Quiz Section #9: Supplementary Exercise Answers

CSE 312: Foundations of Computing II

1. (a) Suppose $x_1, x_2, \ldots, x_n$ are samples from a normal distribution whose mean is known to be zero, but whose variance is unknown. What is the maximum likelihood estimator for its variance?

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2$$

(b) Suppose the mean is known to be $\mu$ but the variance is unknown. How does the maximum likelihood estimator for the variance differ from the maximum likelihood estimator when both mean and variance are unknown?

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

vs.

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

(The former turns out to be unbiased, the latter biased.)

2. Let $f(x \mid \theta) = \theta x^{\theta-1}$ for $0 \leq x \leq 1$, where $\theta$ is any positive real number. Let $x_1, x_2, \ldots, x_n$ be i.i.d. samples from this distribution. Derive the maximum likelihood estimator $\hat{\theta}$.

$$L(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ln L(x_1, x_2, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} (\ln \theta + (\theta - 1) \ln x_i)$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, x_2, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} \left( \frac{1}{\theta} + \ln x_i \right) = 0$$

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

Check that it’s the global maximum:

$$\frac{\partial^2}{\partial \theta^2} \ln L(x_1, x_2, \ldots, x_n \mid \theta) = \sum_{i=1}^{n} \frac{-1}{\theta^2} < 0$$

so $\ln L(x_1, x_2, \ldots, x_n \mid \theta)$ is concave downward everywhere.

3. You are given 100 independent samples $x_1, x_2, \ldots, x_{100}$ from $\text{Ber}(p)$, where $p$ is unknown. These 100 samples sum to 30. You would like to estimate the distribution’s parameter $p$. Give all answers to 3 significant digits.

(a) What is the maximum likelihood estimator $\hat{p}$ of $p$?

$$\hat{p} = \frac{1}{100} \sum_{i=1}^{100} x_i = 30/100 = 0.300$$
(b) Is \( \hat{p} \) an unbiased estimator of \( p \)?

\[
E[\hat{p}] = E\left[ \frac{1}{100} \sum_{i=1}^{100} x_i \right] = \frac{1}{100} \sum_{i=1}^{100} E[x_i] = \frac{1}{100} \cdot 100p = p
\]

so it is unbiased.

(c) Give your best approximation for the 95% confidence interval of \( p \).

\( \hat{p} \) is the average of 100 independent \( \text{Ber}(p) \) random variables. By the Central Limit Theorem, then, \( \hat{p} \) is approximately \( N(p, 0.0021) \), where \( 0.0021 = \hat{p}(1 - \hat{p})/100 \).

\[
P\left( -z < \frac{\hat{p} - p}{\sqrt{0.0021}} < z \right) \approx 2\Phi(z) - 1
\]

\[
P(\hat{p} - z \sqrt{0.0021 < p < \hat{p} + z \sqrt{0.0021}}) \approx 2\Phi(z) - 1 \geq 0.95
\]

\[
\Phi(z) \geq 0.975
\]

\[
z \approx 1.96
\]

\[
z \sqrt{0.0021} \approx 0.0898
\]

That is, the 95% confidence interval is approximately \([0.210, 0.390]\).

(d) Give your best approximation for the 90% confidence interval of \( p \).

\[
P(\hat{p} - z \sqrt{0.0021} < p < \hat{p} + z \sqrt{0.0021}) \approx 2\Phi(z) - 1 \geq 0.90
\]

\[
\Phi(z) \geq 0.95
\]

\[
z \approx 1.65
\]

\[
z \sqrt{0.0021} \approx 0.0756
\]

That is, the 90% confidence interval is approximately \([0.224, 0.376]\).

(e) Give three different reasons why your answers to (c) and (d) are only approximations.

i. \( \hat{p} \) is only approximately normal, but we are using the standard normal table as though it is normal.

ii. The variance of this normal distribution should be \( p(1 - p)/100 \), but we are using \( \hat{p}(1 - \hat{p})/100 \) as an approximation.

iii. The values of \( z \) derived from the standard normal table are approximations.

(f) Explain why it makes sense that the interval in (d) is bigger (or smaller, depending on your answers) than the interval in (c).

The 90% confidence interval should be smaller than the 95% confidence interval. What we have shown in part (d) is that \( P(p \in [0.224, 0.376]) \approx 0.90 \). If we want \( P(p \in [a, b]) \approx 0.95 \), we would need \([a, b]\) to be a larger interval.

4. (a) Suppose that \( \hat{\theta} \) is a biased estimator for \( \theta \) with \( E[\hat{\theta}] = \alpha\theta \), for some constant \( \alpha > 0 \). Find an unbiased estimator for \( \theta \) and prove that it is unbiased.

\( \hat{\theta}' = \hat{\theta}/\alpha \) is unbiased, because

\[
E[\hat{\theta}'] = E[\hat{\theta}/\alpha] = \frac{1}{\alpha} E[\hat{\theta}] = \frac{1}{\alpha} \cdot \alpha\theta = \theta
\]
(b) In lecture, we saw that the maximum likelihood estimator for the population variance \( \theta_2 \) of \( N(\theta_1, \theta_2) \) is the sample variance

\[
\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2
\]

where \( \hat{\theta}_1 \) is the sample mean. It can be shown that \( E[\hat{\theta}_2] = \frac{n-1}{n} \cdot \theta_2 \), so that \( \hat{\theta}_2 \) is biased and always underestimates the variance \( \theta_2 \). Use your result from part (a) to find an unbiased estimator of the variance \( \theta_2 \).

\[
\frac{n}{n-1} \cdot \hat{\theta}_2 = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2
\]

is unbiased, as was stated in lecture.