3. Bias and efficiency

Are the estimators that you got in the previous problem unbiased? If so, find the efficiency.

To answer this question we need an estimator for the last question instead of an estimate. So, I'll do the last question again.

Suppose that in your sample of n, you get $X = 0$ n_0 times and $X = 1$ n_1 times and X_2 n_2 times. (So, $n_0 + n_1 + n_2 = n$). Then the probability of getting this outcome in the order that it occurred is

$$L(\theta) = \theta^{n_0}(2\theta)^{n_1}(1-3\theta)^{n_2}$$

(If you plug in $(n_0, n_1, n_2) = (2, 1, 1)$ you get $2\theta^3 - 6\theta^4$ as before.)

$$\ln L(\theta) = n_0 \ln\theta + n_1 \ln 2 + n_1 \ln\theta + n_2 \ln(1-3\theta)$$

$$\frac{\partial}{\partial\theta}\ln L(\theta) = \frac{n_0}{\theta} + \frac{n_1}{\theta} - \frac{3n_2}{1-3\theta}$$

Set this equal to zero. Multiply by $\theta(1-3\theta)$ (to clear denominators):

$$(n_0 + n_1)(1-3\theta) - 3n_2\theta = 0$$

$$\hat{\theta}_{MLE} = \frac{n_0 + n_1}{3n} \tag{0.1}$$

This is 1/3 of the proportion of time that $X \le 1$. This means that its expected value is

$$\mathbb{E}(\hat{\theta}_{MLE}) = \frac{1}{3}\mathbb{E}\left[\frac{n_0+n_1}{n}\right] = \frac{1}{3}\mathbb{P}(X \le 1) = \frac{1}{3}(\theta + 2\theta) = \theta$$

So, the estimator in (0.1) is an unbiased estimator of θ.

I am using the basic definition of probability and expectation in the discrete case:

$$\mathbb{P}(\text{event } A) = \mathbb{E}\left[\frac{\text{number of times event } A \text{ occurs}}{\text{number of independent trials}}\right]$$

In this case, the event is that $X = 0$ or 1.

find the efficiency.

The ratio

$$\frac{n_0 + n_1}{n}$$

is the proportion of trials for which $X \le 1$. This is a Bernoulli event (either it happens or it doesn't). So, it has expected value $p = \mathbb{P}(X \le 1) = 3\theta$ and variance

$$p(1-p)/n = 3\theta(1-3\theta)/n$$

So, the variance of $\hat{\theta}$ is 1/3 of this

$$\mathrm{Var}(\hat{\theta}_{MLE}) = \frac{1}{3}\frac{3p(1-p)}{n} = \frac{\theta(1-3\theta)}{n}$$

$$\mathrm{Eff}(\hat{\theta}_{MLE}) = \frac{1}{\mathrm{Var}(\hat{\theta}_{MLE})} = \frac{n}{\theta(1-3\theta)}$$

The estimator $\hat{\lambda}_{MOM} = \overline{X} - 1$ is an unbiased estimator of λ because

$$\mathbb{E}(\overline{X} - 1) = \mathbb{E}(\overline{X}) - 1 = \mathbb{E}(X) - 1 = \lambda + 1 - 1 = \lambda$$

To find the variance we need the expected value of X^2:

$$\mathbb{E}(X^2) = \int_\theta^\infty x^2 e^{\theta - x} dx = (-x^2 - 2x - 2)e^{\theta - x}|_\theta^\infty$$

$$= \theta^2 + 2\theta + 2$$

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \theta^2 + 2\theta + 2 - (\theta + 1)^2 = 1$$

$$\text{Var}(\overline{X}) = \text{Var}(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\sum \mathbb{E}(X_i) = \frac{1}{n^2} n \cdot 1 = \frac{1}{n}$$

So,

$$\text{Var}(\hat{\lambda}_{MOM}) = \text{Var}(\overline{X} - 1) = \text{Var}(\overline{X}) = \frac{1}{n}$$

(The rule is that $\text{Var}(Y + \text{constant}) = \text{Var}(Y)$.) So,

$$\text{Eff}(\hat{\lambda}_{MOM}) = \frac{1}{\text{Var}(\hat{\lambda}_{MOM})} = n$$

4. You take a sample of 25 from an exponential distribution and get $\overline{T} = 2.3$. Find the 95% confidence interval for λ. [$\overline{T}$ is approximately normal by the CLT.]

The exponential distribution has mean $1/\lambda$ and variance $1/\lambda^2$. This means that

$$\overline{T} \sim N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$$

Converting to the Z-statistic we get:

$$Z = \frac{\overline{T} - 1/\lambda}{1/\lambda\sqrt{n}} = \overline{T}\lambda\sqrt{n} - \sqrt{n} = 5\overline{T}\lambda - 5$$

The 95% confidence interval is:

$$-1.96 \leq 5\overline{T}\lambda - 5 \leq 1.96$$

Add 5 to everything:

$$3.04 \leq 5\overline{T}\lambda \leq 6.96$$

Divide by $5\overline{T} = 11.5$ to get

$$0.264 \leq \lambda \leq 0.605$$

Just to check, the estimator $\hat{\lambda}$ should be in this interval

$$\hat{\theta} = 1/\overline{T} = 1/2.3 = 0.435$$

which is in the middle of the interval.

5. Suppose that X has density function $f(x) = 2\lambda e^{-\lambda x^2} x$ if $x \geq 0$ and $f(x) = 0$ for $x < 0$. Then find the Cramér-Rao lower bound for the variance of any unbiased estimator for λ.

Fisher information $I(\lambda)$ is the negative of the expected value of the second derivative of the log likelihood function since the support of the density is independent to λ.

$$L(\lambda) = 2x\lambda e^{-\lambda x^2}$$
$$\ln L(\lambda) = \ln(2x) + \ln\lambda - \lambda x^2$$
$$\frac{\partial}{\partial\lambda}\ln L(\lambda) = \frac{1}{\lambda} - x^2$$
$$\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda) = -\frac{1}{\lambda^2}$$
$$I(\lambda) = \mathbb{E}\left[-\frac{\partial^2}{\partial\lambda^2}\ln L(\lambda)\right] = \frac{1}{\lambda^2}$$

So, the Cramér-Rao lower bound for any unbiased estimator $\hat{\lambda}$ of λ is

$$\mathrm{Var}(\hat{\lambda}) \geq \frac{1}{nI(\lambda)} = \frac{\lambda^2}{n}$$