Applied Statistics Exercises: Exercises

Stéphane Rivaud

1 Maxmimum Likelihood Estimation

The exercises in this sections are here for you to practice your mathematical skills to derive formulas for maximum likelihood estimation with common probability distributions. It is not very demanding with statistical thinking, but rather a new playground to derive mathematical results.

1.1 Estimating Parameters of a Bernoulli Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with probability mass function (PMF):

$$f_X(x;p) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}, \quad 0 < p < 1.$$

a) Derive its theoretical mean and variances.

b) Write down the likelihood function $L(p)$ for the sample.

c) Derive the maximum likelihood estimator $\hat{p}$ of p.

d) Given the sample $\{1,0,1,1,0\}$, compute $\hat{p}$.

1.2 Estimating Parameters of a Binomial Distribution

Let $X_1, X_2, \ldots, X_m$ be a random sample from a Binomial distribution Binomial(n,p) with PMF:

$$f_X(x;p) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x \in \{0,1,\ldots,n\}, \quad 0 < p < 1.$$

a) Derive its theoretical mean and variances.

b) Write down the likelihood function $L(p)$ for the sample.

c) Derive the maximum likelihood estimator $\hat{p}$ of p.

d) Given the sample $\{3,4,2,5\}$ from a Binomial$(5,p)$ distribution, compute $\hat{p}$.

1.3 Estimating Parameters of a Geometric Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Geometric distribution with PMF:

$$f_X(x;p) = (1-p)^{x-1}p, \quad x \in \{1,2,3,\ldots\}, \quad 0 < p < 1.$$

a) Derive its theoretical mean and variances.

b) Write down the likelihood function $L(p)$ for the sample.

c) Derive the maximum likelihood estimator $\hat{p}$ of p.

d) Given the sample $\{2,1,3,2,5\}$, compute $\hat{p}$.

1.4 Estimating Parameters of a Poisson Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample from a Poisson distribution with PMF:

$$f_X(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0, 1, 2, \dots\}, \quad \lambda > 0.$$

a) Derive its theoretical mean and variances.

b) Write down the likelihood function $L(\lambda)$ for the sample.

c) Derive the maximum likelihood estimator $\hat{\lambda}$ of λ.

d) Given the sample $\{1, 0, 2, 3, 1\}$, compute $\hat{\lambda}$.

1.5 Estimating Parameters of a Uniform Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample from a Uniform distribution on the interval $[0, \theta]$, with PDF:

$$f_X(x;\theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta.$$

a) Derive its theoretical mean and variances.

b) Write down the likelihood function $L(\theta)$ for the sample.

c) Derive the maximum likelihood estimator $\hat{\theta}$ of θ.

d) Given the sample $\{1, 2, 1.5, 3\}$, compute $\hat{\theta}$.

1.6 Estimating Parameters of a Normal Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$ with PDF:

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

a) Write down the likelihood function $L(\mu, \sigma^2)$ for the sample.

b) Derive the maximum likelihood estimator $\hat{\mu}$ and $\hat{\sigma}^2$ for μ and σ^2.

c) Given the sample $\{1, 2, 1.5, 3\}$, compute $\hat{\mu}$ and $\hat{\sigma}^2$.

2 Transformation of Random Variables

2.1 Transformation $Y = 2X + 1$

Let X be a random variable following a Uniform distribution on the interval $[0, 1]$:

$$f_X(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Define $Y = 2X + 1$.

a) Find the cumulative distribution function (CDF) $F_Y(y)$ of the random variable Y.

b) Find the probability density function (PDF) $f_Y(y)$ of the random variable Y.

c) Calculate the expected value $\mathbb{E}[Y]$.

2.2 Transformation $Z = X^2$

Let X be a random variable following an Exponential distribution with parameter λ:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Define $Z = X^2$.

a) Find the cumulative distribution function (CDF) $F_Z(z)$ of the random variable Z.

b) Find the probability density function (PDF) $f_Z(z)$ of the random variable Z.

c) Calculate the expected value $\mathbb{E}[Z]$.

2.3 Both Transformations

Let X be a random variable following a Normal distribution with mean μ and variance σ^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Define $W = 2X + 1$ and $V = X^2$.

a) Find the expected value $\mathbb{E}[W]$.

b) Find the variance $\mathrm{Var}(W)$.

c) Calculate the expected value $\mathbb{E}[V]$ and the variance $\mathrm{Var}(V)$.

3 Introduction to the Method of Moments

Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with unknown parameters. The **method of moments** provides an alternative way to estimate these parameters based on the moments of the sample and the distribution. The *population* moments are the theoretical moment of the distribution, computed with the expectation operator. The *sample* moments are the moments estimated from the samples by replacing the expectation operator with an average over the samples.

3.1 Step 1: Understanding Moments

The k-th moment of a random variable X is defined as $\mathbb{E}[X^k]$. The first moment corresponds to the mean $\mathbb{E}[X]$, and the second moment is related to the variance.

a) What is the first sample moment, M_1, for a sample $X_1, X_2, \ldots, X_n$?

b) How is the first population moment (the mean μ) defined for a random variable?

3.2 Step 2: Estimating the Parameters Using Moments

The method of moments equates the sample moments to the population moments in order to estimate the parameters. Let's apply this method to the following example.

Consider a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution with the probability density function:

$$f_X(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

a) Derive the population mean (first moment) for the exponential distribution.

b) Write down the equation that sets the sample mean equal to the population mean.

c) Solve this equation for λ, the parameter of the distribution.

3.3 Step 3: Applying the Method of Moments

Given the following sample: $\{2, 3, 1, 4\}$,

a) Compute the sample mean.

b) Use the method of moments to estimate λ.

3.4 Step 4: Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) method finds the value of λ that maximizes the likelihood function.

a) Write down the likelihood function $L(\lambda)$ for the sample.

b) Take the natural logarithm of the likelihood function (the log-likelihood).

c) Differentiate the log-likelihood with respect to λ and solve for the maximum likelihood estimator $\hat{\lambda}_{MLE}$.

d) Using the sample $\{2, 3, 1, 4\}$, compute $\hat{\lambda}_{MLE}$.

3.5 Step 5: Comparison of the Two Estimators

Now compare the two estimators for λ based on the method of moments and the maximum likelihood estimation.

a) Are the two estimators identical? If not, why do you think they differ?

b) Which estimator do you think is more efficient in this case? (Hint: Consider which estimator uses the entire distribution.)

3.6 MoM vs MLE for a Discrete Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with probability mass function (PMF):

$$f_X(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0, 1, 2, \ldots\}, \quad \lambda > 0.$$

a) Use the method of moments to derive an estimator for λ.

b) Derive the maximum likelihood estimator $\hat{\lambda}$ for λ.

c) Given the sample $\{0, 1, 2, 3, 1\}$, compute both the method of moments estimator and the maximum likelihood estimator for λ.

3.7 MoM vs MLE for a Continuous Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Uniform distribution on the interval $[0, \theta]$ with probability density function (PDF):

$$f_X(x; \theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta.$$

a) Use the method of moments to derive an estimator for θ.

b) Derive the maximum likelihood estimator $\hat{\theta}$ for θ.

c) Given the sample $\{1, 2, 1.5, 3\}$, compute both the method of moments estimator and the maximum likelihood estimator for θ.

4 Composite Random Variables

4.1 Inference with a Transformed Random Variable

A manufacturing process produces parts where the diameter X of each part (in cm) follows an Exponential distribution with an unknown parameter λ:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

However, the observed measurement Y is a transformed version of X, given by the equation:

$$Y = 2X + 1.$$

The company needs to estimate λ based on a sample of observed values $Y_1, Y_2, \ldots, Y_n$.

1. Model the transformed random variable: Derive the probability density function (PDF) of Y.
2. Parameter Estimation using Maximum Likelihood:
   - Write down the likelihood function for λ based on the observed data $Y_1, Y_2, \ldots, Y_n$.
   - Derive the maximum likelihood estimator (MLE) for λ.
3. Compute the MLE for a sample: Given the sample $\{4.2, 5.1, 3.9, 4.7, 5.3\}$, compute the maximum likelihood estimate of λ.

4.2 Sum of Independent Random Variables

In this exercise, you will compute the cumulative distribution function (CDF) and the probability density function (PDF) of a random variable Z, which is the sum of two independent continuous random variables X and Y. You will show that the PDF of Z is the convolution of the PDFs of X and Y. Then, you will apply this to the sum of two independent random variables following a uniform distribution on the interval $[0, \theta]$.

4.2.1 CDF of the Sum of Independent Random Variables

Let X and Y be independent random variables with CDFs $F_X(x)$ and $F_Y(y)$, respectively. Define $Z = X + Y$.

1. **Show that** the CDF of Z, denoted by $F_Z(z)$, is given by:

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{\infty} F_Y(z - x) f_X(x)\, dx,$$

where $f_X(x)$ is the PDF of X and $F_Y(y)$ is the CDF of Y.

4.2.2 PDF of the Sum of Independent Random Variables

The PDF of Z, denoted by $f_Z(z)$, is the derivative of its CDF:

$$f_Z(z) = \frac{d}{dz} F_Z(z).$$

2. **Show that** the PDF of Z is the convolution of the PDFs of X and Y, i.e.,

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx.$$

4.2.3 Sum of Two Independent Uniform Random Variables

Let X and Y be independent random variables, both following a uniform distribution on the interval $[0, \theta]$, i.e., $X \sim \text{Uniform}(0, \theta)$ and $Y \sim \text{Uniform}(0, \theta)$. The PDF of a uniform random variable is:

$$f_X(x) = f_Y(y) = \frac{1}{\theta}, \quad 0 \leq x, y \leq \theta.$$

3. **Use the convolution formula** to compute the PDF of $Z = X + Y$, and determine its piecewise form for different intervals of z. Specifically, find $f_Z(z)$ for: - $0 \leq z \leq \theta$, - $\theta \leq z \leq 2\theta$.

5 Estimating the Variance of a Random Variable

In this exercise, you will estimate the variance of a random variable using different estimators and explore the bias that can arise when using the sample mean instead of the theoretical mean. Follow the steps below to answer the questions.

5.1 Estimating Variance using the Theoretical Mean

Let $X_1, X_2, \ldots, X_n$ be a sample of size n drawn from a population with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathbb{E}[(X-\mu)^2]$. The variance estimator using the theoretical mean is defined as:

$$\hat{\sigma}_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

1. **Show that** $\hat{\sigma}_\mu^2$ is an **unbiased estimator** of the population variance σ^2, i.e., show that $\mathbb{E}[\hat{\sigma}_\mu^2] = \sigma^2$.

5.2 Estimating Variance using the Sample Mean

Now, consider replacing the theoretical mean μ by the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. This gives the following estimator:

$$\hat{\sigma}_{\bar{X}}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

2. **Show that** $\hat{\sigma}_{\bar{X}}^2$ is a **biased estimator** of the population variance σ^2, and calculate its bias, i.e., show that:

$$\mathbb{E}[\hat{\sigma}_{\bar{X}}^2] = \sigma^2\left(1 - \frac{1}{n}\right).$$

5.3 Deriving the Unbiased Estimator without the Theoretical Mean

The unbiased estimator of the variance, which does not require the theoretical mean μ, is given by:

$$\hat{\sigma}_{unbiased}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

3. Derive this formula from the biased estimator $\hat{\sigma}_{\bar{X}}^2$ by correcting for the bias. Give (an intuitive) explanation for why replacing the theoretical mean μ by the sample mean $\bar{X}$ causes bias in the estimator.

6 MSE, Bias, and Variance of an Estimator

In this exercise, you will be introduced to the Mean Squared Error (MSE) of an estimator, as well as its bias and variance. You will show that the MSE can be decomposed as a sum of a variance term and a bias term, and you will apply this decomposition to the estimators of the variance from the previous exercise.

6.1 Defining the MSE of an Estimator

Let $\hat{\theta}$ be an estimator of a parameter θ. The Mean Squared Error (MSE) of $\hat{\theta}$ is defined as:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right].$$

1. Show that the MSE of any estimator $\hat{\theta}$ can be decomposed into the sum of its variance and the square of its bias, i.e.,

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \left(\mathrm{Bias}(\hat{\theta})\right)^2,$$

where $\mathrm{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$.

6.2 Applying the Bias-Variance Decomposition

Recall the two estimators of variance from the previous exercise

- The biased estimator $\hat{\sigma}^2_{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- The unbiased estimator $\hat{\sigma}^2_{unbiased} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

2. For each estimator, compute the bias and the variance, and use the bias-variance decomposition to calculate their MSEs.

6.3 Comparing the Two Estimators

3. Compare the MSEs of the biased and unbiased estimators. Under what circumstances might the biased estimator have a smaller MSE than the unbiased one? Discuss how the sample size n affects the comparison between the two estimators.

7 Spam Filtering using Naive Bayes Classifier

7.1 Bag of Words Encoding

Consider a vocabulary $\{w_1, w_2, \ldots, w_k\}$. Each word can appear in an email or not.

1. How many dimensions will the bag of words vector have for a vocabulary of size k?
2. Represent the email "buy now win big" with the vocabulary $\{$"buy", "now", "win", "big", "offer"$\}$.

7.2 Naive Bayes Classifier for Spam

Naive Bayes classifiers assume conditional independence of words given the class label S.

1. How many parameters do you need to store the Naive Bayes model for a vocabulary of size k?
2. Use Bayes' theorem to express $P(S = 1 \mid x)$ given the bag of words vector x.

7.3 Software Stack

Describe the process of classifying the email "limited offer, buy now" using the Naive Bayes classifier. Write pseudocode for the following steps:

1. Tokenize the email string.
2. Encode the email using bag of words.
3. Apply the Naive Bayes classifier.
4. Make a decision.

7.4 Criticism of the Model

Show that permuting the words in an email leads to the same classification result and explain why this happens due to the Naive Bayes assumption.

A Maxmimum Likelihood Estimation

A.1 Estimating Parameters of a Bernoulli Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with probability mass function (PMF):

$$f_X(x;p) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}, \quad 0 < p < 1.$$

1. **Derivation of the theoretical Mean**

   The mean (or expected value) of a discrete random variable X is given by:

   $$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x f_X(x;p),$$

   where $\mathbb{X} = \{0,1\}$ is the event space of the Bernoulli random variable and $f_X(x;p)$ is its probability mass function.

   Step 1: Write the expression for the expectation using the PMF.
   The expected value of X is:

   $$\mathbb{E}[X] = \sum_{x \in \{0,1\}} x f_X(x;p).$$

   Step 2: Substitute the values of x into the expression.
   Now substitute $x = 0$ and $x = 1$ into the sum:

   $$\mathbb{E}[X] = 0 \cdot f_X(0;p) + 1 \cdot f_X(1;p).$$

   This simplifies to:

   $$\mathbb{E}[X] = f_X(1;p).$$

   Step 3: Substitute the PMF into the expression.
   Now, substitute $f_X(1;p) = p$ into the equation:

   $$\mathbb{E}[X] = p.$$

   Thus, the mean of a Bernoulli random variable is:

   $$\mathbb{E}[X] = p.$$

2. **Derivation of the theoretical Variance**

   The variance of a discrete random variable X is defined as:

   $$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

   Step 1: Compute the variance using the alternate definition.
   We know that the variance can also be defined through the following formula:

   $$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

   Since $X^2 = X$ for a Bernoulli random variable (because X takes values in $\{0, 1\}$), we can directly compute $\mathbb{E}[X^2] = \mathbb{E}[X]$.

   Thus,

   $$\mathrm{Var}(X) = \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

   Step 2: Substitute the expression for $\mathbb{E}[X]$.
   We already know that:

   $$\mathbb{E}[X] = p.$$

Now, substitute $\mathbb{E}[X] = p$ into the variance formula:

$$\mathrm{Var}(X) = p - p^2.$$

Step 3: Simplify the expression.
The variance simplifies to:

$$\mathrm{Var}(X) = p(1-p).$$

Thus, the variance of a Bernoulli random variable is:

$$\mathrm{Var}(X) = p(1-p).$$

3. **Likelihood Function:**

   Given a random sample $X_1, X_2, \ldots, X_n$ from the Bernoulli distribution, the likelihood function $L(p)$ is the product of the individual PMFs:

   $$L(p) = \prod_{i=1}^{n} f_X(X_i; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}.$$

   This can be rewritten as:

   $$L(p) = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i}.$$

4. **Maximum Likelihood Estimator (MLE):**

   To find the MLE, we take the log-likelihood function:

   $$\ell(p) = \log L(p) = \sum_{i=1}^{n} X_i \log(p) + \left(n - \sum_{i=1}^{n} X_i\right)\log(1-p).$$

   Differentiate the log-likelihood function with respect to p:

   $$\frac{\partial \ell(p)}{\partial p} = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{n - \sum_{i=1}^{n} X_i}{1-p}.$$

   Setting this equal to zero and solving for p:

   $$\frac{\sum_{i=1}^{n} X_i}{p} = \frac{n - \sum_{i=1}^{n} X_i}{1-p},$$
   $$\Rightarrow p\left(n - \sum_{i=1}^{n} X_i\right) = \left(\sum_{i=1}^{n} X_i\right)(1-p),$$
   $$\Rightarrow pn = \sum_{i=1}^{n} X_i,$$
   $$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   Thus, the maximum likelihood estimator (MLE) for p is:

   $$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

5. **Computation of $\hat{p}$ for the Sample $\{1,0,1,1,0\}$:**

   Given the sample $\{1,0,1,1,0\}$, we have $n = 5$ and $\sum_{i=1}^{5} X_i = 3$. Therefore, the MLE for p is:

   $$\hat{p} = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{3}{5} = 0.6.$$

A.2 Estimating Parameters of a Binomial Distribution

- **Derivation of the Mean**

  The mean (or expected value) of a discrete random variable X is given by:

  $$\mathbb{E}[X] = \sum_{x \in \mathbb{X}} x f_X(x; p),$$

  where $\mathbb{X} = \{0, 1, \ldots, n\}$ is the event space of the Binomial random variable, and $f_X(x; p)$ is its probability mass function. Thus, the expected value is:

  $$\mathbb{E}[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}.$$

  This sum is ugly, and might be solved with smart calculus. However, for this one, we are going to be even smarter.

  Step 1: Model a Binomial variable as a sum of Bernoulli variables.
  If $X = X_1 + \ldots + X_n$ where the X_i are i.i.d. Bernoulli random variables of parameter p, then X follows a Binomial distribution with parameters n and p.

  Step 2: Use the linearity of expectation
  By substituting the above decomposition into the expectation we get:

  $$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

  Since X_i are i.i.d, we have that for all $i \in \{1, \ldots, n\}$

  $$\mathbb{E}[X_i] = p$$

  Thus, the mean of a Binomial random variable X is:

  $$\mathbb{E}[X] = np.$$

- **Derivation of the Variance**

  The variance of a discrete random variable X is defined as:

  $$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

  To compute the variance, we need $\mathbb{E}[X^2]$.

  Step 1: Compute the variance using the sum of Bernoulli variables.
  Since $X = X_1 + \ldots + X_n$, we can write:

  $$\mathbb{E}[X^2] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

  First, let's again use the fact that $X_i^2 = X_i$,

  $$\sum_{i=1}^{n} \mathbb{E}[X_i^2] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$$

  Furthermore, since X_i and X_j are **independent** for $i \neq j$, we have:

  $$\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = p^2$$

  We can thus write:

  $$\sum_{i \neq j} \mathbb{E}[X_i X_j] = n(n-1)p^2$$

and then

$$\mathbb{E}[X^2] = np + n(n-1)p^2$$

Step 2: Substitute the known results. Substitute the known values:

$$\begin{aligned}
\mathrm{Var}(X) &= np + n(n-1)p^2 - n^2p^2 \\
&= np + n^2p^2 - np^2 - n^2p^2 \\
&= np - np^2 \\
&= np(1-p)
\end{aligned}$$

Thus, the variance of a Binomial random variable is:

$$\mathrm{Var}(X) = np(1-p).$$

- **Likelihood Function $L(p)$**

  Given a random sample $X_1, X_2, \ldots, X_m$ from a Binomial distribution, the likelihood function is the product of the probability mass functions (PMFs) for each observation in the sample.

  The PMF for a Binomial distribution is:

  $$f_X(x;p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

  Step 1: Write the likelihood function.
  For a sample $\{x_1, x_2, \ldots, x_m\}$, the likelihood function $L(p)$ is:

  $$L(p) = \prod_{i=1}^{m} f_X(x_i;p).$$

  Substituting the Binomial PMF:

  $$L(p) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}.$$

  Optional: Simplify the likelihood function.
  Since $\binom{n}{x_i}$ is constant with respect to p, we can ignore it for maximization:

  $$L(p) \propto \prod_{i=1}^{n} p^{x_i} (1-p)^{n-x_i}.$$

  This simplifies to:

  $$L(p) \propto p^{\sum_{i=1}^{m} x_i} (1-p)^{m \cdot n - \sum_{i=1}^{m} x_i}.$$

  Such a form is handy for deriving the Maximum Likelihood Estimate.

- **Maximum Likelihood Estimator (MLE) for p**
  Step 1: Write the log-likelihood function.
  The log-likelihood function is easier to work with and is given by:

  $$\ell(p) = \log L(p).$$

  Taking the log of the simplified likelihood:

  $$\ell(p) = \sum_{i=1}^{m} x_i \log p + (m \cdot n - \sum_{i=1}^{m} x_i) \log(1-p).$$

  Step 2: Differentiate the log-likelihood function with respect to p.
  To find the MLE of p, we differentiate $\ell(p)$ with respect to p:

  $$\frac{d\ell(p)}{dp} = \frac{\sum_{i=1}^{m} x_i}{p} - \frac{m \cdot n - \sum_{i=1}^{m} x_i}{1-p}.$$

Step 3: Set the derivative equal to zero and solve for p.
Set $\frac{d\ell(p)}{dp} = 0$:

$$\frac{\sum_{i=1}^{m} x_i}{p} = \frac{m \cdot n - \sum_{i=1}^{n} x_i}{1-p}.$$

Multiplying both sides by $p(1-p)$ and solving for p, we get:

$$p = \frac{\sum_{i=1}^{m} x_i}{m \cdot n}.$$

Thus, the MLE for p is:

$$\hat{p} = \frac{\sum_{i=1}^{m} x_i}{m \cdot n}.$$

- **Compute $\hat{p}$ for the Sample $\{3, 4, 2, 5\}$**

  Given the sample $\{3, 4, 2, 5\}$ from a Binomial$(5, p)$ distribution, we can calculate the MLE for p.

  Step 1: Compute the sum of the sample values.
  The sum of the sample values is:

$$\sum_{i=1}^{n} x_i = 3 + 4 + 2 + 5 = 14.$$

  Step 2: Compute $\hat{p}$.
  Using the formula $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n \cdot n}$, with $n = 5$ (the number of trials for each observation) and the sample size $n = 4$, we get:

$$\hat{p} = \frac{14}{4 \cdot 5} = \frac{14}{20} = 0.7.$$

  Thus, the MLE estimate for p is:

$$\hat{p} = 0.7.$$

A.3 Estimating Parameters of a Geometric Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Geometric distribution with PMF:

$$f_X(x; p) = (1-p)^{x-1} p, \quad x \in \{1, 2, 3, \ldots\}, \quad 0 < p < 1.$$

- **Derive the Theoretical Mean and Variance**

  Step 1: Recall the definition of the expectation for a discrete random variable.

  The theoretical mean $\mathbb{E}[X]$ of a discrete random variable X is given by:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x f_X(x; p).$$

  Substituting the PMF of the Geometric distribution:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x(1-p)^{x-1} p.$$

  Step 2: Rewrite the series

$$\mathbb{E}[X] = p \sum_{x=1}^{\infty} x(1-p)^{x-1}$$

  Now consider the following formal Taylor serie:

$$\sum_{i=1}^{+\infty} k \cdot x^{k-1} = \frac{d}{dx}\left(\sum_{k=0}^{+\infty} x^k\right)$$

Furthermore:

$$\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$$

This can be proven by exploiting the following functional equation:

$$\sum_{k=0}^{+\infty} x^k = 1 + x \cdot \sum_{k=0}^{+\infty} x^k$$

and solving for $\sum_{k=0}^{+\infty} x^k$.
We thus have that:

$$\sum_{i=1}^{+\infty} k \cdot x^{k-1} = \frac{d}{dx}(\frac{1}{1-x}) = \frac{1}{(1-x)^2}$$

We can now substitute into our equation for expectation:

$$\begin{aligned} \mathbb{E}[X] &= p \cdot \frac{1}{(1-(1-p))^2} && (1) \\ &= \frac{1}{p} && (2) \end{aligned}$$

Derive the variance.

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Deriving the variance requires careful manipulation of Taylor series.

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{x=1}^{+\infty} x^2(1-p)^{x-1}p \\ &= \sum_{x=1}^{+\infty} x(x-1)(1-p)^{x-1}p + \sum_{x=1}^{+\infty} x(1-p)^{x-1}p \\ &= (1-p)p\sum_{x=2}^{+\infty} x(x-1)(1-p)^{x-2} + \sum_{x=1}^{+\infty} x(1-p)^{x-1}p \end{aligned}$$

Using similar arguments about Taylor series, we have that

$$\sum_{k=2}^{+\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

while

$$\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

Substituting into the above equations yields:

$$\mathbb{E}[X^2] = \frac{1-p}{p^2} + \frac{1}{p^2}$$

Finally,

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{1-p}{p^2} + \frac{1}{p^2} - \frac{1}{p^2} \\ &= \frac{1-p}{p^2} \end{aligned}$$

- **Write down the likelihood function $L(p)$ for the sample.**

  Given a random sample $X_1, X_2, \ldots, X_n$, the likelihood function is the product of the probability mass functions (PMFs) for each observation in the sample:

  $$L(p) = \prod_{i=1}^{n} f_X(x_i; p).$$

  Substitute the PMF of the Geometric distribution:

  $$L(p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = p^n (1-p)^{\sum_{i=1}^{n}(x_i - 1)}.$$

  Optional: Simplify the likelihood function.

  Simplifying the expression gives:

  $$L(p) = p^n (1-p)^{\sum_{i=1}^{n} x_i - n}.$$

- **Derive the Maximum Likelihood Estimator $\hat{p}$ of p**

  Step 1: Write the log-likelihood function.

  The log-likelihood function $\ell(p)$ is easier to work with and is given by:

  $$\ell(p) = \log L(p) = n \log p + \left(\sum_{i=1}^{n} x_i - n \right) \log(1-p).$$

  Step 2: Differentiate the log-likelihood function with respect to p.

  To find the maximum likelihood estimator (MLE) of p, we differentiate $\ell(p)$ with respect to p:

  $$\frac{d\ell(p)}{dp} = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p}.$$

  Step 3: Set the derivative equal to zero and solve for p.

  Set $\frac{d\ell(p)}{dp} = 0$:

  $$\frac{n}{p} = \frac{\sum_{i=1}^{n} x_i - n}{1-p}.$$

  Multiplying both sides by $p(1-p)$ and solving for p, we get:

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

  Thus, the MLE for p is:

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

- **Compute $\hat{p}$ for the Sample $\{2, 1, 3, 2, 5\}$**

  Given the sample $\{2, 1, 3, 2, 5\}$, we can calculate the MLE for p.

  Step 1: Compute the sum of the sample values.

  The sum of the sample values is:

  $$\sum_{i=1}^{n} x_i = 2 + 1 + 3 + 2 + 5 = 13.$$

  Step 2: Compute $\hat{p}$.

  Using the formula $\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$, where $n = 5$, we get:

  $$\hat{p} = \frac{5}{13} \approx 0.3846.$$

  Thus, the MLE estimate for p is:

  $$\hat{p} \approx 0.3846.$$

A.4 Estimating Parameters of a Poisson Distribution

Let $X_1, X_2, \dots, X_n$ be a random sample from a Poisson distribution with PMF:

$$f_X(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0,1,2,\dots\}, \quad \lambda > 0.$$

a) **Deriving the theoretical mean and variance of a Poisson distribution:**

The expectation (mean) of a Poisson-distributed random variable X is given by:

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}.$$

To compute this, we can break it down as follows:

First, rewrite $x \cdot \frac{\lambda^x}{x!}$ as:

$$x \cdot \frac{\lambda^x}{x!} = \lambda \cdot \frac{\lambda^{x-1}}{(x-1)!}, \quad x \geq 1.$$

Substituting this into the sum:

$$\mathbb{E}[X] = \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}.$$

Now, let $y = x - 1$, so the sum becomes:

$$\mathbb{E}[X] = \lambda \sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!}.$$

Recognize that this is the sum of the Poisson PMF, which equals 1:

$$\mathbb{E}[X] = \lambda \cdot 1 = \lambda.$$

Hence, the theoretical mean of a Poisson-distributed random variable is λ.

Now, for the variance, we use the formula for variance $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. First, compute $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \sum_{x=0}^{\infty} x^2 \cdot \frac{\lambda^x e^{-\lambda}}{x!}.$$

Break this into:

$$\mathbb{E}[X^2] = \sum_{x=0}^{\infty} (x(x-1) + x) \cdot \frac{\lambda^x e^{-\lambda}}{x!}.$$

This splits into two sums:

$$\mathbb{E}[X^2] = \sum_{x=2}^{\infty} x(x-1) \cdot \frac{\lambda^x e^{-\lambda}}{x!} + \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!}.$$

The first sum becomes:

$$\sum_{x=2}^{\infty} x(x-1) \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \lambda^2,$$

and the second sum is just $\mathbb{E}[X] = \lambda$. Therefore:

$$\mathbb{E}[X^2] = \lambda^2 + \lambda.$$

Finally, the variance is:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Hence, the theoretical variance of a Poisson-distributed random variable is also λ.

b) **Writing the likelihood function $L(\lambda)$ for the sample:**

The likelihood function for a sample $X_1, X_2, \ldots, X_n$ from a Poisson distribution is the product of the individual PMFs:

$$L(\lambda) = \prod_{i=1}^{n} f_X(x_i; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}.$$

Simplifying the product:

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}.$$

The likelihood function is:

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{x_1! x_2! \cdots x_n!}.$$

c) **Deriving the maximum likelihood estimator $\hat{\lambda}$ of λ:**

To find the MLE, we first take the natural logarithm of the likelihood function to obtain the log-likelihood:

$$\log L(\lambda) = \sum_{i=1}^{n} x_i \log(\lambda) - n\lambda - \sum_{i=1}^{n} \log(x_i!).$$

Differentiating the log-likelihood with respect to λ:

$$\frac{d}{d\lambda} \log L(\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n.$$

Setting this equal to zero to find the critical point:

$$\frac{\sum_{i=1}^{n} x_i}{\lambda} = n,$$

which simplifies to:

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

Therefore, the maximum likelihood estimator of λ is the sample mean $\hat{\lambda} = \bar{X}$.

d) **Computing $\hat{\lambda}$ for the sample $\{1, 0, 2, 3, 1\}$:**

For the given sample $\{1, 0, 2, 3, 1\}$, we first compute the sample mean:

$$\bar{X} = \frac{1 + 0 + 2 + 3 + 1}{5} = \frac{7}{5} = 1.4.$$

Therefore, the maximum likelihood estimate for λ is:

$$\hat{\lambda} = 1.4.$$

A.5 Estimating Parameters of a Uniform Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Uniform distribution on the interval $[0, \theta]$, with PDF:

$$f_X(x; \theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta.$$

a) **Deriving the theoretical mean and variance of a Uniform distribution:**

The expectation (mean) of a uniformly distributed random variable X over $[0, \theta]$ is given by:

$$\mathbb{E}[X] = \int_0^{\theta} x \cdot \frac{1}{\theta} \, dx.$$

Perform the integration:

$$\mathbb{E}[X] = \frac{1}{\theta} \int_0^{\theta} x \, dx = \frac{1}{\theta} \left[\frac{x^2}{2} \right]_0^{\theta} = \frac{1}{\theta} \cdot \frac{\theta^2}{2} = \frac{\theta}{2}.$$

Hence, the theoretical mean of a uniform distribution is $\frac{\theta}{2}$.

Now, to compute the variance $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, we first calculate $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \int_0^\theta x^2 \cdot \frac{1}{\theta}\, dx.$$

Perform the integration:

$$\mathbb{E}[X^2] = \frac{1}{\theta}\int_0^\theta x^2\, dx = \frac{1}{\theta}\left[\frac{x^3}{3}\right]_0^\theta = \frac{1}{\theta}\cdot\frac{\theta^3}{3} = \frac{\theta^2}{3}.$$

Thus, the variance is:

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

Simplify:

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

Hence, the theoretical variance of a uniform distribution is $\frac{\theta^2}{12}$.

b) **Writing the likelihood function $L(\theta)$ for the sample:**

The likelihood function for a sample $X_1, X_2, \ldots, X_n$ from a uniform distribution is:

$$L(\theta) = \prod_{i=1}^n f_X(x_i;\theta) = \prod_{i=1}^n \frac{1}{\theta}, \quad 0 \le x_i \le \theta.$$

Simplifying the product:

$$L(\theta) = \frac{1}{\theta^n}, \quad \text{for } \theta \ge \max(x_1, x_2, \ldots, x_n).$$

The likelihood function is zero if $\theta < \max(x_1, x_2, \ldots, x_n)$.

c) **Deriving the maximum likelihood estimator $\hat{\theta}$ of θ:**

To find the maximum likelihood estimator (MLE) of θ, note that the likelihood function $L(\theta) = \frac{1}{\theta^n}$ is decreasing in θ, so the likelihood is maximized when θ is as small as possible, but still greater than or equal to all the sample values. Therefore, the MLE of θ is:

$$\hat{\theta} = \max(X_1, X_2, \ldots, X_n).$$

d) **Computing $\hat{\theta}$ for the sample $\{1, 2, 1.5, 3\}$:**

For the given sample $\{1, 2, 1.5, 3\}$, the maximum value is:

$$\hat{\theta} = \max(1, 2, 1.5, 3) = 3.$$

Therefore, the maximum likelihood estimate for θ is:

$$\hat{\theta} = 3.$$

A.6 Estimating Parameters of a Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$ with PDF:

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

a) **Writing the likelihood function $L(\mu, \sigma^2)$ for the sample:**

The likelihood function for a sample $X_1, X_2, \ldots, X_n$ from a Normal distribution is the product of the individual densities:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f_X(x_i; \mu, \sigma^2).$$

Substituting the PDF of the Normal distribution:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

Simplifying the product:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

This is the likelihood function for the parameters μ and σ^2.

b) **Deriving the maximum likelihood estimator $\hat{\mu}$ and $\hat{\sigma}^2$:**

To maximize the likelihood function, we first take the natural logarithm of the likelihood function to obtain the log-likelihood:

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2) = \log\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\right).$$

This simplifies to:

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Next, we differentiate the log-likelihood function with respect to μ and set the derivative equal to zero to find $\hat{\mu}$. Differentiating with respect to μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu).$$

Setting this equal to zero:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0.$$

Solving for $\hat{\mu}$:

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

Thus, the maximum likelihood estimator for μ is the sample mean:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Now, differentiate the log-likelihood with respect to σ^2 and set the derivative equal to zero to find $\hat{\sigma}^2$. Differentiating with respect to σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Setting this equal to zero:

$$-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0.$$

Solving for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

Thus, the maximum likelihood estimator for σ^2 is the sample variance:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

c) **Computing $\hat{\mu}$ and $\hat{\sigma}^2$ for the sample $\{1, 2, 1.5, 3\}$:**

First, compute the sample mean $\hat{\mu}$:

$$\hat{\mu} = \frac{1}{4}(1 + 2 + 1.5 + 3) = \frac{1}{4} \times 7.5 = 1.875.$$

Next, compute the sample variance $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{4}\left((1 - 1.875)^2 + (2 - 1.875)^2 + (1.5 - 1.875)^2 + (3 - 1.875)^2\right).$$

Simplifying each term:

$$(1-1.875)^2 = 0.765625, \quad (2-1.875)^2 = 0.015625, \quad (1.5-1.875)^2 = 0.140625, \quad (3-1.875)^2 = 1.265625.$$

Summing the terms:

$$\hat{\sigma}^2 = \frac{1}{4} \times (0.765625 + 0.015625 + 0.140625 + 1.265625) = \frac{1}{4} \times 2.1875 = 0.546875.$$

Therefore, the MLE for σ^2 is $\hat{\sigma}^2 = 0.546875$.

B Transformation of Random Variables

B.1 Solution: Transformation $Y = 2X + 1$

Let X be a random variable following a Uniform distribution on the interval $[0, 1]$:

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

We are given that $Y = 2X + 1$.

a) **Find the cumulative distribution function (CDF) $F_Y(y)$ of the random variable Y:**

The cumulative distribution function (CDF) of Y is defined as:

$$F_Y(y) = P(Y \leq y) = P(2X + 1 \leq y).$$

To find the probability in terms of X, we first solve the inequality for X:

$$2X + 1 \leq y \quad \Rightarrow \quad X \leq \frac{y-1}{2}.$$

Since X follows a uniform distribution on $[0, 1]$, the CDF $F_Y(y)$ is:

$$F_Y(y) = P\left(X \leq \frac{y-1}{2}\right).$$

At this point, we can simply write:

$$F_Y(y) = F_X\left(\frac{y-1}{2}\right).$$

While this would be an acceptable answer, we will go further to derive the precise numerical values. To do so, we need to consider the range of y:

- If $y < 1$, then $\frac{y-1}{2} < 0$, and since $X \geq 0$, we have $P(X \leq \frac{y-1}{2}) = 0$. Hence, $F_Y(y) = 0$ for $y < 1$.
- If $1 \leq y \leq 3$, then $0 \leq \frac{y-1}{2} \leq 1$, so:

$$F_Y(y) = \frac{y-1}{2}, \quad 1 \leq y \leq 3.$$

- If $y > 3$, then $\frac{y-1}{2} > 1$, and since $X \leq 1$, we have $P(X \leq \frac{y-1}{2}) = 1$. Hence, $F_Y(y) = 1$ for $y > 3$.

Therefore, the CDF of Y is:

$$F_Y(y) = \begin{cases} 0, & y < 1, \\ \frac{y-1}{2}, & 1 \leq y \leq 3, \\ 1, & y > 3. \end{cases}$$

b) **Find the probability density function (PDF) $f_Y(y)$ of the random variable Y:**

To find the PDF $f_Y(y)$, we differentiate the CDF $F_Y(y)$ with respect to y:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

We can use the fact that $F_Y(y) = F_X\left(\frac{y-1}{2}\right)$ to write:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X\left(\frac{y-1}{2}\right) \\ &= \frac{1}{2} f_X\left(\frac{y-1}{2}\right) \end{aligned}$$

We can also start from the numerical values derives above to differentiate each case:

- For $y < 1$, $F_Y(y) = 0$, so $f_Y(y) = 0$.
- For $1 \leq y \leq 3$, $F_Y(y) = \frac{y-1}{2}$, so:

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

- For $y > 3$, $F_Y(y) = 1$, so $f_Y(y) = 0$.

Therefore, the PDF of Y is:

$$f_Y(y) = \begin{cases} 0, & y < 1, \\ \frac{1}{2}, & 1 \leq y \leq 3, \\ 0, & y > 3. \end{cases}$$

c) **Calculate the expected value $\mathbb{E}[Y]$:**

The expected value $\mathbb{E}[Y]$ is given by:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) \, dy.$$

From the PDF of Y, we know that $f_Y(y) = \frac{1}{2}$ for $1 \leq y \leq 3$, and $f_Y(y) = 0$ elsewhere. Thus, the integral simplifies to:

$$\mathbb{E}[Y] = \int_1^3 y \cdot \frac{1}{2} \, dy.$$

We compute the integral:

$$\mathbb{E}[Y] = \frac{1}{2} \int_1^3 y \, dy = \frac{1}{2} \left[\frac{y^2}{2} \right]_1^3 = \frac{1}{2} \left(\frac{9}{2} - \frac{1}{2} \right) = \frac{1}{2} \cdot 4 = 2.$$

Therefore, the expected value of Y is:

$$\mathbb{E}[Y] = 2.$$

B.2 Transformation $Z = X^2$

Let X be a random variable following an Exponential distribution with parameter λ:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Define $Z = X^2$.

a) **Finding the CDF** $F_Z(z)$:

We start by expressing the CDF of Z in terms of the CDF of X:

$$F_Z(z) = P(Z \leq z) = P(X^2 \leq z) = P(X \leq \sqrt{z}) = F_X(\sqrt{z}).$$

The CDF of X, which is exponentially distributed, is:

$$F_X(x) = 1 - e^{-\lambda x}.$$

Therefore, the CDF of Z becomes:

$$F_Z(z) = 1 - e^{-\lambda\sqrt{z}}, \quad z \geq 0.$$

b) **Finding the PDF** $f_Z(z)$:

To find the PDF of Z, we differentiate the CDF $F_Z(z)$ with respect to z:

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \frac{d}{dz}\left(1 - e^{-\lambda\sqrt{z}}\right).$$

Applying the chain rule:

$$f_Z(z) = \lambda \cdot \frac{1}{2\sqrt{z}} \cdot e^{-\lambda\sqrt{z}}, \quad z \geq 0.$$

So the PDF of Z is:

$$f_Z(z) = \frac{\lambda}{2\sqrt{z}}e^{-\lambda\sqrt{z}}, \quad z \geq 0.$$

c) **Calculating the Expected Value** $\mathbb{E}[Z]$:

The boring (and hard) way. The expected value of Z is given by:

$$\mathbb{E}[Z] = \int_0^\infty z f_Z(z) dz = \int_0^\infty z \cdot \frac{\lambda}{2\sqrt{z}}e^{-\lambda\sqrt{z}}dz.$$

Simplifying the integrand:

$$\mathbb{E}[Z] = \frac{\lambda}{2}\int_0^\infty \sqrt{z}e^{-\lambda\sqrt{z}}dz.$$

Now, we perform a substitution. Let $u = \lambda\sqrt{z}$, so that $\sqrt{z} = \frac{u}{\lambda}$ and $dz = \frac{2u}{\lambda^2}du$. The limits of integration remain from 0 to ∞. Substituting these into the integral:

$$\mathbb{E}[Z] = \frac{\lambda}{2}\int_0^\infty \frac{u}{\lambda}e^{-u} \cdot \frac{2u}{\lambda^2}du = \frac{1}{\lambda^2}\int_0^\infty u^2 e^{-u}du.$$

To evaluate $\int_0^\infty u^2 e^{-u}du$, we use integration by parts. Let:

$$I = \int_0^\infty u^2 e^{-u}du.$$

We apply integration by parts twice:

$$
\begin{aligned}
I &= \int_0^\infty u^2 e^{-u} du \\
&= [-u^2 e^{-u}]_0^\infty + \int_0^\infty 2ue^{-u} du \quad \text{(The first term vanishes)} \\
&= \int_0^\infty 2ue^{-u} du \\
&= [-2ue^{-u}]_0^\infty + \int_0^\infty 2e^{-u} du \quad \text{(The first term vanishes)} \\
&= \int_0^\infty 2e^{-u} du \\
&= [-2e^{-u}]_0^\infty \\
&= 2
\end{aligned}
$$

Therefore:

$$I = 2.$$

Finally, substituting back into the expression for $\mathbb{E}[Z]$:

$$\mathbb{E}[Z] = \frac{1}{\lambda^2} \times 2 = \frac{2}{\lambda^2}.$$

Thus, the expected value of Z is:

$$\mathbb{E}[Z] = \frac{2}{\lambda^2}.$$

The smart (probabilistic) way. The expected value of Z is given by:

$$\mathbb{E}[Z] = \int_0^\infty z f_Z(z) dz = \int_0^\infty z \cdot \frac{\lambda}{2\sqrt{z}} e^{-\lambda\sqrt{z}} dz.$$

Let us rewrite the identity $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ to write:

$$\mathbb{E}[X^2] = \text{Var}(X) + (\mathbb{E}[X])^2$$

Now the variance of an exponential random variable is equal to $Var(X) = \frac{1}{\lambda^2}$. This can be proven by writing the expectation as an integral and perform integration by part twice. While this still requires much calculus, we avoided the initial substitution $u = \lambda\sqrt{z}$ in the previous solution, which I strongly prefer. Since the expectation of an exponential random variable is equal to $\frac{1}{\lambda}$, this yields:

$$
\begin{aligned}
\mathbb{E}[X^2] &= \text{Var}(X) + (\mathbb{E}[X])^2 \\
&= \frac{1}{\lambda^2} + \frac{1}{\lambda^2} \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

Conclusions. It is always good to have multiple methods to derive a result to ensure its coherence. If you get different results using different derivation, there is either a mathematical mistake or a reasoning mistake.

B.3 Both Transformations

Let X be a random variable following a Normal distribution with mean μ and variance σ^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Define $W = 2X + 1$ and $V = X^2$.

a) **Finding the Expected Value $\mathbb{E}[W]$:**

The expected value $\mathbb{E}[W]$ can be computed using the linearity of expectation. Since $W = 2X + 1$, we have:

$$\mathbb{E}[W] = \mathbb{E}[2X + 1] = 2\mathbb{E}[X] + \mathbb{E}[1].$$

Since $\mathbb{E}[1] = 1$ and $\mathbb{E}[X] = \mu$ (from the properties of the normal distribution):

$$\mathbb{E}[W] = 2\mu + 1.$$

b) **Finding the Variance $\mathbf{Var}(W)$:**

With smart calculus. Let us recall that $\text{Var}(W) = \mathbb{E}[W^2] - (\mathbb{E}[W])^2$

Let us work out the first expectation:

$$\begin{aligned}
\mathbb{E}[W^2] &= \mathbb{E}[(2X+1)^2] \\
&= \mathbb{E}[4X^2 + 4X + 1] \\
&= 4\mathbb{E}[X^2] + 4\mathbb{E}[X] + 1 \\
&= 4\mathbb{E}[X^2] + 4\mu + 1
\end{aligned}$$

Now let us work out the second expectation:

$$\begin{aligned}
\mathbb{E}[W]^2 &= \mathbb{E}[2X+1]^2 \\
&= (2\mathbb{E}[X] + 1)^2 \\
&= 4\mathbb{E}[X]^2 + 4\mathbb{E}[X] + 1 \\
&= 4\mathbb{E}[X]^2 + 4\mu + 1
\end{aligned}$$

Now substituting into the original equation yields:

$$\begin{aligned}
\text{Var}(W) &= 4\mathbb{E}[X^2] + 4\mu + 1 - (4\mathbb{E}[X]^2 + 4\mu + 1) \\
&= 4(\mathbb{E}[X^2] - \mathbb{E}[X]^2) \\
&= 4\text{Var}(X) \\
&= 4\sigma^2
\end{aligned}$$

With smart properties. We will prove the fact that for any random variable, Var(aX + b) = $a^2\text{Var}(X)$.

Step 1: $\mathbf{Var}(X + b) = \mathbf{Var}(X)$

$$\begin{aligned}
\text{Var}(X+b) &= \mathbb{E}[(X + b - \mathbb{E}[X+b])^2] \\
&= \mathbb{E}[(X + b - (\mathbb{E}[X] + b)^2] \\
&= \mathbb{E}[(X - (\mathbb{E}[X])^2] \\
&= \text{Var}(X)
\end{aligned}$$

Step 2: $\mathbf{Var}(aX) = a^2\mathbf{Var}(X)$

$$\begin{aligned}
\text{Var}(aX) &= \mathbb{E}[(aX)^2] - \mathbb{E}[aX]^2 \\
&= a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2 \\
&= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) \\
&= a^2\text{Var}(X)
\end{aligned}$$

Step 3: Apply step 1 and step 2 to our problem

$$\begin{aligned}
\text{Var}(aX+b) &= \text{Var}(aX) \quad \text{Using step 1.} \\
&= a^2\text{Var}(X) \quad \text{Using step 2.}
\end{aligned}$$

c) **Finding the Expected Value** $\mathbb{E}[V]$:

To compute $\mathbb{E}[V]$ where $V = X^2$, we use the fact that $\mathbb{E}[X^2]$ is related to the variance and the mean of X. Recall the identity:

$$\mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}[X])^2.$$

Substituting $\mathrm{Var}(X) = \sigma^2$ and $\mathbb{E}[X] = \mu$, we get:

$$\mathbb{E}[X^2] = \sigma^2 + \mu^2.$$

Thus:

$$\mathbb{E}[V] = \mathbb{E}[X^2] = \sigma^2 + \mu^2.$$

d) **Finding the Variance Var**(V):

(Disclaimer: deriving this quantity without using any known results is pretty hard. Here is a way to be smart while using known results about the moments of a normal distribution. I will provide the full derivation later, because it takes time to format this on Latex properly.)

To calculate $\mathrm{Var}(V)$, we first need $\mathbb{E}[V^2] = \mathbb{E}[X^4]$. For a normal distribution, we can calculate $\mathbb{E}[X^4]$ by using the closed-form formula for the 4-th moment:

$$\mathbb{E}[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4.$$

Therefore:

$$\mathbb{E}[V^2] = \mathbb{E}[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4.$$

Now, using the variance formula $\mathrm{Var}(V) = \mathbb{E}[V^2] - (\mathbb{E}[V])^2$, we substitute:

$$\mathrm{Var}(V) = \left(3\sigma^4 + 6\sigma^2\mu^2 + \mu^4\right) - (\sigma^2 + \mu^2)^2.$$

Expanding $(\sigma^2 + \mu^2)^2$, we get:

$$(\sigma^2 + \mu^2)^2 = \sigma^4 + 2\sigma^2\mu^2 + \mu^4.$$

Thus:

$$\mathrm{Var}(V) = \left(3\sigma^4 + 6\sigma^2\mu^2 + \mu^4\right) - \left(\sigma^4 + 2\sigma^2\mu^2 + \mu^4\right).$$

Simplifying:

$$\mathrm{Var}(V) = 2\sigma^4 + 4\sigma^2\mu^2.$$

C Composite Random Variables

C.1 Inference with a Transformed Random Variable

A manufacturing process produces parts where the diameter X of each part (in cm) follows an Exponential distribution with an unknown parameter λ:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

However, the observed measurement Y is a transformed version of X, given by the equation:

$$Y = 2X + 1.$$

The company needs to estimate λ based on a sample of observed values $Y_1, Y_2, \ldots, Y_n$.

1. **Derive the probability density function (PDF) of** Y.

   Since the PDF is the derivative of the CDF, let us first derive the CDF.

$$\begin{aligned} F_Y(y) &= P(Y < y) \\ &= P(2X + 1 < y) \\ &= P\left(X < \frac{y-1}{2}\right) \\ &= F_X\left(\frac{y-1}{2}\right) \end{aligned}$$

Now, let's differentiate this function with respect to y to obtain the PDF:

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \frac{d}{dy}F_X\left(\frac{y-1}{2}\right) \\
&= F'_X\left(\frac{y-1}{2}\right)\cdot\frac{d}{dy}\left(\frac{y-1}{2}\right) \\
&= F'_X\left(\frac{y-1}{2}\right)\cdot\frac{1}{2} \\
&= \frac{1}{2}f_X\left(\frac{y-1}{2}\right)
\end{aligned}$$

Hence, $f_Y(y) = \frac{1}{2}f_X\left(\frac{y-1}{2}\right)$.

2. **Parameter Estimation using Maximum Likelihood:**

   - **Writing the likelihood function for λ:**
     Given the observed data $Y_1, Y_2, \ldots, Y_n$, the likelihood function $L(\lambda)$ is the product of the individual densities:
     $$L(\lambda) = \prod_{i=1}^{n} f_Y(Y_i) = \prod_{i=1}^{n}\left(\frac{\lambda}{2}e^{-\frac{\lambda(Y_i-1)}{2}}\right).$$
     Simplifying the likelihood function:
     $$L(\lambda) = \left(\frac{\lambda}{2}\right)^n e^{-\frac{\lambda}{2}\sum_{i=1}^{n}(Y_i-1)} = \left(\frac{\lambda}{2}\right)^n e^{-\frac{\lambda}{2}\left(\sum_{i=1}^{n}Y_i-n\right)}.$$

   - **Deriving the maximum likelihood estimator (MLE) for λ:**
     We will take the natural logarithm of the likelihood function to obtain the log-likelihood function:
     $$\ell(\lambda) = \log L(\lambda) = n\log\left(\frac{\lambda}{2}\right) - \frac{\lambda}{2}\left(\sum_{i=1}^{n}Y_i - n\right).$$
     This simplifies to:
     $$\ell(\lambda) = n\log(\lambda) - n\log(2) - \frac{\lambda}{2}\left(\sum_{i=1}^{n}Y_i - n\right).$$
     Next, we differentiate the log-likelihood function with respect to λ and set the derivative equal to zero:
     $$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \frac{1}{2}\left(\sum_{i=1}^{n}Y_i - n\right) = 0.$$
     Solving for λ:
     $$\frac{n}{\lambda} = \frac{1}{2}\left(\sum_{i=1}^{n}Y_i - n\right),$$
     $$\lambda = \frac{2n}{\sum_{i=1}^{n}Y_i - n}.$$
     Thus, the maximum likelihood estimator for λ is:
     $$\hat{\lambda}_{MLE} = \frac{2n}{\sum_{i=1}^{n}Y_i - n}.$$

3. **Compute the MLE for a sample: Given the sample $\{4.2, 5.1, 3.9, 4.7, 5.3\}$:**

   First, compute the sum of the observed values:
   $$\sum_{i=1}^{5}Y_i = 4.2 + 5.1 + 3.9 + 4.7 + 5.3 = 23.2.$$

Now, substitute this sum into the MLE formula:

$$\hat{\lambda}_{MLE} = \frac{2 \cdot 5}{23.2 - 5} = \frac{10}{18.2} \approx 0.54945.$$

Therefore, the maximum likelihood estimate of λ is approximately:

$$\hat{\lambda}_{MLE} \approx 0.54945.$$

C.2 Sum of Independent Random Variables

C.2.1 CDF of the Sum of Independent Random Variables

To compute the CDF of $Z = X + Y$, we start by expressing the probability:

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z).$$

From this point, there are several equivalent ways to decompose the CDF of Z:

- Conditioning over $X = x$ and using independence of X and Y.
- Using indicator functions.

Conditioning over X. We can use the following formula from probability theory which is valid for any two random variables A and B:

$$P(A) = \mathbb{E}_B[P(A|B)]$$

which in our case would give:

$$\begin{aligned} P(X + Y \leq z) &= \mathbb{E}[P(X + Y \leq z|X)] \\ &= \int_{-\infty}^{+\infty} P(x + Y \leq z|X = x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} P(Y \leq z - x|X = x) f_X(x) dx \end{aligned}$$

However, since X and Y are independent, conditioning Y over X has no impact. This allows us to reach the final steps in our reasoning:

$$\begin{aligned} P(X + Y \leq z) &= \int_{-\infty}^{+\infty} P(Y \leq z - x|X = x) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} P(x + Y \leq z) f_X(x) dx \\ &= \int_{-\infty}^{+\infty} F_Y(z - x) f_X(x) dx \end{aligned}$$

* Where F_Y is the CDF of Y and f_X is the pdf of X.

$$P(X + Y \leq z) = \int_{-\infty}^{+\infty} F_X(z - y) f_Y(y) dy$$

Using indicator functions. We write the probability of an event A as being the expectation of the indicator function $\nVdash_A(X)$:

$$P(A) = \mathbb{E}_X[\nVdash_A(X)]$$

$$\begin{aligned} F_Z(z) &= P(X + Y < z) \\ &= \mathbb{E}_{X,Y}[\nVdash_{X+Y<z}] \\ &= \mathbb{E}_X[\mathbb{E}_Y[\nVdash_{Y<z-X}]] \\ &= \mathbb{E}_X[P(Y < z - X)] \\ &= \mathbb{E}_X[F_Y(z - X)] \\ &= \int_{-\infty}^{+\infty} F_Y(z - x) f_X(x) dx \end{aligned}$$

Conclusion. The resulting function $F_Z(z) = \int_{-\infty}^{+\infty} F_Y(z-x) f_X(x) dx$ is called a convolution between the CDF F_Y and the PDF f_X.

C.2.2 PDF of the Sum of Independent Random Variables

Taking the derivative of $F_Z(z)$ with respect to z, we find the PDF of Z:

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) \\ &= \frac{d}{dz} \int_{-\infty}^{\infty} F_Y(z-x) f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} f_Y(z-x) f_X(x)\, dx. \end{aligned}$$

This is the convolution of the PDFs of X and Y:

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx.$$

C.2.3 Sum of Two Independent Uniform Random Variables

Let X and Y both follow a uniform distribution on $[0, \theta]$ with PDFs $f_X(x) = f_Y(y) = \frac{1}{\theta}$ for $0 \le x, y \le \theta$.
To compute the PDF of $Z = X + Y$, we use the convolution formula:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_0^{\theta} \frac{1}{\theta} \cdot \frac{1}{\theta}\, dx,$$

with limits depending on z.
- For $0 \le z \le \theta$, the limits of integration are from 0 to z:

$$f_Z(z) = \int_0^{z} \frac{1}{\theta^2}\, dx = \frac{z}{\theta^2}.$$

- For $\theta \le z \le 2\theta$, the limits of integration are from $z - \theta$ to θ:

$$f_Z(z) = \int_{z-\theta}^{\theta} \frac{1}{\theta^2}\, dx = \frac{2\theta - z}{\theta^2}.$$

Thus, the PDF of $Z = X + Y$ is:

$$f_Z(z) = \begin{cases} \frac{z}{\theta^2}, & 0 \le z \le \theta, \\ \frac{2\theta - z}{\theta^2}, & \theta \le z \le 2\theta, \\ 0, & \text{otherwise.} \end{cases}$$

You should try to plot such a function to see what it looks like. (Hint: it looks like a triangle)

D Introduction to the Method of Moments

D.1 Step 1: Understanding Moments

a) For a sample $X_1, X_2, \dots, X_n$, the first sample moment is defined as

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$$

b) The first population moment is defined as

$$\mu = \mathbb{E}[X_i]$$

D.2 Step 2: Estimating the Parameters Using Moments

The method of moments equates the sample moments to the population moments in order to estimate the parameters. Let's apply this method to the following example.

Consider a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution with the probability density function:

$$f_X(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

a) **Derivation of the Population Mean of an Exponential Distribution**

Let X be a random variable following an exponential distribution with rate parameter λ. The probability density function (PDF) is given by:

$$f_X(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

We aim to derive the population mean $\mathbb{E}[X]$.

Step 1: Define the expectation of X

The population mean is the expected value of X, defined as:

$$\mathbb{E}[X] = \int_0^\infty x f_X(x;\lambda)\, dx.$$

Substitute the PDF $f_X(x;\lambda)$ into the integral:

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\, dx.$$

Step 2: Perform integration by parts

We need to compute the integral $\int_0^\infty x\lambda e^{-\lambda x}\, dx$. We will use integration by parts. Recall the formula for integration by parts:

$$\int u\, dv = [uv] - \int v\, du.$$

Let:

$$u = x \quad \text{and} \quad dv = \lambda e^{-\lambda x}\, dx.$$

Step 1: Differentiate u and integrate dv.

We compute the derivatives and integrals for the parts:

$$du = dx \quad \text{and} \quad v = -e^{-\lambda x}.$$

Step 2: Apply the integration by parts formula.

Now, apply the integration by parts formula:

$$\mathbb{E}[X] = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx.$$

Step 3: Evaluate the boundary terms

First, evaluate the boundary terms $\left[-xe^{-\lambda x}\right]_0^\infty$:

$$\lim_{x\to\infty} -xe^{-\lambda x} = 0 \quad \text{and} \quad \left(-xe^{-\lambda x}\right)\Big|_{x=0} = 0.$$

Thus, the boundary terms contribute nothing:

$$\left[-xe^{-\lambda x}\right]_0^\infty = 0.$$

Step 4: Compute the remaining integral

The remaining integral is:

$$\int_0^\infty e^{-\lambda x}\, dx.$$

This is a standard exponential integral:

$$\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}.$$

Step 5: Final result

Thus, the expected value of X is:

$$\mathbb{E}[X] = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}.$$

Conclusion: The population mean of an exponential distribution with rate parameter λ is:

$$\mathbb{E}[X] = \frac{1}{\lambda}.$$

b) **Equating sample moments and population moments**
Setting such a constraint introduce the following equation

$$\mathbb{E}[X] \approx \frac{1}{n}\sum_{i=1}^{n} X_i$$

c) **Solve the equation for** λ By substituting the expression of the population moments, we get

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

which yields

$$\lambda = \frac{n}{\sum_{i=1}^{n} X_i}$$

D.3 Step 3: Applying the Method of Moments

Given the following sample: $\{2, 3, 1, 4\}$,

a) The sample mean is equal to $\frac{10}{4} = 2.5$

b) The method of moments thus yields $\hat{\lambda} = \frac{4}{10} = 0.4$

D.4 Step 4: Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with rate parameter λ. The goal is to derive the maximum likelihood estimator (MLE) for λ and apply it to the given sample $\{2, 3, 1, 4\}$.

a) **Write down the likelihood function $L(\lambda)$ for the sample.**

The probability density function (PDF) of an exponential distribution is given by:

$$f_X(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

The likelihood function $L(\lambda)$ for the sample $\{X_1, X_2, \ldots, X_n\}$ is the product of the individual likelihoods:

$$L(\lambda) = \prod_{i=1}^{n} f_X(X_i;\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}.$$

Since the exponential distribution is independent across the samples, we can rewrite the likelihood function as:

$$L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}.$$

This is the likelihood function that we aim to maximize with respect to λ.

b) **Take the natural logarithm of the likelihood function (the log-likelihood).**

To simplify the maximization process, we take the natural logarithm of the likelihood function. The log-likelihood is:

$$\ell(\lambda) = \log L(\lambda) = \log\left(\lambda^n e^{-\lambda\sum_{i=1}^n X_i}\right).$$

Using the properties of logarithms, we expand this expression:

$$\ell(\lambda) = \log(\lambda^n) + \log\left(e^{-\lambda\sum_{i=1}^n X_i}\right).$$

This simplifies to:

$$\ell(\lambda) = n\log(\lambda) - \lambda\sum_{i=1}^n X_i.$$

This is the log-likelihood function.

c) **Differentiate the log-likelihood with respect to λ and solve for the maximum likelihood estimator $\hat{\lambda}_{MLE}$.**

To find the maximum likelihood estimator, we differentiate the log-likelihood function with respect to λ:

$$\frac{d\ell(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i.$$

Setting this derivative equal to zero to find the maximum:

$$\frac{n}{\lambda} - \sum_{i=1}^n X_i = 0.$$

Solving for λ, we get:

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

Therefore, the maximum likelihood estimator for λ is:

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

d) **Using the sample $\{2, 3, 1, 4\}$, compute $\hat{\lambda}_{MLE}$.**

For the given sample $\{2, 3, 1, 4\}$, we first compute the sum of the observed values:

$$\sum_{i=1}^4 X_i = 2 + 3 + 1 + 4 = 10.$$

The sample size is $n = 4$. Substituting these values into the MLE formula:

$$\hat{\lambda}_{MLE} = \frac{4}{10} = 0.4.$$

Therefore, the maximum likelihood estimator for λ based on the sample is:

$$\hat{\lambda}_{MLE} = 0.4.$$

D.5 Step 5: Comparison of the Two Estimators

Now compare the two estimators for λ based on the method of moments and the maximum likelihood estimation.

a) Mathematically, in this case, the two estimators are identical. (Yes I know... I made a calculus mistake when designing the exercises, thus it's not very relevant to compare MLE and MoM. But if my mistake was to be true it would have been a marvelous exercise !)

b) Nevertheless, in general, for a distribution with k parameters, the method of moments selects the set of parameters that allows is to match the population moments. Thus, it only uses knowledge from a finite set of moments. On the other hand, the MLE method uses the value of the PDF (or PMF for discrete variables) evaluated at all X_i and thus makes use of a more fine-grained knowledge about the data distribution. This is why it is often more efficient that MoM.

D.6 MoM vs MLE for a Discrete Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, where the probability mass function is given by:

$$f_X(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \{0, 1, 2, \ldots\}, \quad \lambda > 0.$$

a) **Use the method of moments to derive an estimator for λ.**

The method of moments equates the sample moments to the theoretical moments of the distribution. The mean of a Poisson distribution is known to be:

$$\mathbb{E}[X] = \lambda.$$

The sample mean is given by:

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Using the method of moments, we equate the sample mean to the population mean:

$$\hat{m}_1 = \lambda.$$

Therefore, the method of moments estimator for λ is:

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

This gives us a simple formula for the method of moments estimator of λ, which is the sample mean.

b) **Derive the maximum likelihood estimator $\hat{\lambda}$ for λ.**

The likelihood function $L(\lambda)$ is the product of the probability mass functions for the sample:

$$L(\lambda) = \prod_{i=1}^{n} f_X(X_i; \lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}.$$

Simplifying, we obtain:

$$L(\lambda) = \lambda^{\sum_{i=1}^{n} X_i} e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{X_i!}.$$

The log-likelihood function $\ell(\lambda)$ is:

$$\ell(\lambda) = \log L(\lambda) = \sum_{i=1}^{n} X_i \log(\lambda) - n\lambda + \text{constant}.$$

To find the MLE, we differentiate the log-likelihood with respect to λ and set the derivative equal to zero:

$$\frac{d\ell(\lambda)}{d\lambda} = \frac{\sum_{i=1}^{n} X_i}{\lambda} - n = 0.$$

Solving for λ, we get the maximum likelihood estimator:

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

Conclusion: Both the method of moments and the maximum likelihood estimators for λ are the same for a Poisson distribution. The MLE estimator is equal to the sample mean, just like the method of moments estimator:

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

c) **Given the sample $\{0, 1, 2, 3, 1\}$, compute both the method of moments estimator and the maximum likelihood estimator for λ.**

For the given sample $\{0, 1, 2, 3, 1\}$, we first calculate the sample mean:

$$\sum_{i=1}^{5} X_i = 0 + 1 + 2 + 3 + 1 = 7.$$

The sample size is $n = 5$, so the sample mean is:

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{7}{5} = 1.4.$$

Therefore, both the method of moments estimator and the maximum likelihood estimator for λ are:

$$\hat{\lambda}_{MoM} = \hat{\lambda}_{MLE} = 1.4.$$

D.7 MoM vs MLE for a Continuous Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Uniform distribution on the interval $[0, \theta]$ with probability density function:

$$f_X(x;\theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta.$$

a) **Use the method of moments to derive an estimator for θ.**

For a uniform distribution $U(0, \theta)$, the population mean is given by:

$$\mathbb{E}[X] = \frac{\theta}{2}.$$

The sample mean is:

$$\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Using the method of moments, we equate the sample mean to the population mean:

$$\frac{\theta}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Solving for θ, we get the method of moments estimator for θ:

$$\hat{\theta}_{MoM} = 2 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Thus, the method of moments estimator for θ is twice the sample mean.

b) **Derive the maximum likelihood estimator $\hat{\theta}$ for θ.**

The likelihood function $L(\theta)$ for a random sample from a uniform distribution is the product of the individual probability densities:

$$L(\theta) = \prod_{i=1}^{n} f_X(X_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta}.$$

Simplifying, we get:

$$L(\theta) = \frac{1}{\theta^n}, \quad 0 \le X_i \le \theta \text{ for all } i.$$

The log-likelihood function $\ell(\theta)$ is:

$$\ell(\theta) = \log L(\theta) = -n\log(\theta), \quad 0 \le X_i \le \theta.$$

To find the maximum likelihood estimator, we maximize the log-likelihood function, subject to the constraint $\theta \geq \max(X_1, X_2, \ldots, X_n)$. Since the log-likelihood decreases as θ increases, the maximum likelihood estimator is given by the smallest value of θ that satisfies the constraint:

$$\hat{\theta}_{MLE} = \max(X_1, X_2, \ldots, X_n).$$

c) **Given the sample $\{1, 2, 1.5, 3\}$, compute both the method of moments estimator and the maximum likelihood estimator for θ.**

First, we calculate the sample mean:

$$\hat{m}_1 = \frac{1 + 2 + 1.5 + 3}{4} = \frac{7.5}{4} = 1.875.$$

Using the method of moments estimator formula:

$$\hat{\theta}_{MoM} = 2 \cdot \hat{m}_1 = 2 \cdot 1.875 = 3.75.$$

For the MLE estimator, we take the maximum value of the sample:

$$\hat{\theta}_{MLE} = \max(1, 2, 1.5, 3) = 3.$$

Conclusion: Based on the given sample, the method of moments estimator for θ is $\hat{\theta}_{MoM} = 3.75$, while the maximum likelihood estimator for θ is $\hat{\theta}_{MLE} = 3$.

d) **Further investigation: MSE Comparison for the Uniform Distribution Estimators**

In this section, we compare the Method of Moments (MoM) estimator and the Maximum Likelihood Estimator (MLE) for the parameter θ in the Uniform distribution $U(0, \theta)$ by computing the Mean Squared Error (MSE) for both estimators.

MSE of the Method of Moments Estimator

The MoM estimator for θ is given by:

$$\hat{\theta}_{\text{MoM}} = 2 \cdot \bar{X} = 2 \cdot \frac{1}{n} \sum_{i=1}^{n} X_i,$$

where $\bar{X}$ is the sample mean.

Expectation of $\hat{\theta}_{\text{MoM}}$

For a uniform distribution $U(0, \theta)$, the expectation of the sample mean $\bar{X}$ is:

$$\mathbb{E}[\bar{X}] = \frac{\theta}{2}.$$

Thus, the expectation of $\hat{\theta}_{\text{MoM}}$ is:

$$\mathbb{E}[\hat{\theta}_{\text{MoM}}] = 2 \cdot \frac{\theta}{2} = \theta.$$

This shows that $\hat{\theta}_{\text{MoM}}$ is an **unbiased** estimator.

Variance of $\hat{\theta}_{\text{MoM}}$

The variance of the sample mean $\bar{X}$ is given by:

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\theta^2/12}{n}.$$

Therefore, the variance of $\hat{\theta}_{\text{MoM}}$ is:

$$\text{Var}(\hat{\theta}_{\text{MoM}}) = 4 \cdot \text{Var}(\bar{X}) = 4 \cdot \frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

MSE of $\hat{\theta}_{\text{MoM}}$

Since $\hat{\theta}_{\text{MoM}}$ is unbiased, the MSE is equal to its variance:

$$\text{MSE}(\hat{\theta}_{\text{MoM}}) = \text{Var}(\hat{\theta}_{\text{MoM}}) = \frac{\theta^2}{3n}.$$

MSE of the Maximum Likelihood Estimator

The MLE for θ is given by:

$$\hat{\theta}_{\text{MLE}} = \max(X_1, X_2, \ldots, X_n),$$

i.e., the maximum value in the sample.

Expectation of $\hat{\theta}_{\text{MLE}}$

The expectation of the maximum $\hat{\theta}_{\text{MLE}}$ for a Uniform distribution $U(0, \theta)$ is:

$$\mathbb{E}[\hat{\theta}_{\text{MLE}}] = \frac{n}{n+1}\theta.$$

Thus, the MLE is **biased**, and the bias is:

$$\text{Bias}(\hat{\theta}_{\text{MLE}}) = \theta - \frac{n}{n+1}\theta = \frac{\theta}{n+1}.$$

Variance of $\hat{\theta}_{\text{MLE}}$

The variance of the maximum $\hat{\theta}_{\text{MLE}}$ is given by:

$$\text{Var}(\hat{\theta}_{\text{MLE}}) = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

MSE of $\hat{\theta}_{\text{MLE}}$

The MSE of the MLE is the sum of its variance and the square of the bias:

$$\text{MSE}(\hat{\theta}_{\text{MLE}}) = \text{Var}(\hat{\theta}_{\text{MLE}}) + \text{Bias}(\hat{\theta}_{\text{MLE}})^2.$$

Substituting the values of the variance and bias:

$$\text{MSE}(\hat{\theta}_{\text{MLE}}) = \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{\theta}{n+1}\right)^2.$$

Simplifying:

$$\text{MSE}(\hat{\theta}_{\text{MLE}}) = \frac{\theta^2}{(n+1)^2}\left(\frac{n}{n+2} + 1\right) = \frac{\theta^2}{(n+1)^2} \cdot \frac{n+1}{n+2}.$$

Thus, the MSE of the MLE is:

$$\text{MSE}(\hat{\theta}_{\text{MLE}}) = \frac{\theta^2}{(n+1)(n+2)}.$$

Comparison of the MSEs

The MSEs of the two estimators are:

$$\text{MSE}(\hat{\theta}_{\text{MoM}}) = \frac{\theta^2}{3n}, \quad \text{MSE}(\hat{\theta}_{\text{MLE}}) = \frac{\theta^2}{(n+1)(n+2)}.$$

To compare them, we compute the ratio:

$$\frac{\text{MSE}(\hat{\theta}_{\text{MoM}})}{\text{MSE}(\hat{\theta}_{\text{MLE}})} = \frac{\frac{\theta^2}{3n}}{\frac{\theta^2}{(n+1)(n+2)}} = \frac{(n+1)(n+2)}{3n}.$$

For large n, this ratio approaches:

$$\frac{n^2}{3n} = \frac{n}{3},$$

which indicates that the MSE of the MoM estimator is approximately $n/3$ times larger than the MSE of the MLE for large n.

For $n = 1$ however, the ratio is equal to $\frac{2}{3}$ which indicates that in this case, the MoM estimator should be preferred.

Conclusion

For large sample sizes, the MLE has a much smaller MSE and is therefore a better estimator. For small sample sizes, we can still see that the MSE of the MLE is smaller than that of the MoM estimator. Thus, the **MLE** is generally the preferred estimator based on the MSE criterion.

E Estimating the Variance of a Random Variable

E.1 Estimating Variance using the Theoretical Mean

To show that $\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ is an unbiased estimator of σ^2, we compute its expected value:

$$\mathbb{E}[\hat{\sigma}^2_\mu] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(X_i - \mu)^2] = \frac{1}{n}\sum_{i=1}^n \sigma^2 = \sigma^2.$$

Therefore, $\hat{\sigma}^2_\mu$ is an unbiased estimator of the population variance.

E.2 Estimating Variance using the Sample Mean

The estimator $\hat{\sigma}^2_{\bar{X}} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ is biased. To compute the bias, we expand $(X_i - \bar{X})^2$ as follows:

$$(X_i - \bar{X})^2 = (X_i - \mu + \mu - \bar{X})^2 = (X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu).$$

Taking the expected value:

$$\mathbb{E}[\hat{\sigma}^2_{\bar{X}}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(X_i - \bar{X})^2] = \sigma^2\left(1 - \frac{1}{n}\right).$$

Thus, $\hat{\sigma}^2_{\bar{X}}$ is biased, and the bias is:

$$\text{Bias}(\hat{\sigma}^2_{\bar{X}}) = -\frac{\sigma^2}{n}.$$

E.3 Deriving the Unbiased Estimator without the Theoretical Mean

To correct the bias, we modify the estimator by multiplying by $\frac{n}{n-1}$:

$$\hat{\sigma}^2_{unbiased} = \frac{n}{n-1}\hat{\sigma}^2_{\bar{X}} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

This adjustment ensures that $\mathbb{E}[\hat{\sigma}^2_{unbiased}] = \sigma^2$, making $\hat{\sigma}^2_{unbiased}$ an unbiased estimator of the population variance.

The intuitive explanation for the bias when using the sample mean is that $\bar{X}$ is itself a random variable based on the sample, which introduces additional variability. This reduces the spread of the deviations $X_i - \bar{X}$ compared to using the fixed theoretical mean μ, leading to an underestimation of the true variance.

F MSE, Bias, and Variance of an Estimator

F.1 Defining the MSE of an Estimator

The MSE of an estimator $\hat{\theta}$ is defined as:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right].$$

We decompose this by adding and subtracting $\mathbb{E}[\hat{\theta}]$:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2\right].$$

Expanding the square:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\right] + (\mathbb{E}[\hat{\theta}] - \theta)^2.$$

The first term is the variance of $\hat{\theta}$, and the second term is the square of the bias of $\hat{\theta}$, so we have:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\text{Bias}(\hat{\theta})\right)^2.$$

F.2 Applying the Bias-Variance Decomposition

For $\hat{\sigma}^2_{\bar{X}}$ (biased estimator):

- The bias is $\text{Bias}(\hat{\sigma}^2_{\bar{X}}) = -\frac{\sigma^2}{n}$.
- The variance is $\text{Var}(\hat{\sigma}^2_{\bar{X}}) = \frac{2\sigma^4}{n}$.
- Using the bias-variance decomposition:

$$\text{MSE}(\hat{\sigma}^2_{\bar{X}}) = \frac{2\sigma^4}{n} + \left(-\frac{\sigma^2}{n}\right)^2 = \frac{2\sigma^4}{n} + \frac{\sigma^4}{n^2}.$$

For $\hat{\sigma}^2_{unbiased}$ (unbiased estimator):

- The bias is zero, $\text{Bias}(\hat{\sigma}^2_{unbiased}) = 0$.
- The variance is $\text{Var}(\hat{\sigma}^2_{unbiased}) = \frac{2\sigma^4}{n-1}$.
- The MSE is simply the variance:

$$\text{MSE}(\hat{\sigma}^2_{unbiased}) = \frac{2\sigma^4}{n-1}.$$

F.3 Comparing the Two Estimators

- The MSE of the biased estimator $\hat{\sigma}^2_{\bar{X}}$ includes both a bias term and a variance term. For large n, the bias term $\frac{\sigma^4}{n^2}$ becomes negligible, and the variance dominates.
- The MSE of the unbiased estimator $\hat{\sigma}^2_{unbiased}$ only includes the variance term, which is larger than the variance of the biased estimator for small n.

Conclusion: For small sample sizes n, the biased estimator may have a smaller MSE due to its lower variance. However, as n increases, the unbiased estimator becomes preferable because it has no bias.