X(t)\}$ 也是一 Gauss 过程。下面由布朗桥的定义：

$$\begin{aligned}\mathbb{E}[X(t)] &= (1-t)\mathbb{E}\left[B\left(\frac{t}{1-t}\right)\right] = 0, \\ \mathbb{E}[X(s)X(t)] &= \mathbb{E}\left[(1-t)B\left(\frac{t}{1-t}\right)(1-s)B\left(\frac{s}{1-s}\right)\right] \\ &= (1-t)(1-s)\min\left\{\frac{t}{1-t}, \frac{s}{1-s}\right\} = s(1-t), 0 \leq s \leq t \leq 1.\end{aligned}$$

所以 $\{X(t)\}$ 是 Gauss 过程，均值为 0，协方差为 $s(1-t)$ ，即为 Brown 桥。至于边界条件 $t \rightarrow 1$ 时可通过分布意义或者连续性修正： $\lim_{t \rightarrow 1} \mathbb{V}ar(X(t)) = t(1-t) = 0$.

PROBLEM C.0.9. 设 $\{B_1(t), t \geq 0\}, \{B_2(t), t \geq 0\}$ 为相互独立的标准布朗运动，试证 $\{X(t) = B_1(t) - B_2(t), t \geq 0\}$ 是布朗运动。

SOLUTION. (1) 初始值： $X(0) = B_1(0) - B_2(0) = 0$;

(2) 连续性：由于 B_1, B_2 的样本路径均关于 t 连续，于是 $X(t)$ 也是关于 t 连续的；

(3) 独立平稳增量： $X(t) - X(s) = [B_1(t) - B_1(s)] - [B_2(t) - B_2(s)]$;

(4) 均值与方差： $\mathbb{E}[X(t)] = 0, \mathbb{V}ar(X(t)) = 2t, X(t) \sim N(0, 2t)$;

(5) 还可以验证增量的正态性与平稳性： $X(t) - X(s) \sim N(0, 2(t-s))$.

PROBLEM C.0.10. 设 $\{B(t), t \geq 0\}$ 为标准 Brown 运动，证明 $M(t) = \max_{0 \leq s \leq t} B(s), |B(t)|$ 与 $M(t) - B(t)$ 具有相同的分布，并求出 $m(t) = \min_{0 \leq s \leq t} B(s)$ 的分布。

SOLUTION. (1) 首先证明 $M(t)$ 与 $|B(t)|$ 同分布

根据反射原理，对于任意的 $x > 0$:

$$P(M(t) \geq x) = 2P(B(t) > x) = P(B(t) \geq x) + P(B(t) \leq -x) = P(|B(t)| \geq x),$$

其中第三个等号是因为 Brown 运动的对称性，于是 $M(t) \sim |B(t)|$ 分布函数为

$$P(M(t) \geq x) = \frac{2}{\sqrt{2\pi}} \int_{x/\sqrt{t}}^{\infty} e^{-\frac{y^2}{2}} dy.$$

(2) 证明 $M(t) - B(t)$ 与 $|B(t)|$ 同分布

$$\begin{aligned} P(M(t) - B(t) \geq x) &= P\left(\max_{0 \leq s \leq t} \{B(s) - B(t)\} \geq x\right) \\ &= P\left(\max_{0 \leq s \leq t} \{B(s-t)\} \geq x\right) \\ &= P\left(\max_{0 \leq s \leq t} \{B(s)\} \geq x\right) = P(M(t) \geq x). \end{aligned}$$

对于 $x < 0$ 显然分布也相同。

下面求 $m(t) = \min_{0 \leq s \leq t} B(s)$ 的分布，对于 $x < 0$ ，有

$$P\left(\min_{0 \leq s \leq t} B(s) \leq x\right) = P(M(t) \geq -x) = \frac{2}{\sqrt{2\pi}} \int_{|x|/\sqrt{t}}^{\infty} e^{-\frac{y^2}{2}} dy.$$

PROBLEM C.0.11. 证明 (1) $B_t^2 - t$ 为鞅；(2) 对于任意实数 u , $\exp\left\{uB_t - \frac{u^2}{2}t\right\}$ 是鞅。

SOLUTION. (1) 要证明以下三个鞅的性质：

(a) 可积性：对于所有的 $t \geq 0$, $\mathbb{E}[|B_t^2 - t|] < \infty$.

$$\begin{aligned} B_t &\sim N(0, t) \Rightarrow \mathbb{E}[B_t^2] = \text{Var}(B_t) = t \\ \Rightarrow \mathbb{E}[|B_t^2 - t|] &\leq \mathbb{E}[B_t^2] + t = 2t \leq \infty. \end{aligned}$$

(b) 适应性： $B_t^2 - t$ 是 \mathcal{F}_t 可测的，其中 \mathcal{F}_t 是由 BM B_s 在时间 $s \leq t$ 生成的信息。这是由于 B_t 是 \mathcal{F}_t 可测，于是 $B_t^2 - t$ 也是 \mathcal{F}_t 可测。

(3) 鞅性质：对于所有的 $s \leq t$, $\mathbb{E}[B_t^2 - t | \mathcal{F}_s] = B_s^2 - s$.

$$B_t = B_s + (B_t - B_s)$$

其中 $B_t - B_s$ 独立于 \mathcal{F}_s 且 $B_t - B_s \sim N(0, t-s)$.

$$\begin{aligned} \mathbb{E}[B_t^2 | \mathcal{F}_s] &= \mathbb{E}[(B_s + (B_t - B_s))^2 | \mathcal{F}_s] \\ &= \mathbb{E}[B_s^2 + 2B_s(B_s - B_s) + (B_s - B_t)^2 | \mathcal{F}_s] \\ &= B_s^2 + 2B_s \mathbb{E}[B_t - B_s | \mathcal{F}_s] + \mathbb{E}[(B_t - B_s)^2 | \mathcal{F}_s] \\ &= B_s^2 + (t-s) \end{aligned}$$

因此 $\mathbb{E}[B_t^2 - t | \mathcal{F}_s] = B_s^2 - s$.

(2) $B_t \sim N(0, t)$ 的矩母函数 $\mathbb{E}[e^{uB_t}] = \exp\left\{t\frac{u^2}{2}\right\} < \infty \Rightarrow \exp\{uB_t\}$ 是可积的，并且 $\mathbb{E}\left[\exp\left\{uB_t - \frac{u^2}{2}t\right\}\right] = 1$. 取 $g(x) = e^{ux}$, 由于

$$\mathbb{E}[g(B_{t+s} - B_t)|\mathcal{F}_t] = \mathbb{E}[g(B_{t+s} - B_t)],$$

可得

$$\begin{aligned}\mathbb{E}\left[e^{uB_{t+s}}|\mathcal{F}_t\right] &= \mathbb{E}\left[e^{uB_t+u(B_{t+s}-B_t)}|\mathcal{F}_t\right] \\ &= e^{uB_t}\mathbb{E}\left[e^{u(B_{t+s}-B_t)}|\mathcal{F}_t\right] \\ &= e^{uB_t}\mathbb{E}\left[e^{u(B_{t+s}-B_t)}\right] = e^{uB_t+\frac{u^2}{2}s},\end{aligned}$$

等式两端同时乘以 $e^{-\frac{u^2}{2}(t+s)}$ 即可得证。

PROBLEM C.0.12. 解 SDE: $dX_t = -X_t dt + e^{-t} dB_t$.

SOLUTION. 令 $f(t, X_t) = e^t X_t$

$$\begin{aligned}df(t, X_t) &= e^t X_t dt + e^t dX_t \\ &= e^t X_t dt + e^t(-X_t dt + e^{-t} dB_t) \\ &= dB_t.\end{aligned}$$

因此 $f(t, X_t) = X_0 + B_t = e^t X_t, X_t = (X_0 + B_t)e^{-t}$. 验证解的正确性成立。

PROBLEM C.0.13. Black-Scholes 公式推导

一般地，带漂移和扩散的参数均可以是随机过程 $X(t)$ 以及时间 t 的函数。假设我们令 $a(t, X(t)), b(t, X(t))$ 分别表示漂移和扩散的参数，称满足如下 SDE 的随机过程为伊藤漂移扩散过程，简称伊藤过程：

$$dX(t) = a(t, X(t))dt + b(t, X(t))dB(t).$$

令 $f(t, X(t))$ 为 $X(t)$ 的二阶连续可导函数，并对 t 一阶可导，由伊藤引理可知：

$$df = f_t dt + f_x dX(t) + \frac{1}{2}f_{xx}(dX(t))^2,$$

代入伊藤过程并且略去所有比 dt 更高阶的小量，最终可以得到伊藤引理的一般形式：

$$df = \left(f_t + af_x + \frac{1}{2}b^2 f_{xx}\right) dt + bf_x dB(t).$$

$f(t, X(t))$ 也是伊藤过程。

GBM 求解

对于股票过程，可以用满足如下 SDE 的几何布朗运动来描述

$$dS_t = \mu S_t dt + \sigma S_t dB_t,$$

其中 μ 是股票的期望年收益率, σ 是股票年收益率的标准差, 这显然是一个伊藤过程。为了求解 S_t , 令 $f = \ln S_t$ 并对 f 使用伊藤引理可以得到 $\ln S_t$ 的 SDE:

$$d(\ln S_t) = df = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dB_t.$$

这个式子说明 $\ln S_t$ 是一个带漂移项的布朗运动, 由布朗运动的性质可知, 在任何时间 t , $\ln S_t$ 的变化服从正态分布

$$\ln S_t \sim N \left(\ln S_0 + \left(\mu - \frac{1}{2}\sigma^2 \right) t, \sigma^2 t \right).$$

如果一个随机变量的对数满足正态分布, 我们说这个随机变量本身满足对数正态分布。

通过对 $\ln S_t$ 的 SDE 两边积分再对等式两边取指数, 便可以很容易写出股价随时间变化的解析式

$$S_t = S_0 \exp \left(\left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma B_t \right).$$

Black-Scholes 微分方程

令 C 代表欧式看涨期权的价格, 显然它是标的股票价格 S_t 和时间 t 的函数, 记作 $C(S_t, t)$, 对 C 运用伊藤引理可得:

$$dC = \left(\frac{\partial C}{\partial S} \mu S + \frac{\partial C}{\partial t} + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma^2 S^2 \right) dt + \frac{\partial C}{\partial S} \sigma S dB$$

考虑下面这个投资组合:

$$\begin{aligned} & -1 : \text{option} \\ & + \frac{\partial C}{\partial S} : \text{stock} \end{aligned}$$

该组合做空一份期权并做多 $\frac{\partial C}{\partial S}$ 份股票, 将期权和股票的权重代入 C 和 S 在一个微小时间区间 Δt 内的变化可以验证, 布朗运动 ΔB 被对冲掉了。这种构建投资组合来消除随机性的方法成为 Delta 对冲。

用 P 表示该投资组合的价值, 则它在时间 Δt 内的变化为:

$$\Delta P = \left(-\frac{\partial C}{\partial t} - \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma^2 S^2 \right) \Delta t,$$

这个式子表明我们通过卖出 1 份期权并同时买入 $\frac{\partial C}{\partial S}$ 份股票在 Δt 内对冲了任何随机性的风险, 构建了一个无风险的投资组合。在不存在无风险套利的市场中, 该投资组合在 Δt 内的收益率必须等于无风险利率 r , 即 $\Delta P = r P \Delta t$.

将 ΔP 和 $P = -C + \frac{\partial C}{\partial S} S$ 代入上面的式子可以得到

$$\frac{\partial C}{\partial t} + r S \frac{\partial C}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} = r C,$$

这就是 Black-Scholes 微分方程, 是一个一般的 PDE 不含任何随机项。求解这个微分方程需要给定边界条件, 对于欧式看涨期权, 它的边界条件为当时间 $t = T$ (行权时), 期权的价格 C 必须满足 $C = \max(S(T) - K, 0)$, 这里的 K 是行权价格。

Black-Scholes 期权定价公式

欧式看涨期权在行权日 T 的期望价值为 $\mathbb{E}[\max(S(T) - K, 0)]$, 其中 $S(T)$ 为股票在 T 时刻的价格, K 为行权价。股价 S 满足对数正态分布, 在风险中性定价理论下, S 的期望收益率为无风险收益率 r 且期权的折现率也为 r . 因此, 期权在当前时刻的价格 C 为:

$$C = e^{-rT} \mathbb{E}[\max(S(T) - K, 0)].$$

根据对数正态分布的性质可以计算出 $\mathbb{E}[\max(S(T) - K, 0)]$, 从而得到 BS 期权定价公式

首先, 将 C 写作

$$C = e^{-rT} \mathbb{E}[(S_T - K) \mathbf{1}_{S_T > K}] = e^{-rT} (\mathbb{E}[S_T \mathbf{1}_{S_T > K}] - K \mathbb{E}[\mathbf{1}_{S_T > K}]).$$

计算 $\mathbb{E}[\mathbf{1}_{S_T > K}]$:

$$\ln S_T > \ln K \Leftrightarrow \frac{\ln S_T - \mu}{s} > \frac{\ln K - \mu}{s}, \text{ where } \mu = \ln S_0 + \left(r - \frac{1}{2}\sigma^2\right)T, s = \sigma\sqrt{T}$$

定义:

$$d_2 = \frac{\mu - \ln K}{s} = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$\mathbb{E}[\mathbf{1}_{S_T > K}] = \mathbb{P}(S_T > K) = \mathbb{P}\left(\frac{\ln S_T - \mu}{s} > \frac{\ln K - \mu}{s}\right) = \mathbb{P}(Z < d_2) = N(d_2),$$

其中 N 是标准正态的累积分布函数。

计算 $\mathbb{E}[S_T \mathbf{1}_{S_T > K}]$, 令 $S_T = e^Y, Y = \ln S_T \sim N(\mu, s^2)$.

$$\begin{aligned} \mathbb{E}[S_T \mathbf{1}_{S_T > K}] &= \int_{\ln K}^{\infty} e^y \frac{1}{\sqrt{2\pi}s} e^{-\frac{(y-\mu)^2}{2s^2}} dy = \int_{\ln K}^{\infty} \frac{1}{\sqrt{2\pi}s} e^{y-\frac{(y-\mu)^2}{2s^2}} dy \\ &= \int_{\ln K}^{\infty} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(y-(\mu+s^2))^2}{2s^2} + \mu + \frac{s^2}{2}\right) dy \\ &= S_0 e^{rT} \int_{\ln K}^{\infty} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(y-(\mu+s^2))^2}{2s^2}\right) dy \end{aligned}$$

其中积分项是正态分布 $N(\mu + s^2, s^2)$ 在 $(\ln K, \infty)$ 上的概率, 令

$$d_1 = \frac{(\mu + s^2) - \ln K}{s} = \frac{\mu - \ln K}{s} + s = d_2 + \sigma\sqrt{T} = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

则

$$\int_{\ln K}^{\infty} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(y-(\mu+s^2))^2}{2s^2}\right) dy = \mathbb{P}\left(Z > \frac{\ln K - (\mu + s^2)}{s}\right) = \mathbb{P}(Z < d_1) = N(d_1).$$

因此,

$$C = e^{-rT} (S_0 e^{rT} N(d_1) - K N(d_2)) = S_0 N(d_1) - K e^{-rT} N(d_2).$$

看跌期权价格 P 为

$$P = K e^{-rT} N(-d_2) - S_0 N(-d_1).$$

PROBLEM C.0.14. 令 X_n 表示分支过程中第 n 代的个体数, 每一个体产生后代数量分布具有均值 μ 和方差 σ^2 , 我们已经知道 $\{M_n = \mu^{-n} X_n\}$ 是鞅。

1. 令 \mathcal{F}_n 表示 X_0, \dots, X_n 生成的 σ 代数，证明：

$$\mathbb{E}[X_{n+1}^2 | \mathcal{F}_n] = \mu^2 X_n^2 + \sigma^2 X_n.$$

2. 设 $\mu > 1$, 证明存在 $C < \infty$ 使得对所有的 n , 有

$$\mathbb{E}[M_n^2] < C,$$

且 $\mu \leq 1$ 时上式不成立。

D 机器学习基本方法

PROBLEM D.0.1. 对于高维稀疏变量列举出三种降维方法并说明其原理

SOLUTION.

(1) 主成分分析：PCA 通过线性变换将原始高维数据投影到低维空间，使得投影后的数据方差最大化。具体步骤如下

- 计算数据的协方差矩阵
- 对协方差矩阵进行特征值分解
- 选择前 k 个最大特征值对应的特征向量作为主成分方向
- 将原始数据投影到这些主成分方向上

适用于变量间存在线性相关性的数据，能有效保留数据的主要变异信息。

(2) 线性判别分析：LDA 是一种有监督的降维方法，目标是找到能够最大化类间距离、最小化类内距离的投影方向。具体步骤：

- 计算类间散度矩阵和类内散度矩阵
- 求解广义特征值问题得到最佳投影方向
- 将数据投影到这些方向上

适用于分类问题，需要类别标签信息，能提升分类性能。

(3) 奇异值分解：SVD 将任意矩阵分解为三个矩阵的乘积 $A = U\Sigma V^\top$, 其中 U, V 是正交矩阵， Σ 是对角矩阵。降维时

- 对数据矩阵进行 SVD 分解
- 保留前 k 个最大的奇异值及其对应的左右奇异向量

- 用这些奇异向量构成低维表示

特别适合处理稀疏矩阵，广泛应用于推荐系统、文本挖掘等领域。

对比：*PCA* 是无监督方法，*LDA* 是有监督方法，*SVD* 对稀疏数据更高效。选择时需考虑是否有标签信息、数据稀疏程度和计算效率要求。

PROBLEM D.0.2. (1) 介绍线性可分支持向量机学习算法（最大间隔法），(2) 如何将这个算法扩展到线性不可分训练数据？证明 (3) 中的原始最优化问题等价于合页损失函数最优化问题。

PROBLEM D.0.3.

- 什么是核函数？核函数具有哪些性质？核函数在模型中的应用有哪些？SVM 中核函数的作用。
- 写出三种核函数及其优劣。

1. 核函数 (Kernal Function) 是计算两个样本在某个特征空间中内积的函数，而不需要显式地计算这个特征空间映射。设 $\phi : \mathcal{X} \rightarrow \mathcal{H}$ 是从输入空间到特征空间的映射，则核函数定义为

$$K(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}},$$

其中 \mathcal{H} 是一个再生核 Hilbert 空间 (RKHS)。

核函数的性质：

- 对称性： $K(x, x') = K(x', x)$
- 半正定性：对任意 n 个样本 x_1, \dots, x_n 和任意实系数 c_1, \dots, c_n ，有

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(x_i, x_j) \geq 0.$$

- 再生性：在 RKHS 中， $f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}}$ 。

核函数在模型中的应用：

- 支持向量机：将线性不可分问题转化为高维特征空间中的线性可分问题。
- 核主成分分析：在高维特征空间进行主成分分析
- 核岭回归：非线性回归问题的核化版本。
- 核聚类：如谱聚类中使用核函数

SVM 中核函数的作用主要是为了避免维度灾难：

- 隐式映射：通过核技巧，SVM 可以在高维甚至无限维特征空间中构造最优超平面，无需显示计算 $\phi(x)$ 。
- 计算效率：直接计算 $K(x_i, x_j)$ 通常比先计算 $\phi(x_i), \phi(x_j)$ 再求内积效率更高。
- 使线性 SVM 可以处理非线性分类问题。

(4) 对偶问题: SVM 的对偶问题只涉及样本间的内积, 自然适合核化:

$$\begin{aligned} \max_{\alpha} & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t. } & 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

三种常见核函数及其优劣:

(1) 线性核函数:

$$K(x, x') = x^\top x',$$

优点是计算高效, 时间复杂度 $O(d)$, d 为特征维数, 无需调参, 可解释性强, 决策边界是线性的, 不易过拟合。缺点是只能处理线性可分问题, 对非线性模式表达能力有限, 适用高维场景, 样本数量大, 问题近似线性可分的情况。

(2) 多项式核函数:

$$K(x, x') = (\gamma x^\top x' + r)^d$$

其中 d 为多项式次数, γ 为尺度参数, r 为偏置项。

优点是可以控制非线性程度, 对某些特定问题表现良好, 缺点是当 d 较大时数值不稳定, 且三个参数调参较复杂, 可能产生过拟合。适用场景时问题具有明确的阶数结构。

(3) 高斯径向基 (RBF) 核函数:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2),$$

优点是强大的非线性映射能力, 可以逼近任何连续函数, 调参简单, 数值稳定性好, 缺点是容易产生过拟合, 可解释性差, γ 较大时决策边界复杂可能产生过拟合, γ 较小时可能欠拟合。没有先验知识时默认选择, 特别是处理复杂的非线性模式。

倾向得分匹配

倾向得分匹配是一种用于因果推断的统计方法, 通过构建处理组和对照组的可比样本来估计处理效应。

PSM 主要用于观察性研究 (非随机试验) 中, 当无法进行随机分组时用于估计干预措施 (如政策、治疗、教育项目等) 的因果效应。常见应用场景包括:

- 医学研究: 评估新药疗效 (无法对患者随机分组)
- 政策评估: 分析扶贫政策、教育补贴等政策效果

分析步骤为

1. 定义处理变量 ($D = 1$ 为处理组, $D = 0$ 为对照组); 确定结果变量 Y , 即要评估的效应; 选择协变量 X , 影响处理分配和结果的所有可观测变量。

2. 估计倾向得分：使用 Logit 或 Probit 模型估计每个个体接受处理的概率

$$P(D = 1|X) = e^{\beta x} / (1 + e^{\beta X})$$

其中 X 为协变量向量， β 为估计系数

3. 匹配处理组与对照组：常用匹配方法有最近邻匹配（为每个个处理组个体寻找倾向得分最接近的对照组个体）、卡尺匹配（在倾向得分的一定范围内寻找匹配）、核匹配（使用核函数对多个对照组个体加权）
4. 平衡性检验。检验匹配后处理组与对照组的协变量是否平衡：标准化差异应小于 10%，t 检验或者卡方检验应不显著，否则需要重新选择协变量或者调整匹配方法。
5. 估计处理效应。在匹配样本上计算平均处理效应 (ATE) 或处理组的平均处理效应 (ATT)：

$$ATT = \mathbb{E}[Y(1)|D = 1] - \mathbb{E}[Y(0)|D = 1]$$

其中 $Y(1)$ 为接受处理的结果， $Y(0)$ 为未接受处理的结果。

6. 敏感性分析。检验结果对未观测混杂因素的稳健性，常用方法包括 Rosenbaum 边界分析。

BP 神经网络中的反向传播

反向传播 (Backpropagation) 是通过链式法则计算损失函数对网络参数的梯度，用于神经网络的权重更新。

反向传播参数更新步骤

1. 前向传播：

$$z_j^l = \sum_i w_{ji}^l a_i^{l-1} + b_j^l, \quad a_j^l = \sigma(z_j^l)$$

2. 计算损失函数（以均方误差为例）：

$$L = \frac{1}{2} \sum_k (y_k - a_k^L)^2$$

3. 反向传播误差：

- 输出层误差： $\delta_j^L = \frac{\partial L}{\partial a_j^L} \sigma'(z_j^L)$
- 隐藏层误差： $\delta_j^l = \sum_k w_{kj}^{l+1} \delta_k^{l+1} \sigma'(z_j^l)$

参数更新公式推导

权重更新:

$$\frac{\partial L}{\partial w_{ji}^l} = \frac{\partial L}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{ji}^l} = \delta_j^l a_i^{l-1}$$

偏置更新:

$$\frac{\partial L}{\partial b_j^l} = \frac{\partial L}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} = \delta_j^l$$

梯度下降更新:

$$w_{ji}^l \leftarrow w_{ji}^l - \eta \frac{\partial L}{\partial w_{ji}^l}, \quad b_j^l \leftarrow b_j^l - \eta \frac{\partial L}{\partial b_j^l}$$

其中 η 为学习率。

支持向量机 (SVM)

核心思想: 寻找最优超平面, 使分类间隔最大化。

硬间隔 SVM 原始问题:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1, \quad i = 1, \dots, n \end{aligned}$$

软间隔 SVM (引入松弛变量):

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned}$$

对偶问题 (使用核技巧):

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned}$$

其中 $K(x_i, x_j)$ 为核函数。

随机森林

核心思想: 通过 Bagging 集成多棵决策树, 提高泛化能力。

构建过程:

1. 从原始数据集中进行 B 次 Bootstrap 抽样
2. 对每个 bootstrap 样本构建决策树, 节点分裂时从随机选择的 m 个特征中找最优分裂

3. 集成所有树的预测结果

分类投票公式：

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_B(x)\}$$

回归平均公式：

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

特征重要性计算：

$$\text{Importance}(X_j) = \frac{1}{B} \sum_{b=1}^B \text{DecreaseImpurity}(X_j, T_b)$$

其中特征 j 的杂质减少量为：

$$\text{DecreaseImpurity} = I(\text{node}) - \frac{N_{\text{left}}}{N} I(\text{left}) - \frac{N_{\text{right}}}{N} I(\text{right})$$