MATH 541B, GRADUATE STATISTICS
SELECTED HOMEWORK SOLUTIONS

STEVEN HEILMAN

Contents

1. Homework 1

Exercise 1.1. Estimate the probability that 1000000 coin flips of fair coins will result in more than $501,000$ heads, using the Central Limit Theorem. (Some of the following integrals may be relevant: $\int_{-\infty}^{0} e^{-t^2/2}dt/\sqrt{2\pi} = 1/2$, $\int_{-\infty}^{1} e^{-t^2/2}dt/\sqrt{2\pi} \approx .8413$, $\int_{-\infty}^{2} e^{-t^2/2}dt/\sqrt{2\pi} \approx .9772$, $\int_{-\infty}^{3} e^{-t^2/2}dt/\sqrt{2\pi} \approx .9987$.) (Hint: use Bernoulli random variables.)

Casinos do these kinds of calculations to make sure they make money and that they do not go bankrupt. Financial institutions and insurance companies do similar calculations for similar reasons.

Solution. For any $1 \leq 1$, let $X_i = 1$ if the i^{th} coin flip is heads and $X_i = 0$ otherwise. We assume that $X_1, \ldots$ are iid with $\mathbf{P}(X_1 = 1) = 1/2$, $\mathbf{E}X_1 = 1/2$ and $\mathrm{var}(X_1) = 1/4$. We want to know the probability that

$$X_1 + \cdots + X_{10^7} > 501000.$$

Equivalently, we want the probability of the event

$$\{X_1 + \cdots + X_{10^7} - 10^7/2 > 1000\} = \{\frac{X_1 + \cdots + X_{10^7} - 10^7/2}{\sqrt{10^6}\sqrt{1/4}} > 2\} =$$

Using the Central Limit Theorem as an approximation, we have the approximation

$$\mathbf{P}\left(\frac{X_1 + \cdots + X_{10^7} - 10^7/2}{\sqrt{10^6}\sqrt{1/4}} > 2\right) \approx \int_2^\infty e^{-x^2/2}dx/\sqrt{2\pi}$$
$$= 1 - \int_\infty^2 e^{-x^2/2}dx/\sqrt{2\pi} \approx 1 - .9772 = .0228.$$

□

Date: November 28, 2023

Exercise 1.2 (Numerical Integration). In computer graphics in video games, etc., various integrations are performed in order to simulate lighting effects. Here is a way to use random sampling to integrate a function in order to quickly and accurately render lighting effects. Let $\Omega = [0, 1]$, and let $\mathbf{P}$ be the uniform probably law on Ω, so that if $0 \leq a < b \leq 1$, we have $\mathbf{P}([a, b]) = b - a$. Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbf{P}(X_i \in [a, b]) = b - a$ for all $0 \leq a < b \leq 1$, for all $i \in \{1, \ldots, n\}$. Let $f\colon [0, 1] \to \mathbb{R}$ be a continuous function we would like to integrate. Instead of integrating f directly, we instead compute the quantity

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Show that

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = \int_0^1 f(t)dt.$$

$$\lim_{n\to\infty} \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = 0.$$

That is, as n becomes large, $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ is a good estimate for $\int_0^1 f(t)dt$.

Solution. By definition of X_i we have $\mathbf{E}f(X_i) = \int_0^1 f(t)dt$ for all $i \geq 1$so that $\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = \frac{1}{n}n\int_0^1 f(t)dt = \int_0^1 f(t)dt$. Also, by independence we have

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(f(X_i)) = \frac{1}{n}\mathrm{var}(f(X_1)).$$

This quantity goes to zero as $n \to \infty$. (Since f is continuous on $[0, 1]$, f is bounded by some constant c on $[0, 1]$, i.e. $|f(t)| \leq c$ for all $t \in [0, 1]$, so $|f(X_1)| \leq c$, so $\mathrm{var} f(X_i) \leq \mathbf{E}[f(X_i)]^2 \leq c^2$ for all $i \geq 1$.) □

Exercise 1.3. Let $X := (X_1, \ldots, X_n)$ be a random sample of size n from a binomial distribution with parameters n and p. Here n is a positive (known) integer and $0 < p < 1$ is unknown. (That is, $X_1, \ldots, X_n$ are i.i.d. and X_1 is a binomial random variable with parameters n and p, so that $\mathbf{P}(X_1 = k) = \binom{n}{k}p^k(1-p)^{n-k}$ for all integers $0 \leq k \leq n$.)

You can freely use that $\mathbf{E}X_1 = np$ and $\mathrm{Var}X_1 = np(1-p)$.

- Computer the Fisher information $I_X(p)$ for any $0 < p < 1$. (Consider n to be fixed.)
- Let Z be an unbiased estimator of p^2 (assume that Z is a function of $X_1, \ldots, X_n$). State the Cramér-Rao inequality for Z.
- Let W be an unbiased estimator of $1/p$ (assume that W is a function of $X_1, \ldots, X_n$). State the Cramér-Rao inequality for W.

Solution. Using that the information of independent random variables is the sum of the informations, using the alternate definition of Fisher information using the variance, and using that the variance is unchanged by adding a constant inside the variance,

$$
\begin{aligned}
I_X(p) = nI_{X_1}(p) &= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[\log\Big(\binom{n}{X_1}p^{X_1}(1-p)^{n-X_1}\Big)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[\log\binom{n}{X_1} + X_1\log p + (n-X_1)\log(1-p)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{d}{dp}\Big[X_1\log p + (n-X_1)\log(1-p)\Big]\Big) \\
&= n\mathrm{Var}_p\Big(\frac{1}{p}X_1 - \frac{1}{1-p}(n-X_1)\Big) = n\mathrm{Var}_p\Big(\Big[\frac{1}{p}+\frac{1}{1-p}\Big]X_1\Big) \\
&= n\Big[\frac{1}{p}+\frac{1}{1-p}\Big]^2\mathrm{Var}_pX_1 = n\Big[\frac{1}{p(1-p)}\Big]^2 np(1-p) = \frac{n^2}{p(1-p)}
\end{aligned}
$$

The Cramér-Rao inequality says, if $g(p) := \mathbf{E}_pZ$, then

$$\mathrm{Var}_p(Z) \geq \frac{|g'(p)|^2}{I_X(p)}.$$

If $g(p) = p^2$, then $g'(p) = 2p$, so we get

$$\mathrm{Var}_p(Z) \geq \frac{(2p)^2}{I_X(p)} = \frac{4p^3(1-p)}{n^2}.$$

If $g(p) = 1/p$, then $g'(p) = -p^{-2}$, so we get

$$\mathrm{Var}_p(Z) \geq \frac{p^{-4}}{I_X(p)} = p^{-3}\frac{1-p}{n^2}.$$

□

Exercise 1.4. Let $X_1, \ldots, X_n$ be a random sample of size n from a Poisson distribution with unknown parameter $\lambda > 0$. (So, $\mathbf{P}(X_1 = k) = e^{-\lambda}\lambda^k/k!$ for all integers $k \geq 0$.)

- Find an MLE (maximum likelihood estimator) for λ.
- Is the MLE you found unique? That is, could there be more than one MLE for this problem?

Solution. The MLE of θ is a value of θ maximizing

$$\log\prod_{i=1}^{n}\theta^{X_i}e^{-\theta}/X_i! = \log\Big(\theta^{\sum_{i=1}^n X_i}e^{-n\theta}\prod_{i=1}^{n}[X_i!]\Big) = \sum_{i=1}^{n}\log(X_i!) - n\theta + \log\theta\sum_{i=1}^{n}X_i.$$

Taking a derivative in θ, we get $-n + \frac{1}{\theta}\sum_{i=1}^n X_i$. From the first derivative test, there is a unique maximum value of θ when $\theta = \frac{1}{n}\sum_{i=1}^n X_i$, so the MLE for θ is $\frac{1}{n}\sum_{i=1}^n X_i$. □

Exercise 1.5. Suppose X is a Gaussian distributed random variable with known variance $\sigma^2 > 0$ but unknown mean. Fix $\mu_0, \mu_1 \in \mathbb{R}$. Assume that $\mu_0 - \mu_1 > 0$. We want to test the hypothesis H_0 that $\mu = \mu_0$ versus the hypothesis H_1 that $\mu = \mu_1$. Fix $\alpha \in (0, 1)$. Explicitly describe the UMP test for the class of tests whose significance level is at most α.

Your description of the test should use the function $\Phi(t) := \int_{-\infty}^{t} e^{-x^2/2}dx/\sqrt{2\pi}$, $\Phi\colon \mathbb{R} \to (0,1)$, and/or the function $\Phi^{-1}\colon (0,1) \to \mathbb{R}$. (Recall that $\Phi(\Phi^{-1}(s)) = s$ for all $s \in (0,1)$ and $\Phi^{-1}(\Phi(t)) = t$ for all $t \in \mathbb{R}$.)

Solution. From the Neyman-Pearson Lemma, the UMP is a likelihood ratio test (LRT). Let $k > 0$. (Since $\mathbf{P}_{\theta_0}(f_{\theta_1}(X) = kf_{\theta_0}(X)) = \mathbf{P}_{\theta_1}(f_{\theta_1}(X) = kf_{\theta_0}(X)) = 0$, the UMP is non randomized.) In this case, the LRT has rejection region

$$C := \{x \in \mathbb{R} \colon f_{\theta_1}(x) > kf_{\theta_0}(x)\}.$$

More specifically,

$$\begin{aligned}
C &:= \{x \in \mathbb{R} \colon f_{\theta_1}(x) > kf_{\theta_0}(x)\} \\
&= \{x \in \mathbb{R} \colon \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} > k\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu_0)^2}{2\sigma^2}}\} \\
&= \{x \in \mathbb{R} \colon -\frac{(x-\mu_1)^2}{2\sigma^2} > \log(k) - \frac{(x-\mu_0)^2}{2\sigma^2}\} \\
&= \{x \in \mathbb{R} \colon (x-\mu_1)^2 < -2\sigma^2\log(k) + (x-\mu_0)^2\} \\
&= \{x \in \mathbb{R} \colon (x-\mu_1)^2 - (x-\mu_0)^2 < -2\sigma^2\log(k)\} \\
&= \{x \in \mathbb{R} \colon (2x-\mu_1-\mu_0)(\mu_0-\mu_1) < -2\sigma^2\log(k)\} \\
&= \{x \in \mathbb{R} \colon 2x-\mu_1-\mu_0 < -\frac{2\sigma^2\log(k)}{\mu_0-\mu_1}\} \\
&= \{x \in \mathbb{R} \colon x < -\frac{\sigma^2\log(k)}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2}\}.
\end{aligned}$$

The significance level of this test is

$$\begin{aligned}
\sup_{\theta\in\Theta_0}\beta(\theta) &= \beta(\mu_0) = \mathbf{P}_{\mu_0}(X \in C) = \mathbf{P}_{\mu_0}(X < -\frac{\sigma^2\log(k)}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2}) \\
&= \mathbf{P}_{\mu_0}(X-\mu_0 < -\frac{\sigma^2\log(k)}{\mu_0-\mu_1} + \frac{\mu_0+\mu_1}{2} - \mu_0) \\
&= \mathbf{P}_{\mu_0}\Big(\frac{X-\mu_0}{\sigma} < -\frac{\sigma\log(k)}{\mu_0-\mu_1} + \frac{\mu_1-\mu_0}{2\sigma}\Big) = \Phi\Big(-\frac{\sigma\log(k)}{\mu_0-\mu_1} + \frac{\mu_1-\mu_0}{2\sigma}\Big)
\end{aligned}$$

So, if we want a fixed significance level $\alpha \in (0,1)$, then

$$\Phi^{-1}(\alpha) = -\frac{\sigma\log(k)}{\mu_0-\mu_1} + \frac{\mu_1-\mu_0}{2\sigma}.$$

That is, we choose k such that

$$-\Phi^{-1}(\alpha) + \frac{\mu_1-\mu_0}{2\sigma} = \frac{\sigma\log(k)}{\mu_0-\mu_1}.$$

i.e.

$$k = \exp\Big(\frac{\mu_1-\mu_0}{\sigma}\Phi^{-1}(\alpha) - \frac{(\mu_0-\mu_1)^2}{2\sigma^2}\Big).$$

□

Exercise 1.6. This exercise demonstrates that a UMP might not always exists.

Let $X_1, \ldots, X_n$ be i.i.d. Gaussian random variables with known variance and unknown mean $\mu \in \mathbb{R}$. Fix $\mu_0 \in \mathbb{R}$. Let H_0 denote the hypothesis $\{\mu = \mu_0\}$ and let H_1 denote the hypothesis $\mu \neq \mu_0$. Fix $0 < \alpha < 1$. Let $\mathcal{T}$ denote the set of hypothesis tests with significance level at most α. Show that no UMP class $\mathcal{T}$ test exists, using the following strategy.

- Let $\mu_1 < \mu_0$. You may take as given the following fact (that follows from the Karlin-Rubin Theorem): the power at μ_1 is maximized among class $\mathcal{T}$ tests by the hypothesis test ϕ that rejects H_0 when the sample mean satisfies $\overline{X} < c$ for an appropriate choice of $c \in \mathbb{R}$. Assume for the sake of contradiction that a UMP class $\mathcal{T}$ test ϕ' exists. Then, using the necessity part of the Neyman-Pearson Lemma (i.e. consider testing $\mu = \mu_0$ versus $\mu = \mu_1$), conclude that ϕ' must have the same rejection region as ϕ (just by examining the power of the tests at μ_1.)
- Consider now a test in $\mathcal{T}$ that rejects H_0 when the sample mean satisfies $\overline{X} > c'$ for an appropriate choice of $c' \in \mathbb{R}$. Repeating the previous argument, conclude that ϕ' must reject when $\overline{X} > c'$, leading to a contradiction.

  That is, let $\mu_2 > \mu_0$. You may take as given the following fact (that follows from the Karlin-Rubin Theorem): the power at μ_2 is maximized among class $\mathcal{T}$ tests by the hypothesis test ϕ'' that rejects H_0 when the sample mean satisfies $\overline{X} > c'$ for an appropriate choice of $c' \in \mathbb{R}$. Then, using the necessity part of the Neyman-Pearson Lemma (i.e. consider testing $\mu = \mu_0$ versus $\mu = \mu_2$), conclude that ϕ' must have the same rejection region as ϕ''.

Solution. Since ϕ' is UMP class $\mathcal{T}$ for testing H_0 versus H_1, we have $\beta'(\mu_1) \geq \beta(\mu_1)$. (Here β is the power function of ϕ, and β' is the power function of ϕ'.) From the remark about the Karlin-Rubin Theorem, $\beta'(\mu_1) \leq \beta(\mu_1)$. Therefore, $\beta'(\mu_1) = \beta(\mu_1)$.

Consider $H_1' = \{\mu = \mu_1\}$. Suppose we are testing H_0 versus H_1'. Since $\beta(\mu_1) = \beta'(\mu_1)$, from the Neyman-Pearson Lemma, we must have $\phi' = \phi$ except possibly on a set of probability zero with respect to $\mathbf{P}_{\mu_0}$ and $\mathbf{P}_{\mu_1}$. (Similarly it occurs with probability zero that $f_{\theta_0}(X) = kf_{\theta_1(X)}$ for a constant $k > 0$.) That is, up to probability zero changes to ϕ, both ϕ and ϕ' are nonrandomized hypothesis tests with the same rejection region.

Now, let $\mu_2 > \mu_0$. Since ϕ' is UMP class $\mathcal{T}$ for testing H_0 versus H_1, we have $\beta'(\mu_2) \geq \beta''(\mu_2)$. (Here β' is the power function of ϕ', and β'' is the power function of ϕ''.) From the remark about the Karlin-Rubin Theorem, $\beta'(\mu_2) \leq \beta''(\mu_2)$. Therefore, $\beta''(\mu_2) = \beta'(\mu_2)$.

Consider $H_1'' = \{\mu = \mu_2\}$. Suppose we are testing H_0 versus H_1''. Since $\beta'(\mu_2) = \beta''(\mu_2)$, from the Neyman-Pearson Lemma, we must have $\phi'' = \phi'$ except possibly on a set of probability zero with respect to $\mathbf{P}_{\mu_0}$ and $\mathbf{P}_{\mu_1}$. (Similarly it occurs with probability zero that $f_{\theta_0}(X) = kf_{\theta_1(X)}$ for a constant $k > 0$.) That is, up to probability zero changes to ϕ'', both ϕ'' and ϕ' are nonrandomized hypothesis tests with the same rejection region.

We now have a contradiction, since ϕ' must reject only when $\overline{X} > c$, and ϕ' must reject only when $\overline{X} < c'$.

□

Exercise 1.7. The rejection regions C_α for UMP hypothesis tests of significance level at most $\alpha \in (0, 1)$ are often nested in the sense that $C_\alpha \subseteq C_{\alpha'}$ for all $0 < \alpha < \alpha' < 1$. This exercise demonstrates an example of UMP tests where this nesting behavior does not occur.

Let $\theta_0, \theta_1 \in \mathbb{R}$ be unequal parameters. Let H_0 denote the hypothesis $\{\theta = \theta_0\}$ and let H_1 denote the hypothesis $\{\theta = \theta_1\}$. Suppose $X \in \{1, 2, 3\}$ is a random variable. If $\theta = \theta_0$, assume that X takes the values $1, 2, 3$ with probabilities $.85, .1, .05$, respectively. If $\theta = \theta_1$, assume that X takes the values $1, 2, 3$ with probabilities $.7, .2, .1$, respectively. Let $\mathcal{T}$ denote the set of hypothesis tests with significance level at most α.

- Let $0 < \alpha < .15$. Show that a UMP class $\mathcal{T}$ test is not unique.
- When $\alpha = .05$, show there is a unique nonrandomized hypothesis UMP class $\mathcal{T}$ test.
- When $\alpha = .1$, show there is a unique nonrandomized hypothesis UMP class $\mathcal{T}$ test.
- Show that the $\alpha = .05$ and $\alpha' = .1$ UMP nonrandomized tests from above do not have nested rejection regions.
- However, when $\alpha = .05$ and $\alpha' = .1$, there are randomized UMP tests $\phi, \phi' \colon \mathbb{R}^n \to [0, 1]$ respectively, that are nested in the sense that $\phi \leq \phi'$.

Solution. We have

$$\frac{f_{\theta_1}(1)}{f_{\theta_0}(1)} = \frac{.7}{.85} = \frac{14}{17}, \qquad \frac{f_{\theta_1}(2)}{f_{\theta_0}(2)} = \frac{.2}{.1} = 2, \qquad \frac{f_{\theta_1}(3)}{f_{\theta_0}(3)} = \frac{.1}{.05} = 2.$$

The Neyman-Pearson Lemma says that likelihood ratio tests $\phi \colon \{1, 2, 3\} \to \mathbb{R}$ of the following form are UMP

$$\phi(x) := \begin{cases} 1 & , \text{if } f_{\theta_1}(x) > k f_{\theta_0}(x) \\ 0 & , \text{if } f_{\theta_1}(x) < k f_{\theta_0}(x) \\ ? & , \text{if } f_{\theta_1}(x) = k f_{\theta_0}(x). \end{cases}$$

So, let us examine those tests for all possible $k > 0$. After examining these different tests, we realize that the case $k = 2$ is of particular interest for this problem, so let us focus on that case.

If $k = 2$, then we have two points $x = 2, 3$ such that we can specify the value of ϕ arbitrarily, while maintaining the UMP property. That is,

$$\phi(x) := \begin{cases} 1 & , \text{if } f_{\theta_1}(x) > k f_{\theta_0}(x) \\ 0 & , \text{if } f_{\theta_1}(x) < k f_{\theta_0}(x) \\ ? & , \text{if } f_{\theta_1}(x) = k f_{\theta_0}(x) \end{cases} = \begin{cases} 0 & , \text{if } x = 1 \\ ? & , \text{if } x = 2, 3. \end{cases}$$

More specifically, for any $0 \leq a, b \leq 1$, $\phi \colon \{1, 2, 3\} \to \mathbb{R}$ is UMP where

$$\phi(x) := \begin{cases} 0 & , \text{if } x = 1 \\ a & , \text{if } x = 2 \\ b & , \text{if } x = 3 \end{cases}$$

A test of this form has power function

$$\beta(\theta) = \mathbf{E}_\theta \phi(X).$$

The significance level of this test is

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \beta(\theta_0) = .1\phi(2) + .05\phi(3) = (.1)a + (.05)b.$$

If $0 < \alpha < .15$ is our desired significance level, then any choice of $0 \leq a, b \leq 1$ satisfying

$$(.1)a + (.05)b = \alpha$$

is a UMP with significance level α. (The Neyman-Pearson Lemma guarantees this holds.) For a fixed $0 < \alpha < .15$, infinitely many such a, b exist. So, the UMP tests in this case are non-unique.

If $\alpha = .05$, then the set

$$\{(a,b)\colon (.1)a + (.05)b = .05,\ 0 \leq a, b \leq 1\}$$

has a unique element where one of a, b is zero, occurring when $a = 0$ and $b = 1$. So, when $\alpha = .05$, there is a unique nonrandomized UMP test. This test rejects H_0 when $X = 3$.

If $\alpha = .1$, then the set

$$\{(a,b)\colon (.1)a + (.05)b = .1,\ 0 \leq a, b \leq 1\}$$

has a unique element where one of a, b is zero, occurring when $a = 1$ and $b = 0$. So, when $\alpha = .1$, there is a unique nonrandomized UMP test. This test rejects H_0 when $X = 2$.

The above rejection regions are not nested, since the events $\{X = 3\}$ and $\{X = 2\}$ are disjoint.

However, there is are randomized hypothesis tests ϕ, ϕ' with significance level $\alpha = .05, \alpha' = .1$ respectively, such that $\phi \leq \phi'$. For example, we could use

$$\phi(x) := \begin{cases} 0 & \text{, if } x = 1 \\ 1/4 & \text{, if } x = 2 \\ 1/2 & \text{, if } x = 3 \end{cases}, \qquad \phi'(x) := \begin{cases} 0 & \text{, if } x = 1 \\ 1/2 & \text{, if } x = 2 \\ 1 & \text{, if } x = 3 \end{cases}$$

□

2. Homework 2

Exercise 2.1. Prove the following version of the Karlin-Rubin Theorem, with the inequalities reversed in the definition of the hypotheses.

Let $\{f_\theta\}$ be a family of PDFs with the MLR property, with respect to a real-valued statistic $Y = t(X)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Let $0 \leq \gamma \leq 1$. Fix $\theta_0 \in \Theta$. Consider the hypothesis $H_0 = \{\theta \geq \theta_0\}$ and the hypothesis $H_1 = \{\theta < \theta_0\}$. Let $c \in \mathbb{R}$. Consider the randomized hypothesis test $\phi\colon \mathbb{R}^n \to [0,1]$ defined by

$$\phi(x) := \begin{cases} 0 & \text{, if } t(x) > c \\ 1 & \text{, if } t(x) < c \\ \gamma & \text{, if } t(x) = c. \end{cases}$$

Define $\alpha := \mathbf{E}_{\theta_0}\phi(X)$. Let $\mathcal{T}$ be the class of all randomized hypothesis tests with significance level at most α.

(i) ϕ is UMP class $\mathcal{T}$.
(iii) β, the power function of ϕ, is nonincreasing and strictly decreasing when it takes values in $(0, 1)$.

Proof. We first prove (iii). Let $\theta_1 > \theta_0$ and consider the function $r\colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$r(x) := \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}, \qquad \forall\, x \in \mathbb{R}^n.$$

By assumption, r is a strictly increasing function of $t(x)$. Let $k \in \mathbb{R}$ such that $r(x) = k$ when $t(x) = c$. Since r is a strictly increasing function of $t(x)$, we can rewrite ϕ as

$$\phi(x) = \begin{cases} 0 & \text{, if } r(x) > k \\ 1 & \text{, if } r(x) < k \\ \gamma & \text{, if } r(x) = k. \end{cases}$$

That is, $1 - \phi$ is a likelihood ratio test of the hypothesis $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$. Corollary 3.15 from the notes says $1 - \beta(\theta_1) = \mathbf{E}_{\theta_1}(1 - \phi(X)) > 1 - \alpha = \mathbf{E}_{\theta_0}(1 - \phi(X)) = 1 - \beta(\theta_0)$, if $\mathbf{P}_{\theta_0} \neq \mathbf{P}_{\theta_1}$. (If $\mathbf{P}_{\theta_0} = \mathbf{P}_{\theta_1}$, then $\mathbf{E}_{\theta_1}\phi(X) = \mathbf{E}_{\theta_0}\phi(X) \in \{0, 1\}$ since ϕ is either zero or one with probability one in this case, i.e. $\alpha \in \{0, 1\}$.) Assertion (iii) follows.

We now prove (i). First, note that $\alpha = \mathbf{E}_{\theta_0}\phi(X) = \sup_{\theta \geq \theta_0} \mathbf{E}_{\theta}\phi(X)$ from (iii), so that ϕ is in class $\mathcal{T}$. Now let $\theta_1 < \theta_0$, and let ϕ' be a class $\mathcal{T}$ hypothesis test. By definition of $\mathcal{T}$, $\mathbf{E}_{\theta_0}\phi' \leq \sup_{\theta \geq \theta_0} \mathbf{E}_{\theta}\phi'(X) \leq \alpha$. So, from the Neyman-Pearson Lemma (sufficiency), ϕ is UMP (in the context of that Lemma), i.e. $\mathbf{E}_{\theta_1}\phi(X) \geq \mathbf{E}_{\theta_1}\phi'(X)$. Since this inequality holds for all $\theta_1 < \theta_0$, we conclude that ϕ is UMP class $\mathcal{T}$, i.e. (i) holds. □

Exercise 2.2. Prove the following one-sided version of the Karlin-Rubin Theorem.

Let $\{f_\theta\}$ be a family of PDFs with the MLR property, with respect to a real-valued statistic $Y = t(X)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Let $0 \leq \gamma \leq 1$. Fix $\theta_0 \in \Theta$. Consider the hypothesis $H_0 = \{\theta = \theta_0\}$ and the hypothesis $H_1 = \{\theta > \theta_0\}$. Let $c \in \mathbb{R}$. Consider the randomized hypothesis test $\phi\colon \mathbb{R}^n \to [0, 1]$ defined by

$$\phi(x) := \begin{cases} 1 & \text{, if } t(x) > c \\ 0 & \text{, if } t(x) < c \\ \gamma & \text{, if } t(x) = c. \end{cases}$$

Define $\alpha := \mathbf{E}_{\theta_0}\phi(X)$. Let $\mathcal{T}$ be the class of all randomized hypothesis tests with significance level at most α.

Then ϕ is UMP class $\mathcal{T}$.

Proof. Let $\theta_1 > \theta_0$. From the Karlin-Rubin Theorem itself (part (iii)), we already know that the power function β of ϕ is nondecreasing. Also, as we proved in the Karlin-Rubin Theorem, if

$$r(x) := \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}, \qquad \forall\, x \in \mathbb{R}^n,$$

then by assumption, r is a strictly increasing function of $t(x)$. Let $k \in \mathbb{R}$ such that $r(x) = k$ when $t(x) = c$. Since r is a strictly increasing function of $t(x)$, we can rewrite ϕ as

$$\phi(x) = \begin{cases} 1 & \text{, if } r(x) > k \\ 0 & \text{, if } r(x) < k \\ \gamma & \text{, if } r(x) = k. \end{cases}$$

Note that $\alpha = \mathbf{E}_{\theta_0}\phi(X)$ from (iii), so that ϕ is in class $\mathcal{T}$. Let ϕ' be a class $\mathcal{T}$ hypothesis test. By definition of $\mathcal{T}$, $\mathbf{E}_{\theta_0}\phi' \leq \alpha$. So, from the Neyman-Pearson Lemma (sufficiency), ϕ is UMP (in the context of that Lemma), i.e. $\mathbf{E}_{\theta_1}\phi(X) \geq \mathbf{E}_{\theta_1}\phi'(X)$. Since this inequality holds for all $\theta_1 > \theta_0$, we conclude that ϕ is UMP class $\mathcal{T}$. □

Exercise 2.3. Let $X_1, \ldots, X_n$ be i.i.d. random variables. Let $X = (X_1, \ldots, X_n)$. Let $\theta > 0$. Assume that X_1 is uniformly distributed in the interval $[0, \theta]$. Fix $\theta_0 > 0$. Fix $0 < \alpha < 1$. Let $\mathcal{T}$ denote the set of hypothesis tests with significance level at most α.

- Suppose we test $H_0 = \{\theta \leq \theta_0\}$ versus $H_1 = \{\theta > \theta_0\}$. Identify the set of all UMP class $\mathcal{T}$ hypothesis tests.
- Suppose we test $H_0 = \{\theta = \theta_0\}$ versus $H_1 = \{\theta \neq \theta_0\}$. Show there is a unique UMP class $\mathcal{T}$ hypothesis test in this case.

(Hint: first consider testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$ with $\theta_1 > \theta_0$, and apply the Neyman-Pearson Lemma. That is, mimic the argument of the Karlin-Rubin Theorem.) (As an aside, observe that, if you naïvely apply the Karlin-Rubin Theorem, you will not find all UMP tests, i.e. a non-strict MLR property version of the Karlin-Rubin Theorem will neglect some UMP tests.)

Solution. The joint distribution of $X_1, \ldots, X_n$ satisfies, for any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$f_\theta(x) = \prod_{i=1}^{n} \theta^{-1} 1_{[0,\theta]}(x_i) = \theta^{-n} 1_{0 \leq \max_{1 \leq i \leq n} x_i \leq \theta}.$$

Let $\theta_1 > \theta_0$. Then

$$\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \Big(\frac{\theta_1}{\theta_0}\Big)^{-n} \cdot \frac{1_{0 \leq \max_{1 \leq i \leq n} x_i \leq \theta_1}}{1_{0 \leq \max_{1 \leq i \leq n} x_i \leq \theta_0}}.$$

Since $\theta_1 > \theta_0$, evidently this likelihood ratio has the (non-strict) MLR property with respect to $t(x) := \max_{1 \leq i \leq n} x_i$. (As $t(x)$ increases from 0, the ratio of indicator functions is 1, then ∞, then of the form $0/0$, and the latter case is not considered for the MLR property.)

For the moment, suppose we instead test $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$. Then the Neyman-Pearson Lemma says that any (nontrivial) UMP class $\mathcal{T}$ test is a likelihood ratio test of the form

$$\phi(x) := \begin{cases} 1 & , \text{if } \theta_0 < x_{(n)} < \theta_1 \\ \text{arbitrary} & , \text{if } x_{(n)} \leq \theta_0 \text{ or } x_{(n)} \geq \theta_1. \end{cases}$$

(We find these tests by considering different thresholds k in the likelihood ratio tests that reject when $f_{\theta_1}(x) > k f_{\theta_0}(x)$.) The tests of this form that do not depend on θ_1 are of the form

$$\phi(x) := \begin{cases} 1 & , \text{if } \theta_0 < x_{(n)} \\ \text{arbitrary} & , \text{if } \theta_0 \geq x_{(n)}. \end{cases} \qquad (*)$$

Since this test does not depend on θ_1, we conclude that it is UMP for testing $\{\theta = \theta_0\}$ versus $\{\theta > \theta_0\}$ (as in the proof of the Karlin-Rubin Theorem). Again, as in the proof of the Karlin-Rubin Theorem, we conclude that this test is UMP for testing $\{\theta \leq \theta_0\}$ versus $\{\theta > \theta_0\}$. Conversely, any test that is UMP for $\{\theta \leq \theta_0\}$ versus $\{\theta > \theta_0\}$ must be UMP for testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$ when $\theta_1 > \theta_0$. Consequently, any UMP for $\{\theta \leq \theta_0\}$ versus $\{\theta > \theta_0\}$ must be of the form $(*)$. The first part of the proof is concluded.

We now prove the second part. If ϕ is UMP for $\{\theta = \theta_0\}$ versus $\{\theta \neq \theta_0\}$, then ϕ must be UMP for testing $\{\theta = \theta_0\}$ versus $\{\theta > \theta_0\}$, as in the Karlin-Rubin Theorem. That is, ϕ must be of the form $(*)$.

Moreover, ϕ must be UMP for testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$ with $\theta_1 < \theta_0$. In this case, the Neyman-Pearson Lemma says that any (nontrivial) UMP class $\mathcal{T}$ test is a likelihood ratio test of the form

$$\phi(x) := \begin{cases} 0 & \text{, if } \theta_1 < x_{(n)} < \theta_0 \\ \text{arbitrary} & \text{, if } x_{(n)} \leq \theta_1 \text{ or } x_{(n)} \geq \theta_1. \end{cases} \qquad (**)$$

or

$$\phi(x) := \begin{cases} 1 & \text{, if } x_{(n)} < \theta_1 \\ \text{arbitrary} & \text{, if } x_{(n)} \geq \theta_1. \end{cases} \qquad (***)$$

(We find these tests by considering different thresholds k in the likelihood ratio tests that reject when $f_{\theta_1}(x) > kf_{\theta_0}(x)$. The first type of test occurs when $k = (\theta_0/\theta_1)^n$. The second type of test occurs when $k = 0$.)

If additionally ϕ is of the form $(*)$, (and ϕ does not depend on θ_1) then ϕ must satisfy (for some constant c)

$$\phi(x) := \begin{cases} 1 & \text{, if } \theta_0 < x_{(n)} \text{ or } x_{(n)} < c \\ 0 & \text{otherwise.} \end{cases}$$

(For this particular test ϕ, note that, for any θ_1 satisfying $0 < \theta_1 < \theta_0$, either ϕ is of the form $(**)$ or $(***)$. More specifically, if $c < \theta_1 < \theta_0$, then ϕ is of the form $(**)$ and if $0 < \theta_1 < c$, then ϕ is of the form $(***)$.)

As c changes, so does the significance level α. So, for fixed α, ϕ is unique, as desired.

□

Exercise 2.4. Let $X_1, \ldots, X_n$ be i.i.d. random variables that are uniformly distributed in the interval $[\theta, \theta + 1]$, where $\theta \in \mathbb{R}$ is an unknown parameter. Fix $\theta_0 \in \mathbb{R}$. Suppose we want to test the hypothesis that $\theta \leq \theta_0$ versus $\theta > \theta_0$. For any $0 \leq \alpha \leq 1$, show that there exists a UMP test among tests with significance level at most α, and this test rejects the null hypothesis when $X_{(1)} > \theta_0 + c(\alpha)$ or $X_{(n)} > \theta_0 + 1$.

On the other hand, show that the joint density of $X_1, \ldots, X_n$ does not have the MLR property with respect to any statistic (when $n > 1$). (Hint: if it did have the MLR property, what would the Karlin-Rubin Theorem imply about the UMP rejection regions?)

Solution.

The joint distribution of $X_1, \ldots, X_n$ is

$$f_\theta(x) = \prod_{i=1}^{n} 1_{X_i \in [\theta, \theta+1]} = 1_{X_{(1)}, X_{(n)} \in [\theta, \theta+1]}.$$

Let $\theta_1 > \theta_0$. Then

$$\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} = \frac{1_{X_{(1)}, X_{(n)} \in [\theta_1, \theta_1+1]}}{1_{X_{(1)}, X_{(n)} \in [\theta_0, \theta_0+1]}}.$$

Observe that this ratio can be $0, 1$ or ∞. More specifically, on the set where at least one of these densities is nonzero, we have

- $f_{\theta_1}(x) > f_{\theta_0}(x)$ when $x_{(n)} > \theta_0 + 1$,
- $f_{\theta_1}(x) = f_{\theta_0}(x)$ when $\theta_1 \leq x_{(1)} \leq x_{(n)} \leq \theta_0 + 1$, and
- $f_{\theta_1}(x) < f_{\theta_0}(x)$ when $x_{(1)} < \theta_1$.

For the moment, suppose we instead test $\{\theta=\theta_0\}$ versus $\{\theta=\theta_1\}$. Then the Neyman-Pearson Lemma says that any (nontrivial) UMP class $\mathcal{T}$ test is a likelihood ratio test of the form

$$\phi(x) := \begin{cases} 1 & \text{, if } x_{(n)} > \theta_0+1 \\ \text{arbitrary} & \text{, if } \theta_1 \leq x_{(1)} \leq x_{(n)} \leq \theta_0+1 \\ 0 & \text{, if } x_{(1)} < \theta_1. \end{cases}$$

or

$$\phi(x) := \begin{cases} 1 & \text{, if } \theta_1 \leq x_{(1)} \\ \text{arbitrary} & \text{, if } x_{(1)} < \theta_1. \end{cases}$$

The tests of this form that do not depend on θ_1 are of the following form, where $c \in \mathbb{R}$ is a constant:

$$\phi(x) := \begin{cases} 1 & \text{, if } x_{(1)} > \theta_0 + c \ \text{ or } \ x_{(n)} > \theta_0+1 \\ 0 & \text{, otherwise.} \end{cases}$$

Since this test does not depend on θ_1, we conclude that it is UMP for testing $\{\theta=\theta_0\}$ versus $\{\theta>\theta_0\}$ (as in the proof of the Karlin-Rubin Theorem). Again, as in the proof of the Karlin-Rubin Theorem, we conclude that this test is UMP for testing $\{\theta\leq\theta_0\}$ versus $\{\theta>\theta_0\}$.

When $n>1$, the joint density of $X_1,\ldots,X_n$ does not have the MLR property. If it did, then the Karlin-Rubin Theorem would imply that there is a UMP test defined by a single real-valued statistic, but we just showed this is not true.

□

Exercise 2.5. Let $\{f_\theta\colon \theta\in\mathbb{R}\}$ be a family of positive, single-variable PDFs, i.e. $f_\theta\colon\mathbb{R}\to(0,\infty)$ for all $\theta\in\mathbb{R}$. Assume that $f_\theta(x)$ is twice continuously differentiable in the parameters θ, x.

Show that $\{f_\theta\}$ has the MLR property with respect to the statistic $t(x)=x$ ($x\in\mathbb{R}$) if and only if

$$\frac{\partial^2}{\partial\theta\partial x}\log f_\theta(x) > 0, \qquad \forall\, x,\theta\in\mathbb{R}.$$

Exercise 2.6. Suppose X is a binomial distributed random variable with parameters $n=100$ and $\theta\in[0,1]$ where θ is unknown. Suppose we want to test the hypothesis H_0 that $\theta=1/2$ versus the hypothesis H_1 that $\theta\neq 1/2$. Consider the hypothesis test that rejects the null hypothesis if and only if $|X-50|>10$.

Using e.g. the central limit theorem, do the following:

- Give an approximation to the significance level α of this hypothesis test
- Plot an approximation of the power function $\beta(\theta)$ as a function of θ.
- Estimate p values for this test when $X=50$, and also when $X=70$ or $X=90$.

Solution. We have $\alpha=\beta(1/2)=\mathbf{P}_{1/2}(X\in C)=\mathbf{P}_{1/2}(|X-50|>10)$. From The Central Limit Theorem, we have the approximation

$$\mathbf{P}_{1/2}(|X-50|>10)=\mathbf{P}_{1/2}(\frac{|X-50|}{(1/2)(10)}>2)\approx\mathbf{P}(|Z|>2)\approx .05.$$

Here we used the Matlab command `quad(@(t) (1/sqrt(2*pi))*exp(-t.^2 /2),-2,2)` to get the last probability. So, the significance level of the test is approximately .05. The

p-values for this test are roughly

$$p(50) = \mathbf{P}_{1/2}(|X - 50| > |50 - 50|) \approx 1.$$

$$p(70) = \mathbf{P}_{1/2}(|X - 50| > |70 - 50|) = \mathbf{P}_{1/2}(|X - 50| > 20) = \mathbf{P}_{1/2}(\frac{|X - 50|}{(1/2)(10)} > 4)$$
$$\approx \mathbf{P}(|Z| > 4) \approx 7 \cdot 10^{-5}.$$

Here we used the Matlab command `quad(@(t) (1/sqrt(2*pi))*exp(-t.^2 /2),-4,4)` to get the last probability.

$$p(90) = \mathbf{P}_{1/2}(|X - 50| > |90 - 50|) = \mathbf{P}_{1/2}(|X - 50| > 40) = \mathbf{P}_{1/2}(\frac{|X - 50|}{(1/2)(10)} > 8)$$
$$\approx \mathbf{P}(|Z| > 8) \approx 5 \cdot 10^{-7}.$$

Here we used the Matlab command `quad(@(t) (1/sqrt(2*pi))*exp(-t.^2 /2),-8,8)` to get the last probability. (I think the actual value of $\mathbf{P}(|Z| > 8)$ is much smaller than this, closer to 10^{-13} though.)

More generally, we have the approximation

$$\begin{aligned}
\beta(\theta) &= \mathbf{P}_\theta(X - 50 > 10) + \mathbf{P}_\theta(X - 50 < -10) \\
&= \mathbf{P}_\theta\Big(\frac{X - 100\theta}{10\sqrt{\theta(1-\theta)}} > \frac{1 + 5 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) + \mathbf{P}_\theta\Big(\frac{X - 100\theta}{10\sqrt{\theta(1-\theta)}} < \frac{-1 + 5 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) \\
&\approx \mathbf{P}_\theta\Big(Z > \frac{6 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) + \mathbf{P}\Big(Z < \frac{4 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) \\
&= 1 - \mathbf{P}\Big(Z < \frac{6 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) + \mathbf{P}\Big(Z < \frac{4 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) \\
&= 1 - \Phi\Big(\frac{6 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big) + \Phi\Big(\frac{4 - 10\theta}{\sqrt{\theta(1-\theta)}}\Big)
\end{aligned}$$

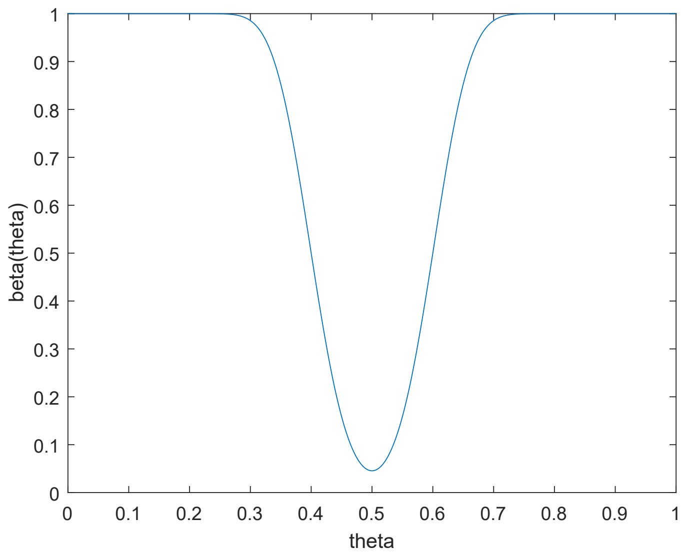
Here we used $\Phi(t) = \mathbf{P}(Z \leq t)$. We can then use following plot in Matlab

```
theta=linspace(0,1,1000);
plot(theta,1-normcdf((6-10*theta)./sqrt(theta.*(1-theta)),0,1)...
+normcdf((4-10*theta)./sqrt(theta.*(1-theta)),0,1));
xlabel('theta');
ylabel('beta(theta)');
```

□

Exercise 2.7. Let $X_1, \ldots, X_n$ be a real-valued random sample of size n from a family of distributions $\{f_\theta \colon \theta \in \Theta\}$. Suppose $\Theta = \mathbb{R}$. Fix $\theta \in \mathbb{R}$. Denote $X := (X_1, \ldots, X_n)$. Consider a set of hypothesis tests $\phi_\alpha \colon \mathbb{R}^n \to [0, 1]$, for any $\alpha \in [0, 1]$. Assume that these tests are nested in the sense that $\phi_\alpha \leq \phi_{\alpha'}$ for all $0 \leq \alpha < \alpha' \leq 1$. Suppose we are testing the hypothesis H_0 that $\{\theta \leq \theta_0\}$ versus H_1 that $\{\theta > \theta_0\}$. Suppose also that $\{f_\theta\}$ has the monotone likelihood ratio property with respect to a statistic $Y = t(X)$ that is a continuous random variable.

- Show that the family of UMP tests with significance level at most α satisfies the nested property mentioned above (for all $\alpha \in [0, 1]$).



- Show that, if $X = x$, then the p-value $p(x)$ satisfies
$$p(x) = \mathbf{P}_{\theta_0}(t(X) > t(x)).$$

Solution. The Karlin-Rubin Theorem implies that the UMP tests with significance level at most α are of the form
$$\phi(x) := \begin{cases} 1 & \text{, if } t(x) > c \\ 0 & \text{, if } t(x) < c \\ \gamma & \text{, if } t(x) = c. \end{cases}$$
Since we assume that $Y = t(X)$ is continuous, $t(X) = c$ occurs with probability zero, i.e. we may assume that
$$\phi(x) := \begin{cases} 1 & \text{, if } t(x) > c \\ 0 & \text{, if } t(x) \leq c. \end{cases}$$
The nested property then follows, since as α increases, c decreases.

Denote c_α as the constant $c = c_\alpha$ in the above definition when $\phi = \phi_\alpha$ has significance level α. Recall that significance level α means that
$$\alpha = \sup_{\theta\in\Theta_0} \mathbf{E}_\theta \phi(X) = \sup_{\theta\leq\theta_0} \mathbf{P}_\theta(t(X) > c_\alpha)$$
Since the Karlin-Rubin Theorem implies that the power function is nondecreasing in θ, we have
$$\alpha = \mathbf{P}_{\theta_0}(t(X) > c_\alpha). \qquad (*)$$
We also have
$$p(x) = \inf\{\alpha \in [0,1] \colon \phi_\alpha(x) = 1\} = \inf\{\alpha \in [0,1] \colon t(x) > c_\alpha\}.$$
The nested property implies that $\{\alpha \in [0,1] \colon t(x) > c\}$ is an interval, so that the infimum of this set is the smaller endpoint of that interval. That is, there exists some $\alpha \in [0,1]$ such that $p(x) = \alpha$ and $t(x) = c_\alpha$. So, from $(*)$,
$$\alpha = p(x) = \mathbf{P}_{\theta_0}(t(X) > c_\alpha) = \mathbf{P}_{\theta_0}(t(X) > t(x)).$$

□

Exercise 2.8. We defined the MLR property so that the likelihood ratio is a strictly increasing function of the statistic. Suppose we instead defined the MLR property so that the likelihood ratio is an increasing function of the statistic. In this case, where does our proof of the Karlin-Rubin Theorem not work correctly? Explain.

Solution. In our proof of the Karlin-Rubin Theorem, we first showed that hypothesis tests that are UMP class $\mathcal{T}$ for testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$ are likelihood ratio tests, by the Neyman-Pearson Lemma (when $\theta_1 > \theta_0$, and $\mathcal{T}$ is the set of tests with significance level at most α). If ϕ is UMP class $\mathcal{T}$ for testing $\{\theta < \theta_0\}$ versus $\{\theta \geq \theta_0\}$, then ϕ is UMP for testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$, when $\theta_1 > \theta_0$. So, if ϕ is UMP class $\mathcal{T}$ for testing $\{\theta < \theta_0\}$ versus $\{\theta \geq \theta_0\}$, then ϕ is a likelihood ratio test for testing $\{\theta = \theta_0\}$ versus $\{\theta = \theta_1\}$, for **any** $\theta_1 > \theta_0$. However, if the likelihood ratio is only an increasing function of $t(x)$, then the set $\{x \in \mathbb{R}^n \colon t(x) > c\}$ can not necessarily be written as $f_{\theta_1}(x) > k f_{\theta_0}(x)$ for **any** $\theta_1 > \theta_0$. (It is still true that $\{t(x) > c\}$ could be written as $f_{\theta_1}(x) > k f_{\theta_0}(x)$ for **some** $\theta_1 > \theta_0$, but it could even occur that no such $\theta_1 \in \Theta$ exists satisfying $\{x \in \mathbb{R}^n \colon t(x) > c\} = \{x \in \mathbb{R}^n \colon f_{\theta_1}(x) > k f_{\theta_0}(x)\}$, in which case the Karlin-Rubin Theorem becomes vacuous.) Likewise, as we observed e.g. in Exercises 2.3 and 2.4, a naïve application of the Karlin-Rubin Theorem would not find all UMP class $\mathcal{T}$ tests. So, although the textbook's proof of the Karlin-Rubin Theorem might technically not be incorrect, having a non-strict MLR assumption in the Karlin-Rubin Theorem might not find all UMP tests, i.e. the Theorem might be deficient in its purpose of finding UMP class $\mathcal{T}$ tests. Then, as in Exercise 2.3, it is best to just repeat the proof, i.e. apply the Neyman-Pearson Lemma to find all UMP class $\mathcal{T}$ tests. □

3. Homework 3

Exercise 3.1. Let $X_1, \ldots, X_n$ be i.i.d. Gaussian random variables with unknown mean and unknown variance.

- Find a real-valued pivotal quantity for $X = (X_1, \ldots, X_n)$.
- Using the pivotal quantity, construct a $1 - \alpha$ confidence interval for the mean μ, for any $0 < \alpha < 1$.

Solution. Recall that $\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ is a mean zero variance one Gaussian, so for any $a > 0$,

$$\mathbf{P}\Big(-a \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\Big) = \int_{-a}^{a} e^{-t^2/2} dt/\sqrt{2\pi} =: \Psi(a).$$

That is,

$$\mathbf{P}\Big(\frac{X_1 + \cdots + X_n}{n} - a\sigma/\sqrt{n} \leq \mu \leq \frac{X_1 + \cdots + X_n}{n} + a\sigma/\sqrt{n}\Big) = \Psi(a).$$

Setting $a := \Psi^{-1}(1 - \alpha)$, we then have

$$\mathbf{P}\Big(\frac{X_1 + \cdots + X_n}{n} - \Psi^{-1}(1-\alpha)\sigma/\sqrt{n} \leq \mu \leq \frac{X_1 + \cdots + X_n}{n} + \Psi^{-1}(1-\alpha)\sigma/\sqrt{n}\Big) = 1 - \alpha.$$

□

Exercise 3.2. Let $X_1, \ldots, X_n$ be a real-valued random sample of size n from a family of distributions $\{f_\theta \colon \theta \in \Theta\}$. Suppose $\Theta = \mathbb{R}$. Fix $\theta \in \mathbb{R}$. Denote $X := (X_1, \ldots, X_n)$. Consider a set of nonrandomized hypothesis tests with rejection regions $C_\alpha \subseteq \mathbb{R}^n$ for all $\alpha \in [0,1]$. Suppose these rejection regions are nested in the sense that $C_\alpha \subseteq C_{\alpha'}$ for all $0 \leq \alpha < \alpha' \leq 1$. As usual, denote $\Theta = \Theta_0 \cup \Theta_1$ with $\Theta_0 \cap \Theta_1 = \emptyset$. Define also the p-valued $p(x) := \inf\{\alpha \in [0,1] \colon x \in C_\alpha\}$, $\forall\, x \in \mathbb{R}^n$.

- Suppose $\sup_{\theta \in \Theta_0} \mathbf{P}_\theta(X \in C_\alpha) \leq \alpha$ for all $0 \leq \alpha \leq 1$. Show that the p-valued satisfies
$$\mathbf{P}_\theta(p(X) \leq c) \leq c, \qquad \forall\, 0 \leq c \leq 1,\ \forall\, \theta \in \Theta_0.$$
- Suppose $\mathbf{P}_\theta(X \in C_\alpha) = \alpha$ for all $0 \leq \alpha \leq 1$. Show that the p-valued satisfies
$$\mathbf{P}_\theta(p(X) \leq c) = c, \qquad \forall\, 0 \leq c \leq 1,\ \forall\, \theta \in \Theta_0.$$
That is, $p(X)$ is uniformly distributed in $[0,1]$.

Solution. Let $0 < c < 1$. By definition of p, if $p(X) \leq c$, then $X \in C_a$ for all $a > c$ (by the nested property). That is, $\mathbf{P}_\theta(p(X) \leq c) \leq \mathbf{P}_\theta(X \in C_a)$. The nested property of the sets together with continuity of the probability law allows us to let a approach c with $a > c$, so that
$$\mathbf{P}_\theta(p(X) \leq c) \leq \lim_{a \to c^+} \mathbf{P}_\theta(X \in C_a) = \mathbf{P}_\theta(X \in C_c).$$
If $\theta \in \Theta_0$, the right probability is at most c by assumption, i.e.
$$\mathbf{P}_\theta(p(X) \leq c) \leq c, \qquad \forall\, \theta \in \Theta_0.$$
In the last case, note that $X \in C_a$ implies that $p(X) \leq a$, so that
$$\mathbf{P}_\theta(p(X) \leq c) \geq \mathbf{P}_\theta(X \in C_c).$$
So, by assumption we have
$$\mathbf{P}_\theta(p(X) \leq c) \geq c, \qquad \forall\, 0 < c < 1.$$
This inequality together with the first part completes the proof. □

Exercise 3.3. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with unknown location parameter $\theta > 0$, i.e. X_1 has density
$$g(x) := 1_{x \geq \theta} e^{-(x-\theta)}, \qquad \forall\, x \in \mathbb{R}.$$
Fix $\theta_0 \in \mathbb{R}$. Suppose we want to test that hypothesis H_0 that $\theta \leq \theta_0$ versus the alternative H_1 that $\theta > \theta_0$. That is, $\Theta = \mathbb{R}$, $\Theta_0 = \{\theta \in \mathbb{R} \colon \theta \leq \theta_0\}$ and $\Theta_0^c = \Theta_1 = \{\theta \in \mathbb{R} \colon \theta > \theta_0$.

- Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis. (Hint: it might be easier to describe the region using $x_{(1)} = \min(x_1, \ldots, x_n)$.)
- Prove that $X_{(1)} := \min(X_1, \ldots, X_n)$ is a sufficient statistic for θ.
- (Optional) If H_0 is true, then does
$$2 \log \frac{\sup_{\theta \in \Theta} f_\theta(X_1, \ldots, X_n)}{\sup_{\theta \in \Theta_0} f_\theta(X_1, \ldots, X_n)}$$
converge in distribution to a chi-squared distribution as $n \to \infty$?

Solution. We can write

$$f_\theta(x) = \prod_{i=1}^n f_\theta(x_i) = 1_{\{\min_{1\leq i\leq n} x_i \geq \theta\}} \cdot \prod_{i=1}^n e^{-(x_i-\theta)} = e^{n\theta} 1_{\{\min_{1\leq i\leq n} x_i \geq \theta\}} \cdot \prod_{i=1}^n e^{-(x_i)}.$$

As a function of θ, this function is strictly increasing, but when $\theta > \min_{1\leq i\leq n} x_i$, we have $f_\theta(x) = 0$. So, the MLE is $\min_{1\leq i\leq n} X_i$, i.e.

$$\sup_{\theta\in\Theta} f_\theta(x) = f_{\min_{1\leq i\leq n} x_i}(x).$$

Similarly since $\Theta_0 = \{\theta \in \mathbb{R} \colon \theta \leq \theta_0\}$, the supremum over Θ_0 occurs at the minimum of θ_0 and $\min_{1\leq i\leq n} x_i$:

$$\sup_{\theta\in\Theta} f_\theta(x) = f_{\min(\theta_0, \min_{1\leq i\leq n} x_i)}(x).$$

So, for any $k \leq 1$, we have

$$\begin{aligned}
C &:= \{x \in \mathbb{R}^n \colon \sup_{\theta\in\Theta} f_\theta(x) \geq k \sup_{\theta\in\Theta_0} f_\theta(x)\} \\
&= \Big\{x \in \mathbb{R}^n \colon \sup_{\theta\in\Theta} f_\theta(x) \geq k \sup_{\theta\in\Theta_0} f_\theta(x), \text{ and } \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \\
&\qquad \cup \Big\{x \in \mathbb{R}^n \colon \sup_{\theta\in\Theta} f_\theta(x) \geq k \sup_{\theta\in\Theta_0} f_\theta(x), \text{ and } \min_{1\leq i\leq n} x_i > \theta_0\Big\} \\
&= \Big\{x \in \mathbb{R}^n \colon 1 \geq k, \text{ and } \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \\
&\qquad \cup \Big\{x \in \mathbb{R}^n \colon f_{\min_{1\leq i\leq n} x_i}(x) \geq k f_{\theta_0}(x), \text{ and } \min_{1\leq i\leq n} x_i > \theta_0\Big\} \\
&= \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \cup \Big\{x \in \mathbb{R}^n \colon f_{\min_{1\leq i\leq n} x_i}(x) \geq k f_{\theta_0}(x), \text{ and } \min_{1\leq i\leq n} x_i > \theta_0\Big\} \\
&= \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \cup \Big\{x \in \mathbb{R}^n \colon e^{n[\min_{1\leq i\leq n} x_i(x) - \theta_0]} \geq k, \text{ and } \min_{1\leq i\leq n} x_i > \theta_0\Big\} \\
&= \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \cup \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i(x) - \theta_0 \geq \frac{1}{n}\log k, \text{ and } \min_{1\leq i\leq n} x_i > \theta_0\Big\} \\
&= \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i \leq \theta_0\Big\} \cup \Big\{x \in \mathbb{R}^n \colon \min_{1\leq i\leq n} x_i(x) \geq \theta_0 + \frac{1}{n}\log k\Big\}
\end{aligned}$$

Theorem 4.7 in the notes answers the last question affirmatively. (If H_0 is true, then $\min(\theta_0, X_{(1)}) = \theta_0$ with probability one, so from the discussion above, $\sup_{\theta\in\Theta_0} f_\theta(X) = f_{\theta_0}(X)$ with probability one, so Theorem 4.7 applies.) □

Exercise 3.4. Let $X_1, \ldots, X_n$ be a random sample from a Gaussian random variable with unknown mean $\mu \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$.

Fix $\mu_0 \in \mathbb{R}$. Suppose we want to test that hypothesis H_0 that $\mu = \mu_0$ versus the alternative H_1 that $\mu \neq \mu_0$.

- Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis.

- Give an explicit formula for the p-value of this hypothesis test. (Hint: If S^2 denotes the sample variance and $\overline{X}$ denotes the sample mean, you should then be able to use the statistic $\frac{(\overline{X}-\mu_0)^2}{S^2}$. Since we have an explicit formula for Snedecor's distribution, you should then be able to write an explicit integral formula for the p-value of this test.)

Solution. For any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

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

Also, $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ and $\Theta_0 = \{(\mu_0, \sigma) : \sigma > 0\}$.

As we computed in 541A, the MLE is the sample mean and sample variance (dividing by n instead of $n-1$), i.e. for any $x \in \mathbb{R}^n$, if we denote $\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$, $v := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

$$\sup_{(\mu,\sigma)\in\Theta} f_{\mu,\sigma^2}(x) = f_{\overline{x},v}(x).$$

Similarly,

$$\sup_{(\mu,\sigma)\in\Theta_0} f_{\mu,\sigma^2}(x) = f_{\mu_0,v'}(x).$$

Here $v' := \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2$. Therefore,

$$\begin{aligned} C &:= \{x \in \mathbb{R}^n : \sup_{(\mu,\sigma^2)\in\Theta} f_{\mu,\sigma^2}(x) \geq k \sup_{(\mu,\sigma^2)\in\Theta_0} f_{\mu,\sigma^2}(x)\} \\ &= \Big\{x \in \mathbb{R}^n : (v/v')^{-n/2} \prod_{i=1}^{n} e^{-\frac{(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j)^2}{2v} + \frac{(x_i-\mu_0)^2}{2v'}} \geq k\Big\} \\ &= \Big\{x \in \mathbb{R}^n : (v/v')^{-n/2} \geq k\Big\} = \Big\{x \in \mathbb{R}^n : v/v' \leq k^{-2/n}\Big\} \end{aligned}$$

Note that

$$\begin{aligned} v' &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x} + \overline{x} - \mu_0)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}[(x_i - \overline{x})^2 + (\overline{x} - \mu_0)^2] = v + \frac{1}{n}\sum_{i=1}^{n}(\overline{x} - \mu_0)^2 = v + (\overline{x} - \mu_0)^2 \end{aligned}$$

Using this fact, and that the inverse of $x \mapsto 1/(1+x)$ is $x \mapsto -1 + 1/x$,

$$\begin{aligned} C &= \Big\{x \in \mathbb{R}^n : \frac{v}{v + (\overline{x} - \mu_0)^2} \leq k^{-2/n}\Big\} = \Big\{x \in \mathbb{R}^n : \frac{1}{1 + \frac{(\overline{x}-\mu_0)^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}} \leq k^{-2/n}\Big\} \\ &= \Big\{x \in \mathbb{R}^n : \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}{(\overline{x} - \mu_0)^2} - 1 \geq k^{2/n} - 1\Big\} = \Big\{x \in \mathbb{R}^n : \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}{(\overline{x} - \mu_0)^2} \geq k^{2/n}\Big\}. \end{aligned}$$

If the null hypothesis holds, then $\frac{1}{n}\sum_{j=1}^{n} X_j - \mu_0$ is a mean zero Gaussian with variance σ^2/n where σ^2 is unknown. Also, $\frac{n-1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a chi squared random variable with

$n-1$ degrees of freedom, that is independent of $\frac{1}{n}\sum_{j=1}^{n} X_j$ by Proposition 2.15 from the notes. So,

$$Y := \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}{\Big(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu_0\Big)^2} = \frac{\frac{n-1}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}{\frac{n-1}{\sigma^2}\Big(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu_0\Big)^2} = \frac{\frac{n-1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2}{(n-1)\Big(\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n} X_j - \mu_0\Big)^2}$$

By Proposition 2.16 in the notes, this is Snedecor's distribution with $n-1$ and 1 degrees of freedom. Since this density does not depend on σ, a p-value for the test is $p(X)$ where

$$t(x) := \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}{\Big(\frac{1}{n}\sum_{j=1}^{n} x_j - \mu_0\Big)^2}$$

$$\begin{aligned} p(x) = \mathbf{P}_{\mu_0}(t(X) \geq t(x)) &= \int_{t(x)}^{\infty} f_Y(y)dy \\ &= \frac{\Gamma(n/2)}{\Gamma((n-1)/2)\Gamma(1/2)}(n-1)^{(n-1)/2}\int_{t(x)}^{\infty} y^{((n-1)/2)-1}(1 + y(n-1))^{-n/2}dy. \end{aligned}$$

□

4. Homework 4

Exercise 4.1. Write down the generalized likelihood ratio estimate for the following alpha particle data, as we did in class for a slightly different data set. The corresponding test treats individual counts of alpha particles as independent Poisson random variables, versus the alternative that the probability of a count appearing in each box of data is a sequence of nonnegative numbers that sum to one. (In doing so, you should need to compute a maximum likelihood estimate using a computer.)

m	0, 1 or 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	≥ 17
# of Intervals	16	26	58	102	125	146	163	164	120	100	72	54	20	12	10	4

Plot the MLE for the Poisson statistic (i.e. plot the denominator of the generalized likelihood ratio test statistic $\frac{\sup_{\theta\in\Theta} f_\theta(X)}{\sup_{\theta\in\Theta_0} f_\theta(X)}$) as a function of λ.

Finally, compute the value s of Pearson's chi-squared statistic S, and compute the probability that $S \geq s$. Does the probability $\mathbf{P}(S \geq s)$ give you confidence that the null hypothesis is true?

Solution.

We compute the value of Pearson's chi-squared statistic to be approximately 10.8, with a p-value of about .7, giving us reasonable confidence in accepting the null hypothesis. □

Exercise 4.2. Let $X_1, \ldots, X_n$ be i.i.d. random variables. Let $0 < \alpha < 1/2$. Define the α-trimmed sample mean to be

$$\overline{X}_n^{(\alpha)} := \frac{1}{n - 2\lfloor n\alpha\rfloor}\sum_{i=\lfloor n\alpha\rfloor+1}^{n-\lfloor n\alpha\rfloor} X_{(i)}.$$

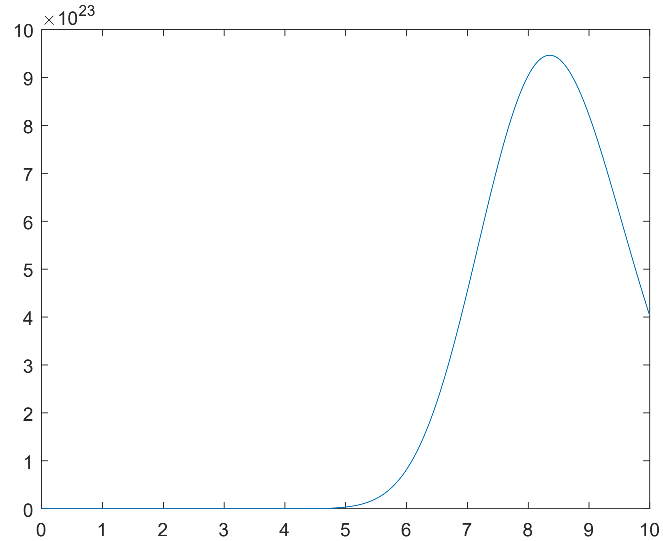


FIGURE 1. Plot of the Poisson-based MLE, to the 1/200 power, demonstrating a maximum value of $\lambda \approx 8.353$

For any $w = (w_1, \ldots, w_n) \in \{1, \ldots, n\}^n$, define the Winsorized sample mean to be

$$\overline{X}_n^{(w)} := \frac{1}{n}\sum_{i=1}^{n} X_{(w_i)}.$$

- Show that the jackknife estimator of $\overline{X}_n^{(\alpha)}$ is
$$\frac{1}{1-2\alpha}(\overline{X}_n^{(w)} - 2\alpha\overline{X}_n^{(\alpha)}),$$
for some vector w.
- Show that the jackknife variance estimator of $\overline{X}_n^{(\alpha)}$ is
$$\frac{1}{n(n-1)(1-2\alpha)^2}\sum_{i=1}^{n}(X_{(w_i)} - \overline{X}_n^{(w)})^2,$$
for some vector w.

Exercise 4.3. Let X_1, X_2, X_3 be i.i.d. continuous random variables such that X_1 has PDF $\{f_\theta \colon \theta \in \Theta\}$. Let W_1, W_2, W_3 be a bootstrap sample from X_1, X_2, X_3. Let Y denote the sample median of X_1, X_2, X_3. (That is, Y is the middle value among X_1, X_2, X_3, which is unique with probability one since the random variables are continuous.)

- Describe the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$.
- Describe the bootstrap estimator of Y.
- Describe the bootstrap estimator of the variance of Y.

Solution. Since X_1, X_2, X_3 are all distinct with probability one, we have

$$\mathbf{P}(W_1 = X_i, W_2 = X_j, W_3 = X_k \mid X_1, X_2, X_3) = (1/3)^3, \qquad \forall\, 1 \le i, j, k \le 3.$$

That is, in describing the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$, we may as well assume that $X_{(1)} = 1, X_{(2)} = 2, X_{(3)} = 3$, and W_1, W_2, W_3 are i.i.d. uniform in $\{1, 2, 3\}$. (We are satisfied with this description of the distribution of $(W_{(1)}, W_{(2)}, W_{(3)})$.)

Now, as covered e.g. in Exercise 2.19 in the notes, by considering Y which is the number of indices $1 \leq j \leq 3$ such that $W_j \leq X_{(i)}$, we have

$$\mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) = \sum_{k=2}^{3} \binom{3}{k} p_i^k (1 - p_i)^{n-k},$$

where $p_i = i/3$ for all $1 \leq i \leq 3$. (This follows since Y is a binomial random variable with parameters 3 and p_i.) That is,

$$\mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) = \sum_{k=2}^{3} \binom{3}{k} (i/3)^k (1 - i/3)^{n-k},$$

Therefore, for all $1 \leq i \leq 3$, we have

$$\begin{aligned}\mathbf{P}(W_{(2)} = X_{(i)} \,|\, X_1, X_2, X_3) &= \mathbf{P}(W_{(2)} \leq X_{(i)} \,|\, X_1, X_2, X_3) - \mathbf{P}(W_{(2)} \leq X_{(i-1)} \,|\, X_1, X_2, X_3) \\ &= \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1 - i/3)^{n-k} - ((i-1)/3)^k (1 - (i-1)/3)^{n-k} \Big).\end{aligned}$$

The bootstrap estimator of Y is then

$$\begin{aligned}\mathbf{E}[W_{(2)} \,|\, X_1, X_2, X_3] &= \sum_{i=1}^{3} X_{(i)} \mathbf{P}(W_{(2)} = X_{(i)} \,|\, X_1, X_2, X_3) \\ &= \sum_{i=1}^{3} X_{(i)} \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1 - i/3)^{n-k} - ((i-1)/3)^k (1 - (i-1)/3)^{n-k} \Big).\end{aligned}$$

And the bootstrap estimator of the variance is

$$\begin{aligned}&\mathrm{Var}(W_{(2)} \,|\, X_1, X_2, X_3) = \mathbf{E}[(W_{(2)} - \mathbf{E}[W_{(2)} \,|\, X_1, X_2, X_3])^2 \,|\, X_1, X_2, X_3] \\ &\quad = \mathbf{E}[W_{(2)}^2 \,|\, X_1, X_2, X_3] - (\mathbf{E}[W_{(2)} \,|\, X_1, X_2, X_3])^2 \\ &\quad = \sum_{i=1}^{3} X_{(i)}^2 \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1 - i/3)^{n-k} - ((i-1)/3)^k (1 - (i-1)/3)^{n-k} \Big) \\ &\qquad - \Big(\sum_{i=1}^{3} X_{(i)} \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1 - i/3)^{n-k} - ((i-1)/3)^k (1 - (i-1)/3)^{n-k} \Big) \Big)^2.\end{aligned}$$

Here we used

$$\begin{aligned}\mathbf{E}[W_{(2)}^2 \,|\, X_1, X_2, X_3] &= \sum_{i=1}^{3} X_{(i)}^2 \mathbf{P}(W_{(2)} = X_{(i)} \,|\, X_1, X_2, X_3) \\ &= \sum_{i=1}^{3} X_{(i)}^2 \sum_{k=2}^{3} \binom{3}{k} \Big((i/3)^k (1 - i/3)^{n-k} - ((i-1)/3)^k (1 - (i-1)/3)^{n-k} \Big).\end{aligned}$$

□

Exercise 4.4. Let $\mu \in \mathbb{R}$ and let $0 < \sigma < \infty$. Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables each with mean μ and variance σ^2. Let $h \colon \mathbb{R} \to \mathbb{R}$ be a function such that h' exists and is continuous. Let $\overline{X}_n := (X_1 + \cdots + X_n)/n$. Let $Y_n := h(\overline{X}_n)$.

Show that the jackknife estimator of the variance of Y_n converges almost surely to the same estimate of the variance you get by applying the Delta Method to Y_n.

Solution. Theorem 5.8 in the notes implies that the jackknife variance estimator V_n is consistent, i.e. (when $d = 1$), we have that

$$\frac{V_n}{\frac{1}{n}(h'(\mu))^2 \mathrm{Cov}(X_1, X_1)} = \frac{V_n}{\frac{1}{n}(h'(\mu))^2 \sigma^2}$$

converges almost surely to 1 as $n \to \infty$.

On the other hand, the Delta Method implies that $\sqrt{n}(h(\overline{X}_n) - h(\mu))$ converges in distribution to a mean zero Gaussian with variance $\sigma^2 (h'(\mu))^2$ as $n \to \infty$. Consequently, the variance of $h(\overline{X}_n)$ is approximately $\sigma^2 (h'(\mu))^2 / n$, asymptotically agreeing with the jackknife variance estimate. □

Exercise 4.5. Suppose $X_1, \ldots, X_n$ is a random sample from a Gaussian random variable X with unknown mean $\mu_X \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. Suppose $Y_1, \ldots, Y_m$ is a random sample from a Gaussian random variable Y with unknown mean $\mu_Y \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$.

Assume that $X_1, \ldots, X_n$ is independent of $Y_1, \ldots, Y_m$, i.e. assume that X, Y are independent.

Assume that $n + m > 2$. Define

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \overline{Y} := \frac{1}{m}\sum_{i=1}^{m} Y_i,$$

$$S_X^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_Y^2 := \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2,$$

$$S^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

Show that

$$\frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has Student's t-distribution with $n + m - 2$ degrees of freedom. Deduce the following confidence intervals for the difference of the means

$$\mathbf{P}\Big(\overline{X} - \overline{Y} - tS\sqrt{\frac{1}{n} + \frac{1}{m}} < \mu_X - \mu_Y < \overline{X} - \overline{Y} + tS\sqrt{\frac{1}{n} + \frac{1}{m}}\Big)$$
$$= \frac{\Gamma(\frac{p+1}{2})}{\sqrt{p}\sqrt{\pi}\,\Gamma(p/2)} \int_{-t}^{t} \Big(1 + \frac{s^2}{p}\Big)^{-(p+1)/2} ds,$$

where $p = n + m - 2$.

5. Homework 5

Exercise 5.1. Let $n > m$ be integers. Let A be an $n \times m$ real matrix of known (deterministic) constants. Let $\beta \in \mathbb{R}^m$ be an unknown vector of (deterministic) constants. And let $\varepsilon \in \mathbb{R}^n$ be a random vector with $\mathbf{E}\varepsilon = 0$ and such that ε is a vector of i.i.d. random variables. Define $Y \in \mathbb{R}^n$ by $Y = A\beta + \varepsilon$. Assume that $A^T A$ is invertible. Define $Z := (A^T A)^{-1} A^T Y$.

Show that the estimator

$$\Big(\frac{1}{n-m}\sum_{i=1}^{n}(Y_i - (AZ)_i)^2\Big)(A^T A)^{-1}$$

is an unbiased estimator of the covariance matrix of $Z := (A^T A)^{-1} A^T Y$.

Solution. As shown in class, $\mathrm{Cov}(Z) = \sigma^2 (A^T A)^{-1}$. We then have

$$\begin{aligned}
\mathbf{E}\sum_{i=1}^{n}(Y_i - (AZ)_i)^2 &= \mathbf{E}\sum_{i=1}^{n} Y_i^2 - 2Y_i(AZ)_i + (AZ)_i^2 \\
&= \mathbf{E}\sum_{i=1}^{n}(A\beta+\varepsilon)_i^2 - 2(A\beta+\varepsilon)_i(A(A^TA)^{-1}A^T(A\beta+\varepsilon))_i + (A(A^TA)^{-1}A^T(A\beta+\varepsilon))_i^2 \\
&= n + \mathbf{E}\sum_{i=1}^{n} -2(A\beta+\varepsilon)_i(A(A^TA)^{-1}A^T(A\beta+\varepsilon))_i + (A(A^TA)^{-1}A^T(A\beta+\varepsilon))_i^2 \\
&= n + \sum_{i=1}^{n} -2[A(A^TA)^{-1}A^T]_{ii} + (A(A^TA)^{-1}A^T)_{ii}^2 \\
&= n - 2\mathrm{Tr}[A(A^TA)^{-1}A^T] + \mathrm{Tr}[(A(A^TA)^{-1}A^T)^2] \\
&= n - 2m + m = n - m.
\end{aligned}$$

Here we used that, since $A^T A$ is an invertible $m \times m$ matrix, we have

$$\mathrm{Tr}[A(A^TA)^{-1}A^T] = \mathrm{Tr}[A^TA(A^TA)^{-1}] = \mathrm{Tr}(I_m) = m.$$

And similarly

$$\mathrm{Tr}[(A(A^TA)^{-1}A^T)^2] = \mathrm{Tr}[A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T] = \mathrm{Tr}[A(A^TA)^{-1}A^T] = \mathrm{Tr}(I_m) = m.$$

□

Exercise 5.2. Assume the one-way ANOVA assumptions. Consider the null hypothesis H_0 that $\beta_1 = \cdots = \beta_p$. Recall that, under this assumption, the F statistic takes the form

$$F = \frac{1}{S^2}\sum_{j=1}^{p} n_i(\overline{Y_j} - \overline{Y})^2.$$

The alternative hypothesis H_1 is that $\beta_i \neq \beta_j$ for some $1 \leq i < j \leq p$. We can therefore reject H_0 when F is large.

Show that the generalized likelihood ratio test of H_0 versus H_1 coincides with the hypothesis test we just described.

Solution. For any $y = (y_1, \ldots, y_{m_p}) \in \mathbb{R}^{m_p}$,

$$f_{\beta_1,\ldots,\beta_p,\sigma^2}(y) = \prod_{j=1}^{p} \prod_{i=m_{j-1}+1}^{m_j} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i-\beta_j)^2}{2\sigma^2}}.$$

Also,

$$\Theta = \{(\beta_1, \ldots, \beta_p, \sigma^2) \colon \beta_i \in \mathbb{R}, \forall\, 1 \le i \le p,\, \sigma > 0\}.$$

$$\Theta_0 = \{(\beta_1, \ldots, \beta_p, \sigma^2) \colon \beta_1 = \cdots = \beta_p\}.$$

As we computed in 541A, the MLE is the sample mean and sample variance (dividing by n instead of $n-1$), i.e. for any $y \in \mathbb{R}^{m_p}$, if we denote $\overline{y}_j := \frac{1}{n_j}\sum_{i=m_{j-1}+1}^{m_j} y_i$, $v := \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(y_i - \overline{y}_j)^2$

$$\sup_{(\beta_1,\ldots,\beta_p,\sigma^2)\in\Theta} f_{\beta_1,\ldots\beta_p}(y) = f_{\overline{y}_1,\ldots,\overline{y}_p,v}(y).$$

Similarly,

$$\sup_{(\beta_1,\ldots,\beta_p,\sigma^2)\in\Theta_0} f_{\beta_1,\ldots\beta_p}(y) = f_{\overline{y},\ldots,\overline{y},v'}(y).$$

Here

$$\overline{y} := \frac{1}{m_p}\sum_{i=1}^{m_p} y_i = \frac{1}{m_p}\sum_{j=1}^{p} n_j \overline{y}_j$$

$$\begin{aligned} v' &:= \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(y_i - \overline{y})^2 = \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(y_i - \overline{y}_j + \overline{y}_j - \overline{y})^2 \\ &= \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}[(y_i - \overline{y}_j)^2 + (\overline{y}_j - \overline{y})^2] = v + \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(\overline{y}_j - \overline{y})^2. \end{aligned}$$

Therefore,

$$\begin{aligned} C &:= \{y \in \mathbb{R}^{m_p} \colon \sup_{(\beta_1,\ldots,\beta_p,\sigma^2)\in\Theta} f_{\beta_1,\ldots\beta_p}(y) \ge k \sup_{(\beta_1,\ldots,\beta_p,\sigma^2)\in\Theta_0} f_{\beta_1,\ldots\beta_p}(y)\} \\ &= \{y \in \mathbb{R}^{m_p} \colon f_{\overline{y}_1,\ldots,\overline{y}_p,v}(y) \ge k f_{\overline{y},\ldots,\overline{y},v'}(y)\} \\ &= \{y \in \mathbb{R}^{m_p} \colon \prod_{j=1}^{p}\prod_{i=m_{j-1}+1}^{m_j} \frac{1}{\sqrt{v2\pi}} e^{-\frac{(y_i-\overline{y}_j)^2}{2v}} \ge k \prod_{j=1}^{p}\prod_{i=m_{j-1}+1}^{m_j} \frac{1}{\sqrt{v'2\pi}} e^{-\frac{(y_i-\overline{y})^2}{2v'}}\} \\ &= \{y \in \mathbb{R}^{m_p} \colon (v/v')^{-m_p/2} \prod_{j=1}^{p}\prod_{i=m_{j-1}+1}^{m_j} e^{-\frac{(y_i-\overline{y}_j)^2}{2v}} \ge k \prod_{j=1}^{p}\prod_{i=m_{j-1}+1}^{m_j} e^{-\frac{(y_i-\overline{y})^2}{2v'}}\} \end{aligned}$$

By definition of v and v', both of the exponential terms simplify to the same constant, i.e.

$$\begin{aligned} C &= \{y \in \mathbb{R}^{m_p} \colon (v/v')^{-m_p/2} \geq k\} \\ &= \{y \in \mathbb{R}^{m_p} \colon \frac{v}{v + \frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(\overline{y}_j - \overline{y})^2} \leq k^{-2/m_p}\} \\ &= \left\{ y \in \mathbb{R}^{m_p} \colon \frac{1}{1 + \frac{\frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(\overline{y}_j - \overline{y})^2}{\frac{1}{m_p}\sum_{j=1}^{p}\sum_{i=m_{j-1}+1}^{m_j}(y_i - \overline{y}_j)^2}} \leq k^{-2/m_p} \right\} \end{aligned}$$

Evidently, the last quantity is a monotone function of F, as desired.

□

Exercise 5.3. In statistics and other applications, we can be presented with data points $(x_1, y_1), \ldots, (x_n, y_n)$. We would like to find the line $y = mx + b$ which lies "closest" to all of these data points. Such a line is known as a linear regression. There are many ways to define the "closest" such line. The standard method is to use least squares minimization. A line which lies close to all of the data points should make the quantities $(y_i - mx_i - b)$ all very small. We would like to find numbers m, b such that the following quantity is minimized:

$$f(m, b) = \sum_{i=1}^{n}(y_i - mx_i - b)^2.$$

Using the second derivative test, show that the minimum value of f is achieved when

$$m = \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right) - n\left(\sum_{k=1}^{n} x_k y_k\right)}{\left(\sum_{i=1}^{n} x_i\right)^2 - n\left(\sum_{j=1}^{n} x_j^2\right)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{j=1}^{n}(x_j - \overline{x})^2}.$$

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - m\sum_{j=1}^{n} x_j\right) = \overline{y} - m\overline{x}.$$

Briefly explain why this is actually the minimum value of $f(m, b)$. (You are allowed to use the inequality $(\sum_{i=1}^{n} x_i)^2 \leq n(\sum_{i=1}^{n} x_i^2)$.)

Solution. Let $X = \sum_{i=1}^{n} x_i$, $Y = \sum_{j=1}^{n} y_j$, $A = \sum_{k=1}^{n} x_k y_k$, $S = \sum_{j=1}^{n} x_j^2$. We have $\nabla f(m, b) = (\sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i), -\sum_{i=1}^{n} 2(y_i - mx_i - b))$. So, $\nabla f(m, b) = (0, 0)$ when $\sum_{i=1}^{n}(y_i - mx_i - b)) = 0$ and $\sum_{i=1}^{n}(y_i - mx_i - b)x_i = 0$. We want to solve for m and b. We have $Y - mX - nb = 0$ and $A - mS - bX = 0$. So, $b = (Y - mX)/n$, and then $A - mS - (Y - mX)X/n = 0$. So, $m(-S + X^2/n) = -A + XY/n$, and $m = (XY - An)/(X^2 - Sn)$.

Finally, strict convexity of f implies that the only critical point we found is the global minimum of f. (If $x_1 = \cdots = x_n$, then the minimum will not be unique. For example, if $n = 1$, then the function $f(m, b) = (1 - m - b)^2$ has a minimum on the line $m + b = 1$.) □

Exercise 5.4. Let

$$h(x) := \frac{1}{1 + e^{-x}}, \qquad \forall\, x \in \mathbb{R}.$$

Fix $x \in \mathbb{R}$ and $y \in [0,1]$. Define $t\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$t(a,b) := \log\Big([h(ax+b)]^y[1-h(ax+b)]^{1-y}\Big), \qquad \forall\, a,b \in \mathbb{R}.$$

Show that t is concave. Conclude that t has at most one global maximum.

6. Homework 6

Exercise 6.1. Let $f\colon \mathbb{R} \to [0,\infty)$ be a PDF. Suppose you can sample (on a computer) any number of i.i.d. real valued random variables $X_1, X_2, \ldots$, each with PDF f. Let $g\colon \mathbb{R} \to [0,\infty)$ be another PDF. Assume there exists some $m > 0$ such that $g(x) \leq mf(x)\ \forall\, x \in \mathbb{R}$.

The goal of this exercise is to sample from a random variable with PDF g.

State a version of accept/reject sampling with these assumptions and goal.

Prove that your version of accept/reject sampling outputs some random variable Z with PDF g.

Solution.

Here is the algorithm. Let $(X_1,Y_1),(X_2,Y_2),\ldots$ be i.i.d random variables uniformly distributed under the curve $\{(x,y) \in \mathbb{R}^2\colon 0 \leq y \leq mf(x)\}$. Define $I := \inf\{n \geq 1\colon Y_n \leq g(X_n)\}$.

Output $Z := X_I$.

Claim. Z has PDF g.

Proof of Claim. Let $z \in \mathbb{R}$. Define $\varepsilon := \frac{\int_{\mathbb{R}} g(x)dx}{\int_{\mathbb{R}} mf(x)dx} = \frac{1}{m}$. (Since $g \leq mf$ and $\int_{\mathbb{R}} g = \int_{\mathbb{R}} f = 1$, it follows that $m \geq 1$.) Then

$$\mathbf{P}(Z \leq z) = \sum_{i=1}^{\infty} \mathbf{P}(Z \leq z,\, Z = X_i) = \sum_{i=1}^{\infty} \mathbf{P}(X_i \leq z,\, Y_i \leq g(X_i)) \prod_{j=1}^{i-1} \mathbf{P}(Y_j > g(X_j))$$

$$= \sum_{i=1}^{\infty} \frac{\int_{-\infty}^{z} g(x)dx}{\int_{\mathbb{R}} mf(x)dx}(1-\varepsilon)^{i-1} = \int_{-\infty}^{z} g(x)dx \sum_{i=1}^{\infty} \varepsilon(1-\varepsilon)^{i-1} = \int_{-\infty}^{z} g(x)dx.$$

□

□

Exercise 6.2. Let $X_1,\ldots,X_n$ be i.i.d. Gaussian random variables with unknown mean $\mu \in \mathbb{R}$ and variance 1. Define $Y_i := \max(X_i, 0)$, for all $1 \leq i \leq n$. Without loss of generality (i.e. by re-ordering the random variables), assume that $Y_1,\ldots,Y_m > 0$ and $Y_{m+1} = \cdots = Y_n = 0$.

In this problem, we assume that we cannot access $X_1,\ldots,X_n$, but we can access $Y_1,\ldots,Y_n$.

- Explain how you could use the EM algorithm to estimate μ from $Y_1,\ldots,Y_m$. Give details about the E and M steps. Let μ_k denote the estimate of μ from the k^{th} iteration of the EM algorithm. Show that

$$\mu_{k+1} = \frac{1}{n}\sum_{i=1}^{m} Y_i \quad + \frac{n-m}{m}\mu_k - \frac{n-m}{m}\frac{\phi(\mu_k)}{\Phi(-\mu_k)}, \qquad \forall\, k \geq 1. \qquad (*)$$

  Here ϕ is the PDF of a standard Gaussian, and Φ is the CDF of a standard Gaussian.
- Find the log-likelihood function $\log \ell(\mu)$ based on the observed data $Y_1,\ldots,Y_n$, and use it to write down a (nonlinear) equation that the MLE Z_n satisfies.

- Use the equation in the previous part to verify that Z_n is a fixed point of the recursion $(*)$.
- Prove that μ_k converges in distribution to μ as $k \to \infty$, for any starting value of $\mu_0 \in \mathbb{R}$, assuming that $m \geq 1$. Hint: Show that $|\mu_k - Y_k|$ decreases as $k \to \infty$. Hint: Use the Mean Value Theorem, and you can also freely use the following inequality

$$0 < \frac{\phi(x)[\phi(x) - x\Phi(-x)]}{[\Phi(-x)]^2} < 1, \qquad \forall\, x \in \mathbb{R}.$$

Exercise 6.3. Let $\theta \in (0, 3/4)$ be unknown. Define

$$(p_1, p_2, p_3, p_4) := ((1+\theta)/2,\ (1-\theta)/4,\ 1/4 - \theta/3,\ \theta/12).$$

Let $X = (X_1, X_2, X_3, X_4)$ be a multinomial distribution, so that

$$\mathbf{P}(X = x) = \binom{200}{x_1, x_2, x_3, x_4} \prod_{i=1}^{4} p_i^{x_i},$$

for any $x = (x_1, x_2, x_3, x_4)$ that are nonnegative integers with $x_1 + x_2 + x_3 + x_4 = 200$.

- Write an equation that would need to be solved in order to obtain an MLE for θ.
- Assume now that instead you have data from a multinomial distribution $Y = (Y_1, \ldots, Y_6)$, but $Y_1, \ldots, Y_4$ are not observed, and $X_1 = Y_1 + Y_2$, $X_2 = Y_3 + Y_4$ are both observed, along with $Y_5 = X_3$ and $Y_6 = X_4$. (Choose convenient probabilities $q_1, \ldots, q_6$ for the multinomial $Y_1, \ldots, Y_6$. For example, consider $q_1 = 1/2$, $q_2 = \theta/2$, $q_3 = 1/4 - \theta/3$, $q_4 = \theta/12$, $q_5 = 1/4 - \theta/3$, $q_6 = \theta/12$.) Write down the E and M steps of the EM algorithm for estimating θ, under the above assumptions.
- Is the EM algorithm simpler than directly finding the MLE of θ?

Exercise 6.4. Let $0 < p, q < 1$. Let $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$. Find the (left) eigenvectors of P, and find the eigenvalues of P. By writing any row vector $x \in \mathbb{R}^2$ as a linear combination of eigenvectors of P (whenever possible), find an expression for xP^n for any $n \geq 1$. What is $\lim_{n\to\infty} xP^n$? Is it related to the vector $\pi = (q/(p+q), p/(p+q))$?

Solution. To find the eigenvalues λ of P, we solve the equation $\det(P - \lambda) = 0$, i.e. $\det\begin{pmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{pmatrix} = 0$, i.e. $(1-p-\lambda)(1-q-\lambda) - pq = 0$, i.e.

$$\lambda^2 - \lambda(2-p-q) + 1 - p - q = 0.$$

Solving form λ, we get $\lambda = 1$ and

$$\begin{aligned}\lambda &= \frac{2-p-q-\sqrt{(2-p-q)^2 - 4(1-p-q)}}{2} \\ &= \frac{2-p-q-\sqrt{(2-p-q)^2 - 4(1-p-q)}}{2} \\ &= \frac{2-p-q-(p+q)}{2} = 1-p-q.\end{aligned}$$

When $\lambda = 1$, the matrix

$$\begin{pmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{pmatrix} = \begin{pmatrix} -p & p \\ q & -q \end{pmatrix}$$

has the row vector $(1, -p/q)$ in its (left) null space. When $\lambda = 1 - p - q$, the matrix

$$\begin{pmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{pmatrix} = \begin{pmatrix} -q & p \\ q & -p \end{pmatrix}$$

has the row vector $(1,1)$ in its (left) null space. As long as $p \neq q$, these vectors form a basis, and $\lim_{n\to\infty} P^n = \begin{pmatrix} \pi \\ \pi \end{pmatrix}$. □

Exercise 6.5. Suppose we have a Markov chain $X_0, X_1, \ldots$ with finite state space Ω. Let $y \in \Omega$. Define $L_y := \max\{n \geq 0 \colon X_n = y\}$. Is L_y a stopping time? Prove your assertion.

Solution. □

Exercise 6.6 (Knight Moves)**.** Consider a standard 8×8 chess board. Let V be a set of vertices corresponding to each square on the board (so V has 64 elements). Any two vertices $x, y \in V$ are connected by an edge if and only if a knight can move from x to y. (The knight chess piece moves in an L-shape, so that a single move constitutes two spaces moved along the horizontal axis followed by one move along the vertical axis (or two spaces moved along the vertical axis, followed by one move along the horizontal axis.) Consider the simple random walk on this graph. This Markov chain then represents a knight randomly moving around a chess board. For every space x on the chessboard, compute the expected return time $\mathbf{E}_x T_x$ for that space. (It might be convenient to just draw the expected values on the chessboard itself.)

Solution. From Example 9.60, the SRW on a graph has stationary distribution $\pi(x) = \deg(x)/(2\,|E|)$. We can write the degrees of vertices as the following matrix

$$\begin{pmatrix} 2 & 3 & 4 & 4 & 4 & 4 & 3 & 2 \\ 3 & 4 & 6 & 6 & 6 & 6 & 4 & 3 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 4 & 6 & 8 & 8 & 8 & 8 & 6 & 4 \\ 3 & 4 & 6 & 6 & 6 & 6 & 4 & 3 \\ 2 & 3 & 4 & 4 & 4 & 4 & 3 & 2 \end{pmatrix}.$$

The sum of degrees is 336. So, $\pi(x) = (deg)(x)/336$ for all $x \in V$. From Corollary 9.47, $\mathbf{E}_x T_x = 1/\pi(x)$, i.e. we can plot $\mathbf{E}_x T_x$ as values in the following matrix

$$\begin{pmatrix} 168 & 112 & 84 & 84 & 84 & 84 & 112 & 168 \\ 112 & 84 & 56 & 56 & 56 & 56 & 84 & 112 \\ 84 & 56 & 42 & 42 & 42 & 42 & 56 & 84 \\ 84 & 56 & 42 & 42 & 42 & 42 & 56 & 84 \\ 84 & 56 & 42 & 42 & 42 & 42 & 56 & 84 \\ 84 & 56 & 42 & 42 & 42 & 42 & 56 & 84 \\ 112 & 84 & 56 & 56 & 56 & 56 & 84 & 112 \\ 168 & 112 & 84 & 84 & 84 & 84 & 112 & 168 \end{pmatrix}.$$

□

USC Department of Mathematics, Los Angeles, CA
Email address: stevenmheilman@gmail.com