PROBLEM 9.1.

(a) Find \( P(X > 4 \mid X > 3) \) when \( X \sim \mathcal{N}(2, 9) \).

Using the definition of conditional probability, \( P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} \), leads to:

\[
P(X > 4 \mid X > 3) = \frac{P(X > 4)}{P(X > 3)} = \frac{P\left( \frac{X - 2}{\sqrt{9}} > \frac{4 - 2}{\sqrt{9}} \right)}{P\left( \frac{X - 2}{\sqrt{9}} > \frac{3 - 2}{\sqrt{9}} \right)}
\]

\[
= \frac{P(Z > \frac{2}{3})}{P(Z > \frac{1}{3})} = \frac{1 - \Phi(2/3)}{1 - \Phi(1/3)} = \frac{1 - 0.7475}{1 - 0.6306} = 0.6834.
\]

(b) Find \( P(3X + 4Y < 5W) \) when \( X \sim \mathcal{N}(1, 1) \), \( Y \sim \mathcal{N}(2, 2) \), and \( W \sim \mathcal{N}(3, 3) \) are independent.

Let \( V = 3X + 4Y - 5W \); since \( X, Y, \) and \( W \) are independent, the mean and variance of \( V \) are easy to compute:

\[
E(V) = E(3X + 4Y - 5W) = 3E(X) + 4E(Y) - 5E(W) = 3(1) + 4(2) - 5(3) = -4;
\]

\[
\text{var}(V) = \text{var}(3X + 4Y - 5W) = 3^2\text{var}(X) + 4^2\text{var}(Y) + (-5)^2\text{var}(W) = 9(1) + 16(2) + 25(3) = 116.
\]

Furthermore, a key property of normal random variables is that any linear combination of independent normals is a normal. In this case, this fact implies that

\( V \sim \mathcal{N}(-4, 116) \).

So the question becomes:
\[ P(3X + 4Y < 5W) = P(V < 0) \]
\[ = P\left( \frac{V - (-4)}{\sqrt{116}} < \frac{0 - (-4)}{\sqrt{116}} \right) \]
\[ = P\left( Z < \frac{2}{\sqrt{29}} \right) \]
\[ = \Phi\left( \frac{2}{\sqrt{29}} \right) = 0.645. \]

(c) A normal RV with variance 4 is positive with probability 0.2. Find its mean.

\[ 0.2 = P(X > 0) \]
\[ = P\left( \frac{X - \mu}{\sqrt{4}} > \frac{0 - \mu}{\sqrt{4}} \right) \]
\[ = P(Z > -\mu/2) \]
\[ = P(-Z < \mu/2) \]
\[ = \Phi(\mu/2) \Rightarrow \mu = 2\Phi^{-1}(0.2) \]
\[ = 2(-0.8416) \]
\[ = -1.6832. \]

(d) A normal RV with mean 6 is bigger than 9 with probability 0.01. Find its variance.

\[ 0.99 = P(X < 9) \]
\[ = P\left( \frac{X - 6}{\sigma} < \frac{9 - 6}{\sigma} \right) \]
\[ = P(Z < 3/\sigma) \]
\[ = \Phi(3/\sigma) \Rightarrow 3/\sigma = \Phi^{-1}(0.99) \Rightarrow \sigma^2 = (3/\Phi^{-1}(0.99))^2 \]
\[ = 1.2896^2 \]
\[ = 1.663. \]
Problem 9.2. Consider the following joint pmf table for a pair of RV’s:

<table>
<thead>
<tr>
<th>x</th>
<th>y 0</th>
<th>y 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.5 - p</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>p</td>
</tr>
</tbody>
</table>

If we define \( p \) as the pmf value when both \( x \) and \( y \) are 1, then it follows that the remaining pmf value must be \( 0.5 - p \), since all values in the table must sum to one. Therefore, as a starting point, we can fill in the table with \( p \) and \( 0.5 - p \), as shown here:

We can now express the marginal pmf’s for \( X \) and \( Y \) in terms of this new variable \( p \), by adding the elements of the table column-wise (for \( X \)) and row-wise (for \( Y \)), yielding:

\[
X \sim \text{Bernoulli}(0.5) \quad \Rightarrow \quad E(X) = 0.5, \quad \text{var}(X) = 0.25
\]
\[
Y \sim \text{Bernoulli}(0.5 + p) \quad \Rightarrow \quad E(Y) = 0.5 + p, \quad \text{var}(Y) = (0.5 + p)(0.5 - p).
\]

Furthermore, the correlation between \( X \) and \( Y \) is easy to compute from the joint pmf:

\[
E(XY) = \sum_x \sum_y xy P_{X,Y}(x, y)
\]
\[
= (0)(0)P_{X,Y}(0, 0) + (0)(1)P_{X,Y}(0, 1)
\]
\[
+ (1)(0)P_{X,Y}(1, 0) + (1)(1)P_{X,Y}(1, 1)
\]
\[
= P_{X,Y}(1, 1)
\]
\[
= p.
\]

We can now express the normalized correlation coefficient as a function of \( p \), namely:

\[
\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}
\]
\[
= \frac{p - (0.5)(0.5 + p)}{\sqrt{(0.25)(0.5 + p)(0.5 - p)}}
\]
\[
= \frac{-(0.5 - p)}{\sqrt{(0.5 + p)(0.5 - p)}}
\]
\[
= \frac{-0.5 - p}{\sqrt{0.5 + p}}.
\]

(a) What is the largest possible value for \( \rho \)?

As \( p \) ranges from its smallest possible value of zero to its largest possible value (0.5), \( \rho = \frac{0.5 - p}{\sqrt{0.5 + p}} \) ranges from \(-1\) to \(0\). Therefore, the maximum possible value is \( \rho = 0 \).
(b) **Complete the table so that** $X$ **and** $Y$ **are uncorrelated** ($\rho = 0$).

We just found in part (a) that a value of $p = 0.5$ leads to $\rho = 0$. Therefore, the completed table looks like this:

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

(c) **Complete the table so that the correlation coefficient between** $X$ **and** $Y$ **is** $\rho = -0.5$.

Setting $\rho = -\frac{0.5 - p}{\sqrt{0.5 + p}} = -0.5$ and solving for $p$ yields $p = 0.3$.

**PROBLEM 9.3.** Find $E(\cos(Z))$ when $Z \sim \mathcal{N}(0, 1)$. [Hint: Euler’s relationship and the mgf for $Z$.]

The mgf for a standard normal is known to be $\phi_Z(s) = e^{s^2/2}$.

Euler’s relationship is $\cos(Z) = \frac{1}{2} e^{jZ} + \frac{1}{2} e^{-jZ}$

$$\Rightarrow E(\cos(Z)) = \frac{1}{2} E(e^{jZ}) + \frac{1}{2} E(e^{-jZ})$$

$$= \frac{1}{2} \phi_Z(j) + \frac{1}{2} \phi_Z(-j)$$

$$= \frac{1}{2} e^{j^2/2} + \frac{1}{2} e^{(-j)^2/2}$$

$$= \frac{1}{2} e^{-1/2} + \frac{1}{2} e^{-1/2}$$

$$= e^{-1/2}$$

**PROBLEM 9.4.**

(a) **Let** $\{U, V, W\}$ **be i.i.d.** $\sim \text{Exp}(1)$ **random variables. Find the normalized correlation coefficient** $\rho$ **between** $X = U + V$ **and** $Y = U + V + W$.

First, observe that $X \sim \text{Erlang}(1, 2)$, which implies that $E(X) = 2$ and $\text{var}(X) = 2$.

Second, observe that $Y \sim \text{Erlang}(1, 3)$, which implies that $E(Y) = 3$ and $\text{var}(Y) = 3$.

Before we can compute $\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$, we also need to know the correlation:

$$E(XY) = E(X(X + W))$$

$$= E(X^2) + E(XW)$$
\[ \begin{align*}
&= \text{var}(X) + (E(X))^2 + E(X) \cdot E(W) \\
&= 2 + 2^2 + 2 \cdot 1 \\
&= 8.
\end{align*} \]

Therefore,
\[ \rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \]
\[ = \frac{8 - (2)(3)}{\sqrt{(2)(3)}} \]
\[ = \frac{2}{\sqrt{6}}. \]

(b) Let \( \{U, V, W\} \) be i.i.d. \( \text{Geom}(0.1) \) random variables. Find the normalized correlation coefficient \( \rho \) between \( X = U + V \) and \( Y = U + V + W \).

First, observe that \( X \sim \text{NegBin}(p = 0.1, k = 2) \), which implies that 
\[ E(X) = k/p = 2/(0.1) = 20, \quad \text{and} \quad \text{var}(X) = kq/p^2 = 2(0.9)/(0.01) = 180. \]

First, observe that \( Y \sim \text{NegBin}(p = 0.1, k = 3) \), which implies that
\[ E(Y) = k/p = 3/(0.1) = 30, \quad \text{and} \quad \text{var}(Y) = kq/p^2 = 3(0.9)/(0.01) = 270. \]

Before we can compute \( \rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \), we also need to know the correlation:
\[ E(XY) = E(X(X + W)) \]
\[ = E(X^2) + E(XW) \]
\[ = \text{var}(X) + (E(X))^2 + E(X) \cdot E(W) \]
\[ = 180 + 20^2 + (20) \cdot (10) \]
\[ = 780. \]

Therefore,
\[ \rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} \]
\[ = \frac{780 - (20)(30)}{\sqrt{(180)(270)}} \]
\[ = \frac{2}{\sqrt{6}}. \]

(c) Let \( \{Z_1, Z_2, \ldots, Z_{90}\} \) be i.i.d. \( \mathcal{N}(0, 1) \). Find the normalized correlation coefficient \( \rho \) between \( X = (Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_{90}^2) \) and \( Y = (Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_{90}^2) \).

First, observe that \( X \sim \chi^2(n = 89) \), which implies that 
\[ E(X) = 89, \quad \text{and} \quad \text{var}(X) = 2(89) = 178. \]
Next, observe that $Y \sim \chi^2(n = 90)$, which implies that $E(Y) = 90$, and $\text{var}(Y) = 2(90) = 180$.

Before we can compute $\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$, we also need to know the correlation:

$$E(XY) = E(X(X + Z^2_{90}))$$

$$= E(X^2) + E(XZ^2_{90})$$

$$= \text{var}(X) + (E(X))^2 + E(X) \cdot E(Z^2_{90})$$

$$= 178 + 89^2 + (89)\cdot(1)$$

$$= 8188.$$ 

Therefore,

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

$$= \frac{8188 - (89)(90)}{\sqrt{(178)(180)}}$$

$$= \frac{178}{\sqrt{(178)(180)}}$$

$$= \frac{\sqrt{89}}{90}.$$ 

(d) **With $Y$ defined in part (c), find $P(Y > 100)$.**

Since $Y \sim \chi^2(n = 90)$ with a paramter $n$ that is significantly bigger than 20, the central-limit theorem kicks in and we can accurately approximate $Y$ using the normal pdf:

$$Y \sim \mathcal{N}(90, 180).$$

Therefore, the question becomes:

$$P(Y > 100) = P\left(\frac{Y - 90}{\sqrt{180}} > \frac{100 - 90}{\sqrt{180}}\right)$$

$$= P\left(Z > \frac{100 - 90}{\sqrt{180}}\right)$$

$$= P\left(Z > \frac{5}{\sqrt{9}}\right)$$

$$= P\left(Z > \frac{5}{3}\right)$$

$$= 1 - \Phi\left(\frac{5}{3}\right)$$

$$= 0.228.$$
**Problem 9.5.** Let $X \sim \mathcal{N}(0, 1)$ and let $Q \sim \text{Uniform}\{-1, 1\}$ be independent random variables. Note that $Q$ is discrete; it is equally likely to be either $-1$ or $+1$. Define a new random variable by the product $Y = QX$.

(a) Find the pdf for $Y$.

Start with the cdf:

$$F_Y(y) = P(Y < y) = P(QX < y) = P(QX < y \mid Q = 1)P(Q = 1) + P(QX < y \mid Q = -1)P(Q = -1)$$

$$= \frac{1}{2} P(X < y) + \frac{1}{2} P(-X < y)$$

$$= \frac{1}{2} \Phi(y) + \frac{1}{2} \Phi(y) = \Phi(y).$$

Since the cdf for $Y$ is $\Phi(y)$, it follows immediately that $Y \sim \mathcal{N}(0, 1)$ is a standard normal:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$ 

In other words, $X$ and $Y$ are identically distributed.

This makes intuitive sense: the factor of $Q$ has no impact with probability 1/2, and it flips the sign with probability 1/2. Since the pdf for $X$ is symmetric about zero to start with, it follows that the pdf for $QX$ will the same as that for $X$.

(b) Are $X$ and $Y$ uncorrelated? Explain.

Yes, by direct calculation:

$$E(XY) = E(XQX) = E(X^2)E(Q) = (1)(0) = 0 = E(X)E(Y) = (0)(0).$$

(c) Are $X$ and $Y$ independent? Explain.

No, since $f(y \mid x)$ is zero for all $y \neq \pm x$, whereas $f(y)$ is the standard normal pdf. The fact that $f(y \mid x) \neq f(y)$ is enough to prove that $X$ and $Y$ are not independent.
Are $X$ and $Y$ jointly normal? Explain.

No. A key property of a jointly normal pair of random variables is that all linear combinations are normal. This property clearly does not hold for $X$ and $Y$, because the sum $U = X + Y$ is clearly not a normal random variable:

$$U = X + Y = X + QX = X(1 + Q):$$

Since $Q = -1$ with probability $1/2$, it follows that $P(U = 0) = 1/2$.

In stark contrast, the probability that any normal random variable is zero is zero. In fact, the probability that any continuous RV is zero is zero. So $U$ is not even a continuous RV, let alone a normal RV.

Since this particular linear combination of $X$ and $Y$ is not a normal RV, it follows that $X$ and $Y$ are not jointly normal.

We thus have a situation where $X$ is normal, by itself, and $Y$ is normal, by itself, but the pair $X$ and $Y$ are not jointly normal.

**Problem 9.6.** You know a random variable $X$ is Uniform$(a, b)$, and you observe $n$ independent realizations $\{X_1, \ldots, X_n\}$, but you don’t know the parameters $a$ and $b$.

(a) Find the ML estimates $\hat{a}_{ML}$ and $\hat{b}_{ML}$ of the parameters $a$ and $b$.

The pdf for $X$ is $f(x) = 1/(b - a)$, when $a \leq x \leq b$. For other values of $x$, the pdf is zero. Therefore, the likelihood function is:

$$L(a, b) = f(x_1)f(x_2)f(x_3) \cdots f(x_n).$$

We want to find the values of $a$ and $b$ that maximize this likelihood function. We see that the likelihood function is the product of $n$ factors. If any of these factors is zero, the entire likelihood function is zero. Therefore, to maximize the likelihood, the values of $a$ and $b$ must ensure that each $f(x_i)$ is be nonzero. In other words, we must ensure that $a \leq x_i \leq b$ for each $i \in \{1, 2, \ldots, n\}$. Subject to this condition, the likelihood function is simple:

$$L(a, b) = 1/(b - a)^n$$

We can maximize this by inspection, without any calculus: Make $b$ as small as possible, and make $a$ as large as possible. Combining this result with the constraint that $a \leq x_i \leq b$ for each $i \in \{1, 2, \ldots, n\}$, we conclude that the ML estimates for $a$ and $b$ are:

$$\hat{a}_{ML} = \min\{X_1, \ldots, X_n\},$$
$$\hat{b}_{ML} = \max\{X_1, \ldots, X_n\}.$$
(b) Are the estimates of part (a) biased? If so, quantify the bias.

Let \( M = \max\{X_1, \ldots, X_n\} \). Start with the cdf:

\[
F_M(m) = P(M \leq m) = P(\max\{X_1, \ldots, X_n\} \leq m) \\
= P(X_1 \leq m) \text{ and } (X_2 \leq m) \ldots \text{ and } (X_n \leq m)) \\
= P(X_1 \leq m)P(X_2 \leq m) \cdots P(X_n \leq m) \\
= P(X_1 \leq m)^n \\
= \left(\frac{m-a}{b-a}\right)^n.
\]

The pdf for \( M \) can be found by differentiating:

\[
f_M(m) = \frac{n}{b-a} \left(\frac{m-a}{b-a}\right)^{n-1}.
\]

The mean is thus:

\[
E(M) = \int_a^b m f_M(m) \, dm = \frac{n}{(b-a)^n} \int_a^b m(b-a)^{n-1} \, dm.
\]

Change of variables: let \( t = m - a \). Then:

\[
E(M) = \frac{n}{(b-a)^n} \int_0^{b-a} (a + t) t^{n-1} \, dt \\
= \frac{na}{(b-a)^n} \int_0^{b-a} t^{n-1} \, dt + \frac{n}{(b-a)^n} \int_0^{b-a} t^n \, dt \\
= a + \frac{n(b-a)}{(n+1)} \\
= b - \frac{b-a}{(n+1)}.
\]

The bias of \( \hat{b}_{\text{ML}} = M \) is thus:

\[
\text{bias}(\hat{b}_{\text{ML}}) = E(\hat{b}_{\text{ML}}) - b \\
= b - \frac{b-a}{(n+1)} - b \\
= -\left(\frac{b-a}{n+1}\right).
\]

Since this is not zero, the ML estimator is biased. We see that the bias is negative, which means that the ML estimate tends to underestimate the true value of \( b \). However, the bias is small when \( n \) is large.

Similarly, let \( W = \min\{X_1, \ldots, X_n\} \). Start with the cdf:

\[
F_W(w) = P(W \leq w) = P(\min\{X_1, \ldots, X_n\} \leq w) \\
= 1 - P(\min\{X_1, \ldots, X_n\} > w)
\]
1 – P((X_1 > w) \text{ and } (X_2 > w) \ldots \text{ and } (X_n > w))
= 1 – P(X_1 > w)P(X_2 > w) \ldots P(X_n > w)
= 1 – P(X_1 > w)^n
= 1 – P(X_1 – a > w – a)^n
= 1 – (1 – \frac{w – a}{b – a})^n
= 1 – \left( \frac{b – w}{b – a} \right)^n.

The pdf for W can be found by differentiating:

\[ f_W(w) = \frac{n}{b – a} \left( \frac{b – w}{b – a} \right)^{n-1}. \]

The mean is thus:

\[ E(W) = \int_a^b w f_W(w)dw = \frac{n}{(b – a)^n} \int_a^b w(b – w)^{n-1}dw. \]

Change of variables: let \( t = b – w \). Then:

\[ E(W) = \frac{n}{(b – a)^n} \int_0^{b-a} (b – t)t^{n-1}dt
= \frac{nb}{(b – a)^n} \int_0^{b-a} t^{n-1}dt – \frac{n}{(b – a)^n} \int_0^{b-a} t^ndt
= b – \frac{n(b – a)}{(b – a)^n}
= a + \frac{b – a}{n + 1}. \]

The bias of \( \hat{\alpha}_{ML} = W \) is thus:

\[ \text{bias}(\hat{\alpha}_{ML}) = E(\hat{\alpha}_{ML}) – a
= a + \frac{b – a}{n + 1} – a
= \frac{b – a}{n + 1}. \]

The bias for \( \hat{\alpha}_{ML} = W \) is positive.
Note that it has the same magnitude as the bias for \( \hat{\delta}_{ML} \).
The method-of-moments (MOM) estimates for \( a \) and \( b \) are found by equating the sample mean with the true mean, and equating the sample variance with the true variance, and then solving these two equations for the two unknowns. (See sect. 7.4). Find the MOM estimates \( \hat{a}_{\text{MOM}} \) and \( \hat{b}_{\text{MOM}} \) of the parameters \( a \) and \( b \).

Equating the sample mean to the true mean yields the equation:

\[
\overline{X} = \frac{(a + b)}{2}
\]

\[
S^2 = \frac{(b - a)^2}{12}
\]

The MOM estimates for \( a \) and \( b \) can be found by solving these two equations for \( a \) and \( b \). The first equation implies that \( a = 2\overline{X} - b \). Substituting this into the second equation yields:

\[
(b - (2\overline{X} - b))^2/12 = S^2.
\]

Solving this for \( b \) yields:

\[
\hat{b}_{\text{MOM}} = \overline{X} + \sqrt{3}S.
\]

Plugging this back into the first equation, \( \overline{X} = \frac{(a + b)}{2} \), and solving for \( a \), yields:

\[
\hat{a}_{\text{MOM}} = \overline{X} - \sqrt{3}S.
\]

(d) Are the estimates from part (c) biased? Don’t try to quantify the bias.

Yes. The bias for \( \hat{b}_{\text{MOM}} \) is:

\[
\text{bias}(\hat{b}_{\text{MOM}}) = E(\hat{b}_{\text{MOM}}) - b
\]

\[
= E(\overline{X} + \sqrt{3}S) - b
\]

\[
= E(\overline{X}) + \sqrt{3}E(S) - b
\]

\[
= \frac{(a + b)}{2} + \sqrt{3}E(S) - b
\]

\[
= -(b - a)/2 + \sqrt{3}E(S).
\]

This bias will be zero if and only if:

\[
E(S) = \frac{\sqrt{\frac{(b-a)^2}{12}}}{n}.
\]

But we recognize this right-hand side as the true standard deviation \( \sigma \) of the Unif\((a, b)\) random variable, since the variance of such a RV is \( \sigma^2 = \frac{(b-a)^2}{12} \). Therefore, the bias will be zero if and only if:

\[
E(S) = \sigma.
\]

In other words, if and only if the sample standard deviation is an unbiased estimator for the standard deviation.

Although \( S^2 \) is an unbiased estimator for \( \sigma^2 \), \( S \) is not an unbiased estimator for \( \sigma \). We already noted this fact in class. But to convince yourself that it must be true, we need only
use the fact that the variance of any nontrivial RV must be positive. (It cannot be zero.) In particular, since $S$ is a nontrivial RV, its variance must be positive. In other words, we must have:

$$E(S^2) - (E(S))^2 > 0.$$ 

Since $S^2$ is an unbiased estimator, this reduces to:

$$\sigma^2 - (E(S))^2 > 0.$$ 

And this implies that:

$$E(S) < \sigma.$$ 

We haven’t quantified the magnitude of the bias, but at least we see that the bias is nonzero and negative.

(e) Compare the ML and MOM estimates for $a$ and $b$ when $n = 3$, and assuming the three observations are $X_1 = 2, X_2 = 4, X_3 = 6$.

The ML estimates are:

$$\hat{a}_{ML} = \min\{2, 4, 6\} = 2,$$

$$\hat{b}_{ML} = \max\{2, 4, 6\} = 6.$$ 

The same mean and sample variance of this sample is:

$$X = \frac{2 + 4 + 6}{3} = 4,$$

$$S^2 = \frac{(2 - 4)^2 + (4 - 4)^2 + (6 - 4)^2}{3 - 1} = \frac{4 + 0 + 4}{2} = 4.$$ 

The ML estimates are therefore:

$$\hat{a}_{MOM} = X - \sqrt{3}S = 4 - 2\sqrt{3} = 0.5359$$

$$\hat{b}_{MOM} = X + \sqrt{3}S = 4 + 2\sqrt{3} = 7.464.$$
Problem 9.7. You know that raindrops are arriving according to a Poisson process, but you don’t know the rate parameter \( \lambda \). In an effort to determine \( \lambda \), you buy an ACME THREE-DROP TIMER, a device with an internal stopwatch that starts at time zero whenever you press the RESET button, and the internal stopwatch automatically stops as soon as it detects 3 successive raindrops. The display shows the captured time until the RESET button is pressed.

Suppose you press RESET \( n \) times and observe the outcomes \( \{X_1, \ldots, X_n\} \).

The random variables \( \{X_1, \ldots, X_n\} \) are i.i.d. \( \sim \text{Erlang}(\lambda, 3) \), i.e., Erlang with parameters \( \lambda \) and \( k = 3 \).

(a) Find the ML estimate \( \hat{\lambda}_{ML} \) for \( \lambda \).

The pdf for \( X \sim \text{Erlang}(\lambda, 3) \) is:

\[
f(x) = \frac{\lambda^3}{2} x^2 e^{-\lambda x}.
\]

The likelihood function is then:

\[
L(\lambda) = f(x_1)f(x_2)f(x_3) \cdots f(x_n) = \frac{\lambda^{3n}}{2^n} (x_1x_2x_3 \cdots x_n)^2 e^{-\lambda(x_1 + x_2 + x_3 + \cdots + x_n)} = \frac{\lambda^{3n}}{2^n} P^2 e^{-\lambda Y},
\]

where we have introduced the product \( P \) and sum \( Y \), defined by:

\[
P = x_1x_2x_3 \cdots x_n,
\]

\[
Y = x_1 + x_2 + x_3 + \cdots + x_n.
\]

The log-likelihood function is then:

\[
\log L(\lambda) = 3n \log(\lambda) - n \log(2) + 2 \log(P) - \lambda Y.
\]

The derivative is:

\[
\frac{d}{d\lambda} \log L(\lambda) = \frac{3n}{\lambda} - Y.
\]

Setting this derivative to zero yields:

\[
\frac{3n}{\lambda} - Y = 0.
\]

Solving for \( \lambda \) yields:

\[
\hat{\lambda}_{ML} = \frac{3n}{Y}.
\]
This makes sense in light of the fact that $X$ is close to $E(X)$, and we know that $E(X) = 3/\lambda$.

(b) **Find the MOM estimate $\hat{\lambda}_{\text{MOM}}$ for $\lambda$.**

The MOM estimate for $\lambda$ is found by equating $X$ with $E(X) = 3/\lambda$:

\[
\bar{X} = \frac{3}{\lambda}.
\]

Solving for $\lambda$ yields:

\[
\hat{\lambda}_{\text{MOM}} = \frac{3}{\overline{X}}.
\]

We see that the MOM estimate is the same as the ML estimate.

(c) **Calculate the bias of each.**

The bias of both the ML and MOM estimates is:

\[
\text{bias}(\hat{\lambda}) = E(\hat{\lambda}) - \lambda
\]

\[
= 3E\left(\frac{1}{\overline{X}}\right) - \lambda.
\]

We need to find the mean of $1/\overline{X}$.

This is not as hard as it seems. Recall that $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

Each $X_i$ is Erlang($\lambda, 3$), which is the sum of 3 i.i.d. Exp($\lambda$) random variables.

Therefore, each $X_i/n$ is the sum of 3 i.i.d. Exp($n\lambda$) random variables. Therefore, the sample mean is the sum of $3n$ i.i.d. Exp($n\lambda$) random variables. In other words, the sample mean is an Erlang random variable with parameters $3n$ and $n\lambda$:

\[
\overline{X} \sim \text{Erlang}(n\lambda, 3n).
\]

Therefore, we find that the bias is:

\[
\text{bias}(\hat{\lambda}) = 3E\left(\frac{1}{\overline{X}}\right) - \lambda
\]

\[
= 3\left(\frac{n\lambda}{3n - 1}\right) - \lambda
\]

\[
= \frac{\lambda}{3n - 1}.
\]
The estimator is biased. The bias is positive, so \( \hat{\lambda} \) tends to overestimate \( \lambda \). However, the bias is small when \( n \) is large.

(d) Compare \( \hat{\lambda}_{ML} \) and \( \hat{\lambda}_{MOM} \) when \( n = 3 \), and \( X_1 = 0.02, X_2 = 0.03, X_3 = 0.04 \).

We saw that both estimators are the same:

\[
\hat{\lambda}_{ML} = \hat{\lambda}_{MOM} = \frac{3(3)}{0.01 + 0.03 + 0.04} = \frac{3}{0.03} = 100.
\]

Problem 9.8. In addition to the THREE DROP TIMER, ACME also offers a wide range of timers at different price points. They manufacture the ACME ONE DROP TIMER\(^\circ\), the ACME TWO DROP TIMER\(^\circ\), the ACME FOUR DROP TIMER\(^\circ\), etc. In fact, there is an ACME \( k \)-DROP TIMER\(^\circ\) for every positive integer \( k \). Suppose you know you have an ACME \( k \)-DROP TIMER\(^\circ\), but you are not sure of the value of \( k \). (The label has worn off.) In an effort to determine which product you own, you place the device in a rain with raindrops that are arriving according to a Poisson process with rate parameter \( \lambda = 60 \) drops/second. You then press \( \text{RESET} \) 3 times and observe the outcomes \( X_1 = 11.1, X_2 = 16.1, X_3 = 17.8 \).

The random variables \( \{X_1, X_2, X_3\} \) are i.i.d. \( \sim \operatorname{Erlang}(60, k) \), i.e., Erlang with parameters \( \lambda = 60 \) and \( k \).

(a) Find the ML estimate for \( k \).

The pdf for \( X \sim \operatorname{Erlang}(60, k) \) is:

\[
f( x ) = \frac{(60)^k}{(k-1)!} x^{k-1} e^{-60x}.
\]

The likelihood function is then:

\[
L(k) = f(x_1) f(x_2) f(x_3)
= \frac{(60)^k}{(k-1)!} x_1^{k-1} e^{-60x_1} \frac{(60)^k}{(k-1)!} x_2^{k-1} e^{-60x_2} \frac{(60)^k}{(k-1)!} x_3^{k-1} e^{-60x_3}
= \frac{(60)^{3k}}{(k-1)!^3} G^{3(k-1)} e^{-180\bar{X}},
\]

where we introduced the “geometric mean” \( G = (x_1 x_2 x_3)^{1/3} \) and the “arithmetic” sample mean \( \bar{X} = (x_1 + x_2 + x_3)/3 \).

The log-likelihood function is therefore:

\[
\ln L(k) = \ln \left( \frac{(60)^{3k}}{(k-1)!^3} G^{3k} e^{-180\bar{X}} \right)
= \ln \left( \frac{(60G)^{3k}}{(k-1)!^3} e^{-180\bar{X}} \right)
\]
\[ = 3k \ln(60G) - 3\ln((k - 1)!) - 3\ln(G) - 180X. \]
\[ = 3k \ln(882.42) - 3\ln((k - 1)!) - 8.06 - 2700. \]
\[ = 20.35k - 3\ln((k - 1)!) - 2708.06. \]

where we used the facts that \( G = (x_1 x_2 x_3)^{1/3} = 14.71 \) and \( \bar{X} = (x_1 + x_2 + x_3)/3 = 15. \)

Now we are faced with the following problem:

Find the value of \( k \) that maximizes the following log-likelihood function:

\[ \ln L(k) = 20.35k - 3\ln((k - 1)!) - 2708.06. \]

We cannot differentiate this with respect to \( k \) and set the result to zero and solve for \( k \), because that would give us most likely a noninteger value for \( k \). And \( k \) is constrained to be an integer in this problem. So we need a different approach. One simple approach is to simply evaluate the quantity \( \ln L(k) \) using a computer or calculator over a wide range of \( k \) values. Here is a plot of \( \ln L(k) \) as a function of \( k \):

The maximum is achieved at \( k = 884. \) So the ML estimate for \( k \) is \( \hat{k}_{\text{ML}} = 884. \)

(b) **Find the MOM estimate for \( k. \)**

Equating the sample mean \( \bar{X} = (1.1 + 16.1 + 17.8)/3 = 15 \) with the true mean \( \mu = k/60 \) for an Erlang(60, \( k) \) random variable yields the following equation:

\[ 15 = k/60. \]

Solving for \( k \) yields:

\[ \hat{k}_{\text{ML}} = 900. \]

We see that the ML and MOM estimators are not too far away in this example.