

Introduction to Statistics for Data Science

omuellerklein@berkeley.edu

August 2024

1 Background

These problems come from tutoring students in the Statistics in Data Science course at SFSU but can be applied to statistics and statistics for Data Science courses at many universities and programs. These problems are inspired by problems from the book **Introduction to Mathematical Statistics, 8th Edition** by Robert Hogg, Joseph McKean, and Allen Craig.

These problems cover Hypothesis Testing, Confidence Interval, joint and marginal probability density functions (PDF), joint and marginal probability mass functions (PMF), Maximum Likelihood Estimation (MLE), and other related topics that are foundational to statistics in Data Science.

If you find this useful, you can reach out to me via email for private tutoring sessions. You can also let me know if you have any questions or concerns about the solutions. -Oliver

2 Hypothesis Testing

Nearly all scientific papers published involve hypothesis testing. The most important statistic from hypothesis testing, the **p-value**, is widely used and also criticized for various reasons. Hypothesis testing revolves around the idea that there is some pattern, difference, or effect that can be captured in the alternative hypothesis that we "*want*" to try and prove or show is happening. The fallback option is the null hypothesis.

There are generally 5 steps to hypothesis testing:

1. Set up null and alternative hypotheses
2. Select significance level (α)
3. Calculate test statistic
4. Determine p-value or critical value and compare to test statistic
5. Either reject or fail to reject the null hypothesis

There are three major scenarios of tests to set up. The test statistics, critical values, and decision rules are specific to each of these type of tests.

2.1 Types of Tests

1. **two-tailed:** this is when you are testing if both group A and B are different or not. This could mean A is larger than B or A is smaller than B. This is the scenario where it doesn't matter how they are different, just that they are different (or not).
2. **left-tailed:** this is a one-tailed test. This is the scenario when you are testing if group A is less than group B.
3. **right-tailed:** this is the other one-tailed test. This is the scenario when you are testing if group B is greater than group A.

2.2 Hypotheses

1. **Two-tailed:**

$$H_0 : \mu_A = \mu_B \quad (\text{no difference in means})$$

$$H_A : \mu_A \neq \mu_B \quad (\text{difference in means})$$

2. **Left-tailed:**

$$H_0 : \mu_A \geq \mu_B \quad (\text{group A is not less than group B})$$

$$H_A : \mu_A < \mu_B \quad (\text{group A is less than group B})$$

3. **Right-tailed:**

$$H_0 : \mu_A \leq \mu_B \quad (\text{group A is not greater than group B})$$

$$H_A : \mu_A > \mu_B \quad (\text{group A is greater than group B})$$

Note: the alternative hypothesis is what we are "*trying*" to prove. The null has to cover everything else that is not what the alternative claims.

2.3 Introduction Questions

These questions are meant to set you up for the first (*and most important*) step of hypothesis testing: **setting up the null and alternative hypotheses**.

1. **Intro-Q1:** A researcher wants to determine if a new medication lowers blood pressure more effectively than an existing medication. A random sample of patients is given the new medication, and their blood pressure is measured. Another random sample of patients is given the existing medication, and their blood pressure is also measured.

(a) Should this be a two-tail, left-tail, or right-tail test?

- i. Two-tail

ii. **Left-tail**

iii. Right-tail

(b) Write the null and alternative hypotheses.

$$H_0 : \mu_{\text{new}} \geq \mu_{\text{existing}}$$

$$H_1 : \mu_{\text{new}} < \mu_{\text{existing}}$$

2. **Intro-Q2:** An agricultural scientist wants to test if a new fertilizer leads to higher crop yields compared to the standard fertilizer. The crop yields from two sets of fields, one treated with the new fertilizer and one with the standard fertilizer, are compared.

(a) Should this be a two-tail, left-tail, or right-tail test?

i. Two-tail

ii. Left-tail

iii. **Right-tail**

(b) Write the null and alternative hypotheses.

$$H_0 : \mu_{\text{new}} \leq \mu_{\text{standard}}$$

$$H_1 : \mu_{\text{new}} > \mu_{\text{standard}}$$

3. **Intro-Q3:** A company claims that its new training program improves the productivity of its employees. The productivity scores of employees before and after the training program are compared.

(a) Should this be a two-tail, left-tail, or right-tail test?

i. Two-tail

ii. Left-tail

iii. **Right-tail**

(b) Write the null and alternative hypotheses.

$$H_0 : \mu_{\text{after}} \leq \mu_{\text{before}}$$

$$H_1 : \mu_{\text{after}} > \mu_{\text{before}}$$

4. **Intro-Q4:** An engineer wants to test if a new type of material has a different tensile strength compared to the current material used in construction. The tensile strengths of samples from both materials are measured and compared.

(a) Should this be a two-tail, left-tail, or right-tail test?

- i. **Two-tail**
- ii. Left-tail
- iii. Right-tail

(b) Write the null and alternative hypotheses.

$$H_0 : \mu_{\text{new}} = \mu_{\text{current}}$$

$$H_1 : \mu_{\text{new}} \neq \mu_{\text{current}}$$

5. **Intro-Q5:** A food scientist wants to determine if a new preservative extends the shelf life of a product compared to the current preservative. The shelf lives of products with both preservatives are measured and compared.

(a) Should this be a two-tail, left-tail, or right-tail test?

- i. Two-tail
- ii. Left-tail
- iii. **Right-tail**

(b) Write the null and alternative hypotheses.

$$H_0 : \mu_{\text{new}} \leq \mu_{\text{current}}$$

$$H_1 : \mu_{\text{new}} > \mu_{\text{current}}$$

2.4 Prompt Questions

1. **HT-Q1:** A study is conducted to determine whether a new teaching method improves the mathematics test scores of high school students. A random sample of 12 students is taught using the new method, while another random sample of 12 students is taught using the traditional method. The hypotheses of interest are $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$ where $\Delta = \mu_{\text{new}} - \mu_{\text{traditional}}$ and μ_{new} and $\mu_{\text{traditional}}$ are the true mean test scores for the new and traditional teaching methods, respectively. The test scores are given below:

New Method	78	85	88	92	76	95	89	83	91	84	87	93
Traditional Method	72	81	79	84	77	80	83	78	85	76	82	75

- (b) Compute the two-sample t-test and obtain the p-value. Are the data significant at the 5% level of significance?
- (c) Obtain a point estimate of Δ and a 95% confidence interval for it.
- (d) Conclude in terms of the problem.

Solution

(b) Two-Sample t-Test and p-value

Conceptual Explanation: The two-sample t-test is used to determine whether there is a significant difference between the means of two independent groups. Here, we compare the means of test scores for two teaching methods.

Technical Steps:

i. Formulate Hypotheses:

$$H_0 : \Delta = 0 \quad (\text{no difference in means})$$

$$H_1 : \Delta \neq 0 \quad (\text{difference in means})$$

ii. Calculate Sample Means and Standard Deviations:

$$\bar{X}_{\text{new}} = \frac{1}{12} \sum_{i=1}^{12} X_{\text{new},i} = \frac{78 + 85 + \cdots + 93}{12} = 86.58$$

$$\bar{X}_{\text{traditional}} = \frac{1}{12} \sum_{i=1}^{12} X_{\text{traditional},i} = \frac{72 + 81 + \cdots + 75}{12} = 79.25$$

$$s_{\text{new}}^2 = \frac{1}{11} \sum_{i=1}^{12} (X_{\text{new},i} - \bar{X}_{\text{new}})^2 = 55.80$$

$$s_{\text{traditional}}^2 = \frac{1}{11} \sum_{i=1}^{12} (X_{\text{traditional},i} - \bar{X}_{\text{traditional}})^2 = 11.34$$

iii. Calculate the Test Statistic:

$$t = \frac{\bar{X}_{\text{new}} - \bar{X}_{\text{traditional}}}{\sqrt{\frac{s_{\text{new}}^2}{12} + \frac{s_{\text{traditional}}^2}{12}}}$$

$$t = \frac{86.58 - 79.25}{\sqrt{\frac{55.80}{12} + \frac{11.34}{12}}} = \frac{7.33}{\sqrt{4.65 + 0.95}} = \frac{7.33}{\sqrt{5.60}} = \frac{7.33}{2.37} = 3.09$$

iv. Degrees of Freedom and p-value:

$$df = \min(12 - 1, 12 - 1) = 11$$

Using a t-table or calculator, find the p-value for $t = 3.09$ with $df = 11$:

$$\text{p-value} = 0.011$$

v. Decision:

Since the p-value $0.011 < 0.05$, we reject the null hypothesis at the 5% significance level.

(c) **Point Estimate and 95% Confidence Interval for Δ**

Conceptual Explanation: The point estimate for Δ is the difference in sample means. A confidence interval provides a range of values that is likely to contain the population parameter with a certain level of confidence (95% here).

Technical Steps:

i. **Point Estimate:**

$$\hat{\Delta} = \bar{X}_{\text{new}} - \bar{X}_{\text{traditional}} = 86.58 - 79.25 = 7.33$$

ii. **Standard Error:**

$$SE = \sqrt{\frac{s_{\text{new}}^2}{12} + \frac{s_{\text{traditional}}^2}{12}} = \sqrt{\frac{55.80}{12} + \frac{11.34}{12}} = \sqrt{4.65 + 0.95} = 2.37$$

iii. **95% Confidence Interval:**

$$CI = \hat{\Delta} \pm t_{0.025,11} \times SE$$

$$t_{0.025,11} = 2.201$$

$$CI = 7.33 \pm 2.201 \times 2.37$$

$$CI = 7.33 \pm 5.22$$

$$CI = [2.11, 12.55]$$

(d) **Conclusion in Terms of the Problem**

Conceptual Explanation: Based on the statistical analysis, we interpret the results in the context of the research question.

Technical Steps:

- i. **Hypothesis Test Conclusion:** Since the p-value is less than 0.05, we reject the null hypothesis. There is statistically significant evidence at the 5% level to suggest a difference in mean test scores between the new and traditional teaching methods.
- ii. **Confidence Interval Interpretation:** The 95% confidence interval for Δ does not include 0, further supporting the conclusion that there is a significant difference between the two teaching methods.

2. **HT-Q2:** A manufacturing company wants to compare the durability of two types of materials used in their products. They randomly select 8 samples of Material A and 8 samples of Material B and measure their breaking strengths. The hypotheses of interest are $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$ where $\Delta = \mu_A - \mu_B$ and μ_A and μ_B are the true mean breaking strengths of Materials A and B, respectively. The breaking strengths (in Newtons) are given below:

Material A	210	215	220	205	230	225	210	220
Material B	200	190	195	205	210	205	200	195

- (a) Compute the two-sample t-test and obtain the p-value. Are the data significant at the 5% level of significance?
- (b) Obtain a point estimate of Δ and a 95% confidence interval for it.
- (c) Conclude in terms of the problem.

Solution

(a) Two-Sample t-Test and p-value

Conceptual Explanation: The two-sample t-test is used to determine whether there is a significant difference between the means of two independent groups. Here, we compare the means of breaking strengths for two materials.

Technical Steps:

i. Formulate Hypotheses:

$$H_0 : \Delta = 0 \quad (\text{no difference in means})$$

$$H_1 : \Delta \neq 0 \quad (\text{difference in means})$$

ii. Calculate Sample Means and Standard Deviations:

$$\bar{X}_A = \frac{1}{8} \sum_{i=1}^8 X_{A,i} = \frac{210 + 215 + 220 + 205 + 230 + 225 + 210 + 220}{8} = 216.875$$

$$\bar{X}_B = \frac{1}{8} \sum_{i=1}^8 X_{B,i} = \frac{200 + 190 + 195 + 205 + 210 + 205 + 200 + 195}{8} = 200$$

$$s_A^2 = \frac{1}{7} \sum_{i=1}^8 (X_{A,i} - \bar{X}_A)^2 = 70.98$$

$$s_B^2 = \frac{1}{7} \sum_{i=1}^8 (X_{B,i} - \bar{X}_B)^2 = 42.85$$

iii. Calculate the Test Statistic:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{8} + \frac{s_B^2}{8}}}$$

$$t = \frac{216.875 - 200}{\sqrt{\frac{70.98}{8} + \frac{42.85}{8}}} = \frac{16.875}{\sqrt{8.873 + 5.356}} = \frac{16.875}{\sqrt{14.229}} = \frac{16.875}{4.09} = 4.474$$

iv. Degrees of Freedom and p-value:

$$df = \min(8 - 1, 8 - 1) = 7$$

Using a t-table or calculator, find the p-value for $t = 4.474$ with $df = 7$:

$$\text{p-value} = 0.003$$

- v. **Decision:** Since the p-value $0.003 < 0.05$, we reject the null hypothesis at the 5% significance level.

(b) **Point Estimate and 95% Confidence Interval for Δ**

Conceptual Explanation: The point estimate for Δ is the difference in sample means. A confidence interval provides a range of values that is likely to contain the population parameter with a certain level of confidence (95% here).

Technical Steps:

i. **Point Estimate:**

$$\hat{\Delta} = \bar{X}_A - \bar{X}_B = 216.875 - 200 = 16.875$$

ii. **Standard Error:**

$$SE = \sqrt{\frac{s_A^2}{8} + \frac{s_B^2}{8}} = \sqrt{\frac{70.98}{8} + \frac{42.85}{8}} = \sqrt{8.873 + 5.356} = 3.77$$

iii. **95% Confidence Interval:**

$$CI = \hat{\Delta} \pm t_{0.025,7} \times SE$$

$$t_{0.025,7} = 2.365$$

$$CI = 16.875 \pm 2.365 \times 3.77$$

$$CI = 16.875 \pm 8.91$$

$$CI = [7.95, 25.80]$$

(c) **Conclusion in Terms of the Problem**

Conceptual Explanation: Based on the statistical analysis, we interpret the results in the context of the research question.

Technical Steps:

- i. **Hypothesis Test Conclusion:** Since the p-value is less than 0.05, we reject the null hypothesis. There is statistically significant evidence at the 5% level to suggest a difference in mean breaking strengths between Materials A and B.
 - ii. **Confidence Interval Interpretation:** The 95% confidence interval for Δ does not include 0, further supporting the conclusion that there is a significant difference between the two materials.
3. **HT-Q3:** A new type of lightbulb is being tested for its average lifespan. The manufacturer claims that the new lightbulb has an average lifespan of over 1200 hours. A random sample of 10 lightbulbs is tested, and the lifespans (in hours) are as follows: 1180, 1190, 1200, 1210, 1195, 1185, 1220, 1215, 1205, 1190. Test the manufacturer's claim at the 5% significance level using the critical region approach.

- (a) State the null and alternative hypotheses.
- (b) Compute the sample mean and standard deviation.
- (c) Determine the critical value and state the decision rule.
- (d) Calculate the test statistic.
- (e) Make a decision and conclude in terms of the problem.

Solution

- (a) **State the null and alternative hypotheses.**

$H_0 : \mu \leq 1200$ (The mean lifespan is less than or equal to 1200 hours)

$H_1 : \mu > 1200$ (The mean lifespan is more than 1200 hours)

- (b) **Compute the sample mean and standard deviation.**

Sample mean:

$$\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = \frac{1180 + 1190 + 1200 + 1210 + 1195 + 1185 + 1220 + 1215 + 1205 + 1190}{10} = \frac{11990}{10}$$

Sample standard deviation:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$s = \sqrt{\frac{1}{9} [(1180 - 1199)^2 + (1190 - 1199)^2 + (1200 - 1199)^2 + (1210 - 1199)^2 + (1195 - 1199)^2 + (1185 - 1199)^2 + (1220 - 1199)^2 + (1215 - 1199)^2 + (1205 - 1199)^2 + (1190 - 1199)^2]}$$

$$s = \sqrt{176.667} \approx 13.291$$

- (c) **Determine the critical value and state the decision rule.**

For a one-tailed t-test at the 5% significance level with $n - 1 = 10 - 1 = 9$ degrees of freedom, the critical value $t_{0.05,9}$ can be found using a t-table:

$$t_{0.05,9} \approx 1.833$$

Decision rule: Reject H_0 if the test statistic t is greater than 1.833.

- (d) **Calculate the test statistic.**

The test statistic for a one-sample t-test is given by:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Substituting the values:

$$t = \frac{1199 - 1200}{13.291/\sqrt{10}} = \frac{-1}{4.203} \approx -0.238$$

(e) **Make a decision and conclude in terms of the problem.**

Since the test statistic $t = -0.238$ is not greater than the critical value 1.833, we fail to reject the null hypothesis.

Conclusion: There is not enough evidence at the 5% significance level to reject the manufacturer's claim that the new lightbulb has an average lifespan greater than 1200 hours.

3 Joint PMFs and Covariances

1. **JC-Q1:** Let X and Y have the following joint pmf:

$$p(x, y) = \begin{cases} \frac{1}{9} & (x, y) = (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2) \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the expected values $E(X)$ and $E(Y)$.
- (b) Determine the covariance $\text{Cov}(X, Y)$.

Solution

(a) **Expected Values $E(X)$ and $E(Y)$**

The expected value $E(X)$ is calculated as:

$$E(X) = \sum_x \sum_y x \cdot p(x, y)$$

Given the joint pmf, we calculate $E(X)$:

$$E(X) = 0 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) + 1 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) + 2 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right)$$

$$E(X) = 0 + \frac{3}{9} + \frac{6}{9} = \frac{9}{9} = 1$$

Similarly, the expected value $E(Y)$ is:

$$E(Y) = \sum_x \sum_y y \cdot p(x, y)$$

Given the joint pmf, we calculate $E(Y)$:

$$E(Y) = 0 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) + 1 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) + 2 \cdot \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right)$$

$$E(Y) = 0 + \frac{3}{9} + \frac{6}{9} = \frac{9}{9} = 1$$

(b) **Determining the Covariance**

The covariance is given by $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. We need to find $E(XY)$.

The expected value $E(XY)$ is:

$$E(XY) = \sum_x \sum_y xy \cdot p(x, y)$$

Given the joint pmf, we calculate $E(XY)$:

$$E(XY) = 0 \cdot 0 \cdot \frac{1}{9} + 0 \cdot 1 \cdot \frac{1}{9} + 0 \cdot 2 \cdot \frac{1}{9} + 1 \cdot 0 \cdot \frac{1}{9} + 1 \cdot 1 \cdot \frac{1}{9} + 1 \cdot 2 \cdot \frac{1}{9} + 2 \cdot 0 \cdot \frac{1}{9} + 2 \cdot 1 \cdot \frac{1}{9} + 2 \cdot 2 \cdot \frac{1}{9}$$

$$E(XY) = 0 + 0 + 0 + 0 + \frac{1}{9} + \frac{2}{9} + 0 + \frac{2}{9} + \frac{4}{9} = \frac{9}{9} = 1$$

Now, calculate the covariance:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 1 - 1 \cdot 1 = 1 - 1 = 0$$

2. **JC-Q2:** Given the joint probability distribution of X and Y :

$$f_{X,Y}(x, y) = \begin{cases} k(2x + 3y) & 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the value of k that makes $f_{X,Y}(x, y)$ a valid joint pdf.
- (b) Calculate $\text{Cov}(X, Y)$.

Solution

- (a) **Finding the Value of k :**

To make $f_{X,Y}(x, y)$ a valid joint pdf, the total probability must be 1. We integrate $f_{X,Y}(x, y)$ over the entire range of X and Y :

$$\int_0^1 \int_0^2 k(2x + 3y) dy dx = 1$$

- (b) **First, integrate with respect to y :**

$$\int_0^2 (2x + 3y) dy = 2x \int_0^2 dy + 3 \int_0^2 y dy$$

$$2x[y]_0^2 + 3 \left[\frac{y^2}{2} \right]_0^2 = 2x \cdot 2 + 3 \cdot 2 = 4x + 6$$

(c) **Next, integrate with respect to x :**

$$\begin{aligned}\int_0^1 k(4x+6) dx &= k \left[4 \int_0^1 x dx + 6 \int_0^1 dx \right] \\ k \left[4 \left(\frac{x^2}{2} \right)_0^1 + 6 (x)_0^1 \right] &= k \left(4 \cdot \frac{1}{2} + 6 \cdot 1 \right) = k(2+6) = 8k\end{aligned}$$

(d) **Set this equal to 1:**

$$8k = 1 \implies k = \frac{1}{8}$$

(e) **Calculating $\text{Cov}(X, Y)$:**

The covariance is given by $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. We need to find $E(X)$, $E(Y)$, and $E(XY)$.

First, find $E(X)$:

$$\begin{aligned}E(X) &= \int_0^1 \int_0^2 x f_{X,Y}(x, y) dy dx = \int_0^1 \int_0^2 x \cdot \frac{1}{8}(2x+3y) dy dx \\ &= \frac{1}{8} \int_0^1 x \left[2x \int_0^2 dy + 3 \int_0^2 y dy \right] dx \\ &= \frac{1}{8} \int_0^1 x(2x \cdot 2 + 3 \cdot 2) dx = \frac{1}{8} \int_0^1 x(4x+6) dx \\ &= \frac{1}{8} \int_0^1 (4x^2 + 6x) dx = \frac{1}{8} \left[\frac{4x^3}{3} + \frac{6x^2}{2} \right]_0^1 \\ &= \frac{1}{8} \left(\frac{4}{3} + 3 \right) = \frac{1}{8} \cdot \frac{13}{3} = \frac{13}{24}\end{aligned}$$

Next, find $E(Y)$:

$$\begin{aligned}E(Y) &= \int_0^1 \int_0^2 y f_{X,Y}(x, y) dy dx = \int_0^1 \int_0^2 y \cdot \frac{1}{8}(2x+3y) dy dx \\ &= \frac{1}{8} \int_0^1 \left[2x \int_0^2 y dy + 3 \int_0^2 y^2 dy \right] dx \\ &= \frac{1}{8} \int_0^1 \left(2x \cdot \frac{y^2}{2} \Big|_0^2 + 3 \cdot \frac{y^3}{3} \Big|_0^2 \right) dx \\ &= \frac{1}{8} \int_0^1 \left(2x \cdot 2 + 3 \cdot \frac{8}{3} \right) dx = \frac{1}{8} \int_0^1 (4x+8) dx \\ &= \frac{1}{8} [2x^2 + 8x]_0^1 = \frac{1}{8}(2+8) = \frac{1}{8} \cdot 10 = \frac{5}{4}\end{aligned}$$

Now, find $E(XY)$:

$$\begin{aligned}
 E(XY) &= \int_0^1 \int_0^2 xy f_{X,Y}(x,y) dy dx = \int_0^1 \int_0^2 xy \cdot \frac{1}{8}(2x+3y) dy dx \\
 &= \frac{1}{8} \int_0^1 x \left[2x \int_0^2 y dy + 3 \int_0^2 y^2 dy \right] dx \\
 &= \frac{1}{8} \int_0^1 x \left(2x \cdot \frac{y^2}{2} \Big|_0^2 + 3 \cdot \frac{y^3}{3} \Big|_0^2 \right) dx \\
 &= \frac{1}{8} \int_0^1 x(4x+8) dx = \frac{1}{8} \left[\frac{4x^3}{3} + 8 \cdot \frac{x^2}{2} \right]_0^1 \\
 &= \frac{1}{8} \left(\frac{4}{3} + 4 \right) = \frac{1}{8} \cdot \frac{16}{3} = \frac{2}{3}
 \end{aligned}$$

Finally, calculate the covariance:

$$\begin{aligned}
 \text{Cov}(X,Y) &= E(XY) - E(X)E(Y) = \frac{2}{3} - \frac{13}{24} \cdot \frac{5}{4} \\
 \text{Cov}(X,Y) &= \frac{2}{3} - \frac{65}{96} = \frac{64}{96} - \frac{65}{96} = -\frac{1}{96}
 \end{aligned}$$

4 MLE

Maximum Likelihood Estimation (MLE) is about finding the largest, most likely value from a specific random variable. The steps involved are random variable distribution specific. In other words, you implement a different series of likelihood and log-likelihood steps depending on what distribution your random variables comes from.

Generally, the steps are:

1. Express the population parameter from the random variable distribution
 2. Likelihood function for that random variable
 3. Log-Likelihood function
 4. Differentiate the log-likelihood
 5. Set the derivative to 0
 6. Express solution as the MLE of the population parameter
1. **MLE-Q1:** Suppose that Y_1, Y_2, \dots, Y_n denote a random sample from an exponentially distributed population with mean β . Find the MLE of the population mean β .

Solution

(a) **Step 1: Understanding the Exponential Distribution**

The probability density function (pdf) of an exponentially distributed random variable Y with parameter β (mean) is given by:

$$f(y; \beta) = \frac{1}{\beta} e^{-y/\beta}$$

for $y \geq 0$.

(b) **Step 2: Likelihood Function**

Given a random sample Y_1, Y_2, \dots, Y_n from the exponential distribution, the likelihood function is the product of the individual pdfs:

$$L(\beta; Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f(Y_i; \beta)$$

Substituting the pdf of the exponential distribution:

$$L(\beta; Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n \left(\frac{1}{\beta} e^{-Y_i/\beta} \right)$$

$$L(\beta; Y_1, Y_2, \dots, Y_n) = \left(\frac{1}{\beta} \right)^n e^{-\sum_{i=1}^n Y_i/\beta}$$

(c) **Step 3: Log-Likelihood Function**

Take the natural logarithm of the likelihood function to obtain the log-likelihood function:

$$\ell(\beta) = \log L(\beta; Y_1, Y_2, \dots, Y_n)$$

$$\ell(\beta) = \log \left(\left(\frac{1}{\beta} \right)^n e^{-\sum_{i=1}^n Y_i/\beta} \right)$$

$$\ell(\beta) = n \log \left(\frac{1}{\beta} \right) + \log \left(e^{-\sum_{i=1}^n Y_i/\beta} \right)$$

$$\ell(\beta) = -n \log \beta - \frac{\sum_{i=1}^n Y_i}{\beta}$$

(d) **Step 4: Differentiating the Log-Likelihood**

Differentiate the log-likelihood function with respect to β :

$$\frac{d\ell}{d\beta} = \frac{d}{d\beta} \left(-n \log \beta - \frac{\sum_{i=1}^n Y_i}{\beta} \right)$$

$$\frac{d\ell}{d\beta} = -n \cdot \frac{1}{\beta} + \frac{\sum_{i=1}^n Y_i}{\beta^2}$$

(e) **Step 5: Setting the Derivative to Zero**

Set the derivative equal to zero to find the critical points:

$$-n \cdot \frac{1}{\beta} + \frac{\sum_{i=1}^n Y_i}{\beta^2} = 0$$

$$-n + \frac{\sum_{i=1}^n Y_i}{\beta} = 0$$

$$\frac{\sum_{i=1}^n Y_i}{\beta} = n$$

$$\beta = \frac{\sum_{i=1}^n Y_i}{n}$$

(f) **Final Answer**

The Maximum Likelihood Estimator (MLE) of the population mean β is:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n Y_i$$

(g) **Conceptual Explanation**

- **Likelihood Function:** Represents the joint probability of the observed data as a function of β .
 - **Log-Likelihood Function:** The logarithm of the likelihood function simplifies the product into a sum, making differentiation easier.
 - **Differentiation and Setting to Zero:** Finding where the slope of the log-likelihood function is zero gives us the MLE.
 - **MLE of β :** The sample mean is the estimator that maximizes the likelihood function for an exponential distribution.
2. **MLE-Q2:** It is known that the probability p of tossing heads on an unbalanced coin is either $\frac{1}{4}$ or $\frac{3}{4}$. The coin is tossed twice and a value for Y , the number of heads, is observed. For each possible value of Y , which of the two values for p ($\frac{1}{4}$ or $\frac{3}{4}$) maximizes the probability that $Y = y$?

Solution

(a) **Step 1: Understanding the Binomial Distribution**

When tossing a coin twice, the number of heads Y follows a binomial distribution with parameters $n = 2$ and p (the probability of getting heads in a single toss):

$$P(Y = y) = \binom{2}{y} p^y (1 - p)^{2-y}$$

where y can be 0, 1, or 2.

(b) **Step 2: Calculate the Probabilities for $p = \frac{1}{4}$**

• **For $Y = 0$:**

$$P(Y = 0|p = \frac{1}{4}) = \binom{2}{0} \left(\frac{1}{4}\right)^0 \left(1 - \frac{1}{4}\right)^2 = 1 \cdot 1 \cdot \left(\frac{3}{4}\right)^2 = \left(\frac{3}{4}\right)^2 = \frac{9}{16}$$

• **For $Y = 1$:**

$$P(Y = 1|p = \frac{1}{4}) = \binom{2}{1} \left(\frac{1}{4}\right)^1 \left(1 - \frac{1}{4}\right)^1 = 2 \cdot \frac{1}{4} \cdot \frac{3}{4} = 2 \cdot \frac{3}{16} = \frac{6}{16} = \frac{3}{8}$$

• **For $Y = 2$:**

$$P(Y = 2|p = \frac{1}{4}) = \binom{2}{2} \left(\frac{1}{4}\right)^2 \left(1 - \frac{1}{4}\right)^0 = 1 \cdot \left(\frac{1}{4}\right)^2 \cdot 1 = \left(\frac{1}{4}\right)^2 = \frac{1}{16}$$

(c) **Step 3: Calculate the Probabilities for $p = \frac{3}{4}$**

• **For $Y = 0$:**

$$P(Y = 0|p = \frac{3}{4}) = \binom{2}{0} \left(\frac{3}{4}\right)^0 \left(1 - \frac{3}{4}\right)^2 = 1 \cdot 1 \cdot \left(\frac{1}{4}\right)^2 = \left(\frac{1}{4}\right)^2 = \frac{1}{16}$$

• **For $Y = 1$:**

$$P(Y = 1|p = \frac{3}{4}) = \binom{2}{1} \left(\frac{3}{4}\right)^1 \left(1 - \frac{3}{4}\right)^1 = 2 \cdot \frac{3}{4} \cdot \frac{1}{4} = 2 \cdot \frac{3}{16} = \frac{6}{16} = \frac{3}{8}$$

• **For $Y = 2$:**

$$P(Y = 2|p = \frac{3}{4}) = \binom{2}{2} \left(\frac{3}{4}\right)^2 \left(1 - \frac{3}{4}\right)^0 = 1 \cdot \left(\frac{3}{4}\right)^2 \cdot 1 = \left(\frac{3}{4}\right)^2 = \frac{9}{16}$$

(d) **Step 4: Compare the Probabilities for Each Y**

• **For $Y = 0$:**

$$P(Y = 0|p = \frac{1}{4}) = \frac{9}{16}$$

$$P(Y = 0|p = \frac{3}{4}) = \frac{1}{16}$$

Maximizes: $p = \frac{1}{4}$

• **For $Y = 1$:**

$$P(Y = 1|p = \frac{1}{4}) = \frac{3}{8}$$

$$P(Y = 1|p = \frac{3}{4}) = \frac{3}{8}$$

Maximizes: Both values of p give the same probability

- **For $Y = 2$:**

$$P(Y = 2|p = \frac{1}{4}) = \frac{1}{16}$$

$$P(Y = 2|p = \frac{3}{4}) = \frac{9}{16}$$

Maximizes: $p = \frac{3}{4}$

(e) **Conceptual Explanation**

- **Likelihoods:** Calculate the likelihood of observing each Y value for the given p values.
- **Comparison:** Compare these likelihoods to determine which p value maximizes the probability for each observed Y .

3. **MLE-Q3:** Given the probability density function:

$$f(y|\theta) = \begin{cases} \frac{1}{\theta^2} y e^{-y/\theta} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the MLE of θ .

Solution

(a) **Step 1: Understanding the Given pdf**

The given pdf is:

$$f(y|\theta) = \begin{cases} \frac{1}{\theta^2} y e^{-y/\theta} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is the probability density function for a gamma distribution with shape parameter $k = 2$ and scale parameter θ .

(b) **Step 2: Likelihood Function**

Given a random sample Y_1, Y_2, \dots, Y_n from this distribution, the likelihood function is the product of the individual pdfs:

$$L(\theta; Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f(Y_i; \theta)$$

Substituting the given pdf:

$$L(\theta; Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n \left(\frac{1}{\theta^2} Y_i e^{-Y_i/\theta} \right)$$

$$L(\theta; Y_1, Y_2, \dots, Y_n) = \left(\frac{1}{\theta^2} \right)^n \prod_{i=1}^n Y_i \prod_{i=1}^n e^{-Y_i/\theta}$$

$$L(\theta; Y_1, Y_2, \dots, Y_n) = \left(\frac{1}{\theta^2} \right)^n \left(\prod_{i=1}^n Y_i \right) e^{-\sum_{i=1}^n Y_i/\theta}$$

(c) **Step 3: Log-Likelihood Function**

Take the natural logarithm of the likelihood function to obtain the log-likelihood function:

$$\ell(\theta) = \log L(\theta; Y_1, Y_2, \dots, Y_n)$$

$$\ell(\theta) = \log \left(\left(\frac{1}{\theta^2} \right)^n \left(\prod_{i=1}^n Y_i \right) e^{-\sum_{i=1}^n Y_i / \theta} \right)$$

$$\ell(\theta) = n \log \left(\frac{1}{\theta^2} \right) + \log \left(\prod_{i=1}^n Y_i \right) + \log \left(e^{-\sum_{i=1}^n Y_i / \theta} \right)$$

$$\ell(\theta) = n \log \left(\frac{1}{\theta^2} \right) + \log \left(\prod_{i=1}^n Y_i \right) - \frac{\sum_{i=1}^n Y_i}{\theta}$$

$$\ell(\theta) = -2n \log \theta + \log \left(\prod_{i=1}^n Y_i \right) - \frac{\sum_{i=1}^n Y_i}{\theta}$$

(d) **Step 4: Differentiating the Log-Likelihood**

Differentiate the log-likelihood function with respect to θ :

$$\frac{d\ell}{d\theta} = \frac{d}{d\theta} \left(-2n \log \theta + \log \left(\prod_{i=1}^n Y_i \right) - \frac{\sum_{i=1}^n Y_i}{\theta} \right)$$

$$\frac{d\ell}{d\theta} = -2n \cdot \frac{1}{\theta} + \frac{\sum_{i=1}^n Y_i}{\theta^2}$$

(e) **Step 5: Setting the Derivative to Zero**

Set the derivative equal to zero to find the critical points:

$$-2n \cdot \frac{1}{\theta} + \frac{\sum_{i=1}^n Y_i}{\theta^2} = 0$$

$$-2n + \frac{\sum_{i=1}^n Y_i}{\theta} = 0$$

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

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

(f) **Final Answer**

The Maximum Likelihood Estimator (MLE) of the parameter θ is:

$$\hat{\theta} = \frac{1}{2n} \sum_{i=1}^n Y_i$$

(g) **Conceptual Explanation**

- **Likelihood Function:** Represents the joint probability of the observed data as a function of θ .
 - **Log-Likelihood Function:** The logarithm of the likelihood function simplifies the product into a sum, making differentiation easier.
 - **Differentiation and Setting to Zero:** Finding where the slope of the log-likelihood function is zero gives us the MLE.
 - **MLE of θ :** The estimator that maximizes the likelihood function for the given distribution.
4. **MLE-Q4:** Suppose that X_1, X_2, \dots, X_n denote a random sample from a Poisson distribution with parameter λ . Find the MLE of the population parameter λ .

Solution

(a) **Step 1: Understanding the Poisson Distribution**

The probability mass function (pmf) of a Poisson distributed random variable X with parameter λ is given by:

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

(b) **Step 2: Likelihood Function**

Given a random sample X_1, X_2, \dots, X_n from the Poisson distribution, the likelihood function is the product of the individual pmfs:

$$L(\lambda; X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i; \lambda)$$

Substituting the pmf of the Poisson distribution:

$$\begin{aligned} L(\lambda; X_1, X_2, \dots, X_n) &= \prod_{i=1}^n \left(\frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \right) \\ L(\lambda; X_1, X_2, \dots, X_n) &= \left(\prod_{i=1}^n \lambda^{X_i} \right) e^{-n\lambda} \left(\prod_{i=1}^n \frac{1}{X_i!} \right) \\ L(\lambda; X_1, X_2, \dots, X_n) &= \lambda^{\sum_{i=1}^n X_i} e^{-n\lambda} \prod_{i=1}^n \frac{1}{X_i!} \end{aligned}$$

(c) **Step 3: Log-Likelihood Function**

Take the natural logarithm of the likelihood function to obtain the log-likelihood function:

$$\ell(\lambda) = \log L(\lambda; X_1, X_2, \dots, X_n)$$

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

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

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

(d) **Step 4: Differentiating the Log-Likelihood**

Differentiate the log-likelihood function with respect to λ :

$$\frac{d\ell}{d\lambda} = \frac{d}{d\lambda} \left(\sum_{i=1}^n X_i \log \lambda - n\lambda - \sum_{i=1}^n \log(X_i!) \right)$$

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

(e) **Step 5: Setting the Derivative to Zero**

Set the derivative equal to zero to find the critical points:

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

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

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

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

(f) **Final Answer**

The Maximum Likelihood Estimator (MLE) of the population parameter λ is:

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

(g) **Conceptual Explanation**

- **Likelihood Function:** Represents the joint probability of the observed data as a function of λ .
 - **Log-Likelihood Function:** The logarithm of the likelihood function simplifies the product into a sum, making differentiation easier.
 - **Differentiation and Setting to Zero:** Finding where the slope of the log-likelihood function is zero gives us the MLE.
 - **MLE of λ :** The sample mean is the estimator that maximizes the likelihood function for a Poisson distribution.
5. **MLE-Q5:** Suppose that X_1, X_2, \dots, X_n denote a random sample from a geometric distribution with parameter p . Find the MLE of the population parameter p .

Solution

(a) **Step 1: Understanding the Geometric Distribution**

The probability mass function (pmf) of a geometrically distributed random variable X with parameter p is given by:

$$P(X = x; p) = (1 - p)^{x-1}p$$

for $x = 1, 2, 3, \dots$

(b) **Step 2: Likelihood Function**

Given a random sample X_1, X_2, \dots, X_n from the geometric distribution, the likelihood function is the product of the individual pmfs:

$$L(p; X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i; p)$$

Substituting the pmf of the geometric distribution:

$$L(p; X_1, X_2, \dots, X_n) = \prod_{i=1}^n (1 - p)^{X_i-1}p$$

$$L(p; X_1, X_2, \dots, X_n) = p^n (1 - p)^{\sum_{i=1}^n (X_i - 1)}$$

(c) **Step 3: Log-Likelihood Function**

Take the natural logarithm of the likelihood function to obtain the log-likelihood function:

$$\ell(p) = \log L(p; X_1, X_2, \dots, X_n)$$

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

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

(d) **Step 4: Differentiating the Log-Likelihood**

Differentiate the log-likelihood function with respect to p :

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

$$\frac{d\ell}{dp} = \frac{n}{p} + \sum_{i=1}^n (X_i - 1) \cdot \frac{-1}{1 - p}$$

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

(e) **Step 5: Setting the Derivative to Zero**

Set the derivative equal to zero to find the critical points:

$$\frac{n}{p} - \frac{\sum_{i=1}^n (X_i - 1)}{1 - p} = 0$$

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

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

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

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

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

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

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

(f) **Final Answer**

The Maximum Likelihood Estimator (MLE) of the population parameter p is:

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

(g) **Conceptual Explanation**

- **Likelihood Function:** Represents the joint probability of the observed data as a function of p .
 - **Log-Likelihood Function:** The logarithm of the likelihood function simplifies the product into a sum, making differentiation easier.
 - **Differentiation and Setting to Zero:** Finding where the slope of the log-likelihood function is zero gives us the MLE.
 - **MLE of p :** The estimator that maximizes the likelihood function for a geometric distribution.
6. **MLE-Q6:** Suppose that X_1, X_2, \dots, X_n denote a random sample from a gamma distribution with shape parameter α and rate parameter β . The pdf is given by:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

Find the MLE of the shape parameter α and rate parameter β .

Solution

(a) **Step 1: Understanding the Gamma Distribution**

The gamma distribution with shape parameter α and rate parameter β has the following pdf:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $\Gamma(\alpha)$ is the gamma function.

(b) **Step 2: Likelihood Function**

Given a random sample X_1, X_2, \dots, X_n from the gamma distribution, the likelihood function is the product of the individual pdfs:

$$L(\alpha, \beta; X_1, X_2, \dots, X_n) = \prod_{i=1}^n f(X_i; \alpha, \beta)$$

Substituting the pdf of the gamma distribution:

$$L(\alpha, \beta; X_1, X_2, \dots, X_n) = \prod_{i=1}^n \left(\frac{\beta^\alpha}{\Gamma(\alpha)} X_i^{\alpha-1} e^{-\beta X_i} \right)$$

$$L(\alpha, \beta; X_1, X_2, \dots, X_n) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \prod_{i=1}^n X_i^{\alpha-1} e^{-\beta \sum_{i=1}^n X_i}$$

(c) **Step 3: Log-Likelihood Function**

Take the natural logarithm of the likelihood function to obtain the log-likelihood function:

$$\ell(\alpha, \beta) = \log L(\alpha, \beta; X_1, X_2, \dots, X_n)$$

$$\begin{aligned}\ell(\alpha, \beta) &= \log \left(\left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \prod_{i=1}^n X_i^{\alpha-1} e^{-\beta \sum_{i=1}^n X_i} \right) \\ \ell(\alpha, \beta) &= n \log \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right) + \log \left(\prod_{i=1}^n X_i^{\alpha-1} \right) + \log \left(e^{-\beta \sum_{i=1}^n X_i} \right) \\ \ell(\alpha, \beta) &= n (\alpha \log \beta - \log \Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^n \log X_i - \beta \sum_{i=1}^n X_i\end{aligned}$$

(d) **Step 4: Differentiating the Log-Likelihood**

Differentiate the log-likelihood function with respect to α and β :

$$\begin{aligned}\frac{d\ell}{d\alpha} &= n \left(\log \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) + \sum_{i=1}^n \log X_i \\ \frac{d\ell}{d\beta} &= \frac{n\alpha}{\beta} - \sum_{i=1}^n X_i\end{aligned}$$

(e) **Step 5: Setting the Derivatives to Zero**

Set the derivatives equal to zero to find the critical points:

$$\begin{aligned}n \left(\log \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) + \sum_{i=1}^n \log X_i &= 0 \\ \log \beta - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} &= -\frac{1}{n} \sum_{i=1}^n \log X_i \\ \beta &= \frac{n\alpha}{\sum_{i=1}^n X_i}\end{aligned}$$

(f) **Final Answer**

The Maximum Likelihood Estimators (MLE) of the shape parameter α and rate parameter β are found by solving the above equations numerically since they involve the digamma function $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$:

$$\hat{\alpha} \quad (\text{solved numerically})$$

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

(g) **Conceptual Explanation**

- **Likelihood Function:** Represents the joint probability of the observed data as a function of α and β .
- **Log-Likelihood Function:** The logarithm of the likelihood function simplifies the product into a sum, making differentiation easier.

- **Differentiation and Setting to Zero:** Finding where the slope of the log-likelihood function is zero gives us the MLE.
- **MLE of α and β :** The estimators that maximize the likelihood function for a gamma distribution.

5 Feedback and Tutoring

Looking to master these and other statistics topics? I offer private one-on-one tutoring. I specialize in AI and Data Science. I also offer tutoring in all aspects of statistics, mathematics, Bayesian statistics and modeling, Python programming, JavaScript programming, R programming, Machine Learning, Deep Learning, Reinforcement Learning, Transformers, Large Language Models, and all other related AI topics.

Email me and we can setup a time to start you on a journey to mastery!