信息学中的概率统计：期末考试答案

第一题

(1)

$$\frac{dQ}{d\beta} = 2\sum_{i=1}^{n} x_i(y_i - \beta x_i)。$$

令 $\frac{dQ}{d\beta} = 0$，得

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}。$$

(2)

$$E(\hat{\beta}) = \frac{\sum_{i=1}^{n} x_i E(y_i)}{\sum_{i=1}^{n} x_i^2} = \beta。$$

因此 $\hat{\beta}$ 的偏差为 0。对于无偏估计量，有

$$\mathrm{MSE}(\hat{\beta}) = \mathrm{Var}(\hat{\beta}) = \frac{\sum_{i=1}^{n} x_i^2\sigma^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}。$$

第二题

(1) $E(X_i) = \frac{1}{\lambda}$，$\mathrm{Var}(X_i) = \frac{1}{\lambda^2}$。因此 $E(Y) = \frac{n}{\lambda}$，$\mathrm{Var}(Y) = \frac{n}{\lambda^2}$。

(2) 由切比雪夫不等式，

$$P(Y \geq a) \leq P(|Y - n/\lambda| \geq a - n/\lambda) \leq \frac{\mathrm{Var}(Y)}{(a - n/\lambda)^2} = \frac{n}{(\lambda a - n)^2}。$$

(3) 注意到 $Y \sim \Gamma(n, \lambda)$。根据作业三第六题，当 $t \geq \lambda$，$M_Y(t)$ 不存在。当 $t < \lambda$，

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n。$$

另一种做法：注意到

$$M_{X_i}(t) = \int_0^\infty e^{-\lambda x} \cdot e^{xt} dx = \int_0^\infty e^{-(\lambda - t)x} dx。$$

当 $t < \lambda$，有 $M_{X_i}(t) = \frac{\lambda}{\lambda - t}$。当 $t \geq \lambda$，期望不存在。因此，当 $t < \lambda$，

$$M_Y(t) = e^{Yt} = e^{t \cdot \sum_{i=1}^{n} X_i} = \prod_{i=1}^{n} e^{X_i t} = \prod_{i=1}^{n} M_{X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n。$$

当 $t \geq \lambda$，$M_Y(t)$ 不存在。

(4) 根据 Chernoff bound，当 $t > 0$，

$$P(Y \geq a) = P(e^{Yt} \geq e^{at}) \leq M_Y(t)e^{-at} = \left(\frac{\lambda}{\lambda - t}\right)^n \cdot e^{-at}。$$

带入 $t = \lambda - n/a$，注意到 $0 < t < \lambda$，有

$$P(Y \geq a) \leq (\lambda a/n)^n e^{-(\lambda a - n)}。$$

第三题

(1) 与作业五第二题第一问一致，对于任意 $0 < u < 1$，

$$f_U(u) = \int_0^1 f_X(x) f_Y(u - x) dx = u。$$

对于任意 $1 < u < 2$，

$$f_U(u) = \int_0^1 f_X(x) f_Y(u - x) dx = 2 - u。$$

(2) 注意到 $x = v/2$，$y = u - v/2$。因此 $\frac{\partial x}{\partial u} = 0$，$\frac{\partial x}{\partial v} = 1/2$，$\frac{\partial y}{\partial u} = 1$，$\frac{\partial y}{\partial v} = -1/2$。

$$|J| = \left| \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right| = 1/2。$$

因此，

$$f_{U,V}(u, v) = \begin{cases} 1/2 & u \in (0, 2), v \in [\max\{0, 2(u - 1)\}, \min\{2, 2u\}] \\ 0 & \text{其余情况} \end{cases}。$$

(3) 当 $0 < u < 1$，

$$f(v \mid u) = \frac{f_{U,V}(u, v)}{f_U(u)} = \begin{cases} \frac{1}{2u} & v \in (0, 2u) \\ 0 & \text{其他情况} \end{cases}。$$

当 $1 < u < 2$，

$$f(v \mid u) = \frac{f_{U,V}(u, v)}{f_U(u)} = \begin{cases} \frac{1}{2(2-u)} & v \in (2(u - 1), 2) \\ 0 & \text{其他情况} \end{cases}。$$

(4)

$$\mathrm{Cov}(U, V) = E(UV) - E(U)E(V) = E(2X^2) + E(2XY) - 2(E(X) + E(Y))E(X) = 2E(X^2) - 2(E(X))^2 = \frac{1}{6}。$$

$$\sigma(U) = \sqrt{\mathrm{Var}(X) + \mathrm{Var}(Y)} = \frac{1}{\sqrt{6}}, \sigma(V) = \sqrt{4\mathrm{Var}(X)} = \frac{1}{\sqrt{3}},$$

$$\mathrm{Corr}(U, V) = \frac{\mathrm{Cov}(U, V)}{\sigma(U)\sigma(V)} = \frac{1}{\sqrt{2}}。$$

第四题

(1) 根据提示，以及服从自由度为 2 的卡方分布的随机变量的期望为 2，有 $E(X^2)=\theta$。因此

$$\hat{\theta}_1 = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,\ E(\hat{\theta}_1) = \frac{1}{n}\cdot n\cdot\theta = \theta。$$

因此 $\hat{\theta}_1$ 是无偏估计量以及渐进无偏估计量。

(2)

$$\ln L(\theta) = \sum_{i=1}^{n}\ln(2x_i) - n\ln\theta - \frac{\sum_{i=1}^{n} x_i^2}{\theta}。$$

求导，得

$$\frac{dL}{d\theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i^2}{\theta^2}。$$

因此，

$$\hat{\theta}_2 = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\theta}_1。$$

(3) 由于 $\hat{\theta}_2$ 是无偏估计量，有 $\mathrm{MSE}(\hat{\theta}_2)=\mathrm{Var}(\hat{\theta}_2)$。根据提示，以及服从自由度为 2 的卡方分布的随机变量的方差为 4，有 $\mathrm{Var}(X_i^2)=\theta^2\cdot 4/4=\theta^2$。因此，

$$\mathrm{MSE}(\hat{\theta}_2) = \mathrm{Var}(\hat{\theta}_2) = \frac{\theta^2}{n}。$$

由于 $\lim_{n\to\infty}\mathrm{MSE}(\hat{\theta}_2)\to 0$，因此 $\hat{\theta}_2$ 为一致估计量。也可以使用大数定律判断一致性。

(4) 令枢轴量

$$G = \frac{2\sum_{i=1}^{n} X_i^2}{\theta},$$

根据提示有 $G\sim\chi^2(2n)$。令 $F(x)$ 为 $\chi^2(2n)$ 分布的分布函数，则有

$$P(F^{-1}(\alpha/2)\le G\le F^{-1}(1-\alpha/2)) = 1-\alpha,$$

因此

$$P\left(\frac{2\sum_{i=1}^{n} X_i^2}{F^{-1}(1-\alpha/2)}\le\theta\le\frac{2\sum_{i=1}^{n} X_i^2}{F^{-1}(\alpha/2)}\right) = 1-\alpha。$$

第五题

(1) 令 $f(x)$ 为 Y_n 的概率密度函数。

$$\begin{aligned}
&\int_0^{\infty} P(Y_n > x)dx - \int_{-\infty}^{0} P(Y_n < x)dx \\
=&\int_0^{\infty}\int_x^{\infty} f(y)dydx - \int_{-\infty}^{0}\int_{-\infty}^{x} f(y)dydx \\
=&\int_0^{\infty} f(y)\int_0^{y} dxdy - \int_{-\infty}^{0} f(y)\int_y^{0} dxdy \\
=&\int_0^{\infty} f(y)ydy + \int_{-\infty}^{0} f(y)ydy \\
=&E(Y_n)
\end{aligned}$$

(2) 根据 Chernoff bound 或作业三第五题，

$$P(X_i \geq 10\sqrt{\ln n}) \leq e^{-100\ln n/2} \leq \frac{1}{n^2}。$$

根据 Union bound，

$$P(\max X_i \geq 10\sqrt{\ln n}) \leq 1/n。$$

(3) 根据第一问

$$E(Y_n) \leq \int_0^{10\sqrt{\ln n}} P(Y_n > x)dx + \int_{10\sqrt{\ln n}}^{\infty} P(Y_n > x)dx。$$

注意到，对于任意 $x > 0$，$P(Y_n > x) < 1$，因此第一段最多为 $10\sqrt{\ln n}$。对于第二段积分，根据 Union bound，$P(Y_n > x) \leq nP(X_i > x)$，因此有

$$\int_{10\sqrt{\ln n}}^{\infty} P(Y_n > x)dx \leq n \cdot \int_{10\sqrt{\ln n}}^{\infty} P(X_i > x)dx。$$

根据 Chernoff bound 或作业三第五题，

$$\int_{10\sqrt{\ln n}}^{\infty} P(X_i > x)dx \leq \int_{10\sqrt{\ln n}}^{\infty} e^{-x^2/2}dx = P(X_i \geq 10\sqrt{\ln n}) \cdot \sqrt{2\pi}。$$

根据第二问，有

$$\int_{10\sqrt{\ln n}}^{\infty} P(Y_n > x)dx \leq n \cdot \frac{\sqrt{2\pi}}{n^2} = \frac{\sqrt{2\pi}}{n}。$$

因此，

$$E(Y_n) \leq 10\sqrt{\ln n} + \frac{\sqrt{2\pi}}{n}。$$

另一种（不在课程范围内的）做法，

$$E(e^{Y_n t}) = E(e^{t\cdot\max\{X_1,X_2,\ldots,X_n\}}) \leq E\left(\max_{1\leq i\leq n} e^{tX_i}\right) \leq E\left(\sum_{i=1}^{n} e^{tX_i}\right) \leq n \cdot E\left(e^{tX_i}\right) = ne^{t^2/2}。$$

根据琴生不等式，

$$e^{E(Y_n t)} \leq ne^{t^2/2},$$

因此，当 $t > 0$，有

$$E(Y_n) \leq \frac{\ln n + t^2/2}{t}。$$

取 $t = 10\sqrt{\ln n}$ 即可。

(4) 注意到

$$E(Y_n) \geq \int_0^{0.1\sqrt{\ln n}} P(Y_n > x)dx - \int_{-\infty}^{0} P(Y_n < x)dx。$$

对于第一项，可以使用下面的粗糙估计

$$P(X_i \geq 0.1\sqrt{\ln n}) = \frac{1}{\sqrt{2\pi}} \int_{0.1\sqrt{\ln n}}^{0.1\sqrt{\ln n}+1} e^{-x^2/2} \geq \frac{1}{\sqrt{2\pi}} \cdot e^{-\ln n} = \frac{1}{\sqrt{2\pi}n} \geq \frac{1}{3n}。$$

(上述不等式的另一种证明方法是使用作业三第五题，得到

$$P(X_i \geq 0.1\sqrt{\ln n}) \geq \frac{0.1\sqrt{\ln n}}{0.01\ln n + 1} \cdot e^{-0.005\ln n}/\sqrt{2\pi},$$

而对于充分大的 n，上式不等号右边至少是 $\frac{1}{3n}$。）

因此，

$$P(Y_n \geq 0.1\sqrt{\ln n}) \geq 1 - \left(1 - \frac{1}{3n}\right)^n \geq 1 - 1/e^{1/3},$$

也即

$$\int_0^{0.1\sqrt{\ln n}} P(Y_n > x)dx \geq 0.1\sqrt{\ln n} \cdot P(Y_n > 0.1\sqrt{\ln n}) \geq (1 - 1/e^{1/3}) \cdot 0.1\sqrt{\ln n}。$$

另外，对于第二段积分，

$$\int_{-\infty}^0 P(Y_n < x)dx = \int_{-\infty}^0 (P(X_i < x))^n \, dx \leq \int_{-\infty}^0 P(X_i < x)dx = \int_0^\infty P(X_i > x)dx \leq \int_0^\infty e^{-x^2/2}dx = \sqrt{\frac{\pi}{2}}。$$

（注意上面的估计非常粗糙，事实上可以证明

$$\int_{-\infty}^0 P(Y_n < x)dx \leq \sqrt{\frac{\pi}{2n}},$$

但即使上面的粗糙估计也足够证明本题需要的结论。）

因此

$$E(Y_n) \geq (1 - 1/e^{1/3}) \cdot 0.1\sqrt{\ln n} - \sqrt{\frac{\pi}{2}}。$$