Practice questions

1. The distance (in metres) of an archer's target from the bull's eye is a random variable with PDF $f(x) = \theta^2 x e^{-\theta x}$ for $x \geq 0$, where the parameter $\theta \geq 0$ measures the archer's skill.

   (a) Bob's first hit is 40 cm away from the bull's eye. What is the maximum likelihood estimate (MLE) of θ?

   Solution: The likelihood function is $f_{X_1}(\theta) = \mathrm{P}(0.4|\theta) = 0.4\theta^2 e^{-0.4\theta} \propto \theta^2 e^{-0.4\theta}$. It is maximized when $df_{X_1}(\theta)/d\theta = 0$, namely when $2\theta e^{-0.4\theta} - 0.4\theta^2 e^{-0.4\theta} = 0$. The only extremal point is $\theta = 2/0.4 = 5$ and it is a maximum. The MLE is 5.

   (b) Bob's second hit is 20 cm away from the bull's eye. What is the new MLE of θ?

   Solution: The likelihood function is now

   $$f_{X_1,X_2}(0.4, 0.2|\theta) = 0.4\theta^2 e^{-0.4\theta} \cdot 0.2\theta^2 e^{-0.2\theta} \propto \theta^4 e^{-0.6\theta}.$$

   Its derivative is zero when $4\theta^3 e^{-0.6\theta} - 0.6\theta^4 e^{-0.6\theta} = 0$. The unique extremal point is $\theta = 4/0.6 = 6.5$ and it is again a maximum. The new MLE is 6.5.

   (c) Alice's two hits are 15cm and 50cm away from the bull's eye. Does the MLE predict that Alice is a more skilled archer than Bob?

   Solution: In general, the likelihood for hits at distance x_1 and x_2 is $f_{X_1,X_2}(x_1, x_2|\theta) = \theta^4 x_1 x_2 e^{-(x_1+x_2)\theta}$, and its maximum occurs at $4/(x_1+x_2)$. Thus the player with a smaller sum of distances from the bull's eye is the more skilled archer. As $0.15 + 0.5 > 0.4 + 0.2$ the MLE predicts that Bob is more skilled.

2. In Lecture 4 we showed that the maximum likelihood estimator for the mean λ of a Poisson(λ) random variable given that N occurrences are observed within a time unit is N.

   (a) Now subdivide the time unit into 10 equal intervals and suppose that N_i occurrences are observed in the i-th interval. The N_i are then independent samples of a Poisson($\lambda/10$) random variable. What is the maximum likelihood estimator for λ?

   Solution: The joint PMF of $N_1, \dots, N_{10}$ is

   $$\begin{aligned}\mathrm{P}(N_1 = x_1, \dots, N_{10} = x_{10}|\lambda) &= e^{-\lambda/10} \cdot \frac{(\lambda/10)^{x_1}}{x_1!} \cdot e^{-\lambda/10} \frac{(\lambda/10)^{x_2}}{x_2!} \cdots e^{-\lambda/10} \frac{(\lambda/10)^{x_{10}}}{x_{10}!} \\ &\propto e^{-\lambda} \lambda^{x_1+\cdots+x_{10}}.\end{aligned}$$

   The maximum likelihood estimator is a critical point of λ so it should occur either at 0 or where $(d/d\lambda)e^{-\lambda}\lambda^{x_1+\cdots+x_{10}} = 0$. The only solution of this equation is $\lambda = x_1 + \cdots + x_{10}$ and this is a maximum, therefore $MLE = N_1 + \cdots + N_{10}$.

   (b) Is the maximum likelihood estimator in part (a) biased or not?

   Solution: $\mathrm{E}[MLE] = \mathrm{E}[N_1 + \cdots + N_{10}] = \mathrm{E}[N_1] + \cdots + \mathrm{E}[N_{10}] = 10(\lambda/10) = \lambda$, so it is unbiased.

   (c) **(for ESTR)** Can you come up with a sufficient statistic for n samples of a Poisson(λ) random variable?

Solution: Their sum $N = N_1 + \cdots + N_n$ is a sufficient statistic. Given N, N_i can be taken as the number of samples that take values between $i-1$ and i among N independent samples of a Uniform$(0, n)$ random variable. You then need to verify that $N_1, \ldots, N_n$ are (unconditionally) independent Poisson(λ) random variables.

3. You have a coin that is either always heads ($\theta = 1$) or fair ($\theta = 0$).

   (a) What is the maximum likelihood estimator for θ from n independent coin flips?

   Solution: When $\theta = 1$, the all-heads outcome has probability 1 and all others have probability zero. When $\theta = 0$, all outcomes have probability 2^{-n}. Therefore the maximum likelihood estimate is 1 for an all-heads sequence and 0 for all other outcomes.

   (b) What is *the* unbiased estimator for θ from one coin flip? (There is only one.)

   Solution: Let h and t be the outputs of the estimator when observing a head and a tail, respectively. When a head is observed, the estimator must output 1 with probability one, so $h = 1$. If it didn't, the estimator would be biased in the case $\theta = 1$. Then the bias in case $\theta = 0$ is $\frac{1}{2}h + \frac{1}{2}t$. For the estimator to be unbiased in this case also, t must equal -1. Therefore the estimator should output 1 when observing a head and -1 when observing a tail.

   (c) **(Optional)** Among all unbiased estimators for θ from n coin flips, which one has the smallest variance?

   Solution: Let H be the event of observing n heads. By the same reasoning as in part (a), the estimator $\hat{\Theta}$ must output 1 conditioned on H. Therefore the estimator has zero-variance when $\theta = 1$, i.e., $\mathrm{Var}_1[\hat{\Theta}] = 0$.
   When $\theta = 0$, by the total expectation theorem, the bias is

$$\mathrm{E}_0[\hat{\Theta}] = \mathrm{E}_0[\hat{\Theta}|H] \cdot 2^{-n} + \mathrm{E}_0[\hat{\Theta}|H^c] \cdot (1 - 2^{-n}),$$

   so if the bias is zero, we must have $\mathrm{E}_0[\hat{\Theta}|H^c] = -2^{-n}/(1-2^{-n})$. The variance is

$$\mathrm{Var}_0[\hat{\Theta}] = \mathrm{E}_0[\hat{\Theta}^2] = \mathrm{E}_0[\hat{\Theta}^2|H] \cdot 2^{-n} + \mathrm{E}_0[\hat{\Theta}^2|H^c] \cdot (1 - 2^{-n}),$$

   which is minimized when $\mathrm{E}_0[\hat{\Theta}^2|H^c]$ is. Since

$$\mathrm{E}_0[\hat{\Theta}^2|H^c] = \mathrm{Var}_0[\hat{\Theta}|H^c] + \mathrm{E}_0[\hat{\Theta}|H^c]^2 = \mathrm{Var}_0[\hat{\Theta}|H^c] + \frac{2^{-2n}}{(1-2^{-n})^2}$$

   the minimum is attained when $\mathrm{Var}_0[\hat{\Theta}|H^c]$ is zero, namely when $\hat{\Theta}$ takes the same value $-2^{-n}/(1-2^{-n})$ on all sequences that have at least one tail. In conclusion, the desired estimator is

$$\hat{\Theta} = \begin{cases} 1, & \text{if all } n \text{ flips are heads,} \\ -2^{-n}/(1-2^{-n}), & \text{if there is at least one tail.} \end{cases}$$

   The indicator of the event H is in fact a sufficient statistic for θ.

4. You are given three samples of a Zig(θ) random variable, which has PDF

$$f(x) = \begin{cases} 2(x - \theta), & \text{when } \theta \le x \le \theta + 1 \\ 0, & \text{otherwise.} \end{cases}$$

   (a) What is the expected value μ of a Zig(θ) random variable?

   Solution: $\mu = \int_\theta^{\theta+1} 2(x-\theta)x\,dx = \theta + \frac{2}{3}$.

(b) Come up with an unbiased estimator for θ that depends only on the sample mean $\overline{X}$.

Solution: Since $\overline{X}$ is an unbiased estimator for μ, $\mathrm{E}_\theta[\overline{X}] = \theta + \frac{2}{3}$, so $\overline{X} - \frac{2}{3}$ is an unbiased estimator for θ.

(c) Repeat part (b) for the sample maximum MAX.

Solution: The CDF of MAX is

$$\mathrm{P}(MAX \le t|\theta) = P(X_1 \le t, X_2 \le t, X_3 \le t) = P(X_i \le t)^3 = \left(\int_\theta^t 2(x-\theta)dx\right)^3 = (t-\theta)^6$$

when $\theta \le t \le \theta + 1$. The PDF is $f_{MAX}(t) = 6(t-\theta)^5$ and the expected value is

$$\mathrm{E}_\theta[MAX] = \int_\theta^{\theta+1} t \cdot 6(t-\theta)^5 dt = \int_0^1 (\theta + t) \cdot 6t^5 dt = \theta + \frac{6}{7}.$$

Therefore $MAX - \frac{6}{7}$ is an unbiased estimator of θ.

(d) (**Optional**) What are the variances of your estimators in (b) and (c)? (**Hint:** Argue that the variance should not depend on θ and assume $\theta = 0$ in the calculation.)

Solution: The shifted samples $X_1 - \theta, X_2 - \theta, X_3 - \theta$ are Zig(0) random variables with sample mean $\overline{X} - \theta$ and sample maximum $MAX - \theta$. Since shifting by a constant does not change the variance, i.e.

$$\begin{aligned} \mathrm{Var}_\theta[\overline{X} - \tfrac{2}{3}] &= \mathrm{Var}_0[\overline{X} - \theta - \tfrac{2}{3}] = \mathrm{Var}_0[\overline{X}] \\ \mathrm{Var}_\theta[MAX - \tfrac{6}{7}] &= \mathrm{Var}_0[MAX - \theta - \tfrac{6}{7}] = \mathrm{Var}_0[MAX], \end{aligned}$$

we may assume $\theta = 0$ and ignore the constant shift when calculating the sample variances.

$$\mathrm{Var}(\overline{X}) = \frac{1}{3}\mathrm{Var}(X) = \frac{1}{3}(E[X^2] - E[X]^2) = \frac{1}{54}$$

because $\mathrm{E}[X] = \mu = \frac{2}{3}$ and $\mathrm{E}[X^2] = \int_0^1 x^2 \cdot 2x dx = \frac{1}{2}$. As for the sample max,

$$\mathrm{Var}[MAX] = \mathrm{E}[MAX^2] - \mathrm{E}[MAX]^2 = \int_0^1 t^2 \cdot 6t^5 dt - \left(\frac{6}{7}\right)^2 = \frac{3}{4} - \left(\frac{6}{7}\right)^2 = \frac{3}{196}$$

so the part (c) estimator has lower variance.