

机器学习导论作业 1

2020 年 10 月 4 日

1

(1)

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2}x, & 0 < x \leq 1 \\ \frac{1}{2}, & 1 < x \leq 2 \\ \frac{1}{6}(x+1), & 2 < x \leq 5 \\ 1, & x > 5 \end{cases} \quad (1)$$

(2)

$$\begin{aligned} F_Y(y) &= P(\frac{1}{X^2} \leq y) = P(X > \frac{1}{\sqrt{y}}) + P(X < -\frac{1}{\sqrt{y}})(y > 0) \\ &= F_X(-\frac{1}{\sqrt{y}}) + 1 - F_X(\frac{1}{\sqrt{y}}) = 1 - F_X(\frac{1}{\sqrt{y}}) \end{aligned}$$

$$F_Y(y) = \begin{cases} 1 - \frac{1}{6}(\frac{1}{\sqrt{y}} + 1), & \frac{1}{25} \leq y < \frac{1}{4} \\ \frac{1}{2}, & \frac{1}{4} \leq y < 1 \\ 1 - \frac{1}{2\sqrt{y}}, & y \geq 1 \\ 0, & \text{else} \end{cases} \quad (2)$$

$$f_Y(y) = F_Y(y)' = f_X\left(\frac{1}{\sqrt{y}}\right) * \left(\frac{1}{2}\right)y^{-\frac{3}{2}}$$

$$f_Y(y) = \begin{cases} \frac{1}{4}y^{-\frac{3}{2}}, & y > 1 \\ \frac{1}{12}y^{-\frac{3}{2}}, & \frac{1}{25} \leq y \leq \frac{1}{4} \\ 0 & else \end{cases} \quad (3)$$

(3)

$$\begin{aligned} E(Z) &= \int_{z=0}^{\infty} Pr\{Z \geq z\}dz = \int_{z=0}^{\infty} (1 - F_Z(z))dz \\ E(Z) &= \int_{z=0}^{\infty} zf(z)dz = \int_{z=0}^{\infty} z dF_Z(z) \\ &= zF_Z(z)|_{z=0}^{\infty} - zF_Z(z)|_{z=0} - \int_{z=0}^{\infty} F_Z(z)dz \\ &= z|_{z=\infty} - \int_{z=0}^{\infty} F_Z(z)dz = \int_{z=0}^{\infty} dz - \int_{z=0}^{\infty} F_Z(z)dz \\ &= \int_{z=0}^{\infty} (1 - F_Z(z))dz \end{aligned}$$

得证

$$\begin{aligned} \text{验证: } E(X) &= \int_0^{\infty} (1 - F_X(x))dx = \int_0^1 (1 - \frac{1}{2}x)dx + \int_1^2 \frac{1}{2}dx + \int_2^5 \frac{1}{6}(5 - x)dx \\ &= \frac{3}{4} + \frac{1}{2} + \frac{5}{2} - \frac{7}{4} = 2 \\ E(X) &= \int_0^{\infty} xf(x)dx = \int_0^1 \frac{1}{2}xdx + \int_2^5 \frac{1}{6}xdx = \frac{1}{4} + \frac{7}{4} = 2 \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_0^{\infty} (1 - F_Y(y))dy = \int_{\frac{1}{25}}^{\frac{1}{4}} \frac{1}{6}\left(\frac{1}{\sqrt{y}} + 1\right)dy + \int_{\frac{1}{4}}^1 \frac{1}{2}dy + \int_1^{\infty} \frac{1}{2}y^{-\frac{1}{2}}dy + \frac{1}{25} \\ &= \frac{27}{200} + \frac{3}{8} + \int_1^{\infty} \frac{1}{2}y^{-\frac{1}{2}}dy + \frac{1}{25} = \frac{11}{20} + \int_1^{\infty} \frac{1}{2}y^{-\frac{1}{2}}dy \\ &= y^{\frac{1}{2}}|_{y=\infty} - \frac{9}{20} \\ E(Y) &= \int_0^{\infty} yf_Y(y)dy = \int_1^{\infty} \frac{1}{4}y^{-\frac{1}{2}}dy + \int_{\frac{1}{25}}^{\frac{1}{4}} \frac{1}{12}y^{-\frac{1}{2}}dy \\ &= \frac{1}{2}y^{\frac{1}{2}}|_{y=\infty} - \frac{9}{20} \end{aligned}$$

E(Y) 不存在, 无法验证

2

(1)

A=rain today,B=rain tomorrow

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.25}{0.3} = \frac{5}{6}$$

(2)

$$\begin{aligned} P(G|\neg H) &= \frac{P(G\neg H)}{P(\neg H)} = \frac{P(\neg H|G)P(G)}{1-P(H)} = \frac{(1-P(H|G))P(G)}{1-P(H)} \\ &= \frac{P(G)-P(HG)}{1-P(H)} \end{aligned}$$

(3)

令 A 代表第一个球是白色,B 代表第一个球是黑色,C 代表第二个球是白色

$$\begin{aligned} P(C) &= P(C|A)P(A) + P(C|B)P(B) = \frac{w}{w+b} * \frac{w+d-1}{w+b+d-1} + \frac{b}{w+b} * \frac{w}{w+b+d-1} = \\ &\frac{w}{w+b} \end{aligned}$$

得证

3

(1)

由于 x_{\perp} 是在 y 上的投影, 可以令 $x_{\perp} = (a, \sqrt{3}a)^T$

由 $xy = x_{\perp}y$

$$2\sqrt{3} = a + 3a \rightarrow a = \frac{\sqrt{3}}{2}$$

$$x_{\perp} = (\frac{\sqrt{3}}{2}, \frac{3}{2})^T$$

(2)

$$x - x_{\perp} = (\frac{\sqrt{3}}{2}, -\frac{1}{2})^T$$

$$y(x - x_{\perp}) = 0$$

因此 $y \perp (x - x_{\perp})$

(3)

$$\|x - x_{\perp}\| = 1$$

$$\begin{aligned} \|x - \lambda y\| &= |(\sqrt{3} - \lambda, 1 - \sqrt{3}\lambda)^T| = \sqrt{3 - 2\sqrt{3}\lambda + \lambda^2 + 1 - 2\sqrt{3}\lambda + 3\lambda^2} \\ &= 2\sqrt{\lambda^2 - \sqrt{3}\lambda + 1} = 2\sqrt{(\lambda - \frac{\sqrt{3}}{2})^2 + \frac{1}{4}} \geq 1 \end{aligned}$$

得证

4

原假设 $H_0: \mu_0 \leq 0.5$

令 $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} t(n-1)$

拒绝域为 $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_\alpha(n-1)$

$$S = \sqrt{\frac{1}{49}(0.7^2 * 15 + 0.3^2 * 35)} = 0.463$$

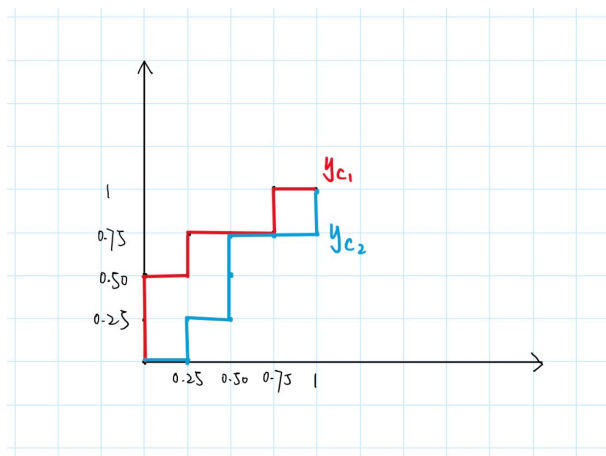
$$\text{而 } T = \frac{0.7 - 0.5}{0.463/\sqrt{7}} = 3.024, t_{0.05}(49) = 1.676$$

T 落在拒绝域内, 原假设被拒绝

因此可以认为 $\mu > 0.5$, 也即硬币不均匀偏向人头的一边

5

(1)



$$AUC_{C_1} = 0.75, AUC_{C_2} = 0.4375$$

(2)

$C_1:$	真实类别/预测类别	1	0
	1	3	1
	0	1	3

$$P = \frac{TP}{TP+FP} = \frac{3}{4}, R = \frac{TP}{TP+FN} = \frac{3}{4}$$

$$F_1 = \frac{2PR}{P+R} = \frac{3}{4}$$

	真实类别/预测类别	1	0
$C_2:$	1	3	1
	0	3	1

$$P = \frac{TP}{TP+FP} = \frac{1}{2}, R = \frac{TP}{TP+FN} = \frac{3}{4}$$

$$F_1 = \frac{2PR}{P+R} = \frac{3}{5}$$

$$P = \frac{TP}{TP+FP} = \frac{1}{2}, R = \frac{TP}{TP+FN} = \frac{3}{4}$$

$$F_1 = \frac{2PR}{P+R} = \frac{3}{5}$$