Fall 2012

1. a) Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$ be an iid sample. Find the method of moments estimate $\hat{\lambda}_{MOM}$ and the maximum likelihood estimate $\hat{\lambda}_{MLE}$ of λ.

Solution. The first moment of X_1 is

$$\mathbb{E}X_1 = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1} \frac{\lambda^{k-1}}{(k-1)!} = \lambda,$$

and so, equating the first moment with the first sample moment gives us the moment estimate

$$\boxed{\hat{\lambda}_{MOM} = \frac{1}{n} \sum_{i=1}^{n} X_i}.$$

The likelihood function is

$$\mathcal{L}(\lambda; \boldsymbol{x}) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!},$$

and so, the log-likelihood function is

$$\log \mathcal{L}(\lambda; \boldsymbol{x}) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log \prod_{i=1}^{n} x_i!.$$

Since

$$\frac{d}{d\lambda} \log \mathcal{L} = -n + \frac{\sum x_i}{\lambda},$$

and

$$\frac{d^2}{d\lambda^2} \log \mathcal{L} = -\frac{\sum x_i}{\lambda^2} \leq 0,$$

we have that

$$\boxed{\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

is the MLE of λ.

b) Is $\hat{\lambda}_{MLE}$ unbiased? Is it efficient?

Solution. The Fisher information is

$$\mathcal{I}(\lambda) = -\mathbb{E}\left(-\frac{\sum X_i}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda},$$

and so, the Cramer-Rao lower bound is $\frac{\lambda}{n}$. On the other hand,

$$\text{Var}\left(\hat{\lambda}_{MLE}\right) = \frac{1}{n}^2 n\text{Var}(X_1) = \frac{\lambda}{n},$$

and so, our MLE achieves the Cramer-Rao lower bound and is efficient. □

c) Give an example of a distribution where the MOM estimate and the MLE are different.

Solution. Let $X_1, \ldots, X_n \sim \mathcal{U}(0, \theta)$ be an iid sample, where θ is unknown. Then, the first moment of X_1 is $\theta/2$, and so, the method of moment estimator is $\hat{\theta}_{MOM} = \frac{2}{n} \sum_{i=1}^{n} X_i$. On the other hand, the likelihood function is

$$\mathcal{L}(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n},$$

which is decreasing in θ, and so the MLE of θ is the minimum possible value of θ:

$$\hat{\theta}_{MLE} = X_{(n)}.$$

We see that $\hat{\theta}_{MOM} \neq \hat{\theta}_{MLE}$. □

2. a) Prove that, for any (possibly correlated) collection of random variables $X_1, \ldots, X_k$,

$$\operatorname{Var}\left(\sum_{i=1}^{k} X_i\right) \leq k \sum_{i=1}^{k} \operatorname{Var}(X_i).$$

Solution. First note that if $\boldsymbol{a} = \langle a_1, a_2, \ldots, a_n \rangle$, and $\boldsymbol{b} = \langle 1, 1, \ldots, 1 \rangle$, then Cauchy-Schwarz inequality gives us that

$$\left(\sum_{i=1}^{n} a_i\right)^2 = \langle \boldsymbol{a}, \boldsymbol{b} \rangle^2 \leq \langle \boldsymbol{a}, \boldsymbol{a} \rangle \langle \boldsymbol{b}, \boldsymbol{b} \rangle = n \sum_{i=1}^{n} a_i^2.$$

In particular,

$$\left(\sum_{i=1}^{k} \sqrt{\operatorname{Var}(X_i)}\right)^2 \leq k \sum_{i=1}^{k} \operatorname{Var}(X_i).$$

Another application of Cauchy-Schwarz gives us

$$\operatorname{Cov}(X, Y) \leq \sqrt{\operatorname{Var}(X)} \sqrt{\operatorname{Var}(Y)}.$$

Hence,

$$\begin{aligned} \operatorname{Var}\left(\sum_{i=1}^{k} X_i\right) &= \sum_{i=1}^{k} \sum_{j=1}^{k} \operatorname{Cov}(X_i, X_j) \leq \sum_{i=1}^{k} \sum_{j=1}^{k} \sqrt{\operatorname{Var}(X_i)} \sqrt{\operatorname{Var}(X_j)} \\ &= \left(\sum_{i=1}^{k} \sqrt{\operatorname{Var}(X_i)}\right)^2 \leq k \sum_{i=1}^{k} \operatorname{Var}(X_i), \end{aligned}$$

and so, we are done. □

b) Construct an example with $k \geq 2$ where equality holds above.

Solution. Let us consider the case where $X_1 = \cdots = X_k$. Then,

$$\operatorname{Var}\left(\sum_{i=1}^{k} X_i\right) = \operatorname{Var}(kX_1) = k^2 \operatorname{Var}(X_1),$$

and

$$k \sum_{i=1}^{k} \operatorname{Var}(X_i) = k^2 \operatorname{Var}(X_1),$$

and so, equality is achieved. □