Solutions for Homework 2

Problems

7a, 16a, 30, 48 (ignore bias question), 51, 53, 58abc

For 53c, also ask "is the probability density of the mle found in c) consistent with the theoretical results announced in class? Where do you think the problem might come from?"

Comment on Notation and the Exponential Distribution

If we write $exp(3)$ we usually mean e^3. There is little reason when using 3, but the former is clearer when you have a lot of stuff in the exponent. The symbol $exp(3)$ can also refer to the exponential distribution with parameter 3, but context should make it clear.

If we write $X \sim \exp(\lambda)$ what we mean that X follows an exponential distribution with parameter λ. ie)

$$X \quad \text{has density} \quad f(x|\lambda) = \lambda * exp(-\lambda x) \quad \text{for} \quad x > 0$$

We also recall that if $X \sim exp(\lambda)$ then $\quad \mathbb{E}(X) = \frac{1}{\lambda} \quad$ and $\quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$

Problem 7a

In order to use the method of moments, we first want to calculate $\mu_1 = \mathbb{E}(X)$ and see if we can express p in terms of μ_1. Referring to p. 117 of the text or the argument below, we see that for a geometric distribution we have

$$\mu_1 = \mathbb{E}(X) = \frac{1}{p}.$$

We solve for p and find that

$$p = \frac{1}{\mu_1}.$$

We then calculate our estimate of μ_1 and get $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$ and plug this in to get our method of moments estimate for p:

$$\hat{p}_{mom} = \frac{1}{\bar{X}}.$$

Alternative Argument to Page 117

Suppose $X \sim geo(p)$. We condition on the event $X = 1$

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(X|X=1) * \mathbb{P}(X=1) + \mathbb{E}(X|X \neq 1) * \mathbb{P}(X \neq 1) \\
&= 1 * p + (\mathbb{E}(X) + 1) * (1-p) \\
&= \mathbb{E}(X) * (1-p) + 1
\end{aligned}$$

Therefore $\mathbb{E}(X) = \frac{1}{p}$

Problem 16a

Let X have density $f(x|\sigma) = \frac{1}{2\sigma} * exp(-\frac{|x|}{\sigma})$

First, note that $\mathbb{E}(X) = 0$ since our density function is symmetric. This means that we will consider the second moment instead.

If we let $Z = |X|$, then we have $X \sim exp(\frac{1}{\sigma})$. Therefore

$$E(X^2) = E(|X|^2) = E(Z^2) = \text{Var}(Z) + (\mathbb{E}(Z))^2 = \sigma^2 + \sigma^2 = 2\sigma^2$$

We have shown that $\mu_2 = 2\sigma^2$. Hence the method of moment estimator of σ, $\hat{\sigma}_{mom}$, is chosen to satisfy

$$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 2\hat{\sigma}_{mom}^2 \; .$$

and therefore $\left(\frac{1}{2n}\sum_i X_i^2\right)^{1/2}$ is the method of moments estimator for σ.

Problem 30

We have that,

$$\mathbb{P}(X > d|\lambda) = e^{-d\lambda}$$
$$f(X = c|\lambda) = \lambda e^{-c\lambda} \quad for \quad \lambda < 10$$

So out joint hybrid function is

$$\lambda e^{-5\lambda} * \lambda e^{-3\lambda} e^{-10\lambda} = \lambda^2 exp(-18\lambda)$$

So our log-likelihood is

$$l(\lambda) = 2ln(\lambda) - 18\lambda$$

Taking the derivative and setting it equal to zero gives the MLE

$$\hat{\lambda}_{MLE} = 1/9$$

Problem 48 (ignore bias part)

Direct Calculation of Variance

Note that we can calculate this variance exactly. Recall that Y is the sum of n iid bernoulli random variables. A single bernoulli random variable has variance $p_0 - p_0^2$ (as I did in section), so $\text{Var}(Y) = n(p_0 - p_0^2)$, and therefore $\text{Var}(\frac{Y}{n}) = \frac{p_0 - p_0^2}{n}$

A longer argument to find the (asymptotic) Variance

Considering that $Y \sim \text{bin}(n, p_0)$, we know that we can write the likelihood function for Y as

$$L(Y, p_0) = \binom{n}{Y} p_0^Y (1 - p_0)^{n-Y}$$

and the log-likelihood as

$$\ell(p_0) = \log \binom{n}{Y} + Y \log p_0 + (n - Y) \log(1 - p_0).$$

Taking the derivative with respect to p_0 and setting it equal to 0, we get our maximum likelihood estimate of p_0 to be $\hat{p_0} = \frac{Y}{n}$. NOTE THE HAT!!!

To find the asymptotic variance of this maximum likelihood estimate, we calculate

$$-\mathbb{E}(\ell''(p_0)) = \frac{Y}{p_0^2} + \frac{n - Y}{(1 - p_0)^2} = \frac{np_0}{p_0^2} + \frac{n(1 - p_0)}{(1 - p_0)^2} = \frac{n}{p_0(1 - p_0)}$$

where we have used the fact that $\mathbb{E}(Y) = np_0$. This means that the asymptotic variance of our maximum likelihood estimate is $\frac{p_0 - p_0^2}{n}$.

Note that we only have one observation on Y. The MLE and the asymptotics refer to the i.i.d. variables that Y is a sum of.

Delta Method

Consider the random variable $Z = \frac{Y}{n}$ and note that it has mean p_0 and variance $\frac{p_0(1-p_0)}{n}$.

Consider the function: $h(z) = -\log z$. Then $h'(z) = 1/z$, so $h'(\mu_Z) = \frac{1}{p_0}$ Therefore,

$$Var(h(Z)) \approx (1/p_0)^2 * var(Z) = \frac{1}{p_0^2} * \frac{p_0(1 - p_0)}{n} = \frac{1 - p_0}{np_0}$$

Relative Efficiency

As in 8.5.3 Example B, the asymptotic variance of the MLE is $\frac{\lambda}{n}$.

CAUTION: Here I am talking about the MLE for n iid observations from a poisson distribution.

The relative efficiency of the MLE to the estimator proposed in the problem is approximately

$$\frac{1-p_0}{np_0} \div \frac{\lambda}{n} = \frac{1-p_0}{\lambda p_0} = \frac{e^{\lambda}-1}{\lambda} > 0 \quad since \quad \lambda > 0$$

Note that this function goes to 1 as λ goes to zero. This means that if the exponential had very high mean $1/\lambda$, then the two procedures have about the same variability.

However, if the mean is small, so λ is large, then the MLE does much better.

Problem 8.51a

First, given our i.i.d. sample, we can write the likelihood function as

$$f(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} \frac{1}{2} \exp -|x_i - \theta|$$

and our log-likelihood function as

$$l(\theta) = -n \log 2 - \sum_{i=1}^{n} |X_i - \theta|.$$

We notice that this function is maximized with respect to θ when the term $\sum_{i=1}^{n} |X_i - \theta|$ is minimized. That is, we want to find the value of $\hat{\theta}$ such that the sum of the distances from each point in our sample to $\hat{\theta}$ is a minimum.

Without loss of generality, imagine that our sample values are indexed in order, from lowest to highest. (In the case of ties, just number tied values consecutively.) Now, consider the value of $\sum_{i=1}^{n} |X_i - \hat{\theta}|$ where we set $\hat{\theta} = \text{med}\ \{X_i\} = X_{m+1}$. In this case, we can think of the sample values as appearing in corresponding pairs: the two outermost values, X_1 and X_{2m+1} form the first pair, the next pair in from the outside, X_2 and X_{2m} form the next pair, with only the median, X_{m+1} without a match. We can think of the sum of the distances between the members of each pair and the median as being simply the distance between the members of the pair, i.e.

$$|X_{2m+1} - X_{m+1}| + |X_1 - X_{m+1}| = (X_{2m+1} - X_{m+1}) + (X_{m+1} - X_1) = (X_{2m+1} - X_1).$$

This means we can write the following for the expression we are trying to minimize, when we set $\hat{\theta} = \text{med}\ \{X_i\} = X_{m+1}$:

$$\sum_{i=1}^{n} |X_i - X_{m+1}| = (X_{2m+1} - X_1) + (X_{2m} - X_2) + \ldots + (X_{2m+2} - X_{2m}).$$

Now I will show that for any $\hat{\theta} \neq X_{m+1}$ we have

$$\sum_{i=1}^{n} |X_i - \hat{\theta}| > (X_{2m+1} - X_1) + (X_{2m} - X_2) + \ldots + (X_{2m+2} - X_{2m}).$$

First, consider $\hat{\theta} > X_{m+1}$. There is some maximum value i for which $X_i \leq \hat{\theta} \leq X_{i+1}$, unless $\theta \geq X_{2m+1}$. Regardless, thinking again of our pairs of values, we can see that we can write $\sum_{i=1}^{n} |X_i - \hat{\theta}|$ as follows. Any pairs which have upper value greater than or equal to $\hat{\theta}$ contribute $X_{upper} - X_{lower}$ as in the case of the median above. However, any pairs which have upper limit less than $\hat{\theta}$ contribute not only $X_{upper} - X_{lower}$, but also $\hat{\theta} - X_{upper}$ to the sum as well. If no pairs have upper limit less than $\hat{\theta}$, but $\hat{\theta} > X_{m+1}$ we still have the addition of $\hat{\theta} - X_{m+1}$ to the sum, this proving that for any $\hat{\theta} > X_{m+1}$, we have

$$\sum_{i=1}^{n} |X_i - \hat{\theta}| > (X_{2m+1} - X_1) + (X_{2m} - X_2) + \ldots + (X_{2m+2} - X_{2m}).$$

A symmetric argument works for $\hat{\theta} < X_{m+1}$, thus completing the proof that $\hat{\theta} = \text{med}\,\{X_i\} = X_{m+1}$ is the maximum likelihood estimate of θ.

[Note: If this proof is not clear, try drawing a picture, using five sample points and reconstruct the argument referring to your picture.]

Problem 8.53

First, a couple of results about the uniform distribution on the interval $[0, \theta]$. The density function of the uniform distribution on $[0, \theta]$ is given by $f(x|\theta) = 1/\theta$ for any x between 0 and θ, and $f(x|\theta) = 0$ for any x not in this range. We see this by observing that we need $f(x|\theta)$ to be a constant c such that

$$\int_0^\theta c dx = 1.$$

Clearly, $c = 1/\theta$ works to make this true.

If X has uniform distribution on $[0, \theta]$ then we know that

$$\mathbb{E}(X) = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta}{2}.$$

Finally, we can find the variance of X by computing

$$\mathbb{E}(X^2) = \int_0^\theta \frac{x^2}{\theta} dx = \frac{\theta^2}{3}$$

and taking

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}.$$

This implies that if $X_1, \ldots, X_n$ are i.i.d uniform on $[0, \theta]$, then if we call $\bar{X} = \sum_{i=1}^{n} X_i/n$, $\operatorname{Var} \bar{X} = \frac{\theta^2}{12n}$.

a

From what was shown above, we know that $\mu_1 = \mathbb{E}(X) = \frac{\theta}{2}$, and plugging in the sample estimate of μ_1, $\hat{\mu}_1 = \bar{X}$, we find that the method of moments estimate for θ is

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

Since $\mathbb{E}(\bar{X}) = \theta/2$, we have $\mathbb{E}(\hat{\theta}_{MM}) = \theta$ (i.e. the method of moments estimate is unbiased) and $\operatorname{Var}(\hat{\theta}_{MM}) = 4 \operatorname{Var}(\bar{X}) = \theta^2/3n$.

b

There is a small trick to writing the joint density function in this case. While it may seem at first that we can just write

$$f(X_1, \ldots, X_n|\theta) = \frac{1}{\theta^n}$$

we have to remember that this is only the density if all the X_i fall between 0 and θ. Otherwise, the joint density will be 0. This means that the correct way of writing the likelihood function L is

$$L(X_1, \ldots, X_n, \theta) = \begin{cases} 1/\theta^n & \text{if } \forall i : \theta \geq X_i \\ 0 & \text{otherwise} \end{cases}.$$

Since $1/\theta^n$ increases as θ decreases, we'd should pick θ as small as possible to maximize the likelihood — without, of course, picking it below any of the X_i, since this would turn the likelihood to 0. Thus $\hat{\theta}_{MLE} = X_{(n)}$, the n^{th} order statistic of our data set, i.e. the highest value in our data set.

c

In order to find the probability density of the MLE, we will start by finding its cumulative density. Observe that for non-negative t,

$$\begin{aligned} F(t) &\equiv \mathbb{P}(X_{(n)} \leq t) \\ &= \mathbb{P}(X_1 \leq t, \ldots, X_n \leq t) \\ &= \prod_{i=1}^{n} P(X_i \leq t) \quad \text{(b/c the } X_i \text{ are independent.)} \\ &= \left(\frac{t}{\theta}\right)^n \quad \text{(b/c the } X_i \text{ are identically distributed.).} \end{aligned}$$

We have used the facts that in order for $X_{(n)}$ to be less than or equal to t, we need all our observed values to be less than or equal to t, and that the probability that any one of the variables is less than or equal to t is given by t/θ.

Differentiating $F(t)$ gives the MLE's density: $dF(t)/dt = f(t) = (n/\theta)(t/\theta)^{n-1}$ for $t \in [0, \theta]$, and $f(t) = 0$ otherwise.

Simple integration using $f(t)$, and then the identity $\operatorname{Var} X = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, gives

$$\mathbb{E}(\hat{\theta}_{MLE}) = \frac{n}{n+1}\theta \qquad \mathbb{E}(\hat{\theta}^2_{MLE}) = \frac{n}{n+2}\theta^2 \qquad \operatorname{Var}(\hat{\theta}_{MLE}) = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

This means that the bias is given by $\mathbb{E}(\hat{\theta}_{MLE}) - \theta = \theta/(n+1)$.

Mean squared error is variance plus the square of the bias, so

$$\begin{aligned}
\operatorname{MSE}(\hat{\theta}_{MLE}) &= \frac{n}{(n+2)(n+1)^2}\theta^2 + \left(\frac{\theta}{n+1}\right)^2 \\
&= \theta^2\left(\frac{n}{(n+2)(n+1)^2} + \frac{1}{(n+1)^2}\right) \\
&= \theta^2\left(\frac{2}{(n+2)(n+1)}\right).
\end{aligned}$$

Note that the MSE for the MLE is $O(n^{-2})$.

Since $\hat{\theta}_{MM}$ is unbiased, we know that its MSE is equal to its variance, which as we showed above, is $\theta^2/3n$. Note that it vanishes at rate $O(n^{-1})$, more slowly that the MSE for the MLE.

d

Simply correct $\hat{\theta}_{MSE}$ by multiplying it by $(n+1)/n$.

Additional Question

First, we will explicitly show that the asymptotic distribution of the MLE is exponential, not normal as we would expect. We have the following:

$$\begin{aligned}
\mathbb{P}(X_{(n)} \leq \theta - x) &= \left(\frac{\theta - x}{\theta}\right)^n \\
&= \left(1 - \frac{x}{\theta}\right)^n
\end{aligned}$$

Replacing x by x/n and rearranging terms gives

$$
\begin{aligned}
\mathbb{P}(\theta - X_{(n)} \geq x/n) &= \left(1 - \frac{x}{n\theta}\right)^n \\
&\simeq e^{-x/\theta} \quad (\text{ for large } n)
\end{aligned}
$$

This means that $n(\theta - X_{(n)})$ has an asymptotic distribution that is not normal, but is instead exponential, contrary to the theory of maximum likelihood estimators presented in class. Also, in this case, a scaling factor of n, and not $\sqrt{n}$ determines the rate of convergence.

The reason that our asymptotic theory breaks down in this case is that first, the support (the values on which our density function is non-zero) depends on our parameter θ in the sense that if x takes a value outside of $[0, \theta]$ the value of the density function at x is 0. Also, the smoothness conditions required of the density are violated by the uniform density function.

Problem 8.58abc

a

This question is very similar to Example A in Section 8.5.1. We have a multinomial distribution with three cells and with cell probabilities given by $(1-\theta)^2, 2\theta(1-\theta)$ and θ^2. This means we can write the likelihood function in terms of our observed cell counts, call them X_1, X_2 and X_3, with $X_1 + X_2 + X_3 = n$, as

$$
L(X_1, X_2, X_3, \theta) = \frac{n!}{X_1!X_2!X_3!}[(1-\theta)^2]^{X_1}[2\theta(1-\theta)]^{X_2}[\theta^2]^{X_3}
$$

and the log-likelihood function as

$$
\begin{aligned}
\ell(\theta) &= \log n! - \log(X_1!X_2!X_3!) + 2X_1 \log(1-\theta) + X_2 \log 2\theta(1-\theta) + 2X_3 \log \theta \\
&= \log n! - \log(X_1!X_2!X_3!) + (2X_1 + X_2)\log(1-\theta) + (X_2 + 2X_3)\log \theta + X_2 \log 2.
\end{aligned}
$$

Differentiating this with respect to θ and setting it equal to 0 gives

$$
-\frac{2X_1 + X_2}{1-\theta} + \frac{2X_3 + X_2}{\theta} = 0
$$

and solving for θ gives

$$
\begin{aligned}
\hat{\theta}_{MLE} &= \frac{2X_3 + X_2}{2X_1 + 2X_2 + 2X_3} \\
&= \frac{2X_3 + X_2}{2n}.
\end{aligned}
$$

Plugging in our values $n = 180, X_1 = 10, X_2 = 68$ and $X_3 = 112$, we get $\hat{\theta}_{MLE} = 0.768$.

b

In order to find the asymptotic variance of the MLE, we will compute $-\mathbb{E}(\ell''(\theta))$ and then find the asymptotic variance by taking its reciprocal.

$$\ell''(\theta) = -\frac{2X_1 + X_2}{(1-\theta)^2} - \frac{2X_3 + X_2}{\theta^2}$$

To take the expectation of this, we note that each X_i is binomial with parameters n and the respective cell probability. This means that we have

$$\begin{aligned}
\mathbb{E}(X_1) &= n(1-\theta)^2 \\
\mathbb{E}(X_2) &= n2\theta(1-\theta) \\
\mathbb{E}(X_3) &= n\theta^2.
\end{aligned}$$

Using these facts and performing some algebra, we find that

$$-\mathbb{E}(\ell''(\theta)) = \frac{2n}{\theta(1-\theta)}.$$

This means that the asymptotic variance of $\hat{\theta}_{MLE}$ is $\theta(1-\theta)/2n$.

c

We construct our confidence interval by taking

$$[\hat{\theta}_{MLE} - z(1-\alpha/2)s_{\hat{\theta}}, \hat{\theta}_{MLE} + z(1-\alpha/2)s_{\hat{\theta}}].$$

In this case, we have $z(1-\alpha/2) = z(0.995) = 2.56$ and $s_{\hat{\theta}} = 0.216$. This means our 99% confidence interval is given by $[0.713, 0.824]$.