Review Session 7 – Hypothesis Testing

References/suggested reading

(i) Casella & Berger, chapter 8.

1 Introduction

In hypothesis testing, we are concerned with assessing the validity of a statement about an unknown population parameter, θ, using an observed sample $X_1, \ldots, X_n \sim f_\theta$. For example, we might want to know whether the mean of the underlying distribution is large or small, or even just positive or negative. As in said example, it is typical to consider two complementary hypotheses which are formally represented by sets Θ_0 and Θ_1 partitioning the space of possible parameter values $\Theta = \Theta_0 \sqcup \Theta_1$ (e.g., $\Theta_0 = (-\infty, 0]$ and $\Theta_1 = (0, \infty)$). The goal is then to decide, based on our sample, which of the two hypotheses "$\theta \in \Theta_0$" or "$\theta \in \Theta_1$" is true. Of course, this is not a determination we can make correctly based on every sample, due to the inherent randomness in the problem. However, we'd like to minimize the "probability of making an error", or the probability that one hypothesis has been chosen given that the other one is in fact true..

Up to this point, we've treated the two hypotheses on an equal basis. In practice, though, we make an epistomological distinction by letting one of the hypotheses, the *null hypothesis* denoted H_0, be a default or initial claim (e.g., there is no treatment effect in an experiment or there is no difference between two populations) which stands to be disproved in a hypothesis test if we choose the other *alternative hypothesis*, denoted H_1. The null hypothesis H_0 typically follows the general principle of parsimony, or Occam's razor, that the underlying reality is "simple" or explained by conventional wisdom. Thus, the onus is on us to provide evidence that disproves or rejects the simpler null hypothesis.

2 Terminology

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$ and we consider hypotheses $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$. A hypothesis test, or decision procedure, for choosing one of H_0 or H_1 is specified in terms of the value of a test statistic $W(X_1, \ldots, X_n)$. For example, if $\theta = \mathbb{E}_{X \sim f_\theta}[X]$ and we were testing $H_0 : \theta \le 0$ vs $H_1 : \theta > 0$, we might use the test statistics $W(X_1, \ldots, X_n) = \overline{X}_n$ and choose to reject H_0 if $W(X_1, \ldots, X_n)$ is large, and accept H_0 if $W(X_1, \ldots, X_n)$ is small.

The *critical region*, or rejection region denoted R, is the set of values of $W(X_1, \ldots, X_n)$ for which we would reject H_0, so that its complement is the set of values for which we would accept H_0. Thus, a decision procedure is completely defined by a test statistic and critical region $(W(X_1, \ldots, X_n), R)$.

If H_0 is true (i.e., $\theta \in \Theta_0$), but a hypothesis test $(W(X_1, \ldots, X_n), R)$ incorrectly decides to reject H_0, then the test has made a *Type I Error*. On the other hand, if $\theta \in \Theta_1$, but the test decides to accept H_0, a *Type II Error* has occurred. We can speak about the probability of making a Type I error for each $\theta \in \Theta_0$ by computing $\mathbb{P}_\theta(W(X_1, \ldots, X_n) \in R)$. Similarly, for each $\theta \in \Theta_1$, the probability of committing a Type II error will be given by $1 - \mathbb{P}_\theta(W(X_1, \ldots, X_n) \in R)$.

The *power function* of a test $(W(\cdot), R)$ is the function β of θ defined by $\beta(\theta) = \mathbb{P}_\theta(W(X_1, \ldots, X_n) \in R)$. Ideally, we would like the power function to be 0 for all $\theta \in \Theta_0$ and 1 for all $\theta \in \Theta_1$ (thus, making both Type I and II error probabilities zero), but this cannot be attained in most situations of interest.

For $\alpha \in [0, 1]$ we say a test with power function $\beta(\theta)$ is *size* α if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$, and that it is *level* α if $\sup_{\theta \in \Theta_0} \beta(\theta) \le \alpha$.

Going back to our distinction of the null vs. the alternative hypotheses, generally one of the two errors (Type I or II) is more costly or less palatable in practice, and the hypotheses should be arranged so that the Type I error is to be avoided. Thus, a decent test is one which is level α for some small α. Note that a test which never rejects (or such that $R = \emptyset$) will always be level α, but is a very uninformative test of course. So, the goal should be to find a test which is level α and also achieves a small Type II error probability, if possible.

Example 2.1 (two-sided Z-test)

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$ with $\theta_0 = \{0\}$ and $\Theta_1 = \mathbb{R}\backslash\{0\}$. Then, the hypotheses are $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$. One natural approach is to estimate θ by $\overline{X}_n$. We reject H_0 if $|\overline{X}_n| > c$ for some cutoff c. Suppose H_0 is true so that we have $\overline{X}_n \sim \mathcal{N}(0, 1/n)$. The probability of making a Type I error is then

$$\mathbb{P}(|\mathcal{N}(0, 1/n)| > c).$$

So, we see that as we increase c, the probability of Type I error decreases. However, since $\overline{X}_n \sim \mathcal{N}(\theta, 1/n)$ for $\theta \in \Theta_1$ under H_1, we see that the probability of Type II error increases (as our rejection region becomes smaller and smaller).

To make our test level α, we should choose $c = \Phi^{-1}(1-\alpha/2) \cdot n^{-1/2}$, where $\Phi(\cdot)$ is the standard normal cdf.

In the above example, we can change α in the formula $c = \Phi^{-1}(1-\alpha/2) \cdot n^{-1/2}$ to get hypothesis tests at different levels α. Smaller α will correspond to smaller rejections regions. At the same time, an observed test statistic $\overline{X}_n$ will be rejected for some choices of α's (i.e., larger α's) and accepted for others (i.e., smaller α's). One way of quantifying the strength of the decision made for a particular sample $(X_1, \ldots, X_n)$ is to report the ***p-value***, or the smallest level α_0 for which we would reject the null hypothesis at level α_0 with the observed data. This allows us to capture the continuum of all values α for which H_0 would have been rejected based on the data at level α, namely $\alpha > \alpha_0$.

Following the above example, we see that by solving for α in $\Phi^{-1}(1-\alpha/2) \cdot n^{-1/2}$, the p-value for the two-sided Z-test is:

$$\alpha_0 = 2 \cdot (1 - \Phi(|\overline{X}_n| \cdot \sqrt{n})) = 2 \cdot \mathbb{P}_{Z \sim \mathcal{N}(0,1)}(Z > |\overline{X}_n| \cdot \sqrt{n} | X_1, \ldots, X_n).$$

From this, we see that one interpretation of the p-value is that it is the probability that we would have observed a test statistic at least as extreme, or deviant from the null, as the one we observed from the sample.

Example 2.2 (two sided t-test)

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ where σ is this time unknown, and we are again testing $H_0 : \theta = 0$ vs. $H_1 : \theta \neq 0$. Then, we can estimate σ with the sample standard deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2}$ and use the test statistic $\overline{X}_n/s$ which is distributed as $1/\sqrt{n}$ times a Student's t distribution T_{n-1} under the null. Thus, a level α test with rejection region $\{|\overline{X}_n/s| > c\}$ would then choose c such that $\mathbb{P}(|T_{n-1}/\sqrt{n}| > c) \leq \alpha$, our p-value would then be, similar to before, $2 \cdot \mathbb{P}(T_{n-1} \geq |\overline{X}_n|/(s/\sqrt{n}) | X_1, \ldots, X_n)$.

3 Equivalence of hypothesis tests and confidence sets

Again, consider the normal setup $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$, with σ known where this time we are interested in testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Similar to before, we can consider the two-sided Z-test with acceptance region

$$|\overline{X}_n - \theta_0| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma \cdot n^{-1/2},$$

which we know is size α. The above translates to

$$\overline{X}_n - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma \cdot n^{-1/2} < \theta_0 < \overline{X}_n + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma \cdot n^{-1/2}.$$

We can think of the lower and upper bounds above as bounds on the possible null hypothesis values θ_0 for which we would not reject the given data $\{X_1, \ldots, X_n\}$ at level α.

In fact, this gives us a *confidence set* or interval for θ, i.e. a random interval which contains the true value of θ with probability $1 - \alpha$. In particular, since we know the test is size α, we know that under the null (i.e., $\mathcal{N}(\theta_0, \sigma^2)$ is the true data-generating distribution) the above inequalities will be true with probability $1 - \alpha$.

Thus, the interval

$$\left(\overline{X}_n - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma \cdot n^{-1/2}, \overline{X}_n + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \sigma \cdot n^{-1/2}\right),$$

covers the true value of θ with probability $1 - \alpha$, i.e. is a confidence set at level $1 - \alpha$.

We can also go in the reverse direction: given a level γ confidence interval for θ, we can derive a level $1 - \gamma$ hypothesis test. This will typically amount to doing the reverse of the above, or solving for some critical value in the confidence interval bounds.

Example 3.1 (deriving a test from a confidence interval)

Consider again the t-test setup where $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ where σ is unknown. By standardizing the sample mean $\overline{X}_n$ via the sample standard deviation s, we can establish that

$$\left(\overline{X}_n - T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \cdot \frac{s}{n^{1/2}}, \overline{X}_n + T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right) \cdot \frac{s}{n^{1/2}}\right),$$

is a γ confidence interval for θ, where $T_{n-1}^{-1}(\cdot)$ is the quantile function of the Student's t distribution with $n-1$ degrees of freedom. But, θ lying in the above interval translates to

$$\left|n^{1/2} \cdot \frac{\overline{X}_n - \theta}{s}\right| \le T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right).$$

Thus, we obtain the two-sided t-test from before.

The general result is as follows:

Theorem 3.2 (deriving confidence sets from tests)

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a distribution with parameter θ. Suppose for each value θ_0, there is a level α test δ_{θ_0} of the hypotheses $H_{0,\theta_0} : \theta = \theta_0$ and $H_{1,\theta_0} : \theta \neq \theta_0$. For each possible value $\mathbf{x}$ of $\mathbf{X}$, define

$$C(\mathbf{x}) = \{\theta_0 : \text{test } \delta_{\theta_0} \text{ does not reject } H_{0,\theta_0} \text{ if } \mathbf{X} = \mathbf{x} \text{ is observed}\}.$$

We can think of $C(\mathbf{x})$ as just "inverting" the acceptance region of test δ_{θ_0} to solve for θ_0, as we did in the previous example. Then, the random set $C(\mathbf{X})$ satisfies

$$\mathbb{P}_{\theta=\theta_0}(\theta_0 \in C(\mathbf{X})) \ge 1 - \alpha.$$

Proof. The proof follows from the definition of δ_{θ_0} being a level α_0 test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. ■

Theorem 3.3 (deriving tests from confidence sets)

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a distribution with parameter θ. Let $C(\mathbf{X})$ be a $1 - \alpha$ confidence set for θ. For each parameter value θ_0, the test which does not reject $H_0 : \theta = \theta_0$ (vs. $H_1 : \theta \neq \theta_0$) if and only if $\theta_0 \in C(\mathbf{X})$ is a level α test.

4 Likelihood ratio test

The likelihood ratio test is a popular type of hypothesis test related to maximum likelihood estimation. Recall that if $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta$, then the likelihood function is defined as, for observations $x_1, \ldots, x_n$,

$$L(\theta|x_1, \ldots, x_n) = L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta).$$

Definition 4.1 (likelihood ratio test statistic)**.** The *likelihood ratio test statistic* (LRT) for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}$$

where $L(\theta|\mathbf{x})$ is the likelihood function and where $\Theta = \Theta_0 \sqcup \Theta_0^c$. A *likelihood ratio test* is any test that has a rejection region of the form

$$\{\mathbf{x} : \lambda(\mathbf{x}) \le c\}$$

where $c \in [0, 1]$.

Intuitively, the LRT is small if there are parameter points θ in the alternative hypothesis Θ_0^c for which the observed sample is much more likely than for any parameter in the null hypothesis. On the other hand, the LRT is large and close to 1 if the likelihood is maximized at a $\theta \in \Theta_0$. Thus, we should reject H_0 for small values of $\lambda(\mathbf{x})$. Likelihood ratio tests enjoy certain theoretical properties which make them more favorable over other tests (cf. the Neyman-Pearson lemma in the next section) and also a simple asymptotic form, which we will cover in the final review session.

From the definition, we see that computing the LRT amounts to solving two MLE problems, (1) an unrestricted MLE over the entire parameter space Θ in the denominator and (2) a restricted MLE over Θ_0 in the numerator. The LRT can then be computed by plugging in the respective MLE's into the likelihood function.

Example 4.2 (normal LRT)

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$ population. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Then, we have

$$\lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\overline{x}_n|\mathbf{x})}$$

since the observed sample mean $\overline{x}_n$ is the MLE of θ. This becomes

$$\lambda(\mathbf{x}) = \exp\left(\left(-\sum_{i=1}^{n}(x_i - \theta_0)^2 + \sum_{i=1}^{n}(x_i - \overline{x}_n)^2\right)/2\right) = \exp(-n(\overline{x}_n - \theta_0)^2/2)$$

whence the rejection region can be written as

$$\{\mathbf{x} : |\overline{x}_n - \theta_0| \geq \sqrt{-2(\log c)/n}\}.$$

Thus, the normal LRT is the same as the two-sided Z-test we saw before.

Example 4.3 (one-sided normal LRT)

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ with σ^2 known. An LRT of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ is in fact a test that rejects H_0 if $(\overline{X} - \theta_0)/(\sigma/\sqrt{n}) > c$ (the details are similar to the previous example) where $c > 0$. The power function is then:

$$\beta(\theta) = \mathbb{P}_\theta\left(\frac{\overline{X}_n - \theta_0}{\sigma/\sqrt{n}} > c\right) = \mathbb{P}_\theta\left(\frac{\overline{X}_n - \theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right) = \mathbb{P}_\theta\left(Z > c + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right),$$

where Z is a standard normal random variable since $(\overline{X}_n - \theta)/(\sigma/\sqrt{n}) \sim N(0, 1)$. From this, we see that the power function $\beta(\theta)$ is increasing in θ with $\lim_{\theta\to-\infty}\beta(\theta) = 0$, $\lim_{\theta\to\infty}\beta(\theta) = 1$, and $\beta(\theta_0) = \mathbb{P}_{\theta_0}(Z > c)$.

Example 4.4 (binomial two-sided LRT)

Suppose we observe $Y \sim$ Binomial(n, θ) where θ is unknown and we wish to test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. The likelihood function is just the pmf

$$f(y|\theta) = \binom{n}{y}\theta^y(1 - \theta)^{n-y}.$$

Then, since $\Theta_0 = \{\theta_0\}$ and $\Theta = [0, 1]$, the LRT is

$$\lambda(y) = \frac{\theta_0^y(1 - \theta_0)^{n-y}}{\sup_{\theta\in[0,1]}\theta^y(1 - \theta)^{n-y}}.$$

Meanwhile, the MLE of θ is $\hat{\theta}_{\text{MLE}} = y/n$. Thus, plugging this into the denominator, we find that

$$\lambda(y) = \left(\frac{n\theta_0}{y}\right)^y\left(\frac{n(1 - \theta_0)}{n - y}\right)^{n-y}.$$

Example 4.5 (size of LRT)

A size α LRT is constructed by choosing c such that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\lambda(\mathbf{X}) \le c) = \alpha$. For example, consider the case of the normal family from before with $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$. We have that Θ_0 consists of a single point $\theta = \theta_0$ and $\sqrt{n}(\overline{X} - \theta_0) \sim \mathcal{N}(0,1)$ if $\theta = \theta_0$. So the test

$$\text{reject } H_0 \text{ if } |\overline{X} - \theta_0| \ge z_{\alpha/2}/\sqrt{n}$$

where $z_{\alpha/2}$ is the standard normal critical value, is the size α LRT. This corresponds to choosing $c = \exp(-z_{\alpha/2}^2/2)$.

5 Methods of comparing tests

Definition 5.1 (uniformly most powerful). Let $\mathcal{C}$ be a class of tests for testing $H_0 : \theta \in \theta_0$ versus $H_1 : \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a *uniformly most powerful* (UMP) class $\mathcal{C}$ test if $\beta(\theta) \ge \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$. Typically, the class $\mathcal{C}$ will be the class of all level α tests so that the UMP level α test, while it may not always exist for a generic problem, is a notion of an optimal or best test among the level α tests.

In a simple setting, it turns out the likelihood ratio test is UMP.

Theorem 5.2 (Neyman-Pearson Lemma)

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf or pmf corresponding to θ_i is $f(\mathbf{x}|\theta_i)$ for $i = 0, 1$ using a test with rejection region R that satisfies

$$\mathbf{x} \in R \text{ if } f(\mathbf{x}|\theta_1) > c f(\mathbf{x}|\theta_0)$$

and

$$\mathbf{x} \in R^c \text{ if } f(\mathbf{x}|\theta_1) < c f(\mathbf{x}|\theta_0)$$

for some $c \ge 0$, and

$$\alpha = \mathbb{P}_{\theta_0}(\mathbf{X} \in R)$$

Then,

1. (Sufficiency) Any test that satisfies both these conditions is an UMP level α test.
2. (Necessity) If there exists a test satisfying both these conditions with $c > 0$, then every UMP level α test is a size α test and has the same rejection region as the test above, up to a measure zero set.

Note that the ratio of likelihoods used in Neyman-Pearson $f(\mathbf{x}|\theta_0)/f(\mathbf{x}|\theta_1)$ is slightly different from the way we defined LRT previously. Namely, the denominator is the maximum likelihood over the alternative hypothesis set Θ_0^c rather than the entire parameter space. These two definitions of likelihood ratio only differ when the unrestricted MLE $\hat{\theta}_{\text{MLE}}$ lies in Θ_0; in this case, we wouldn't want to reject H_0 anyways under either definition since this is strong evidence in support of H_0. So, for most reasonable levels α, the two tests are equivalent.

Example 5.3 (UMP binomial test)

Let $X \sim$ binomial$(2, \theta)$. We want to test $H_0 : \theta = 1/2$ versus $H_1 : \theta = 3/4$. Calculating the ratios of the pmfs gives

$$\frac{f(0|\theta = 3/4)}{f(0|\theta = 1/2)} = 1/4, \frac{f(1|\theta = 3/4)}{f(1|\theta = 1/2)} = 3/4, \frac{f(2|\theta = 3/4)}{f(2|\theta = 1/2)} = 9/4$$

If we choose the cutoff/threshold $c \in (3/4, 9/4)$, the Neyman-Pearson Lemma says the test that reject H_0 if $X = 2$ is the UMP level $\alpha = \mathbb{P}(X = 2|\theta = 1/2) = 1/4$ test. If we choose $c \in (1/4, 3/4)$, the Neyman-Pearson Lemma says the test that rejects H_0 if $X = 1, 2$ is the UMP level $\alpha = \mathbb{P}(X = 1 \text{ or } 2|\theta = 1/2) = 3/4$ test. Choosing $c < 1/4$ or $c > 9/4$ yields the UMP level $\alpha = 1$ or level $\alpha = 0$ test.

Unfortunately, outside of simple hypotheses, an UMP test often does not exist in general. The general difficulty lies in the fact that a test cannot be simultaneously UMP for different $\theta \in \Theta_1$ in most setups.

Example 5.4 (nonexistence of UMP test)

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ with σ^2 known. Consider testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. We claim no UMP level α test exists for this problem. Let's first consider testing a specific alternative parameter $\theta_1 < 0$. The Neyman-Pearson lemma tells us that (1) a test that rejects H_0 if $\overline{X}_n < \sigma \cdot z_\alpha/\sqrt{n}$ has the highest possible power at θ_1 and (2) any other level α test for $H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_1$ that has as high a power as this test at θ_1 must have the same rejection region $\{\overline{X}_n < \sigma \cdot z_\alpha/\sqrt{n}\}$, up to a set of measure zero.

So, we know that if an UMP level α test exists for testing $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$, it must essentially be the test which rejects small values of $\overline{X}_n$. However, we can repeat these considerations supposing $\theta_1 > 0$ and testing $H_0 : \theta = 0$ vs. $H_1 : \theta = \theta_1$. In this case, the Neyman-Pearson lemma tells us that the test with rejection region $\{\overline{X}_n > \sigma \cdot z_\alpha/\sqrt{n}\}$ is level α. Furthermore, we can show this test has a higher power at this positive θ_1 than the test $\{\overline{X}_n < \sigma \cdot z_\alpha/\sqrt{n}\}$. Thus, the first test could not have been UMP level α for the composite alternate hypothesis $H_1 : \theta \neq 0$, and so there is no UMP level α test for this problem.

6 Problems

6.1 Previous Core Competency Problems

Problem 1 (September 2018, # 5)**.** We obtain observations $Y_1, \dots, Y_n$ which can be described by the relationship

$$Y_i = i \times \theta + \epsilon_i,$$

where $\epsilon_1, \dots, \epsilon_n$ are i.i.d. $N(0, \sigma^2)$; $\sigma^2 > 0$. Here θ and σ^2 are unknown.

(i) Find the least squares estimator $\hat{\theta}$ of θ; i.e., $\hat{\theta} = \operatorname{argmin}_{\theta \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - i\theta)^2$.

(ii) Is $\hat{\theta}$ unbiased?

(iii) Find the exact distribution of $\hat{\theta}$.

(iv) Find the asymptotic (non-degenerate) distribution of $\hat{\theta}$ (properly normalized).

(v) How would you test the hypothesis $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ (at level $(\alpha \in (0,1))$? Describe the test statistic and the critical value.

Problem 2 (2019 May, # 1)**.** Let $X \sim$ Poisson(λ) and $Y \sim$ Poisson(μ), where $\lambda, \mu > 0$ and assume that X and Y are independent.

(i) Find the conditional distribution of X given that $X + Y = n$.

(ii) Use the above, or otherwise, to test the hypothesis (at level $\alpha \in (0,1)$)

$$H_0 : \lambda = \mu \qquad \text{versus} \qquad \lambda > \mu.$$

Problem 3 (2020 May, # 5)**.** Suppose that $X_1, \dots, X_n$ are i.i.d. observations from the exponential distribution with parameter λ (recall that $\mathbb{E}(X_1) = \lambda^{-1}$). Consider the following testing problem:

$$H_0 : \lambda = \lambda_0 \qquad \text{versus} \qquad H_1 : \lambda = \lambda_1,$$

where $0 < \lambda_1 < \lambda_0$. Let $f_0(X_1, \dots, X_n)$ be the likelihood of the data under H_0 and $f_1(X_1, \dots, X_n)$ that under H_1.

(i) Show that $\log \frac{f_1(X_1,\dots,X_n)}{f_0(X_1,\dots,X_n)}$ is an increasing function of $\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$.

(ii) Suppose that $c_{\alpha,n}$ is such that $\mathbb{P}_{\lambda_0}(\overline{X}_n \geq c_{\alpha,n}) = \alpha$, for $\alpha \in (0,1)$. Relate $c_{\alpha,n}$ to $q_k(\beta)$ – the β-th quantile of the χ^2_k distribution (for some k).

(iii) How would you test the hypothesis

$$H_0 : \lambda = \lambda_0 \qquad \text{versus} \qquad H_1 : \lambda < \lambda_1,$$

Derive an expression for the power function of the test.

Problem 4 (2020 May, # 7)**.** Consider the random variable $X = \mu + \sigma Z$, where $\mu \in \mathbb{R}$, $\sigma > 0$ and Z is a random variable with density f. Suppose that μ and σ are unknown parameters and that the density f is known (completely specified). We have a random i.i.d sample $X_1, \dots, X_n$ with the same distribution as X. *You may assume for this problem that* $\mathbb{E}[Z] = 0$*,* $\mathbb{E}[|Z|] < \infty$*, and* $\operatorname{Var}(Z) \in (0, \infty)$.

(i) Propose unbiased estimators, $\hat{\mu}$ and $\hat{\sigma}^2$, of μ and σ^2.

(ii) Does the joint distribution of $(X_i - \hat{\mu})/\hat{\sigma}$ ($i = 1, \dots, n$) depend on μ and σ? Explain your answer.

(iii) For a given level $\alpha \in (0,1)$, describe a way to construct a confidence interval for μ with exact coverage probability $1 - \alpha$.

Note: I added some extra assumptions to this problem (in *italics*) since I don't think the problem is solvable without them in general.

Problem 5 (2020 September, # 1)**.** Researchers notice that a mutation in a gene predisposes individuals to a kind of radiation-induced cancer. The researchers theorize that the gene is involved in repairing damage from radiation, and that the mutation disables the gene. To explore their theory, the researchers obtain cells growing in a laboratory that have the mutation. They take eight different clumps of the cells, and randomize the clumps to treatment with radiation or no radiation (four in each group). They then examine a marker of damage from radiation in each cell in each clump, recording whether or not there appears to be damage. The researchers run the same experiment in clumps of cells that do not have the mutation. They explain that the cells that do not have the mutation are a "control". The researchers ask you to analyze the results.

(a) Propose a "reasonable" model to analyze the data.

(b) Propose how you plan to conclude whether mutation plays a role in repairing radiation damage.

Problem 6 (2020 September, # 3)**.** Suppose that $(X_{1i}, X_{2i}) \overset{i.i.d.}{\sim} N_2(\theta, I_2)$ for $1 \le i \le n$, where the parameter space is restricted to $\Theta := \{\theta = (\theta_1, \theta_2) : \theta_1, \theta_2 \ge 0\}$. Consider the following hypothesis testing problem:

$$H_0 : \theta = (0,0) \qquad \text{versus} \qquad H_1 : \theta \in \Theta \backslash \{(0,0)\}.$$

(a) Find the MLE of θ (when $\theta \in \Theta$).

(b) Find an expression for the likelihood ratio statistic $\Lambda_n \in (0,1]$ in this case.

(c) Find the asymptotic distribution of $-2\log\Lambda_n$, under H_0 [Hint: You may want to consider the cases where $(\overline{X}_1, \overline{X}_2)$ belongs to each of the four quadrants separately.]

Problem 7 (2021 May, # 2)**.** Let $X_1, X_2, \ldots, X_n$ be from an i.i.d. random sample Uniform$(0,\theta)$, where $\theta > 0$ is an unknown parameter. Suppose that we want to test the following hypothesis:

$$H_0 : 3 \le \theta \le 4 \qquad \text{versus} \qquad H_1 : \theta < 3 \text{ or } \theta > 4. \tag{1}$$

Let $Y_n = \max\{X_1, \ldots, X_n\}$. Consider the following two tests:

$$\delta_1 : \text{Reject } H_0 \text{ if } Y_n \le 2.9 \text{ or } Y_n \ge 4$$

and

$$\delta_2 : \text{Reject } H_0 \text{ if } Y_n \le 2.9 \text{ or } Y_n \ge 4.5.$$

(i) Find the power functions of δ_1 and δ_2, when $\theta \le 4$

(ii) Find the power functions of δ_1 and δ_2, when $\theta > 4$.

(iii) Which of the two tests seems better for testing the hypothesis (1)?

6.2 Additional Practice

Problem 8 (Casella & Berger, Exercise 8.8)**.** A special case of a normal family is one in which the mean and variance are related, the $\mathcal{N}(\theta, a \cdot \theta)$ family.

(a) Find the LRT of $H_0 : a = 1$ versus $H_1 : a \neq 0$ based on a sample $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} \mathcal{N}(\theta, a \cdot \theta)$, where θ is unknown.

(b) Now consider the $\mathcal{N}(\theta, a \cdot \theta^2)$ family. If $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, a \cdot \theta^2)$, where θ is unknown, find the LRT of $H_0 : a = 1$ versus $H_1 : a \neq 1$.

Problem 9 (Casella & Berger, Exercise 8.13)**.** Let X_1, X_2 be i.i.d. Unif$[(\theta, \theta+1)]$. For testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$, consider two competing tests:

$$\phi_1(X_1) : \text{Reject } H_0 \text{ if } X_1 > .95.$$
$$\phi_2(X_1, X_2) : \text{Reject } H_0 \text{ if } X_1 + X_2 > C.$$

(i) Find the value of C so that ϕ_2 has the same size (i.e., probability of committing a Type 1 error) as ϕ_1.

(ii) Calculate the power function of each test.

(iii) Which test ϕ_1 or ϕ_2 is more powerful (i.e., has larger power function)? Does it depend on the value of θ?