Introduction to Statistics Solutions 1

---

1. Let X denote a random variable with the Rayleigh distribution with parameter σ, and with density f given in the question. Then, it's mean is given by

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} \frac{x^2}{\sigma^2} \exp\Big(-\frac{x^2}{2\sigma^2}\Big) dx \\ &= \frac{1}{2}\int_{-\infty}^{\infty} \frac{x^2}{\sigma^2} \exp\Big(-\frac{x^2}{2\sigma^2}\Big) dx \\ &= \frac{\sqrt{2\pi}}{2\sigma}\int_{-\infty}^{\infty} x^2 \phi(x) dx \ = \ \sigma\sqrt{\frac{\pi}{2}},\end{aligned}$$

where the third equality holds by symmetry, and the last equality follows from the hint.

Hence, a method of moments estimator for the parameter σ is

$$\hat{\sigma}_{\text{MoME}} = \sqrt{\frac{2}{\pi}}\overline{X} = \sqrt{\frac{2}{\pi}}\frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Next, to find the maximum likelihood estimator, we first compute the log-likelihood of the data. Using the density f given in the question, we get that the log-likelihood function is

$$\ell(\sigma; x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n}\Big(\log\frac{x_k}{\sigma^2} - \frac{x_k^2}{2\sigma^2}\Big) = -2n\log\sigma - \frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n}\log x_k.$$

To find the value of σ that maximises this function, we set the derivative to zero:

$$\frac{d\ell}{d\sigma} = \frac{-2n}{\sigma} + \frac{1}{\sigma^3}\sum_{k=1}^{n} x_k^2 = 0.$$

Solving this, we get the maximum likelihood estimator to be

$$\hat{\sigma}_{\text{MLE}} = \sqrt{\frac{1}{2n}\sum_{k=1}^{n} x_k^2}$$

2. (a) The mean of a Geom(p) distribution is $1/p$. Equating this to the sample mean gives us the method of moments estimator:

$$\frac{1}{\hat{p}_{MoME}} = \frac{1}{k}\sum_{i=1}^{k} X_i = \frac{N}{k}, \text{ i.e., } \hat{p}_{MoME} = \frac{k}{N},$$

which is the same as our first estimator.

To obtain the MLE, we first compute the log-likehood of the data as a function of the parameter p:

$$\ell(p; X_1, X_2, \ldots, X_k) = \sum_{i=1}^{k} \Big(\log p + (X_i - 1) \log(1-p) \Big) = k \log p + (N-k) \log(1-p).$$

Now we can obtain the MLE by setting the derivative to zero and solving the resulting equation:

$$\frac{d\ell}{dp} = \frac{k}{p} - \frac{N-k}{1-p} = 0 \; \Rightarrow \; \hat{p}_{MLE} = \frac{k}{N}.$$

Thus, all three estimators are the same, and equal to k/N.

(b) The method of moments estimator depends only on the sample mean of the X_i's, which is N/k, rather than on the individual values. In the case of the MLE, the likelihood of any sequence depends only on the number of heads and tails in the sequence, and not on where those heads and tails are located; this is a consequence of the coin tosses being iid.

(c) Please ignore this part of the question, as it is quite a bit harder than I thought! It involves computing the mean and second moment of $1/N$, which is not straightforward. If you want to solve this, you can look up the mean and moments of the negative binomial distribution.

3. The mean of U is $a/2$ (by symmetry or calculation); equating this to the sample mean gives the MoME,

$$\hat{a}_{MoME} = \frac{2}{n} \sum_{i=1}^{n} x_i.$$

To compute the likelihood function, note that the density of the uniform distribution is given by

$$f(x) = \begin{cases} 1/a, & 0 \le x \le a, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood function is given by

$$L(a; x_1, x_2, \ldots, x_n) = \begin{cases} \frac{1}{a^n}, & \text{if } a \ge \max_{i=1}^n x_i, \\ 0, & \text{otherwise.} \end{cases}$$

Now, a^n is an increasing function of a and $1/a^n$ a decreasing function. Hence, to maximise it, we need to choose a as small as possible. However, if we choose it so that $a < \max_{i=1}^n x_i$, i.e., if any of the x_i are bigger than a, then the likelihood is zero (as the event $x_i > a$ is impossible if all the X_i are distributed uniformly between 0 and a).

Hence, the likelihood is maximised by choosing,

$$\hat{a}_{MLE} = \max_{i=1}^{n} x_i.$$

We now compute the mean and variance of each of these estimators. Firstly, if U is uniformly distributed on $[0, a]$, then $\mathbb{E}[U] = a/2$, the density of U is given by f above, and

$$\text{Var(U)} = \int_{x=0}^{a} \left(x - \frac{a}{2}\right)^2 f(x)dx = \int_{y=-a/2}^{a/2} y^2 \frac{1}{a} dy = \frac{a^2}{12}.$$

As $\hat{a}_{MoME} = 2(X_1 + X_2 + \ldots + X_n)/n$, and $X_1, X_2, \ldots, X_n$ are iid with the same distribution as U, we get

$$\mathbb{E}[\hat{a}_{MoME}] = \frac{2}{n} \sum_{i=1} \mathbb{E}[X_i] = \frac{2}{n} n\mathbb{E}[U] = a,$$

which means that $\hat{a}_{MoME}$ is unbiased (its mean is equal to the true value of the parameter). Using the independence of the X_i, the variance of the estimator is

$$\text{Var}(\hat{a}_{MoME}) = \frac{4}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{4n}{n^2} \text{Var}(U) = \frac{a^2}{3n}.$$

Hence, the mean-squared error (MSE) of the estimator is

$$MSE(\hat{a}_{MoME}) = (\text{bias})^2 + \text{variance} = \frac{a^2}{3n}.$$

Next, we compute the density of the MLE. It is easier to compute the cdf. We have

$$\mathbb{P}(\hat{a}_{MLE} \leq y) = \mathbb{P}\left(\max_{i=1}^{n} X_i \leq y\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq y),$$

as the X_i are independent, and as their maximum is smaller than y if and only if each of them is smaller. Hence,

$$F(y) = \mathbb{P}(\hat{a}_{MLE} \leq y) = \begin{cases} 0, & y < 0, \\ (y/a)^n, & 0 \leq y \leq a, \\ 1, & y > 1. \end{cases}$$

Differentiating this, we obtain the density of the maximum-likelihood estimator to be

$$f_{MLE}(x) = nx^{n-1}/a^n,$$

on the interval $[0, a]$, and zero outside this interval. Consequently, its mean and second moment are given by

$$\mathbb{E}[\hat{a}_{MLE}] = \int_0^a x f_{MLE}(x)dx = \int_0^a \frac{nx^n}{a^n} dx = \frac{n}{n+1} a,$$

and

$$\mathbb{E}[\hat{a}_{MLE}^2] = \int_0^a x^2 f_{MLE}(x)dx = \int_0^a \frac{nx^{n+1}}{a^n} dx = \frac{n}{n+2} a^2.$$

Consequently, $\text{Var}(\hat{a}_{MLE}) = \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)a^2 = \frac{n}{(n+1)^2(n+2)} a^2$. Thus, we get,

$$\text{bias}(\hat{a}_{MLE}) = \mathbb{E}[\hat{a}_{MLE}] - a = -\frac{1}{n+1} a,$$

while

$$\mathrm{MSE}(\hat{a}_{MLE}) = \text{bias}^2 + \text{variance} = \frac{2n+2}{(n+1)^2(n+2)}a^2 = \frac{2}{(n+1)(n+2)}a^2.$$

Note that for large n, the MSE of the maximum likelihood estimator is approximately $2a^2/n^2$, i.e., it decays as $1/n^2$, whereas the MSE of the method of moments estimator only decays as $1/n$. Thus, in this problem, the MLE is asymptotically a much better estimator.

4. Following the hint and writing $X = Y + \alpha$, where Y has an Exp(λ) distribution, we note that the mean and variance of X are given by

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[Y] + \alpha = \frac{1}{\lambda} + \alpha, \\ \mathrm{Var}(X) &= \mathrm{Var}(Y) = \frac{1}{\lambda^2}. \end{aligned}$$

Equating these to the sample mean and variance, and solving the equations, we get the method-of-moments estimators,

$$\begin{aligned} \hat{\lambda}_{MoME} &= \Big(\frac{1}{n}\sum_{i=1}^n x_i^2 - \big(\frac{1}{n}\sum_{i=1}^n x_i\big)^2\Big)^{-1/2}, \\ \hat{\alpha}_{MoME} &= \frac{1}{n}\sum_{i=1}^n x_i - \frac{1}{\hat{\lambda}_{MoME}}. \end{aligned}$$

In order to compute the MLE, we write down the log-likelihood of the data as a function of the parameters (α, λ):

$$\ell(\alpha, \lambda; x_1, x_2, \ldots, x_n) = \begin{cases} n\log\lambda + n\lambda\alpha - \lambda\sum_{i=1}^n x_i, & \text{if } \alpha \le \min_{i=1}^n x_i, \\ -\infty, & \text{otherwise.} \end{cases}$$

Note that the latter case corresponds to the zero probability event that some x_i is smaller than α; as the likelihood is zero in this case, the log-likelihood is negative infinity.

Now, as λ is positive, the term $n\lambda\alpha$ is increasing. Hence, to maximise the log-likelihood with respect to α, we need to take α as large as possible, subject to the constraint $\alpha \le \min_{i=1}^n x_i$. This is clearly achieved by taking $\alpha = \min_{i=1}^n x_i$, irrespective of the value of λ. Thus, the MLE for α is given by

$$\hat{\alpha}_{MLE} = \min_{i=1}^n x_i.$$

Next, for this value of α, we carry out the maximisation over λ by setting the derivative of the log-likelihood function to zero. We get the equation

$$\frac{n}{\lambda} + n\min_{i=1}^n x_i - \sum_{i=1}^n x_i = 0,$$

whose solution is

$$\hat{\lambda}_{MLE} = \Big(\frac{1}{n}\sum_{i=1}^n x_i - \min_{i=1}^n x_i\Big)^{-1}.$$

5. Let us first consider the counter. The log-likelihood of observations $x_1, x_2, \ldots, x_n$ is given by

$$\ell(\lambda; x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} \big(x_i \log \lambda - \lambda + \log(x_i!)\big).$$

Setting the derivative with respect to λ and solving, we obtain the MLE for the counter, which we denote by $\hat{\lambda}_c$. We have

$$\hat{\lambda}_c = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

This is for a given data sequence $x_1, \ldots, x_n$. We model the data as iid random variables $X_1, \ldots, X_n$ with a Poisson(λ) distribution, which has mean λ and variance λ. Hence, the estimator $\hat{\lambda}_c$ has mean and variance given by

$$\begin{aligned} \mathbb{E}[\hat{\lambda}_c] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \lambda, \\ \text{Var}(\hat{\lambda}_c) &= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\lambda}{n}. \end{aligned}$$

In particular, it is unbiased, and so the MSE is the same as the variance, and equal to λ/n.

Next, we consider the detector. Let $y_1, y_2, \ldots, y_n$ denote the data seen by the detector, where the y_i are realisations of iid random variable Y_i, and $Y_i = 1$ if $X_i \geq 1$ and $Y_i = 0$ if $X_i = 0$. Since X_i are Poisson(λ), we see that $\mathbb{P}(Y_i = 0) = e^{-\lambda} = 1 - \mathbb{P}(Y_i = 1)$; in other words, Y_i is Bernoulli($1 - e^{-\lambda}$). We can now write down the log-likelihood function:

$$\ell(\lambda; y_1, y_2, \ldots, y_n) = \sum_{k=1}^{n} \Big(y_k \log(1 - e^{-\lambda}) - (1 - y_k)\lambda\Big).$$

Hence,

$$\frac{d\ell}{d\lambda} = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{k=1}^{n} y_k - \sum_{k=1}^{n} (1 - y_k).$$

Setting this equal to zero and simplifying, we get

$$e^{-\lambda} = \frac{1}{n} \sum_{k=1}^{n} (1 - y_k).$$

Thus, the MLE for the detector, which we denote by $\hat{\lambda}_d$, is given by

$$\hat{\lambda}_d = -\log\Big(\frac{1}{n} \sum_{k=1}^{n} (1 - y_k)\Big).$$

Please ignore the remainder of the question. The above estimator fails if the data is such that all y_i are equal to 1. This can happen with non-zero probability. But in this event, $\hat{\lambda}_d = -\log 0 = +\infty$, which is not a valid estimate. In this situation, we simply have no useful information on λ, other than that it has to be large, so there is no good estimation procedure using only the detector. It is true that the probability of this occuring decays rapidly as n increases, but it is always non-zero.