1. Problem 1: Alice's Study Habits — Stochastic Processes

Adapted from Bertsekas & Tsitsiklis, Ch. 6, Example 6.4

Alice is taking a probability class and in each week she can be either **up-to-date** (state 1) or **behind** (state 2). If she is up-to-date in a given week, the probability that she will be up-to-date the next week is 0.8, and behind is 0.2. If she is behind in a given week, the probability that she will be up-to-date the next week is 0.6, and behind is 0.4.

The transition probability matrix is:

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}$$

(a) Write the balance equations for the stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2)$.

(b) Solve for the stationary distribution $\boldsymbol{\pi}$.

(c) In the long run, what fraction of weeks is Alice up-to-date?

(d) If Alice starts behind (state 2), what is the probability that she is up-to-date two weeks later? (Compute $r_{21}(2)$ using the Chapman-Kolmogorov equation or by computing P^2.)

$-4\pi_1 + \rho\pi_1 = 0$

$\rho - 4\pi_1 = 0?$

3. Problem 3: MLE for an Exponential Distribution — Maximum Likelihood Estimation

Adapted from Matloff, Ch. 18, Section 18.1.3

Suppose $X_1, X_2, \ldots, X_n$ are a random sample from an exponential distribution with parameter $\lambda > 0$ (rate), so the density is:

PDF

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

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

(b) Write down the log-likelihood function $\ell(\lambda)$.

(c) Find the MLE $\hat{\lambda}$ by differentiating the log-likelihood, setting equal to zero, and solving.

(d) Verify that this is a maximum (not a minimum).

(e) Show that $\hat{\lambda}$ is an unbiased estimator of λ.

a) $L(\lambda) = \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdot \ldots \cdot \lambda e^{-\lambda x_n}$

$= \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$

$e^2 e^3 = e^5$

$$\boxed{\lambda^n e^{-\sum_{i=1}^{n} \lambda x_i}}$$

b) $\ell(\lambda) = \lg(L(\lambda))$

$= \log\left(\lambda^n e^{-\sum_{i=1}^{n} \lambda x_i}\right)$

$\lg(a^b) = b\lg(a)$

$\log(ab) = \log(a) + \lg(b)$

$= \lg(\lambda^n) + \lg\left(e^{-\sum_{i=1}^{n} \lambda x_i}\right)$

$= n \lg \lambda + -\sum_{i=1}^{n} \lambda x_i \cdot \lg(e)$

$$\boxed{= n \lg \lambda - \sum_{i=1}^{n} \lambda x_i}$$

$$\frac{\partial}{\partial\lambda} \ell(\lambda) = 0$$

$$\frac{\partial}{\partial\lambda} \ln\lambda = \frac{1}{\lambda}$$

$$\frac{\partial}{\partial\lambda}\left[n \ln\lambda - \sum_{i=1}^{n} \lambda x_i\right] = 0$$

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

$$\rightarrow \lambda x_1 + \lambda x_2 + \lambda x_3 \cdots$$

$$\lambda \sum x_i$$

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

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

$$\frac{\partial^2}{\partial^2\lambda} = \frac{\partial}{\partial\lambda} n\lambda^{-1} = -n\lambda^{-2} = \boxed{\frac{-n}{\lambda^2} \text{ always neg}}$$

✓

e) Show $\hat\lambda$ is <u>unbiased estimate</u> for λ

$$\mathbb{E}[\hat\lambda] = \lambda$$

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

3. Problem 3: MLE for an Exponential Distribution — Maximum Likelihood Estimation

Adapted from Matloff, Ch. 18, Section 18.1.3

Suppose $X_1, X_2, ..., X_n$ are a random sample from an exponential distribution with parameter $\lambda > 0$ (rate), so the density is:

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

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

(b) Write down the log-likelihood function $\ell(\lambda)$.

(c) Find the MLE $\hat{\lambda}$ by differentiating the log-likelihood, setting equal to zero, and solving.

(d) Verify that this is a maximum (not a minimum).

(e) Show that $\hat{\lambda}$ is an unbiased estimator of λ.

4. Problem 4: MLE for a Discrete Distribution — Maximum Likelihood Estimation

Adapted from Matloff, Ch. 18, Sections 18.1.6–18.1.7

Suppose a random variable X takes values 1, 2, and 3 with probabilities c, c, and $1 - 2c$, respectively, where $0 < c < 1/2$.

We have a random sample $X_1, ..., X_n$ from this distribution. Let N_1, N_2, N_3 denote the number of observations equal to 1, 2, and 3, respectively (so $N_1 + N_2 + N_3 = n$).

(a) Find the Method of Moments estimator $\hat{c}_{\text{MoM}}$ of c.

(b) Show that $\hat{c}_{\text{MoM}}$ is unbiased.

(c) Write down the log-likelihood function for c.

(d) Find the MLE $\hat{c}_{\text{MLE}}$.

5. Problem 5: Simple Linear Regression — Linear Regression

Adapted from Matloff, Ch. 22, Section 22.6 (Baseball Data example)

A simple linear regression of baseball players' Weight (Y, in pounds) on Height (X, in inches) yields the following R output (based on $n = 1033$ players):

Coefficient	**Estimate**	**Std. Error**
Intercept ($\hat{\beta}_0$)	-155.092	17.699
Height ($\hat{\beta}_1$)	4.841	0.240

Additionally: Residual standard error: 17.78, $R^2 = 0.2829$.

(a) Write down the fitted regression equation.

(b) Predict the weight of a player who is 72 inches tall.

(c) Construct an approximate 95% confidence interval for the true slope β_1. Interpret it.

(d) A player who is 72 inches tall actually weighs 210 lbs. What is the residual for this observation?

(e) Interpret $R^2 = 0.2829$ in context.

(f) Suppose we test $H_0 : \beta_1 = 0$. The test statistic is $t = \hat{\beta}_1 / \text{ s.e.}\left(\hat{\beta}_1\right)$. Compute it and state whether we reject H_0 at the 5% level.

6. Problem 6: Least Squares Derivation and Residuals — Linear Regression

Adapted from Matloff, Ch. 22, Sections 22.5 and 22.10

Consider n paired observations $(x_1, y_1), \ldots, (x_n, y_n)$. We wish to fit the linear model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ by minimizing the sum of squared residuals:

$$\text{SSR} = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 \quad ①$$

(a) Take partial derivatives of SSR with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, set them equal to zero, and show that the normal equations are:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i$$

$$\hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i y_i$$

(b) Solve these to show that $\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.

(c) A small dataset has $n = 5$ points: $(1, 2), (2, 4), (3, 5), (4, 4), (5, 5)$. Compute $\hat{\beta}_1$ and $\hat{\beta}_0$.

(d) Compute the five residuals and verify that they sum to zero.

$$\hat{y}_i = \hat{\beta}_1 x_i + \hat{\beta}_0 \qquad e_i = y_i - \hat{y}_i$$

$$SSR = \sum_{i=1}^{\hat{n}} e_i^2 = \sum_{i=1}^{\hat{n}} (y_i - \hat{\beta}_1 x_i - \hat{\beta}_0)^2 = (y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_0)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_0)^2 +$$

$$\frac{\partial}{\partial \hat{\beta}_0} SSR = 0 =$$

$$[y_i - \hat{\beta}_1 x_1 - \hat{\beta}_0][y_i - \hat{\beta}_1 x_i - \hat{\beta}_0]$$

$$\cancel{y_i^2} - \cancel{y_i \hat{\beta}_1 x_i} - y_i \hat{\beta}_0$$

$$- \cancel{y_i \hat{\beta}_1 x_i + (\hat{\beta}_1 x_i)^2} + \hat{\beta}_0 \hat{\beta}_1 x_i$$

$$- \hat{\beta}_0 y_i + \hat{\beta}_0 \hat{\beta}_1 x_i + \hat{\beta}_0^2$$

$$-y_i + \hat{\beta}_1 x_i - y_i + \hat{\beta}_1 x_i + 2\hat{\beta}_0$$

$$-2y_i + 2\hat{\beta}_1 x_i + 2\hat{\beta}_0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = 0 = \sum_{i=1}^{\hat{n}} [-2y_i + 2\hat{\beta}_1 x_i + 2\hat{\beta}_0]$$

$$\rightarrow \frac{\partial SSR}{\partial \hat{\beta}_1} = 0 = \sum_{i=1}^{\hat{n}} [-2 x_i y_i + 2\hat{\beta}_0 x_i + 2\hat{\beta}_1 x_i^2]$$

$$\frac{\partial SSR}{\partial \hat{B}_0} = -2\sum_{i=1}^{n}\left[y_i - \hat{B}_1 x_i - \hat{B}_0\right]$$

$$= \left[\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{B}_1 x_i - \sum_{i=1}^{n}\hat{B}_0\right] = 0$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{B}_1 x_i - n\hat{B}_0 = 0$$

$$n\hat{B}_0 \quad \hat{B}_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \Rightarrow \boxed{\text{b)}\ \hat{B}_0 = \bar{y} - \hat{B}_1\bar{x}}$$

$$\frac{\partial SSR}{\partial \hat{B}_1} = -2\sum_{i=1}^{n} x_i y_i + 2\cdot B_0\sum x_i + 2\hat{B}_1\sum x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i = \hat{B}_0\sum x_i + \hat{B}_1\sum x_i^2$$

$$\sum_{i=1}^{n} x_i y_i = (\bar{y} - \hat{B}_1\bar{x})\sum x_i + \hat{B}_1\sum x_i^2$$

$$\sum x_i y_i = \frac{1}{n}\sum y_i - \hat{B}_1\frac{1}{n}\sum x_i\sum x_i + \hat{B}_1\sum x_i^2$$

$$= \bar{y} - \hat{B}_1\bar{x}\cdot n\bar{x} + \hat{B}_1\sum x_i^2$$

$$\hat{B}_1\left(\sum x_i^2 - \bar{x}\cdot n\bar{x}\right) =$$

(b) From the first normal equation: $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$.

Substituting into the second:

$$\left(\overline{y} - \hat{\beta}_1\overline{x}\right)\sum x_i + \hat{\beta}_1\sum x_i^2 = \sum x_i y_i$$

$$\overline{y}\cdot n\overline{x} - \hat{\beta}_1\overline{x}\cdot n\overline{x} + \hat{\beta}_1\sum x_i^2 = \sum x_i y_i$$

$$\hat{\beta}_1\left(\sum x_i^2 - n\overline{x}^2\right) = \sum x_i y_i - n\overline{x}\overline{y}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - n\overline{x}^2} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum\left(x_i - \overline{x}\right)^2}$$

7. Problem 7: Confidence Interval for Difference of Means — Hypothesis Testing

Adapted from Matloff, Ch. 16, Section 16.6.2 (Network Security example)

In a network security study, researchers compare round-trip packet travel times for wired and wireless networks. The summary statistics are:

Sample	**Mean ($\overline{X}$)**	**Std. Dev. (s)**	**Size (n)**
Wired	2.00 ms	6.30 ms	436
Wireless	11.52 ms	9.94 ms	344

(a) State the null and alternative hypotheses for testing whether the population mean round-trip times differ between wired and wireless networks.

(b) Construct an approximate 95% confidence interval for $\mu_{\text{wireless}} - \mu_{\text{wired}}$.

(c) Compute the test statistic Z for testing $H_0 : \mu_{\text{wireless}} - \mu_{\text{wired}} = 0$.

(d) What is your conclusion at the 5% significance level? Interpret in context.

(e) The reported p-value is approximately 2.2×10^{-16}. Does this tiny p-value mean the difference is *large*? Explain carefully.

8. Problem 8: Hypothesis Test for a Proportion — Hypothesis Testing

Adapted from Matloff, Ch. 17, Sections 17.1 and 17.4

A coin will be flipped at the Super Bowl to determine the first kickoff. To assess fairness, you toss the coin $n = 100$ times and observe 62 heads. Let p denote the true probability of heads.

(a) State the null and alternative hypotheses for testing whether the coin is fair.

(b) Under H_0, write down the distribution of $\hat{p}$ (the sample proportion of heads) and the test statistic Z.

(c) Compute Z for the observed data.

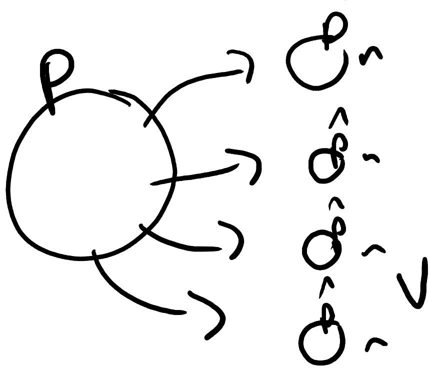
(d) At the $\alpha = 0.05$ significance level, do we reject H_0? Explain.

(e) Compute the p-value for this test. Interpret it.

(f) Now suppose we only toss the coin 10 times and decide to reject H_0 if we get 8 or more heads (using a one-sided test $H_A : p > 0.5$). What is the exact significance level α?

a) $H_0 : p = 0.5$ $H_A : p \neq 0.5$

b)



$\hat{p} = \frac{\sum x_i}{n}$ $x_i \sim \text{Bern}(p)$

$E(\hat{p}) = p$ $Y = \sum x_i$

$Var(\hat{p}) = \frac{Var(x_i)}{n}$ $Y \sim \text{Binom}(n, p)$

CLT $\hat{p} \sim N\left(p, \frac{p\cdot(1-p)}{n}\right)$ $Y \sim \text{Binom}(n, p)$

$E[Y] = n \cdot p$

fair:

$p = 0.5$ $Var(Y) = n \cdot p \cdot (1-p)$

$\hat{p} \sim N\left(0.5, \frac{0.5 \cdot 0.5}{n}\right)$

$Std(\hat{p}) = \sqrt{\frac{0.5^2}{n}}$ $n = 100$

Standardize:

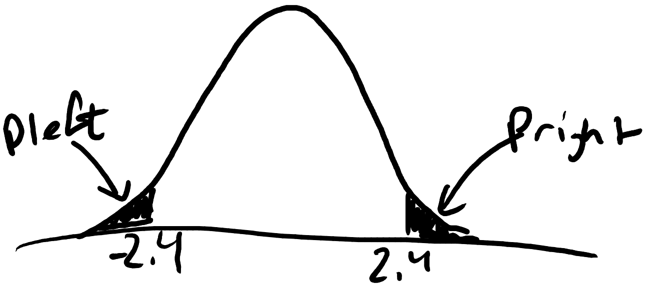
$$Z = \frac{\hat{p} - E[\hat{p}]}{Std(\hat{p})} = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5^2}{n}}} = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.25}{100}}} = \frac{\hat{p} - 0.5}{.05}$$

$\frac{.62 - .5}{.05} = 2.4$ $|Z| \leq 1.96$

$\hat{p} \sim N(0.5, .0025)$

$z \sim N(0, 1)$

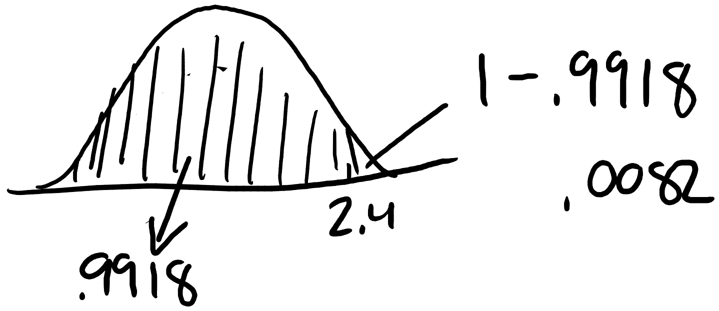
$Z = 2.4$



$p_{left} = p_{right} = .0082$

$p = p_{left} + p_{right}$

$= .0164$ $\alpha = .05$



$\boxed{p = .0164}$ $p < \alpha$

reject H_0

1.64% chance fair coin

f) reject if 8 or more heads out of 10 tosses

reject if $p < \alpha$

$$P(\geq \underbrace{8 \text{ H in 10 tosses}}_{\text{Binom}})$$

$$= \left.\begin{matrix} \binom{10}{8} p^8 (1-p)^2 + \\ \binom{10}{9} p^9 (1-p)^1 + \\ \binom{10}{10} p^{10} (1-p)^0 \end{matrix}\right\} \xrightarrow{p = 0.5} \begin{matrix} \binom{10}{8} 0.5^{10} + \\ \binom{10}{9} 0.5^{10} + \\ \binom{10}{10} 0.5^{10} \end{matrix} \approx \boxed{0.055 = \alpha}$$

t vs. z:

95% CI: $(\hat{p} - 1.96\,\text{s.e.}, \hat{p} + 1.96\,\text{s.e.})$

$\hat{p}$ 95% CI: $(-1.96, 1.96)$

$Z(\hat{p})$

t: need to compute p-value & have sample means, variances t-test : test for diffs in sample means

Z-score : testing proportions

~~X_1 ↑ #umbrellas @ home t=1~~ ~~X_2 ↑ #umbrellas @ home at t=2~~

X_1 #umb. @ home X_2 #umb @ office

2. Problem 2: The Absent-Minded Professor — Stochastic Processes

Adapted from Bertsekas & Tsitsiklis, Ch. 6, Example 6.5

An absent-minded professor has two umbrellas that she uses when commuting from home to office and back. If it rains and an umbrella is available at her current location, she takes it. If it is not raining, she always forgets to take an umbrella. Suppose it rains with probability p each time she commutes, independently of other times.

Model this as a Markov chain where state i means i umbrellas are available at her current location ($i = 0, 1, 2$).

(a) Write down the transition probability matrix.

(b) Find the stationary distribution in terms of p.

(c) What is the long-run probability that the professor gets wet on a given commute?

(d) Evaluate your answer for $p = 0.5$.

$$T = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & & p_{01} & \\ 1 & & & \\ 2 & & & \end{array}$$

$$p_{01} = P(X_{t+1} = 1 \mid X_t = 0)$$

office

a) $$\begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ \text{home } 1 \\ 2 \end{array} & \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix} \end{array} = T$$

$$\pi P = \pi \qquad \pi = [\pi_0 \quad \pi_1 \quad \pi_2] \qquad \sum_{i=0}^{2} \pi_i = 1$$

$$\pi_0 + \pi_1 + \pi_2 = 1$$

$$[\pi_0 \quad \pi_1 \quad \pi_2] \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix} = [\pi_0 \quad \pi_1 \quad \pi_2]$$

$$0 \cdot \pi_0 + 0 \cdot \pi_1 + (1-p)\pi_2 = \pi_0 \rightarrow \pi_0 = (1-p)\pi_2$$

$$0 \cdot \pi_0 + (1-p)\pi_1 + p\pi_2 = \pi_1 \rightarrow \pi_1 = (1-p)\pi_1 + p\pi_2$$

$$1 \cdot \pi_0 + p\pi_1 + 0 \cdot \pi_2 = \pi_2 \rightarrow \pi_2 = \pi_0 + p\pi_1$$

$$\pi_0 = (1-p)\pi_2$$

$$(1-p)\pi_2 + \pi_1 + \pi_2 = 1$$

$$\pi_1 = (1-p)\pi_1 + p\pi_2$$

$$(\cancel{1 - 1} + \cancel{p})\pi_1 = \cancel{p}\pi_2$$

$$\pi_1 = \pi_2$$

$$(1-p)\pi_2 + \pi_2 + \pi_2 = 1$$

$$(3-p)\pi_2 = 1$$

$$\boxed{\begin{aligned} \pi_2 &= \frac{1}{3-p} \\ \pi_1 &= \frac{1}{3-p} \\ \pi_0 &= \frac{1-p}{3-p} \end{aligned}}$$

$$P(\text{wet}) = P(0 \text{ umbrellas} \cap \text{rain})$$

$$= P(\underset{\pi_0}{0 \text{ umb.}}) \cdot P(\underset{p}{\text{rain}})$$

$$= \frac{1-p}{3-p} \cdot p$$

d) $p = 0.5$

$$\frac{1-0.5}{3-0.5} = \frac{0.5}{2.5} = 0.2$$

$$\frac{(1-0.5)\,0.5}{3-0.5} = 0.1$$

$\boxed{\text{10 \% chance}}$