# Challenge Problems — Statistics

For enrichment. Harder than the problem sets; no time limit.

A problem marked with a star ($\star$) uses a theorem from the appendix; the rest need nothing beyond the course so far.

**1.** A needle of length $\ell$ is dropped onto a plane ruled with parallel lines a distance $d$ apart, with $\ell \le d$. Model the drop by letting the distance $X$ from the midpoint of the needle to the nearest line be uniform on $[0, d/2]$, and the acute angle $\Theta$ between the needle and the lines be uniform on $[0, \pi/2]$, with $X$ and $\Theta$ independent. Find the probability that the needle crosses a line.

**2.** Let $X_1, \dots, X_n$ be iid Exponential($\lambda$), with rate $\lambda > 0$, and let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics. Define the successive spacings

$$D_1 = X_{(1)}, \qquad D_k = X_{(k)} - X_{(k-1)} \quad (k = 2, \dots, n).$$

Show that $D_1, \dots, D_n$ are mutually independent, and that

$$D_k \sim \text{Exponential}\big((n-k+1)\lambda\big), \qquad k = 1, \dots, n.$$

**3.** A partial converse to the previous problem. Let $X_1, \dots, X_n$ be iid from a continuous distribution, and let $X_{(1)} \le X_{(2)}$ be the two smallest order statistics. Define the first two spacings

$$D_1 = X_{(1)}, \qquad D_2 = X_{(2)} - X_{(1)}.$$

Assume the common density is twice differentiable. Show that if $D_1$ and $D_2$ are independent, then the common distribution of the $X_i$ must be exponential.

**4.** Give an example of a model and a sample for which an entire interval of parameter values are all maximum likelihood estimators.

**5.** Let $X_1, \dots, X_n$ be iid with density

$$f(x \mid \alpha, x_m) = \frac{\alpha\, x_m^\alpha}{x^{\alpha+1}}, \qquad x \ge x_m,$$

where $x_m > 0$ and $\alpha > 1$ are both unknown.

(a) Find the joint maximum likelihood estimator of $(x_m, \alpha)$.

(b) Show that $Y_i = \log(X_i/x_m)$ has an exponential distribution with rate parameter $\alpha$.

(c) Show that the maximum likelihood estimator of $\alpha$ is consistent.

(d) Find the distribution of $X_{(1)} = \min_i X_i$, the maximum likelihood estimator of $x_m$.

**6.** Let $X_1, \ldots, X_n$ be iid Binomial$(N, p)$, where $N \in \{1, 2, \ldots\}$ and $0 < p < 1$ are both unknown. For the observations

$$X_1 = 0, \qquad X_2 = 2,$$

does a finite maximum likelihood estimator of $(N, p)$ exist?

**7.** Give an example of a family of discrete random variables for which being uncorrelated implies independence.

**8.** Let $(X_i, Y_i)$, $i = 1, \ldots, n$, be iid, where

$$X_i \sim \text{Poisson}(\theta), \qquad Y_i \sim \text{Poisson}(\theta^2),$$

independently, with $\theta > 0$.

(a) Find the maximum likelihood estimator of $\theta$.

(b) Show that the maximum likelihood estimator is consistent.

(*c*) $\star$ Consider the two estimators

$$T_{1n} = \bar{X}, \qquad T_{2n} = \sqrt{\bar{Y}}.$$

Show that both are consistent and compare their asymptotic variances. *[Hint: use the delta method from the appendix.]*

**9.** Let $X_1, \ldots, X_n$ be iid Exponential$(\lambda)$, with rate $\lambda > 0$.

(a) Show that

$$\sqrt{n}\left(\bar{X} - \tfrac{1}{\lambda}\right) \xrightarrow{d} N\!\left(0, \tfrac{1}{\lambda^2}\right).$$

(*b*) $\star$ The maximum likelihood estimator of $\lambda$ is $\hat{\lambda} = 1/\bar{X}$. Show that

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

*[Hint: use the delta method from the appendix.]*

**10.** Let $(X, Y)$ be a random vector whose coordinates are independent standard normals. Write $(X, Y)$ in polar coordinates as

$$X = R\cos\Theta, \qquad Y = R\sin\Theta, \qquad R \geq 0, \quad \Theta \in [0, 2\pi).$$

Compute the joint density of $(R, \Theta)$.

**11.** Let $X$ have density $f(x; \theta)$, $\theta \in \mathbb{R}$, under the usual regularity conditions (smoothness in $\theta$, support free of $\theta$, and differentiation under the integral sign permitted). The *score* is $U(\theta) = \dfrac{\partial}{\partial\theta} \log f(X; \theta)$, and the *Fisher information* is $I(\theta) = \text{Var}\big(U(\theta)\big)$.

(a) Show that $E[U(\theta)] = 0$, so that $I(\theta) = E\big[U(\theta)^2\big]$.

(b) Show that $I(\theta) = -E\left[\dfrac{\partial^2}{\partial\theta^2} \log f(X;\theta)\right]$.

(c) Show that the Fisher information contained in $n$ iid observations $X_1, \ldots, X_n$ equals $n\,I(\theta)$.

**12.** Compute the Fisher information $I(\theta)$ from a single observation for each of the following.

(a) Bernoulli($p$), for $p$.

(b) Poisson($\lambda$), for $\lambda$.

(c) Exponential($\lambda$), for $\lambda$.

(d) $N(\mu, \sigma^2)$ with $\sigma^2$ known, for $\mu$.

**13.** $\star$ Let $X_1, \ldots, X_n$ be iid with density $f(x;\theta)$, $\theta \in \mathbb{R}$, satisfying the usual regularity conditions (the density is smooth in $\theta$, the support does not depend on $\theta$, and derivatives in $\theta$ may be taken under the integral sign). Let $\hat{\theta}_n$ be the maximum likelihood estimator, a root of the score equation

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i;\theta) = 0,$$

and define the Fisher information $I(\theta) = E\Big[\big(\frac{\partial}{\partial\theta}\log f(X;\theta)\big)^2\Big]$. By expanding the score about the true value $\theta$, show that

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{d} N\Big(0, \tfrac{1}{I(\theta)}\Big).$$

*[Hint: Slutsky.]*

**14.** Passengers arrive at a bus stand at an average rate of 8 per minute; the number who have arrived $t$ minutes after opening is distributed as Poisson($8t$). The bus arrives after a time that is exponentially distributed with mean 5 minutes. It has 40 seats, of which the number already occupied is equally likely to be any integer from 1 to 40. Find the expected number of passengers still waiting when the bus departs.

**15.** $N$ people line up to board an aeroplane, each holding a boarding pass with an assigned seat. The first passenger has lost his pass and takes a seat at random. Thereafter each passenger takes his assigned seat if it is free, and otherwise sits in a uniformly random free seat. Find the probability that the last passenger sits in his assigned seat, for $N = 3$ and for $N = 100$.

**16.** A mathematician keeps a matchbox in each of two pockets, each box initially holding ten matches. Whenever a match is needed a pocket is chosen at random (each equally likely) and one match is taken from it. Eventually a chosen box is found empty. Let $X$ be the number of matches in the *other* box at that moment. Find the distribution $p_X(k)$, $k = 0, 1, \ldots, 10$.

**17.** The ten digits 0–9 are written one per card and placed face down in random order. A psychic guesses the cards one at a time; after each guess the card is turned over and shown to her, and she uses everything she has seen. If she in fact has no ability, what is the expected number of correct guesses?

**18.** A homogeneous good is offered by many sellers; each seller's price is an independent draw from a distribution with cdf $F$ and pdf $f$. A buyer cannot learn a price without asking, and each quotation costs $c > 0$. He fixes in advance the number $n$ of sellers to canvass, then buys at the lowest of the $n$ quoted prices. Find the $n$ that minimises his expected total expenditure (price paid plus cost of search), and evaluate it when prices are Uniform$(0, 1)$.

**19.** Let $Z_1, Z_2, \ldots$ be iid, and for each $n$ let $X_{(n)}$ be the sample median of the first $2n + 1$ observations $Z_1, \ldots, Z_{2n+1}$. Does $X_{(n)}$ converge in probability? If so, identify its limit.

**20.** Chernoff bounds and large deviations.

(a) Let $X$ have moment generating function $M_X(t) = E(e^{tX})$ finite for $t$ near 0, and let $a \in \mathbb{R}$. Derive the Chernoff bound

$$P(X \geq a) \;\leq\; \min_{t>0}\; e^{-ta}\, M_X(t).$$

(b) Let $X_1, \ldots, X_n$ be independent copies of $X$ and $S_n = \sum_{i=1}^{n} X_i$. Deduce that

$$P(S_n \geq na) \;\leq\; e^{-n\,I(a)}, \qquad I(a) = \max_{t>0} \big(ta - \ln M_X(t)\big).$$

(c) Now let $X \sim$ Bernoulli$(p)$. For $p < a < 1$, evaluate $I(a)$ in closed form, and state the resulting large-deviation bound on $P(S_n \geq na)$ that holds for the whole Bernoulli family.

---

## Appendix: results used by starred ($\star$) problems

Stated without proof.

**R1.** *(Slutsky.)* If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for a constant $c$, then

$$X_n + Y_n \xrightarrow{d} X + c, \qquad X_n Y_n \xrightarrow{d} cX, \qquad X_n/Y_n \xrightarrow{d} X/c \;\; (c \neq 0).$$

**R2.** *(Delta method.)* If $\sqrt{n}\,(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$ and $g$ is differentiable at $\theta$ with $g'(\theta) \neq 0$, then

$$\sqrt{n}\,\big(g(T_n) - g(\theta)\big) \xrightarrow{d} N\big(0,\, \sigma^2 g'(\theta)^2\big).$$