

Probability and Statistics

PhD Qualification Exam

Mock Examination No. 3 – Solutions

Problem 1

10 points

由标准正态分布关于零对称, $\varepsilon Z \stackrel{d}{=} Z$, 所以 X, Y 的边缘均为 $N(0, 1)$ 。且

$$\text{Cov}(X, Y) = \mathbb{E}(\varepsilon Z^2) = \mathbb{E}\varepsilon \mathbb{E}Z^2 = 0.$$

然而 $X^2 = Y^2$ a.s.; 若两个连续变量独立, 则事件 $X^2 = Y^2$ 的概率为零。因此它们不独立。若 (X, Y) jointly Gaussian, 则零协方差会蕴含独立, 产生矛盾, 所以该向量不 jointly Gaussian。

给定 $Y = y$, 由正态对称性, 两种可能 $(\varepsilon, Z) = (1, y)$ 与 $(-1, -y)$ 的条件权重相同。因此

$$\text{Law}(X | Y = y) = \frac{1}{2}\delta_y + \frac{1}{2}\delta_{-y},$$

从而

$$\mathbb{E}(X | Y) = 0, \quad \mathbb{E}(X^2 | Y) = Y^2.$$

Scoring guide (10 points). 两个边缘律 2 分、协方差 1 分; 不独立 2 分、非联合高斯 1 分; 条件二点律 2 分、两个条件矩 2 分。

Problem 2

10 points

由 i.i.d. central limit theorem,

$$\frac{\sum_i X_i}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1).$$

由 strong law of large numbers (弱大数律也足够),

$$\frac{1}{n} \sum_i X_i^2 \xrightarrow{a.s.} \mathbb{E}X_1^2 = \sigma^2.$$

故分母为零的概率趋于零, 并且

$$T_n = \frac{\sum_i X_i / (\sigma\sqrt{n})}{\sqrt{\sum_i X_i^2 / (n\sigma^2)}} \xrightarrow{d} N(0, 1)$$

由 Slutsky theorem 成立。

若 $\mathbb{E}X_1 = \mu \neq 0$, 则两个 LLN 分别给出

$$\frac{T_n}{\sqrt{n}} = \frac{n^{-1} \sum_i X_i}{\sqrt{n^{-1} \sum_i X_i^2}} \xrightarrow{a.s.} \frac{\mu}{\sqrt{\mathbb{E}X_1^2}}.$$

Scoring guide (10 points). CLT 2 分、平方项 LLN 2 分、分母零事件 1 分、Slutsky 及结论 2 分; 非零均值下两个 LLN 与极限 3 分。

Problem 3**15 points**

令 $q = 1 - p$ 。样本 likelihood 为

$$L(p; x) = p^n q^{\sum_i x_i},$$

所以 factorization theorem 给出 $T = \sum_i X_i$ 充分。又

$$\mathbb{P}_p(T = t) = \binom{n+t-1}{t} p^n q^t, \quad t = 0, 1, \dots$$

若 $\mathbb{E}_p g(T) = 0$ 对所有 $0 < p < 1$ 成立, 则除以 p^n 后,

$$\sum_{t=0}^{\infty} g(t) \binom{n+t-1}{t} q^t = 0, \quad 0 < q < 1.$$

在绝对可积所保证的收敛域内, 幂级数唯一性迫使每个系数为零, 故 $g(T) = 0$ a.s., 即 T 完全。定义

$$\delta(T) = \frac{n-1}{n+T-1}.$$

利用恒等式

$$\frac{n-1}{n+t-1} \binom{n+t-1}{t} = \binom{n+t-2}{t},$$

可得

$$\mathbb{E}_p \delta(T) = p^n \sum_{t \geq 0} \binom{n+t-2}{t} q^t = p^n (1-q)^{-(n-1)} = p.$$

δ 是 complete sufficient statistic 的无偏函数, 故由 Lehmann–Scheffé theorem, 它是 p 的 UMVU estimator。

当 $0 < p < 1$ 时, log-likelihood 为 $n \log p + T \log(1-p)$ 。在题设扩展参数空间 $(0, 1]$ 上,

$$\hat{p} = \frac{n}{n+T},$$

其中 $T = 0$ 时 $\hat{p} = 1$, 对应样本全为零的退化边界模型。对固定 $0 < p < 1$, 因为 $\bar{X} \rightarrow q/p$, 它相合。又 $\text{Var}(X_1) = q/p^2$, 对 $g(u) = 1/(1+u)$ 使用 CLT 和 Delta method:

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p^2 q).$$

Scoring guide (15 points). factorization 2 分、 T 的分布 1 分、幂级数 completeness 2 分; 候选函数 2 分、组合恒等式与无偏性 3 分、Lehmann–Scheffé 1 分; MLE 2 分、CLT/Delta 极限 2 分。

Problem 4**15 points**

对固定 $0 < p < 1$, Bernoulli CLT 给出

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{d} N(0, p(1-p)).$$

在 $p = 0$ 时, $X_i = 0$ a.s., 故 $\hat{p}_n = 0$ a.s. 对每个 n 都成立, 任何中心化缩放仍退化于零。通常的 MLE asymptotic-normality theorem 要求真参数是参数空间内点且 Fisher information 有良好、有限的局部表示; 这里 $p = 0$ 是边界, 邻域也不能在两侧展开, 故不能套用该定理。

当 $p_n = \lambda/n$ 时,

$$n\hat{p}_n = \sum_{i=1}^n X_i \sim \text{Binomial}(n, \lambda/n) \xrightarrow{d} \text{Poisson}(\lambda)$$

由 law of small numbers。因此

$$n(\hat{p}_n - p_n) \xrightarrow{d} \text{Poisson}(\lambda) - \lambda.$$

而

$$\sqrt{n}(\hat{p}_n - p_n) = \frac{\sum_i X_i - \lambda}{\sqrt{n}} \xrightarrow{p} 0,$$

因为分子是 tight (事实上依分布收敛)。所以 $\sqrt{n}$ 标度在此局部边界情形给出退化极限。

Scoring guide (15 points). 内点 CLT 4 分; 边界精确分布 2 分、正则条件说明 2 分; Poisson 极限 3 分、中心化极限 2 分、 $\sqrt{n}$ 退化解释 2 分。

Problem 5**15 points**

log-likelihood 为 $\ell(\theta) = \sum_i \log f_\theta(X_i)$ 。直接微分得

$$\ell'(\theta) = \sum_{i=1}^n \{2F_\theta(X_i) - 1\}, \quad \ell''(\theta) = -2 \sum_{i=1}^n f_\theta(X_i) < 0.$$

当 $\theta \rightarrow -\infty$ 时 score 趋于 n , 当 $\theta \rightarrow +\infty$ 时趋于 $-n$ 。故连续且严格递减的 score 有唯一零点, 该点是唯一 MLE。

若 $U = F_\theta(X)$, 则 probability integral transform 给出 $U \sim \text{Uniform}(0, 1)$, 而单个 score 为 $2U - 1$ 。因此

$$I(\theta) = \mathbb{E}(2U - 1)^2 = \frac{1}{3}.$$

该 family 有共同支持, 关于位置参数光滑, 信息有限且与 θ 无关, score 的邻域矩可积; 标准 regular MLE theorem 适用:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 3).$$

总体 median 是 θ , 且 $f_\theta(\theta) = 1/4$ 。Sample quantile CLT 给出

$$\sqrt{n}(M_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{4f_\theta(\theta)^2}\right) = N(0, 4).$$

在 squared-error 的 asymptotic variance 比较下,

$$\text{ARE}(M_n, \hat{\theta}) = \frac{3}{4}.$$

Scoring guide (15 points). score 2 分、严格凹性及两端极限 3 分; PIT 与信息 2 分、正则条件 1 分、MLE 极限 2 分; median CLT 条件与方差 3 分、ARE 2 分。

Problem 6

15 points

若 $d^\top Y$ 对 $\tau = c^\top \beta$ 无偏, 则

$$d^\top X\beta = c^\top \beta \quad \text{对所有 } \beta,$$

等价于 $X^\top d = c$, 故 $c \in \mathcal{C}(X^\top)$ 。反之, 若 $c = X^\top d$, 则 $d^\top Y$ 即为线性无偏估计量。这证明了充要性。

令

$$\hat{\tau} = c^\top X^+ Y, \quad d_0 = (X^+)^\top c.$$

因为 $X^+ X$ 是到 $\mathcal{C}(X^\top)$ 的正交投影, $c^\top X^+ X = c^\top$, 所以 $\hat{\tau}$ 无偏。其方差为

$$\text{Var}(\hat{\tau}) = \sigma^2 \|d_0\|^2 = \sigma^2 c^\top (X^\top X)^+ c.$$

任意另一个无偏系数 d 满足 $X^\top d = c$ 。令 $h = d - d_0$, 则 $X^\top h = 0$; 同时 $d_0 \in \mathcal{C}(X)$, 故 $d_0 \perp h$ 。于是 $\|d\|^2 = \|d_0\|^2 + \|h\|^2$, 证明 $\hat{\tau}$ 为 BLUE。

令 $P_X = X X^+$ 。Cochran theorem 给出

$$\frac{(n-r)s^2}{\sigma^2} \sim \chi_{n-r}^2,$$

且 s^2 与 $\hat{\tau}$ 独立。因此精确区间为

$$\hat{\tau} \pm t_{n-r, 1-\alpha/2} s \sqrt{c^\top (X^\top X)^+ c}.$$

Scoring guide (15 points). 必要性 3 分、充分性 2 分; Moore-Penrose 构造与无偏 2 分、方差 1 分、正交最优性 3 分; RSS 分布/独立性 2 分、区间 2 分。

Problem 7

10 points

令 $T = \sum_i X_i^2$ 。似然为

$$L(v) = (2\pi v)^{-n/2} \exp\{-T/(2v)\},$$

故 $\hat{v} = T/n$ ($T=0$ 是零概率例外)。令 $a = T/(nv_0)$, 则

$$-2 \log \Lambda = 2\{\ell(\hat{v}) - \ell(v_0)\} = n\{a - 1 - \log a\}.$$

在 H_0 下 $T/v_0 \sim \chi_n^2$ 。因此选择 c_α 满足

$$\mathbb{P}\left(n\left\{\frac{Q}{n} - 1 - \log \frac{Q}{n}\right\} > c_\alpha\right) = \alpha, \quad Q \sim \chi_n^2,$$

并在统计量大于 c_α 时拒绝, 即为 exact level- α GLR rule。也可将拒绝域写成 $T/v_0 < q_L$ 或 $T/v_0 > q_U$, 其中两端由同一 likelihood-ratio level set 与总概率 α 共同确定。

$v_0 > 0$ 是开参数空间内点, support 固定, normal scale likelihood 光滑且信息非奇异, 故 Wilks theorem 适用:

$$-2 \log \Lambda \xrightarrow{d} \chi_1^2.$$

渐近规则为 $-2 \log \Lambda > \chi_{1,1-\alpha}^2$ 。

Scoring guide (10 points). likelihood/MLE 2 分、GLR 公式 3 分; 精确 pivot 与 exact rule 3 分; Wilks 条件和渐近规则 2 分。

Problem 8

10 points

posterior kernel 为

$$\pi(\theta | y) \propto \exp \left\{ -\frac{(y - \theta)^2}{2\sigma^2} - \lambda|\theta| \right\}.$$

因此 posterior mode 等价于最小化严格凸函数

$$Q(\theta) = \frac{(\theta - y)^2}{2\sigma^2} + \lambda|\theta|.$$

在 $\theta > 0$ 时一阶条件给出 $\theta = y - \lambda\sigma^2$; 在 $\theta < 0$ 时给出 $\theta = y + \lambda\sigma^2$ 。在零点用 subgradient 条件 $0 \in -y/\sigma^2 + \lambda[-1, 1]$ 。所以

$$\hat{\theta}_{\text{MAP}} = \text{sign}(y)(|y| - \lambda\sigma^2)_+,$$

且 $|y| \leq \lambda\sigma^2$ 时唯一 mode 为零 (包括等号)。

Squared-error posterior risk 的唯一极小点是 posterior mean $\mathbb{E}(\theta | y)$, 不是 posterior mode。该 posterior 通常不对称于 mode, 故两者一般不同; 不能把 MAP 机械称作平方损失下的 Bayes estimator。

Scoring guide (10 points). posterior kernel 2 分; 凸目标 1 分、正负两区一阶条件 2 分、subgradient 阈值 2 分、完整 soft-threshold 1 分; 平方损失下 posterior mean 及区别 2 分。