SOLUTION FOR HOMEWORK 5, STAT 6331

Welcome to your fifth homework. It is the beginning of our point-estimation part.

Reminder: if you find a mistake/misprint, do not e-mail or call me. Write it down on the first page of your solutions and you may give yourself a partial credit — but keep in mind that the total for your homeworks cannot exceed 20 points.

Now let us look at your problems.

1. Exerc.7.1. Note that $\hat{\theta}_{MLE} = \psi(X)$ meaning that it depends on a single outcome. Then for each value of X we choose the value of parameter that yields the largest likelihood (here $f(X|\hat{\theta}_{MLE}) := \max_{\theta \in \{1,2,3\}} f(X|\theta)$).

Answer: $\hat{\theta}_{MLE}$ is 1,1,(2 or 3),3,3 for X being 0,1,2,3,4, respectively.

2. Exerc. 7.3. In short, this is because likelihood is nonnegative and $\log(x)$ is a monotonically increasing function.

Nonetheless, let us present a rigorous solution. We want to show that if

$$\hat{\theta} := \text{argmax}_{\theta \in \Omega} L(\theta|\mathbf{X}), \quad \hat{\mu} := \text{argmax}_{\theta \in \Omega} \log(L(\theta|\mathbf{X})),$$

then $\hat{\theta} = \hat{\mu}$ a.s. Assume that the latter is wrong. Then

$$L(\hat{\theta}|\mathbf{X}) > L(\hat{\mu}|\mu), \quad \log(L(\hat{\theta}|\mathbf{X})) < \log(L(\hat{\mu}|\mu)).$$

But $\log(z)$ is the monotonically increasing function for $z > 0$ meaning that $z_1 < z_2$ iff $\log(z_1) < \log(z_2)$. The contradictory proves that $\hat{\theta} = \hat{\mu}$, and we can search after the MLE using either likelihood or loglikelihood.

3. Exers. 7.6. Given: X is distributed according to the pdf

$$f(x|\theta) = \theta x^{-2} I(0 < \theta \leq x < \infty).$$

Note how I wrote the pdf using the indicator function! This is a very important step!

Solution: (a) Write for a sample of size n,

$$f(\mathbf{x}|\theta) = \theta^n \prod_{l=1}^{n} x_l^{-2} I(0 < \theta \leq x_{(1)}).$$

By the Factorization Theorem $X_{(1)}$ is a SS.

(b) To find MLE we need to use graphic (this is always the case when parameter defines the support). From the graphic it is clear that $\hat{\theta}_{MLE} = X_{(1)}$. Note that it is our sufficient statistic!

(c) Let us calculate the expectation of X:

$$\mu = E(X) = \theta \int_{\theta}^{\infty} x x^{-2} dx = \infty.$$

As a result, there is no standard/classical method of moments estimator for the problem.

Remark: If you desire to get a generalized method of moments estimator, you can choose, for instance $E(X^{1/2})$ or something like this.

4. Exerc. 7.7 Let us look at where $1^n \geq \prod_{l=1}^n 1/[2x_l^{1/2}]$. This is equivalent to $\prod_{l=1}^n x_l \geq 2^{-2n}$. As a result, the MLE is 0 if $\prod_{l=1}^n X_l \geq 2^{-2n}$ and the MLE is 1 otherwise.

5. Exerc. 7.8. Given: $X \sim N(0, \sigma^2)$. The sample size $n = 1$.

(a). Clearly $\hat{\theta} = X^2$ is unbiased estimator of σ^2 because

$$E(\hat{\theta}) = E(X^2) = \mathrm{VarX} + E^2(X) = \sigma^2 + 0 = \sigma^2.$$

(b) Here

$$L(\sigma|X) = (2\pi\sigma^2)^{-1/2} e^{-X^2/(2\sigma^2)}.$$

It is more convenient to work with loglikelihood,

$$l(\sigma|X) = -(1/2)\log(2\pi) - (1/2)\log(\sigma^2) - X^2/(2\sigma^2) = -(1/2)\log(2\pi) - \log(\sigma) - X^2/(2\sigma^2).$$

Before taking the derivative, let us look at a graphic.

Now take derivative with respect to σ and get

$$dl(\sigma|X)/d\sigma = -\frac{1}{\sigma} + \frac{X^2}{\sigma^3}.$$

Equate to zero, remember that $\sigma > 0$, and the solution is $\sigma_{MLE} = |X|$. Then check, via graphic or second derivative, that this solution is indeed the maximum of the likelihood (do not forget about this step on your exams — you may not get a credit for solution if you do not check that the extreme point is indeed the maximum!).

(c) Let us begin with the first moment. We have $E_\sigma(X) = 0$, and because the population mean does not depend on σ, we cannot use the equality of the population and sample mean for estimation of σ. For the second moment the situation is better, and we have

$$E(X^2) = \sigma^2$$

which should be equated to X^2. Again, remember that $\sigma > 0$ so solution of

$$\sigma^2 = X^2$$

yields $\hat{\sigma}_{MME} = |X|$.

6. Exerc. 7.9. Here we have a sample of size n from a distribution with the pdf $f(x|\theta) = \theta^{-1}I(0 \le x \le \theta)$ and $\theta \in \Omega := (0, \infty)$.

(a) Method of Moments. Write

$$\mu_1 = E(X) = \theta^{-1}\int_0^\theta x dx = \theta/2.$$

As a result, the MME is defined as the solution of

$$\hat{\theta}_{MME}/2 = \bar{X}.$$

We get $\hat{\theta}_{MME} = 2\bar{X}$.

(b) Let us calculate the MLE. I may use $\underline{X} := \mathbf{X} := X^n$ to denote a sample. These are more popular notations used in the literature. Write

$$L(\theta|X^n) = \theta^{-n}I(0 \le X_{(1)})I(X_{(n)} \le \theta).$$

Look at the corresponding graphic.

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

(c) Now let us compare statistical properties of these two estimators. For the method of moments we have

$$E(\hat{\theta}_{MME}) = \theta,$$

(note that this we have by definition of the used MME). For iid observations,

$$\mathrm{Var}(\hat{\theta}_{MME}) = \mathrm{Var}(2\bar{X}) = 4n^{-2}\mathrm{Var}(\sum_{l=1}^n X_l)$$

$$= (4/n)\mathrm{Var}(X) = (4/n)[E(X^2) - (E(X))^2] = (4/n)[\theta^{-1}\int_0^\theta x^2 dx - (\theta/2)^2]$$

$$= (4/n)\theta^2[(1/3) - (1/4)] = n^{-1}\theta^2/3.$$

We conclude that

$$E(\hat{\theta}_{MME}) = \theta; \quad \mathrm{Var}(\hat{\theta}_{MME}) = n^{-1}[\theta^2/3]. \tag{1}$$

For the MLE we must begin with the distribution of $X_{(n)}$. Its cdf is, for $0 < x < \theta$,

$$F(x|\theta) = \mathrm{P}(X_{(n)} \le x) = \mathrm{P}(X_1 \le x, \dots, X_n \le x) = \prod_{l=1}^{n} \mathrm{P}(X_l \le x < \theta)$$

$$= [\theta^{-1} \int_0^x du]^n I(x < \theta) = \theta^{-n} x^n I(0 < x < \theta).$$

This yields the pdf

$$f_{X_{(n)}}(x) = dF(u|\theta)/du\Big|_{u=x} = n\theta^{-n} x^{n-1} I(0 < x < \theta)$$

Now we are in a position to make all calculations for the MLE. Write,

$$E(X_{(n)}) = \int_0^\theta n\theta^{-n} x^{n-1} x dx = n\theta^{-n} \int_0^\theta x^n dx$$

$$= n\theta^{-n}(n+1)^{-1}\theta^{n+1} = [n/(n+1)]\theta.$$

Further,

$$E(X_{(n)}^2) = \int_0^\theta n\theta^{-n} x^{n-1} x^2 dx = n\theta^{-n} \int_0^\theta x^{n+1} dx = [n/(n+2)]\theta^2.$$

We conclude that

$$\mathrm{Var}(\hat\theta_{MLE}) = E(\hat\theta_{MLE}^2) - [E(\hat\theta_{MLR})]^2 = [n/(n+2)]\theta^2 - [n/(n+1)]^2\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Let us combine for the MLE the results

$$E(\hat\theta_{MLE}) = \frac{n}{n+1}\theta, \quad \mathrm{Var}(\hat\theta_{MLE}) = \frac{n}{(n+2)(n+1)^2}\theta^2. \tag{2}$$

Now we are comparing (1) and (2). The MME is unbiased while the MLE is only asymptotically unbiased. On the other hand, the MLE has smaller variance. As a result, the choice is based on practical interests. To avoid such a dilemma, typically another risk function, the mean squared error, $E(\tilde\theta - \theta)^2$, is used to choose the better estimator.

7. Exerc. 7.10(a,b). Note that the cdf is given here, which can be written as

$$F(x|\alpha,\beta) = \min(1, (x/\beta)^\alpha) I(x > 0), \quad (\alpha,\beta) \in (0,\infty)^2.$$

Remark: can α be negative?

This is a continuous RV with the pdf (take accurately derivative of the cdf)

$$f(x|\alpha,\beta) = \beta^{-\alpha}\alpha x^{\alpha-1} I(0 \le x \le \beta).$$

Now we can solve the problems.

(a) To find a SS, via using the Factorization Theorem, we write

$$f(\mathbf{x}|\alpha,\beta) = \beta^{-n\alpha}\alpha^n e^{(\alpha-1)\sum_{l=1}^n \ln(x_l)} I(x_{(1)} \ge 0) I(x_{(n)} \le \beta).$$

By the FT the statistic $(T_1, T_2) := (\sum_{l=1}^n \ln(X_l), X_{(n)})$ is a two-dimensional SS.

(b) To find the MLE we write using the above-written pdf,

$$L(\alpha, \beta|\mathbf{X}) = \beta^{-n\alpha}\alpha^n e^{(\alpha-1)\sum_{l=1}^n \ln(X_l)} I(X_{(1)} \geq 0) I(X_{(n)} \leq \beta).$$

For a given $\alpha = \alpha_0$, we can graph $L(\alpha_0, \beta|\mathbf{X})$ as a function in β. Note that $\beta \geq 0$ by the assumption.

It follows from the graphic that, regardless of α_0, the MLE of β is $\hat{\beta}_{MLE} = X_{(n)}$. Now let $\beta = \beta_0$. The log-likelihood is

$$l(\alpha, \beta_0|\mathbf{X}) := \ln(L(\alpha, \beta_0|\mathbf{X}))$$
$$= \alpha[-n\ln(\beta_0) + T_1] + n\ln(\alpha) + g$$

where the term g does not depend on α.

Note that we are interested in a specific $\beta_0 = \hat{\beta}_{MLE} = X_{(n)}$, and this yields that

$$[-n\ln(X_{(n)}) + T_1] = [n(-\ln(X_{(n)}) + n^{-1}\sum_{l=1}^n \ln(X_l)] \leq 0$$

with the equality iff $\alpha = 0$. Thus we have the following graphic:

We have a global maximum (you can also get this by checking the second derivative). Let us find the extreme point via taking the derivative,

$$dl(\alpha, \beta_0|\mathbf{X})/d\alpha = -n\ln(\beta_0) + T_1 + n/\alpha.$$

Thus, the extreme point is the solution of the equation $-n\ln(\beta_0) + T_1 + n/\alpha = 0$ which yields

$$\hat{\alpha}_{MLE} = [\ln(\beta_0) - T_1/n]^{-1}.$$

Answer:

$$(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE}) = (\frac{1}{Z_{(n)} - \bar{Z}},\ X_{(n)}),$$

where $Z_l := \ln(X_l)$.

Remark: Look at how nicely the problem would look if we had taken the logarithmic transformation of the random variable X. Do you know why? Let us consider U with the pdf $f(u|\alpha) = \beta^{-\alpha}\alpha u^{\alpha-1}I(0 \le u \le \beta)$. Then $V := \ln(U)$ has the cdf

$$F_V(v) = F_U(e^v) = \int_0^{e^v} \beta^{-\alpha}\alpha u^{\alpha-1}du = \beta^{-\alpha}[e^{\alpha v} - 1]$$

which is the exponential cdf with the pmf

$$f_V(v) = \alpha\beta^{-1}e^{\alpha v}I(-\infty < v < \ln(\beta)).$$

(c) Here $\hat{\beta}_{MLE} = \ln(25)$, and you need a calculator to find $\hat{\alpha}_{MLE}$. I skip it.

8. Exerc. 7.11. Here

$$f(x|\theta) = \theta x^{\theta-1}I(0 \le x \le 1), \quad \theta \in \Omega := (0, \infty).$$

(a). Write

$$l(\theta|\mathbf{x}) = n\ln(\theta) + (\theta - 1)\sum_{l=1}^{n}\ln(X_l).$$

Taking derivative gives us $l'(\theta|\mathbf{x}) = n/\theta + \sum_{l=1}^{n}\ln(X_l)$, so the extrema is $\theta^* = -[n^{-1}\sum_{l=1}^{n}\ln(X_l)]^{-1}$.

Is this the point of maximum? Note that $-n^{-1}\sum_{l=1}^{n}\ln(X_l)) > 0$ because $X_l \in [0, 1]$. Thus the graphic of the log-likelihood looks like this:

Or, note that $l''(\theta|\mathbf{X}) = -n/\theta^2 < 0$ so we do have the maximum.

We conclude that $\hat{\theta}_{MLE} = -[n^{-1}\sum_{l=1}^{n}\ln(X_l)]^{-1}$.

(b) Method of moments:

$$\mu_1 = E(X) = \theta\int_0^1 x x^{\theta-1}dx = \frac{\theta}{\theta+1}x^{\theta+1}\Big|_0^1 = \frac{1}{1+1/\theta}.$$

This yields that

$$\frac{1}{1+\hat{\theta}_{MME}} = \bar{X}$$

and solving the equation yields $\hat{\theta}_{MME} = \bar{X}/[1-\bar{X}]$.

Look at how different the MLE and the MME are!

9. Exerc. 7.12. Given the pmf

$$p(x|\theta) = \theta^x (1-\theta)^{1-x)I(x\in\{0,1\}}), \quad \theta \in \Omega := [0, 1/2].$$

Note that this is Bernoulli pmf with the information that the probability of success is at most 1/2. We know that, in a general case of $\theta \in [0,1]$, the sample mean $\bar{X}$ is the MLE and MME. Let us see what a difference, if any, that prior makes.

(a). (i) We are searching after MME. Because $\mu_1 = E(X) = \theta$, we get $\hat{\theta}_{MME} = \bar{X}$ (no change). Note that this is always the case with a MME because Ω does not affect the estimator.

(ii) Consider MLE:

$$\mathrm{P}(\mathbf{X} = \mathbf{x}^n|\theta) = e^{\ln(\theta)\sum_{l=1}^n x_l + \ln(1-\theta)\sum_{l=1}^n (1-x_l)} = e^{n[\ln(\theta/(1-\theta))\bar{x} + \ln(1-\theta)]}.$$

Further, by taking logarithm we simplify a bit,

$$l(\theta|\mathbf{x}) = n\bar{x}\ln(\theta) + n\ln(1-\theta)(1-\bar{x}).$$

To find extreme points, let us take the derivative,

$$l'(\theta|\mathbf{x}) = n\bar{x}\theta^{-1} - n(1-\bar{x})(1-\theta)^{-1}.$$

This yields a single extreme point $\theta^* = \bar{x}$. Is it the maximum?

The second derivative is

$$l''(\theta|\mathbf{x}) = -n[\bar{x}\theta^{-2} + (1-\bar{x})(1-\theta)^{-2}] < 0.$$

Now is time for graphics.

Note that $\hat{\theta}_{MLE} \in [0, 1/2]$ by its definition because

$$\hat{\theta}_{MLE} := \mathrm{argmax}_{\theta\in\Omega} l(\theta|\mathbf{X}).$$

Answer: $\hat{\theta}_{MLE} = \min(0.5, \bar{X})$. Note how Ω affects the MLE. Also, do you see that the obtained MLE is a projection of our general MLE estimator $\bar{X}$ on Ω?

Clearly the MLE is better than the MME whenever $\bar{X} > 1/2$ because it closer to an underlying θ.

b) By definition, the mean squared error is

$$MSE = E_\theta(\hat{\theta} - \theta)^2. \tag{3}$$

(i) For the MME, because $n\bar{X}$ is a Binomial RV, namely $Binom(\theta, n)$, we get

$$E(\hat{\theta}_{MME} - \theta)^2 = E(\bar{X} - \theta)^2 = n^{-2}\mathrm{Var}(n\bar{X}) = n^{-2}[n\theta(1-\theta)] = n^{-1}\theta(1-\theta).$$

(ii) For the MLE the situation is more complicated,

$$E(\hat{\theta}_{MLE} - \theta)^2 = E\{(\bar{X} - \theta)^2 I(\bar{X} \le .5)\} + (.5 - \theta)^2 E\{I(\bar{X} > .5)\}.$$

Now I use the fact that $Y := n\bar{X}$ is $Binom(\theta, n)$. Write (I denote by $\lfloor n/2 \rfloor$ the rounded down $n/2$)

$$E_\theta(\hat{\theta}_{MLE} - \theta)^2 = E_\theta\{(Y/n - \theta)^2 I(Y \le n/2)) + (.5 - \theta)^2 \mathrm{P}_\theta(Y > n/2)$$

$$= \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}(k/n - \theta)^2 + (.5-\theta)^2 \sum_{k=\lfloor n/2 \rfloor + 1}^{n} \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}.$$

Here we can stop and say that a software should be used to find the number.
As I explained earlier, the MME is better than the MMR whenever $\bar{X} > .5$.

10. Exerc. 7.16. (e)(i). Given a sample of size n from the pdf

$$f(x|\theta) = e^{\theta h(x) - H(\theta)} g(x),$$

where $h(x) = dH(x)/dx$ and $h(x)$ is increasing in x.
To find the MLE, note that

$$l(\theta|\mathbf{x}) = \ln(f(\mathbf{x}|\theta) = \theta \sum_{l=1}^{n} h(x_l) - nH(\theta) + \sum_{l=1}^{n} \ln(g(x_l)).$$

Then

$$l'(\theta|\mathbf{x}) = \sum_{l=1}^{h}(X_l) - nh(\theta).$$

Because $h(z)$ is increasing, only one extrema θ^* exists, and

$$h(\theta^*) = n^{-1}\sum_{l=1}^{n} h(x + l).$$

Further, the inverse function h^{-1} exists, and we can continue that

$$\theta^* = h^{-1}(n^{-1}\sum_{l=1}^{n} h(x_l).$$

The only remaining issue is to verify that the extrema is the maximum. Note that you cannot take the second derivative because $h(\theta)$ may not be differentiable (this is not assumed). On the other hand, if h is differentiable then $l''(\theta|\mathbf{x}) = -nh'(\theta) < 0$ because $h(\theta)$ is increasing.

In general, please look at the above-written expression for $l'(\theta|\mathbf{x})$ and note that $l'(\theta|\mathbf{x})$ is negative for $\theta > \theta^*$ and $l'(\theta|\mathbf{x})$ is positive for $\theta < \theta^*$ because $h(\theta)$ is increasing function. This yields that the extrema is the maximum. Two graphics below highlight this conclusion.