Unit 2 Cont'd Foundations of Inference

Introduction to Hypothesis Testing

Modeling Clinical Trials I

1/1 point (graded)

In a clinical trial, a pharmaceutical company wants to determine the efficacy of a cold remedy. To do so, they recruit $2n$ individuals to participate in a study, (randomly) placing n individuals in the **treatment group** and n individuals in the **control group**. Throughout the study, the treatment group will receive the actual drug, while the control group will only receive a placebo (for example, a sugar pill).

To statistically model this scenario, we let

- $X_1, \ldots, X_n$ be random variables that denote the number of coughs per hour of individuals $1, \ldots, n$, respectively in the treatment group, and
- $Y_1, \ldots, Y_n$ be random variables that denote the number of coughs per hour of individuals $1, \ldots, n$, respectively, in the control group.

Let's assume that the individuals participating in the trial are separated throughout the trial, so that it's reasonable to expect the coughs per hour of one individual in the study will not affect the coughs per hour of some other individual in the study. Moreover, we expect the administered drug to induce the same distribution of coughs for each individual in the treatment group. We will also assume that the distribution of coughs in the control group is the same for each individual.

What collection of mathematical assumption(s) below would capture exactly all of the assumptions stated in the previous paragraph, but nothing more? (Choose all that apply.)

- [ ] $X_1, \ldots, X_n$ are independent, but may not all have the same distribution. The same holds for $Y_1, \ldots, Y_n$.
- [ ] $X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated. The same holds for $Y_1, \ldots, Y_n$.
- [x] The random variables $X_1, \ldots, X_n$ are iid and the random variables $Y_1, \ldots, Y_n$ are iid (though perhaps from a different distribution from $X_1, \ldots, X_n$).
- [ ] The random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are all iid (in particular, the X_i's and Y_i's are sampled from the *same* distribution).
- [x] The random variable X_i for any i is independent of Y_j for any j.

Solution:

The third choice "The random variables $X_1, \ldots, X_n$ are iid and the random variables $Y_1, \ldots, Y_n$ are iid (though perhaps from a different distribution from $X_1, \ldots, X_n$)." and the last choice "The random variable X_i for any i is independent of Y_j for any j." *together* capture all the assumptions we need. Since, intuitively speaking, we do not expect individuals in the study to affect one another, this translates to imposing that all random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are all mutually independent. We also assumed that $X_1, \ldots, X_n$ have the same distribution induced by the drug and that the control group $Y_1, \ldots, Y_n$ has a common distribution of coughs. Thus, the assumptions that $X_1, \ldots, X_n$ are iid, $Y_1, \ldots, Y_n$ are iid, and the two groups of random variables are mutually independent capture all of the information described. It is important to note, however, that X_i and Y_i may be sampled from **different** distributions.

We now look at the incorrect choices in order.

- The first and second choices, "$X_1, \ldots, X_n$ are independent, but may not all have the same distribution. The same holds for $Y_1, \ldots, Y_n$." and "$X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated. The same holds for $Y_1, \ldots, Y_n$.", respectively, are incorrect because each directly contradicts the iid assumption.
- The fourth choice "The random variables $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ are all iid (in particular, the X_i's and Y_i's are sampled from the *same* distribution.)" is incorrect. The paragraph mentioned does not assume anywhere that the X_i's should have the same distribution as the Y_i's. Since we are mainly interested in deciding, based on the data, whether or not the X_i's and Y_i's have the same (or differing) distribution, for the purpose of modeling, it would not make sense to impose that they have the same distribution.

3. Statistical Model of a Two Sample Experiment

Bookmark this page

Preparation: Statistical Model of a Two Sample Experiment

2/2 points (graded)

The observed outcome of a statistical experiment consists of two samples:

$$X_1, X_2, \ldots X_n \overset{\text{i.i.d.}}{\sim} X \sim \text{Ber}(p_1)$$
$$Y_1, Y_2, \ldots Y_m \overset{\text{i.i.d.}}{\sim} Y \sim \text{Ber}(p_2).$$

where in addition, X, Y and the two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent.
An associated statistical model is $\left(E, \{P_\theta\}_{\theta\in\Theta}\right)$ where E is the smallest sample space of the pair (X, Y), and P_θ is the joint distribution of (X, Y) with parameter θ. Because X and Y are independent, their joint distribution is the product of their respective distributions.

Identify the sample space E and the parameter space Θ:
(Choose one per column. The notation (x, y) denotes the ordered pair; notation $]x, y[$ denotes an open interval.)

Sample space E:	Parameter space Θ:
◯ $\{0, 1\}$	◯ $\{0, 1\}$
◉ $\{0, 1\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$	◯ $\{0, 1\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$
◯ $]0, 1[$	◯ $]0, 1[$
◯ $]0, 1[\times]0, 1[\subset \mathbb{R}^2$	◉ $]0, 1[\times]0, 1[\subset \mathbb{R}^2$
✔	✔

Solution:

Since $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$, the pair (X, Y) takes value in the sample space $E = \{0,1\} \times \{0,1\} = \{(0,0), (0,1), (1,0), (1,1)\}$.

Since X, Y are independent, the joint distribution of (X, Y) is the product $\text{Ber}(p_1) \times \text{Ber}(p_2)$. Hence, the family $\{P_\theta\}_{\theta \in \Theta}$ of joint distributions is parametrized by $\theta = (p_1, p_2)$ and the parameter space is

$$\Theta = \{(p_1, p_2) : p_1 \in]0,1[, p_2 \in]0,1[\} =]0,1[\times]0,1[\subset \mathbb{R}^2.$$

Preparation: Statistical Model of a Two Sample Experiment II

0/2 points (graded)

Recall the statistical experiment from the lecture: to test whether boarding times by the Window-Middle-Aisle boarding method is shorter than boarding times by the rear-to-front method, we collect a sample of boarding times of each method. We model these boarding times as the following two sets of normal variables:

$$\begin{aligned} X_1, X_2, \ldots X_n \text{ are } i.i.d. \text{ copies of } X \sim \mathcal{N}(\mu_1, \sigma_1^2) &\qquad \text{boarding times of rear-to-front} \\ Y_1, Y_2, \ldots Y_m \text{ are } i.i.d. \text{ copies of } Y \sim \mathcal{N}(\mu_2, \sigma_2^2) &\qquad \text{boarding times of window-middle-aisle} \end{aligned}$$

where X and Y are also independent.

Let $\left(E, \{P_\theta\}_{\theta \in \Theta}\right)$ be the statistical model associated with this experiment where

- E is the sample space of the pair of random variables (X, Y);
- $\{P_\theta\}_{\theta \in \Theta}$ is the family of joint distributions of (X, Y).

For simplicity, **assume the two standard deviations σ_1 and σ_2 are some known, fixed quantities σ_1^* and σ_2^***.

Choose a valid candidate for the parametrization θ, which describes the family of joint probability distributions of (X, Y).

- ◯ $\mu_1 - \mu_2$
- ◉ $\left(\mu_1, (\sigma_1)^2, \mu_2, (\sigma_2)^2\right)$ where $(\sigma_1)^2$ and $(\sigma_2)^2$ can each take on more than a single value
- ◯ (μ_1, μ_2) ✔
- ◯ (μ_2, μ_1)

✖

Which of the following are legitimate choice(s) of the parameter space Θ?
(Choose all that apply)

- [x] $\Theta = \mathbb{R}$
- [x] $\Theta = [0, \infty)$
- [ ] $\Theta = \mathbb{R}^2$ ✔
- [ ] $\Theta = [0, \infty) \times [0, \infty)$ ✔

✖

Solution:

Since X, Y are independent, the joint distribution of (X, Y) is the product $\mathcal{N}\left(\mu_1, (\sigma_1)^2\right) \times \mathcal{N}\left(\mu_2, (\sigma_2)^2\right)$

Since the variances σ_1 and σ_2 are fixed and known, the only parameters determining the joint distribution is μ_1 and μ_2. Hence, a choice of the parameter θ is the 2-dimensional vector $\begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix}$. (We could also have chosen to construct the statistical model using the pair (Y, X) instead. The family of joint distributions in that case would be parametrized by $\begin{pmatrix} \mu_2 & \mu_1 \end{pmatrix}$).

This gives the parameter space

$$\Theta = \{(\mu_1, \mu_2) : \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}\} = \mathbb{R}^2.$$

Because μ_1 and μ_2 model average boarding times, we can further restrict to

$$\Theta = \{(\mu_1, \mu_2) : \mu_1 \in [0, \infty), \mu_2 \in [0, \infty)\} = [0, \infty) \times [0, \infty)\}.$$

Modeling Clinical Trials II

2/2 points (graded)

Let's use the same statistical set-up as in an earlier question. Recall that X_i denotes the **number of coughs per hour** for individual i in the treatment group, and Y_i denotes the number of coughs per hour for individual i in the control group. Assume the distributions on coughs per hour to be $X_1, \ldots, X_n \sim \text{Poiss}(\mu_{\text{drug}})$ for the treatment group and $Y_1, \ldots, Y_n \sim \text{Poiss}(\mu_{\text{control}})$ for the control group.

What is(are) the unknown parameter(s) in this example?

- ◯ Only μ_{drug}
- ◯ Only μ_{control}
- ◉ Both μ_{drug} and μ_{control}
- ◯ Neither μ_{drug} nor μ_{control}

✔

Which of the following statement about the efficacy of the cold remedy corresponds to $\mu_{\text{drug}} < \mu_{\text{control}}$?

- ◯ This drug is less effective than the placebo.
- ◉ This drug is more effective than the placebo.
- ◯ This cold remedy is more effective than the most commonly used one in the US
- ◯ None of the above

✔

Solution:

Consider the first question. Since a priori (*i.e.*, before running the clinical trial), we do not know what the true mean of the control group or treatment group will be, this implies that μ_{drug} and μ_{control} are unknown parameters. Since there are two unknown parameter corresponding to two *different* samples, this is an example of a **two-sample hypothesis test** .

Now consider the second question. We examine the choices in order.

- "This drug is more effective than the placebo." is correct. If we knew the true parameters μ_{control} and μ_{drug}, we could just compare their values to determine if the drug was more effective than the placebo. And if $\mu_{\text{drug}} < \mu_{\text{control}}$, this implies that the number of coughs per hour is lower when the drug is administered vs. the placebo. Thus, it is reasonable to conclude that the drug is more effective than the placebo.

  Remark: In actual clinical trials, we do not have access to the true parameters, which is why we need to employ the methods of hypothesis testing to determine whether the treatment or placebo is more effective.

- "This drug is less effective than the placebo." is incorrect. See the explanation of the previous choice to understand why this is not a reasonable interpretation.
- "This cold remedy is more effective than the most commonly used one in the US" is incorrect. We have only compared this drug to the placebo, not to any other drug. Thus this is not a reasonable conclusion.

Certainty of a Two-Sample Hypothesis Test

1/1 point (graded)

Let's use the same statistical set-up as above. Recall that $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poiss}(\mu_{\text{drug}})$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Poiss}(\mu_{\text{control}})$ where X_i denotes the number of coughs per hour of the i-th individual in the treatment group and Y_i denotes the number of coughs per hour of the i-th individual in the control group. The parameters μ_{drug} and μ_{control} are unknown. You would like to determine from the two samples if $\mu_{\text{drug}} < \mu_{\text{control}}$.

To do so, you compute the sample mean corresponding to each group:

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

and observe that $\overline{X}_n < \overline{Y}_n$.

Can you conclude with 100% certainty that $\mu_{\text{drug}} < \mu_{\text{control}}$?

Choose the correct answer that also has a correct explanation.

- ◯ Yes, because we do not expect the placebo effect to factor in to this trial.
- ◯ Yes, because we have carefully chose the treatment and control group so their sample means match the true means: $\overline{X}_n = \mu_{\text{drug}}$ and $\overline{Y}_n = \mu_{\text{control}}$.
- ◉ No, we cannot conclude $\mu_{\text{drug}} < \mu_{\text{control}}$. Since there are possible fluctuations in $\overline{X}_n$ and $\overline{Y}_n$ about their respective means μ_{drug} and μ_{control}, there is some positive probability that $\overline{X}_n < \overline{Y}_n$ while at the same time $\mu_{\text{drug}} > \mu_{\text{control}}$.
- ◯ No, because the sample means $\overline{X}_n$ and $\overline{Y}_n$ are biased estimators of their true means, μ_{drug} and μ_{control}, respectively.

Solution:

First we examine the correct choice and then look at the incorrect responses in order.

- The third response "No, we cannot conclude $\overline{X}_n < \overline{Y}_n$. Since there are significant fluctuations in $\overline{X}_n$ and $\overline{Y}_n$ about their respective means μ_{drug} and μ_{control}, there is some positive probability that $\overline{X}_n < \overline{Y}_n$ while at the same time $\mu_{\text{drug}} > \mu_{\text{control}}$." is the correct response. Using $\overline{X}_n < \overline{Y}_n$ to predict $\mu_{\text{drug}} < \mu_{\text{control}}$ is only a heuristic, and this may fail at times. For example, perhaps by chance we chose a treatment group that responds extremely well to the drug (*i.e.* $X_1, \ldots, X_n$ are outliers), but the vast majority of the population will not see a significant effect. In this case it is possible that $\overline{X}_n < \overline{Y}_n$ while $\mu_{\text{drug}} > \mu_{\text{control}}$.

- The first response "Yes, because we do not expect the placebo effect to factor in to this trial." is incorrect. On the contrary, the goal of this trial is to compare the placebo effect to the effect of the treatment so we can determine if the drug is useful for treating the cold.

- The second response "Yes, because we have carefully chose the treatment and control group so their sample means match the true means: $\overline{X}_n = \mu_{\text{drug}}$ and $\overline{Y}_n = \mu_{\text{control}}$." is incorrect. We have no way of selecting the participants of the trial so that the sample means match the true means, and this is the case for a couple reasons. First, we do not know the true means, so even if we were given the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ in advance, this would be not possible. And moreover, we do not even have access to the observations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ until after the clinical trial has completed. Hence, we have no way of controlling the sample mean in advance, and doing so would actually defeat the purpose of running a clinical trial.

- The fourth response "No, because the sample means $\overline{X}_n$ and $\overline{Y}_n$ are biased estimators of their true means, μ_{drug} and μ_{control}, respectively" is incorrect: this choice gives the right answer "No" but for a reason which is false.
  Both $\overline{X}_n$ and $\overline{Y}_n$ are unbiased estimators of μ_{drug} and μ_{control} respectively.
  In addition, it is possible for a simple comparison test to yield an incorrect answer even if the estimators are unbiased (again, highlighting why the third choice is correct).

Another Example: Modeling the Height of the U.S. Population I

1/1 point (graded)

You have access to U.S. census data for the height of individuals from the year 1920. The dataset shows that the average height of the U.S. was 5.5 feet. For simplicity, let's assume that the 1920 dataset included the heights of *all* people residing in the U.S. at that time.

Your goal as a statistician is to provide a response to the **question of interest**:

"**Were people in the U.S. taller in 2018 than in 1920?**".

The company that you work for has limited resources, so you will not be able to survey the entire U.S. population, but you still would like to assess the heights of individuals in the U.S. Therefore, you decide to take the following sampling approach:

Pick 1 million people (with replacement, for simplicity) randomly from the U.S. population and record their heights. Let X_i denote the random variable equal to the height of the i-th person chosen. Assume that any particular individual's height does not influence anyone else's and that there is a common underlying distribution which describes the random variables $X_1, \ldots, X_n$.

Which mathematical property of $X_1, \ldots, X_n$ most accurately captures all assumptions made in the previous paragraph?

- ◯ $X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated.
- ◯ $X_1, \ldots, X_n$ are independent, but may not all have the same distribution.
- ◉ The random variables $X_1, \ldots, X_n$ are iid.

✔

Solution:

We first examine the correct choice and then look at the incorrect choices in order.

- The third choice "The random variables $X_1, \ldots, X_n$ are iid." is correct. Since we are assuming that a person's height will not affect any other peron's height, it makes sense to impose that the X_i's are mutually independent. Moreover, since we stated that there is an underlying distribution describing $X_1, \ldots, X_n$, this is the same as saying that the X_i's are identically distributed.

- The first and second choices "$X_1, \ldots, X_n$ are independent, but may not all have the same distribution." and "$X_1, \ldots, X_n$ all have the same distribution, but some of them are correlated.", respectively, are incorrect because either would contradict the iid assumption.

Modeling the Height of the U.S. Population II

0/1 point (graded)

Continuing from the problem above, your goal is to answer the question of interest

"**Were people in the U.S. taller in 2018 than in 1920?**"

You do so by sampling 10^6 individuals labeled $1, 2, \ldots, 10^6$ chosen randomly from the U.S. population. Let X_i denote the height of the i-th individual. We will treat X_i as a random variable, and use the sample $X_1, \ldots, X_n$ to answer the question of interest.

In addition to the initial modeling assumptions on $X_1, \ldots, X_n$ discussed in the previous problem, we further assume:

- X_i is Gaussian;
- $\text{Var}(X_i) = 1.3$.

These assumptions were derived by fitting the data from the 1920 census.

Having established these assumptions, we decide on the following protocol for answering the question of interest. If $\mu = \mathbb{E}[X_i] > 5.5$ (and the goal of this lecture is to tackle the question "Is $\mu > 5.5$?"), then we respond by "Yes, the 2018 U.S. population was taller as a whole than the 1920 population". Otherwise, we respond by "No."

Which of the following are **true** statements regarding the two additional assumptions above? (Choose all that apply.)

- [x] They place restrictions on the different possible distributions that $X_1, \ldots, X_n$ could follow. ✔
- [x] For the purposes of hypothesis testing, they allow us to interpret the question of interest as a very specific mathematical question about the mean of X_i. ✔

Solution:

We examine the choices in order.

- "They place restrictions on the different possible distributions that $X_1, \ldots, X_n$ could follow." is correct. There are many possible distributions that $X_1, \ldots, X_n$ could follow, but we have specifically assumed that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1.3)$ where the parameter μ is unknown.

- "For the purposes of hypothesis testing, it allows us to interpret the question of interest as a very specific mathematical question about the mean of X_1." is correct. Originally, the question "**Were people in the U.S. taller in 1920 or 2018?**" is not precise enough to be well-posed mathematically. However, in making the above assumptions, we were able to focus on some specific property of $X_1, \ldots, X_n$ that can be rigorously tested. The new, more specific question that we have to answer is now:
  "Is the true, unknown parameter μ that describes the mean height of the 2018 U.S. population larger than 5.5 or smaller than 5.5?"
  Remark: We will not be able to answer this question directly because, from practicality constraints, we cannot sample the *entire* U.S. population. Rather, we will use our sample of 1 million individuals to statistically infer, with quantified error, what the answer to the above question should be.

Certainty of a One-Sample Hypothesis Test

1/1 point (graded)

As above, the question of interest is "**Were people in the U.S. taller in 1920 or 2018?**".

As above, you decided to answer this question using the following strategy and assumptions:

- Sample 1 million individuals labeled $1, 2, \ldots, 10^6$ randomly from the 2018 U.S. population.
- Model the height of the i-th individual as a random variable X_i and make the assumption, based on 1920 data, that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1.3)$ where μ is an unknown parameter.

This allowed us to specify a precise way to answer the initial question of interest:

- If $\mu = \mathbb{E}[X_i] > 5.5$, then you would conclude that the U.S. population is taller in 2018 than it was in 1920 and report "Yes" to the question of interest. Otherwise, you would say "No."

Suppose you access samples $X_1, \ldots, X_{10^6}$ from the 2018 U.S. population and observe that the **sample mean** $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is much larger than 5.5.

Can you conclude with 100 $\%$ certainty that $\mu > 5.5$? (Equivalently, can you know for sure that the answer to the question of interest is "Yes" ?)

Choose the correct answer that also has a correct explanation.

- ○ Yes, because there are only 10^6 people in the 2018 population to begin with.
- ○ Yes, because we have carefully chosen the 10^6 individuals so that their sample mean agrees with the true mean μ.
- ◉ No, because if, by chance, we chose a 'bad sample' (for example, the million tallest individuals in the U.S.), then the true parameter μ may be much smaller than $\overline{X}_n$ and even much smaller than 5.5.
- ○ No, because the sample mean $\overline{X}_n$ is a biased estimator of the true mean.

Solution:

We handle the choices in order.

- "Yes, because there are only 10^6 people in the 2018 population to begin with." is incorrect. The U.S. population is currently roughly 325 million, which is significantly larger than the number of samples we will access.
- "Yes, because we have carefully chosen the 10^6 individuals so that their sample mean agrees with the true mean μ." is also incorrect. Since we are sampling individuals randomly from the entire U.S. population, we did not use any specific information about population when choosing our sample.
- "No, because if, by chance, we chose a 'bad sample' (for example, the million tallest individuals in the U.S.), then the true parameter μ may be much smaller than $\overline{X}_n$ and even much smaller than 5.5." is correct. In general, the sample mean may have large fluctuations about the true mean, so it is entirely possible that $\overline{X}_n > 5.5$ while $\mu < 5.5$.
- "No, because the sample mean $\overline{X}_n$ is a biased estimator of the true mean." is not the best choice, because the reason it gives is false. By linearity of expectation, the sample mean is an unbiased estimator of the true mean: $\mu = \mathbb{E}[\overline{X}_n]$.

Remark: In general, it is not possible to answer hypothesis testing questions with 100 $\%$ certainty. However, you will see later in this lecture how to quantify this inherent uncertainty.

Aside: Accessing a Global Data-Set

1/1 point (graded)

As above,the **question of interest** is "**Were people in the U.S. taller in 2018 than in 1920?**".

In the problem, you consider the following two approaches to answer this question.

Approach 1:

- Access the entire 2018 U.S. population of ≈ 325 million people.
- Compute the average μ of the entire data set.
- If $\mu > 5.5$ ", then the answer to the question of interest is "Yes". Otherwise, the answer is "No".

Approach 2:

- Sample 10^6 people labeled $1, 2, \ldots, 10^6$ at random from the 2018 U.S. population.
- Model the heights of the i-th individual as a random variable X_i and make the assumption, based on 1920 data, that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1.3)$ where μ is an unknown parameter.
- If $\mu = \mathbb{E}[X_1] > 5.5$, then you would conclude that the U.S. population from 2018 is taller as a whole than the 1920 population and report "Yes" to the question of interest. Otherwise, you would say "No."

Which of the following is a potential **disadvantage** of using **Approach 1** vs. **Approach 2**?

- [ ] In Approach 1, we obtain μ exactly.
- [x] In Approach 1, we have to invest the time, money, and overall resources to assess the heights of the entire 2018 U.S. population of ≈ 325 million people.
- [ ] In Approach 1, we are not working with a restricted data set (e.g. a limited sample of the population), so we don't have to worry about errors stemming from choosing a 'bad sample' (for example, a sample consisting entirely of outliers)

Hypothesis Testing vs. Parameter Estimation (Optional)

0 points possible (ungraded)

As above, your goal is to answer the question of interest "**Were people in the U.S. taller in 2018 than in 1920?**".

As in previous problems, you know that in 1920, the heights of the U.S. population were distributed (approximately) like a Gaussian with mean 5.5 and variance 1.3. In addition to imposing that $X_1, \ldots, X_{10^6}$ are iid, you also made the assumptions that

- the heights $X_1, \ldots, X_{10^6}$ 2018 are also distributed like a Gaussian, and
- the variance of X_1 is 1.3

Since we made no assumptions about the mean $\mu := \mathbb{E}[X_1]$, we will treat μ as an unknown parameter.

The goal of this unit is to learn how to answer questions similar to the following:
Is $\mu > 5.5$, or is $\mu \leq 5.5$?

This is a basic example of a **hypothesis testing** question.

Which of the following are **true statements** regarding **hypothesis testing** as exemplified above and **parameter estimation** as discussed in previous lectures?
(Choose all that apply.)

- [x] In the above hypothesis testing set-up and in the models in the previous lectures on parameter estimation, we make the assumption that our data is iid from some unknown distribution.
- [x] When carrying out parameter estimation, we are interested in coming up with an estimator $\hat{\mu}$ that we want to be close to the true parameter μ.
- [x] When performing hypothesis testing (as above), we are **not** necessarily interested in finding an estimator for μ. Rather, our goal is to decide whether or not the true parameter μ lies in a certain region.
- [ ] When performing hypothesis testing, our main goal is to come up with a good approximation of the true parameter.

Solution:

We examine the choices in order.

- "In the above hypothesis testing set-up and in the models considered in the previous unit on parameter estimation, we make the assumption that our data is iid from some unknown distribution." is correct. In the parameter estimation unit, for all statistical models we assumed that our sample consisted of iid random variables. In the U.S. heights example, we have also made the assumption that the heights $X_1, \ldots, X_n$ are iid.

- "When carrying out parameter estimation, we are interested in coming up with an estimator $\hat{\mu}$ that we want to be close to the true parameter μ." is correct. The main goal of parameter estimation is to come up with some approximation for the unknown true parameter using the sample $X_1, \ldots, X_n$.

- "When performing hypothesis testing (as above), we are **not** necessarily interested in finding an estimator for μ. Rather, our goal is to decide find out if mu has some particular property (for example, whether or not mu lies in a certain region)." is correct. The question stated above is: **Is $\mu > 5.5$ OR is $\mu \leq 5.5$?**.To answer this question, we do not necessarily need to come up with an estimator for the true parameter. It would be enough to decide whether or not μ lies in some particular region (in this case, the interval $(5.5, \infty)$).

- "When performing hypothesis testing, our main goal is to come up with a good approximation of the true parameter." is incorrect. As elaborated upon in the previous bullet, this is not the goal of hypothesis testing. Rather, we want to decide if the true parameter has some particular property (e.g. whether it lies in a particular region or not).

Two Sample vs. One Sample Tests

1/1 point (graded)

A **one-sample test** is a hypothesis test where an unknown parameter μ is to be compared to a known reference value. For example, the U.S. heights example was a one-sample test because we wanted to compare the unknown mean μ from 2018 to the known average height from 1920, which was 5.5.

A **two-sample test** is a hypothesis test where two unknown parameters are compared to each other. For example, the clinical trial example is a two-sample test because we want to compare the unknown μ_{drug} to the unknown μ_{control} to quantify the drug effect.

Which of the following is a two-sample hypothesis testing question? (choose all that apply)

- [ ] Recall the kiss example of Unit 1. The question is: **Do the majority of people turn their head to the left or right?**.
- [ ] We collect data in a college. We find that out of 824 sampled students, 487 prefer nacho cheese flavored chips and 337 prefer the cool ranch flavor. The question is: **Do students prefer nacho cheese or cool ranch?**
- [x] James has to choose between two routes to go to work: either by the subway or by the bus. To decide he samples 112 persons at his workplace who also live in the same neighborhood as him and asks them for two pieces of information: (1) method used to commute AND (2) commute time. He will use this data to answer the question: **Is it faster to travel to work by the bus or by the subway?**.

✔

Solution:

We examine the choices in order.

- "Recall the kiss example of Unit 1. The question is: **Do the majority of people turn their head to the left or right?**." is incorrect because it is a one-sample test. Namely, we treat turning to the right as 1 and turning to the left as 0, and we modeled this as $\mathrm{Ber}\,(p)$ for some unknown parameter p. Then we can rephrase the above question as: "**Is** $p > 1/2$ **OR is** $p < 1/2$ **?**". Hence, $1/2$ takes the role of the reference value, and we are only one unknown parameter is involved, so this is indeed a one-sample test.

- "We collect data in a college. We find that out of 824 sampled students, 487 prefer nacho cheese flavored chips and 337 prefer the cool ranch flavor. The question is: **do students prefer nacho cheese or cool ranch?**" is incorrect because it is a one-sample test. If we encode preferring nacho cheese flavor as 1 and preferring cool ranch as 0, then we can model this as a Bernoulli random variable, just like in the kiss example. You are encouraged to fill in the details to show that this is a one-sample test. (e.g. what is the reference value?)

- "James has to choose between two routes to go to work: either by the subway or by the bus. To decide he samples 112 persons at work who live in the same neighborhood as him and asks them for two pieces of information: (1) method used to commute AND (2) commute time. He will use this data to answer the question: **Is it faster to travel to work by the bus or by the subway?**." is correct because it is a two-sample test. The two unknown parameters of interest are μ_{subway}, the average commute time via subway, and μ_{bus} the average commute time via bus. We can rephrase the above question as: "**Is** $\mu_{\text{bus}} > \mu_{\text{subway}}$ **OR is** $\mu_{\text{subway}} > \mu_{\text{bus}}$ **?**". Indeed this is a two-sample test.

Intuition for Hypothesis Testing

1/1 point (graded)

The purpose of this question is not to formally outline the procedure of hypothesis testing, but rather to illustrate some of the intuition involved in answering a hypothesis testing question.

Your friend claims to you that a random variable X has the distribution $\mathcal{N}(0,1)$, and your goal is to decide whether or not this claim is true. You observe a single realization this random variable, which comes out to be $X = 3.514$.

Which of the following is the most plausible assessment of the experiment?

- [ ] It is **not** very unlikely for a standard Gaussian random variable to be at least 3.514 (*i.e.*, the event has probability larger than 5%), so you are not able to refute your friend's claim that $X \sim \mathcal{N}(0,1)$.
- [ ] It is **not** very unlikely for a standard Gaussian random variable to be at least 3.514 (*i.e.*, the event has probability larger than 5%), so you can affirm with 100% certainty your friend's claim that $X \sim \mathcal{N}(0,1)$.
- [x] It is very unlikely for a standard Gaussian random variable to be at least 3.514 (*i.e.*, the event has probability less than 0.1%), so if indeed $X \sim \mathcal{N}(0,1)$, then you just observed a very rare event. Intuitively, it seems unlikely that your friend's claim is true.
- [ ] It is very unlikely for a standard Gaussian random variable to be at least 3.514 (*i.e.*, the event has probability less than 0.1%), so you can conclude with 100% certainty that X is **not** distributed like a Gaussian.

✔

Solution:

The third choice is correct. We can compute using computational tools or a table that if $X \sim \mathcal{N}(0,1)$, then

$$P(X > 3.514) = \int_{3.514}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx .00022$$

which is smaller than 0.1%. Indeed this is a very rare event, so based on this heuristic argument, it seems unlikely that your friend's claim is true. We examine the incorrect choices in order:

- The first two choices are both incorrect. As above, $P(X \geq 3.514)$ is much smaller than 5%, so X being larger than the given observation is **not** a likely event.
  Remark: Note how the language between these two choices differs: the first one says "you are not able to refute your friend's claim," and the second says "you can affirm with 100% certainty your friend's claim". The logic of the two statements are very different. For statistical analysis, we almost always stick with the first one.

- The fourth choice is incorrect. While the observation $X \geq 3.514$ would be a rare event given that $X \sim \mathcal{N}(0,1)$, there is still some positive probability (roughly 0.02%) of it happening. Rare events can still occur, so we cannot rule out with 100% certainty that the distribution of X is $\mathcal{N}(0,1)$.

Review: Central Limit Theorem

1/1 point (graded)

Recall the central limit theorem states that if

- $X_1, \ldots, X_n$ are i.i.d.;
- $\mathbb{E}[X_1] = \mu < \infty$, and $\text{Var}(X_1) = \sigma^2 < \infty$,

then a shift and a rescaling of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to a standard Gaussian $\mathcal{N}(0,1)$ in distribution as $n \to \infty$:

$$\sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1).$$

Suppose $\mu = 0$ and $\sigma^2 = 1$. Given this assumption, which of the following limits is **strictly** between 0 and 1?

- $\lim_{n\to\infty} P(\overline{X}_n \in (-1, 1))$
- $\lim_{n\to\infty} P\left(\overline{X}_n \in \left(-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right)\right)$
- $\lim_{n\to\infty} P\left(\overline{X}_n \in \left(-\frac{1}{n}, \frac{1}{n}\right)\right)$

Solution:

Let $Z \sim \mathcal{N}(0,1)$ and let a_n, b_n denote sequences depending on n. By the central limit theorem (CLT),

$$\begin{aligned}\lim_{n\to\infty} P(\overline{X}_n \in (a_n, b_n)) &= \lim_{n\to\infty} P(\sqrt{n}\,\overline{X}_n \in (\sqrt{n}a_n, \sqrt{n}b_n)) \\ &= P(Z \in (\lim_{n\to\infty} \sqrt{n}a_n, \lim_{n\to\infty} \sqrt{n}b_n))\end{aligned}$$

Now let's examine the choices in order.

- $\lim_{n\to\infty} P(\overline{X}_n \in (-1, 1)) = 1$, so this choice is incorrect. Setting $a_n = -1$ and $b_n = 1$, we see that

  $$\lim_{n\to\infty} \sqrt{n}a_n = -\infty, \quad \lim_{n\to\infty} \sqrt{n}b_n = \infty.$$

  Hence, by the above calculation,

  $$\lim_{n\to\infty} P(\overline{X}_n \in (a_n, b_n)) = P(Z \in (-\infty, \infty)) = 1.$$

- $\lim_{n\to\infty} P\left(\overline{X}_n \in \left(-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right)\right)$ lies strictly between 0 and 1, as we will show below. Setting $a_n = -\frac{1}{\sqrt{n}}$ and $b_n = \frac{1}{\sqrt{n}}$, we see that

$$\sqrt{n}a_n = -1, \quad \sqrt{n}b_n = 1.$$

Hence, by the above calculation,

$$\lim_{n\to\infty} P(\overline{X}_n \in (a_n, b_n)) = P(Z \in (-1, 1))$$

Since Gaussian variables have a positive probability of being inside $(-1, 1)$ and also a positive probability of being outside $(-1, 1)$, we can also conclude without doing any computation that $0 < P(Z \in (-1, 1)) < 1$.

Remark: Alternatively we can compute, using computational tools or a table that

$$P(Z \in (-1, 1)) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx 0.6827.$$

- $\lim_{n\to\infty} P\left(\overline{X}_n \in \left(-\frac{1}{n}, \frac{1}{n}\right)\right) = 0$, so this choice is incorrect. Setting $a_n = -\frac{1}{n}$ and $b_n = \frac{1}{n}$, we see that

$$\lim_{n\to\infty} \sqrt{n}a_n = \lim_{n\to\infty} -\frac{1}{\sqrt{n}} = 0, \quad \lim_{n\to\infty} \sqrt{n}b_n = \lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0$$

Hence, by the above calculuation,

$$\lim_{n\to\infty} P(\overline{X}_n \in (a_n, b_n)) = P(Z \in (0, 0)) = 0.$$

Remark: This exercise emphasizes the heuristic interpretation of the CLT which states that the sample mean $\overline{X}_n$ lives inside an interval of radius $Constant \times \frac{1}{\sqrt{n}}$ around its expectation. This heuristic will be useful for designing hypothesis tests.

CLT Concept Check

0/1 point (graded)

In the next few questions, we will flip a coin 200 times in order to try and answer the hypothesis testing **question of interest**:

"**Is this coin fair?**."

As in lecture, we model the i'th flip as X_i where $X_i = 1$ for a heads and $X_i = 0$ for a tails. Since the flips should not interact with each other and we always flip the same coin, we make the familiar modeling assumption $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p)$ where p is an unknown parameter. Then our original question of interest can be rephrased:

"**Does $\mathbf{p} = 0.5$ or does $\mathbf{p} \neq 0.5$?**".

Note that this is a very specific question. In particular, we do not care so much about the particular value of p– we just want to test whether or not it is equal to 0.5.

To answer this question, we consider the statistic

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ denotes the sample mean.

Recall that we do not know the true value of p. Assume that n is very large. Can we conclude that the distribution of $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is very close to the distribution of a standard Gaussian $\mathcal{N}(0,1)$?

Choose the correct response that also has the correct explanation.

- ○ Yes, because $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is a shift and rescaling of a binomial distribution. We know that for n large enough, the binomial distribution $\mathrm{Bin}(n,p)$ provides a good approximation to the distribution of a standard Gaussian $\mathcal{N}(0,1)$.
- ○ Yes, because the central limit theorem (CLT) guarantees that for n sufficiently large, $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \approx \mathcal{N}(0,1)$ (in distribution).
- ◉ No. Since we do not know for sure that $p = 0.5$, we cannot conclude that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. (e.g. If $p = 0.6$, then this estimator will not converge to $\mathcal{N}(0,1)$.)
- ○ No. Even if $p = 0.5$, it is not true that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}(0,1)$ in distribution. Hence, even in the case of a fair coin, we do not expect this estimator be close in distribution to $\mathcal{N}(0,1)$.

Solution:

We examine the choices in order.

- "Yes, because $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is a shift and rescaling of a binomial distribution. We know that for n large enough, the binomial distribution $\text{Bin}\,(n, p)$ provides a good approximation to the distribution of a standard Gaussian $\mathcal{N}\,(0, 1)$." is incorrect. The explanation is wrong: $\text{Bin}\,(n, p)$ does **not** provide a good approximation for the distribution $\mathcal{N}\,(0, 1)$.
  Remark: However, by the CLT, if $X \sim \text{Bin}\,(n, p)$, then for as $n \to \infty$,

$$\sqrt{n}\left(\frac{\frac{X}{n} - p}{\sqrt{p\,(1-p)}}\right) \to \mathcal{N}\,(0, 1)$$

  in distribution.

- "Yes, because the central limit theorem (CLT) guarantees that for n sufficiently large, $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \approx \mathcal{N}\,(0, 1)$ (in distribution)." is incorrect. We can only apply the CLT to the given estimator if the mean is 0.5 and the variance is $0.5\,(1-0.5)$. This is only the case if the coin is fair, *i.e.*, $p = 0.5$.

- "No. Since we do not know for sure that $p = 0.5$, we cannot conclude that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}\,(0, 1)$ in distribution. (e.g. If $p = 0.6$, then this estimator will not converge to $\mathcal{N}\,(0, 1)$.)" is the correct response. We can only apply the CLT to conclude $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}\,(0, 1)$ in distribution if $p = 0.5$, as discussed in the previous bullet.

- "No. Even if $p = 0.5$, it is not true that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}\,(0, 1)$ in distribution. Hence, even in the case of a fair coin, we do not expect this estimator be close in distribution to $\mathcal{N}\,(0, 1)$." is incorrect. Though the answer is "No", the explanation is incorrect: the case where $p = 0.5$ is the **only** situation in which we can apply the CLT to say that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \to \mathcal{N}\,(0, 1)$ in distribution.

In the next two problems, we will illustrate some of the basic steps behind hypothesis testing.

The set up is the same as in the problem above:

Let $X_1, \ldots, X_{200} \stackrel{iid}{\sim} \text{Ber}(p)$, and we are interested in determining from the sample whether or not $p = 0.5$. The hypothesis testing question of interest is then

$$\textbf{Does p = 0.5} \qquad \textbf{or} \qquad \textbf{does p} \neq \textbf{0.5 ?}$$

To answer this question, we introduced the statistic, which is also an estimator:

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}}.$$

The reason for considering this estimator is that, **if $\mathbf{p = 0.5}$, then the CLT applies** (check this!), so that for n very large we may assume

$$\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5\,(1-0.5)}} \approx \mathcal{N}(0,1).$$

In other words, **if $\mathbf{p = 0.5}$**, then the above estimator distributed approximately as a standard Gaussian when n is large enough.

Our strategy will be to evaluate this estimator on the data set. Supposing that $p = 0.5$, then the value of our statistic should resemble the typical value of a single observation of a standard Gaussian random variable. Hence, if the value $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ lies deep in the tails of the standard normal distribution, we would logically conclude that it is **unlikely** that $p = 0.5$. Otherwise, we will not be able to refute that $p = 0.5$.

Hypothesis Testing: A Sample Data Set of Coin Flips I

1/3 points (graded)

We use the statistical set-up from the previous problem. Consider a statistical experiment where you flip the coin 200 times. In one run of this experiment, you observe **80 heads**. We will use this data and the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ (as in the previous problem) to provide an answer to the hypothesis testing **question of interest**: "**Does $\mathbf{p = 0.5}$ or does $\mathbf{p \neq 0.5}$?**".

Let D_1 denote the value of the realization of the statistic $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ on the given data set. (Here $n = 200$, the number of flips.) What is D_1?

$D_1 =$ -2.8284 ✔ **Answer:** -2.82842

Let $Z \sim \mathcal{N}(0,1)$. What is $\mathbf{P}(Z < D_1)$?
(You are welcome to use table or any computational tools e.g. R, or this online normal table calculator.)

$P(Z < D_1) =$ 0.00730744335082954 ✖ **Answer:** 0.00234

Since $n = 200$ is fairly large, we may assume that if $p = 0.5$ that $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}} \sim \mathcal{N}(0,1)$.

Suppose that $p = 0.5$ and you ran the experiment above (consisting of 200 coin flips) a total of 1000 times (i.e. a total 200×1000 coin flips). What is the expected number of experiments such that the estimator $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is smaller than the value D_1 attained in the first experiment? (Round your answer to the nearest integer.)

7 ✖ **Answer:** 2

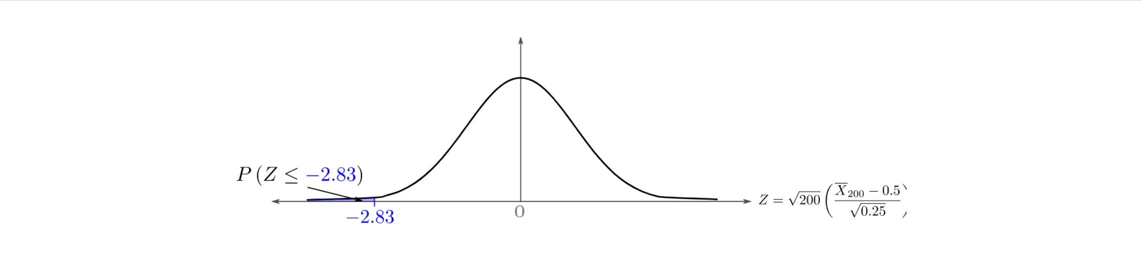
Solution:

First,

$$D_1 = \sqrt{200}\left(\frac{\frac{80}{200} - 0.5}{\sqrt{0.25}}\right) \approx -2.82842.$$

Using a table or computational software, we can also compute that if $Z \sim \mathcal{N}(0, 1)$,

$$P(Z < D_1) = \int_{-\infty}^{-2.82842} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx .00234$$



Hence, for a single experiment, if $p = 0.5$, then there is (approximately) a 0.23% chance of seeing an observation smaller than $D_1 \approx -2.82842$. Thus if we run 1000 experiments, we would expect to see

$$1000 * (.00234) \approx 2.33907$$

experiments where $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is smaller than $D_1 \approx -2.82842$.

Hypothesis Testing: Another Sample Data Set of Coin Flips

1/3 points (graded)

We repeat the above exercise with a different data set.

As above, consider a **statistical experiment** where you flip the coin 200 times. However, in this run of the experiment, you observe **106 heads**. We will use this data and the statistic $\sqrt{n}\frac{(\overline{X}_n-0.5)}{\sqrt{0.5(1-0.5)}}$ from the previous problem to provide an answer to the hypothesis testing question of interest:

"Does $\mathbf{p} = 0.5$ or does $\mathbf{p} \neq 0.5$?"

Let D_2 denote the value of the realization of the estimator $\sqrt{n}\frac{(\overline{X}_n-0.5)}{\sqrt{0.5(1-0.5)}}$ on the given data set. (Here $n = 200$, the number of flips.) What is D_2?

$D_2 =$ 0.8485 ✔ **Answer:** 0.8485

Let $Z \sim \mathcal{N}(0,1)$. What is $\mathbf{P}(Z > D_2)$?
(You are welcome to use any tables or any computational tools e.g. R or this online normal table calculator.)

$\mathbf{P}(Z > D_2) =$ 0.802 ✖ **Answer:** 0.19808

Since $n = 200$ is fairly large, we may assume if $p = 0.5$ that $\sqrt{n}\frac{(\overline{X}_n-0.5)}{\sqrt{0.5(1-0.5)}} \sim \mathcal{N}(0,1)$.

Suppose that $p = 0.5$ and you ran the experiment above (consisting of 200 coin flips) a total of 1000 times. What is the expected number of experiments such that the estimator $\sqrt{n}\frac{(\overline{X}_n-0.5)}{\sqrt{0.5(1-0.5)}}$ is larger than the value D_2 attained in the first experiment? (Round your answer to the nearest integer.)

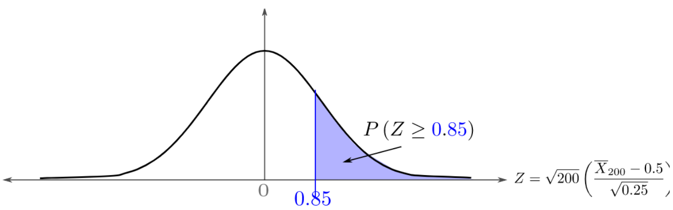
802 ✖ **Answer:** 198

First,

$$D_2 = \sqrt{200}\left(\frac{\frac{106}{200} - 0.5}{\sqrt{0.25}}\right) \approx 0.8485.$$

Using a table or computational software, we can also compute that if $Z \sim \mathcal{N}(0, 1)$,

$$\mathbf{P}(Z > D_2) = \int_{0.8485}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \approx 0.19808.$$



Hence, for a single experiment, if $p = 0.5$, then there is (approximately) a 19.8% chance of seeing an observation larger than $D \approx 0.8485$. Thus if we run 1000 experiments, we would expect to see

$$1000 * (0.19808) \approx 198.08$$

experiments where $\sqrt{n}\frac{\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5(1-0.5)}}$ is larger than $D_2 \approx 0.8485$.

Remark 1: By the previous result, from a heuristic perspective, we would be unable to refute the hypothesis that $p = 0.5$ (Note that this is a **different** conclusion than saying "We may conclude that $p = 0.5$"). Indeed, if $p = 0.5$, observing a value larger than $D_2 \approx 0.8485$ would be **not** be a rare event, intuitively speaking. In practice, one has to set the threshold of what determines a "rare" event, and this will be studied later in this lecture.

Remark 2: Though we are considering a very specific example and applying a very specific test, the steps taken in this problem and the previous one are illustrative of the general principles of hypothesis testing. In general, we will transform our data into a given statistic whose distribution we know well that does **not** depend on the true parameter (e.g., as in this problem, the standard Gaussian). Such a distribution is known as **pivotal** . Then we can reduce our hypothesis testing question to a problem of deciding whether or not a given observation is likely (or not) for this pivotal distribution.

Properties of the Null and Alternative Hypothesis

1/1 point (graded)

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} P_{\theta^*}$ for some unknown parameter θ^*. The associated statistical model is $(E, \{P_\theta\}_{\theta \in \Theta})$.

Which of the following are true statements regarding the **null hypothesis** $H_0 : \theta^* \in \Theta_0$ and the **alternative hypothesis** $H_1 : \theta^* \in \Theta_1$? (Choose all that apply.)

- [x] Θ_0 and Θ_1 must be subsets of the parameter space Θ.
- [x] Θ_0 and Θ_1 must be disjoint, i.e. $\Theta_0 \cap \Theta_1 = \emptyset$.
- [ ] Θ_0 and Θ_1 must make up all of the parameter space Θ, i.e. $\Theta_0 \cup \Theta_1 = \Theta$.
- [ ] The null and alternative hypotheses play symmetric roles: *i.e.* if we set $H_0' : \theta \in \Theta_1$ and $H_1' : \theta \in \Theta_0$ then we are still doing the exact same hypothesis test as in the problem statement.

✔

Solution:

We examine the answer choices in order.

- "Θ_0 and Θ_1 must be subsets of the parameter space Θ." is correct. We are trying to decide if θ^* is in a particular region of the parameter space, so Θ_0 and Θ_1 must be subsets of Θ.

- "Θ_0 and Θ_1 must be disjoint, i.e. $\Theta_0 \cap \Theta_1 = \emptyset$" is correct. For hypothesis testing, we are trying to determine, based on observations, whether or not it is likely that $\theta^* \in \Theta_0$. Since we only want to test whether or not the parameter lies in some region (or not), it makes sense to impose that the region Θ_0 determined by the null hypothesis and Θ_1, the region determined by the alternative hypothesis are disjoint.

- "Θ_0 and Θ_1 must make up the entire parameter space Θ" is **not** correct. For example, recall in the two sided test comparing the boarding times of the rear-to-front and the WILMA from the beginning of this lecture, the parameter space is $\Theta = \{(\mu_1, \mu_2) : \mu_1 \in \mathbb{R},\ \mu_2 \in \mathbb{R}\} = \mathbb{R}^2$, while the null and alternative hypotheses are given by

$$\begin{aligned} \Theta_0 &= \{(\mu_1, \mu_2) : \mu_1 = \mu_2\} && \text{a line in } \mathbb{R}^2 \\ \Theta_1 &= \{(\mu_1, \mu_2) : \mu_1 > \mu_2\} && \text{the region on one side of the line } \mu_1 = \mu_2 \text{ in } \mathbb{R}^2, \end{aligned}$$

  The region $\Theta_0 \cup \Theta_1$ does not make up the all of $\mathbb{R}^2$ because the region on the other side of the line $\mu_1 = \mu_2$ is not included. A simpler example is when Θ_0 and Θ_1 both consists of single values of θ when the parameter space Θ is for example $\mathbb{R}^2$.

- "The null and alternative hypotheses play symmetric roles: *i.e.* if we set $H_0' : \theta \in \Theta_1$ and $H_1' : \theta \in \Theta_0$ then we are still doing the exact same hypothesis test as in the problem statement." is incorrect. Actually, H_0 and H_1 play asymmetric roles. Our only goal in hypothesis testing is to use the data to determine whether or not we can **reject** H_0. This is a different statistical objective than using the data to determine whether or not we can reject H_0'.

  Remark: Regardless of the data, our conclusion will never be to *accept* the null. On observing the data, we will either **reject** the null in favor of the alternative OR we will **fail to reject** the null. In the latter case, we are not claiming that the null is true, rather we are stating that the data does not provide us with enough evidence to refute the null hypothesis.

Formulating the Null and Alternative Hypothesis: Is This Coin Fair?

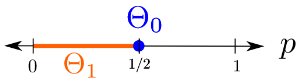
1/1 point (graded)

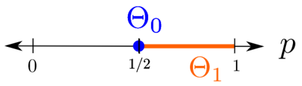
Refer back to the statistical experiment where you flip the coin 200 times in order to answer the question of interest: "**Is this coin fair?**."

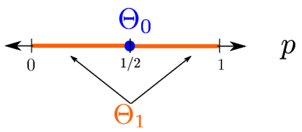
You take as the **status quo** that the coin is fair. Hence, the data must show strong evidence to the contrary in order for this status quo to be rejected. In other words, the coin is considerd fair until "proven" otherwise.

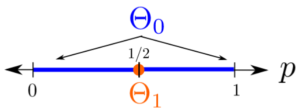
You model the coin flips by $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Ber}(p)$ where p is an unknown parameter and rephrase the question as: "**Does p $= 0.5$ or does p $\neq 0.5$?**".

Formulate the null and alternate hypothesis for this test. Which of the following depicts Θ_0 and Θ_1?









Solution:

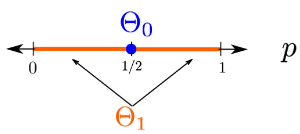
In this test, we are looking for evidence in the data to show that the coin is **not** fair. Hence, the null hypothesis is

$$H_0 : p \in \Theta_0 = \{1/2\},$$

and the alternate hypothesis is

$$H_1 : p \in \Theta_1 = (0, 1/2) \cup (1/2, 1)\,,$$

depicted by the figure



Remark: This is called a **two sided test** since Θ_1 lies on both sides of Θ_0.

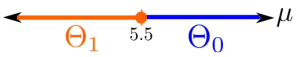
As in a previous problem, you try to answer the question "**Were people in the U.S. taller in 2018 than in 1920, when the average height was 5.5 feet?**" by sampling 1 million individuals labeled $1, 2, \ldots, 10^6$ randomly from the 2018 U.S. population.

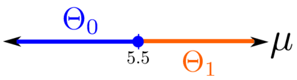
You take as the **status quo** that the people are **not** taller in 2018, and look for evidence in the data to reject this status quo. In other words, you assume people are **not** taller in 2018, until "proven" otherwise.

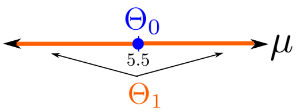
You model the height of the i-th individual as a random variable X_i and make the assumption, based on 1920 data, that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1.3)$ where μ is an unknown parameter.

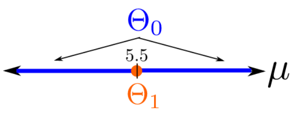
You rephrase the question of interest as: **Is $\mu > 5.5$? or is $\mu \leq 5.5$?**

Formulate the null and alternate hypothesis for this test. Which of the following depicts Θ_0 and Θ_1?









Solution:

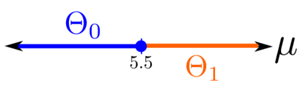
In this test, we are looking for evidence in the data to show that $\mu > 5.5$. Hence, the null hypothesis is

$$H_0 : \mu \in \Theta_0 = (-\infty, 5.5],$$

and the alternate hypothesis is

$$H_1 : p \in \Theta_1 = (5.5, \infty),$$

depicted by the figure



Remark: This is called a **one sided test** since Θ_1 lies on only one side of Θ_0.

Null and Alternative Hypotheses for testing Drug Effect

1/1 point (graded)

As in the two-sample test in the first example in this lecture sequence, you are running a clinical trial to determine the effectiveness of a drug for treating an illness. You administer the drug to a **treatment group** and give a placebo to the **control group**.

However, in this problem, you will consider a different data set: at the end of the trial, you survey all participants with the yes or no question: "Did you recover from this illness?"

You model a "Yes" response as 1 and a "No" response as 0. Thus, we can model:

- the treatment group's responses as $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p_{\text{drug}})$;
- the control group's responses as $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Ber}(p_{\text{control}})$.

Your goal is to use this data to answer the **question of interest**:
"**Is this drug effective in treating the illness?**"

It is standard practice in a clinical trial to take as the **status quo** (which represents a prior assumption) that the drug is no more effective than placebo. Hence, the data must show strong evidence to the contrary in order for this status quo to be rejected. Note that the placebo is considered to have *no* effect (*i.e.*, it is not possible for the drug to be *less* effective than the placebo).

Given this standard, that the drug is considered to be no more effective than the placebo until "proven" otherwise, how should the **null hypothesis** H_0 and **alternative hypothesis** H_1 be defined?

- ( ) $H_0 : p_{\text{drug}} > p_{\text{control}}, \; H_1 : p_{\text{drug}} = p_{\text{control}}$
- ( ) $H_0 : p_{\text{drug}} \leq p_{\text{control}}, \; H_1 : p_{\text{drug}} \leq p_{\text{control}}$
- (•) $H_0 : p_{\text{drug}} = p_{\text{control}}, \; H_1 : p_{\text{drug}} > p_{\text{control}}$
- ( ) $H_0 : p_{\text{drug}} = p_{\text{control}}, \; H_1 : p_{\text{drug}} < p_{\text{control}}$

✔

Solution:

We examine the choices in order.

- The choice "$H_0 : p_{\text{drug}} > p_{\text{control}}, H_1 : p_{\text{drug}} = p_{\text{control}}$" is incorrect because it does not align with the status quo. Namely, we do **not** take as a prior assumption that the drug is more effective than placebo.

  Remark: From a practical standpoint, it makes sense to be skeptical and assume the status quo $p_{\text{drug}} = p_{\text{control}}$ because this will make it **harder** for scams or ineffective drugs to make it through clinical trials. Concretely, it seems like a bad idea to allow a drug to pass trial which does not show strong evidence of being an effective treatment.

- The choice $H_0 : p_{\text{drug}} \leq p_{\text{control}}, H_1 : p_{\text{drug}} \leq p_{\text{control}}$ is incorrect because the regions defined by H_0 and H_1 are not disjoint.

- The correct choice is $H_0 : p_{\text{drug}} = p_{\text{control}}, H_1 : p_{\text{drug}} > p_{\text{control}}$. In general, the status quo should be taken to be the **null hypothesis**. Since the status quo is that the drug is no more effective than the placebo and we have stated that it is not possible for the drug to be *less* effective than the placebo, the hypothesis $p_{\text{drug}} \leq p_{\text{control}}$ captures our prior assumptions. Moreover, to reject the null hypothesis, we would need to use the data to show that our observations are very unlikely under the assumption $p_{\text{drug}} = p_{\text{control}}$. In this situation, we would deem that $p_{\text{drug}} > p_{\text{control}}$ is more likely, so moreover the **alternative hypothesis should be** $p_{\text{drug}} > p_{\text{control}}$.

- The choice $H_0 : p_{\text{drug}} = p_{\text{control}}, H_1 : p_{\text{drug}} < p_{\text{control}}$ is incorrect. While the null hypothesis is consistent with our prior assumptions, it is not possible for the drug to be less effective than the placebo. Thus the alternative hypothesis is incorrectly stated.

Identify Null and Alternative Hypotheses Regions for a Two Sample Test

1 point possible (graded)

As above, you are testing for whether a drug is effective using data comprising of yes or no responses from both the treatment and control groups to the question “did you recover from the illness at the end of this clinial trial?"

As above, you model

- the treatment group's responses as $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}\,(p_{\mathrm{drug}})$;
- the control group's responses as $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{Ber}\,(p_{\mathrm{control}})$.

where $X_i = 1$ means the response is “Yes", and $X_i = 0$ mean “No", and similarly for Y_i. You assume the two sets of responses are independent of one another.

The statistical model for the example in the drug testing is $\left(\{0,1\}^2, \{P_{(p_{\mathrm{control}}, p_{\mathrm{drug}})}\}_{(p_{\mathrm{control}}, p_{\mathrm{drug}}) \in (0,1)^2}\right)$.

In the problem above, you formulated the null and alternative hypotheses. Which of the following depicts the regions Θ_0 (corresponding to the null hypothesis) and Θ_1 (corresponding to the alternative hypothesis)?

Solution:

First, the parameter space is

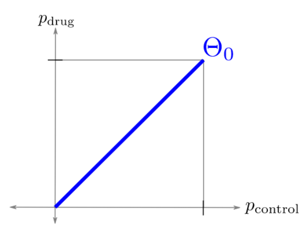
$$\Theta = \{(p_{\mathrm{control}}, p_{\mathrm{drug}}) : p_{\mathrm{control}} \in (0,1)\,, p_{\mathrm{drug}} \in (0,1)\} = (0,1)^2.$$

Since $\Theta_0, \Theta_1 \subset \Theta$, only the figures in which the shaded regions Θ_0 and Θ_1 are within the unit square can be correct.

The null hypothesis is $H_0 : p_{\mathrm{drug}} = p_{\mathrm{control}}$, hence

$$\Theta_0 = \{(p_{\mathrm{control}}, p_{\mathrm{drug}}) \in (0,1)^2 : p_{\mathrm{drug}} = p_{\mathrm{control}}\}$$

which defines the diagonal line in the unit square:

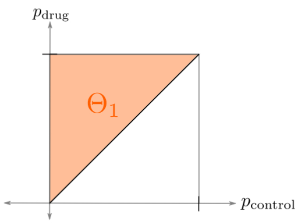


The alternative hypothesis is $H_1 : p_{\mathrm{drug}} > p_{\mathrm{control}}$, hence

The alternative hypothesis is $H_1 : p_{\text{drug}} > p_{\text{control}}$, hence

$$\Theta_1 = \{(p_{\text{control}}, p_{\text{drug}}) \in (0,1)^2 : p_{\text{drug}} > p_{\text{control}}\}$$

which defines the region above the diagonal line in the unit square:



Which Statistics are Tests?

1/1 point (graded)

Recall that a **statistic** is, intuitively speaking, a function that can be computed from the data.

A **(statistical) test** is an **statistic** whose output is **always** either 0 or 1, and like an estimator, does not depend explicitly on the value of true unknown parameter.

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Ber}(\theta)$ for some unknown parameter $\theta \in (0,1)$. Which of the following statistics are also tests?

(Recall that $\mathbf{1}(A)$ is the indicator defined as follows: $\mathbf{1}(A) = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$.

(Choose all that apply.)

- [ ] $\overline{X}_n$
- [x] $\mathbf{1}(\overline{X}_n > 0.5)$
- [x] $\mathbf{1}(|\overline{X}_n - 0.5| > 0.01)$
- [ ] $\mathbf{1}(|\overline{X}_n - \theta| > 0.5)$
- [x] $\mathbf{1}(\overline{X}_n \text{ is a rational number})$

✔

Solution:

We examine the choices in order.

- $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ is **not** a statistical test, because the sample average is not **always** either 0 or 1.
- $\mathbf{1}(\overline{X}_n > 0.5)$ is a statistical test. Its expression only depends on the sample (and not the true parameter), and since it is an indicator, it takes values only in $\{0, 1\}$.
- $\mathbf{1}(|\overline{X}_n - 0.5| > 0.01)$ is a statistical test. Its expression only depends on the sample (and not the true parameter), and since it is an indicator, it takes values only in $\{0, 1\}$.
- $\mathbf{1}(|\overline{X}_n - \theta| > 0.5)$ is **not** a statistical test because its output depends on the unknown parameter θ.
- $\mathbf{1}(\overline{X}_n \text{is a rational number})$ is a statistical test. Its expression only depends on the sample (and not the true parameter), and since it is an indicator, it takes values only in $\{0, 1\}$. This is a rather bizarre test, but it does satisfy all required properties.

Applying a Statistical Test on a Data Set

1/1 point (graded)

Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$ where μ is an unknown parameter. You are interested in answering the **question of interest**: "**Does $\mu = 0$?**". To do so you construct the **null hypothesis** $H_0 : \mu = 0$ and the **alternative hypothesis** $H_1 : \mu \neq 0$.

You design the test

$$\psi = \mathbf{1}\left(\sqrt{n}\left|\overline{X}_n\right| > 0.25\right).$$

If $\psi = 1$, you will **reject** the null hypothesis, and if $\psi = 0$, you will **fail to reject**. For simplicity, we will set the sample size to be $n = 7$.

On which of the following data sets would you reject the null hypothesis?
(Choose all that apply. Feel free to use computational tools.)

- [x] $-1.0, -0.8, -2.9, 1.4, 0.3, -0.8, 1.4$
- [x] $-1.7, -0.1, -0.2, 0.3, 0.3, -0.9, -0.03$
- [ ] $-0.2, 0.6, 1.1, -0.9, 0.1, -1.2, 1.1$

✔

Solution:

We examine the choices in order.

- The first choice is correct. For this data set, we compute $\sqrt{7}\,\overline{X}_7 \approx -0.9072$. Since $|-0.9072| > 0.25$, we reject.
- The second choice is correct. For this data set, we compute $\sqrt{7}\,\overline{X}_7 \approx -0.8768$. Since $|-0.8768| > 0.25$, we reject.
- The third choice is incorrect. For this data set, we compute $\sqrt{7}\,\overline{X}_7 \approx 0.2267$. Since $|0.2267| \leq 0.25$, we fail to reject.

Remark 1: It is useful to keep in mind the following mnemonic,

$$\begin{aligned} \psi = 0 &\Rightarrow H_0 \\ \psi = 1 &\Rightarrow H_1. \end{aligned}$$

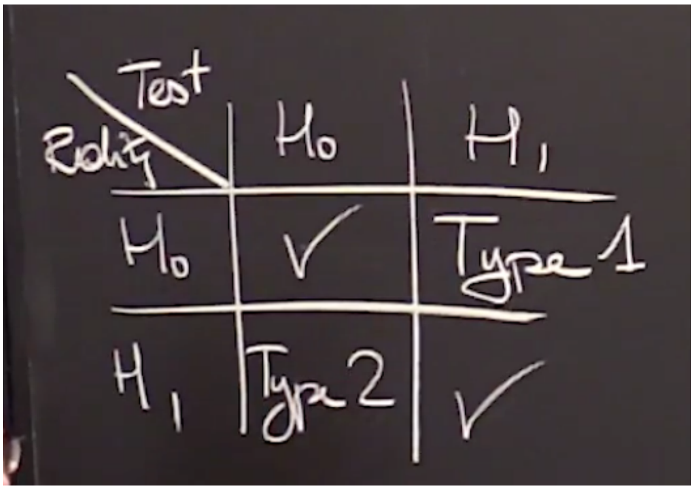
Of course, the implications above are informal and should not be taken literally. To be precise, we say that if $\psi = 0$, we fail to reject H_0, and if $\psi = 1$, then we reject H_0 in favor of H_1.

Remark 2: If we assume the null hypothesis $H_0 : \mu = 0$, then since the variance is known to be 1, the CLT guarantees that

$$\sqrt{n}\,\overline{X}_n \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, 1).$$

The quantiles of $\mathcal{N}(0, 1)$ can be understood using tables or computational software, so if n is very large, then we can approximate the probability of our test ψ **rejecting** or **failing to reject** under the null hypothesis. This concept will be further explored in the next page where we explore the "type 1" and "type 2 error" of a test.

slide 52, types of errors



An Analogy to the U.S. Justice System:Type 1 and Type 2 Errors

3/3 points (graded)

In a criminal court in the U.S., the goal is to decide between the following null and alternative hypotheses:

$$\begin{aligned} H_0 &: \text{The defendant is innocent.} \\ H_1 &: \text{The defendant is guilty.} \end{aligned}$$

In the U.S. criminal justice system, the informal principle “innocent until proven guilty" is the status quo, so this is the rationale for the choice of null hypothesis above. While this example is not, strictly speaking, a statistical hypothesis test, it provides some intuition about the meaning of type 1 and type 2 errors.

Suppose we have a defendant X who will be tried by a jury in the U.S. If guilty, X will go to jail, and otherwise is free to go.

In this example, let's say that the jury makes a **type 1 error** if the suspect satisfies H_0 while the jury rules in favor of H_1. Let's say the jury makes a **type 2 error** if the suspect satisfies H_1 while the jury rules in favor of H_0.

If the jury commits a type 1 error, the defendant is...

- ○ Innocent in reality, and will walk away free.
- ○ Guilty in reality, and will go to jail.
- ◉ Innocent in reality, but still will go to jail.
- ○ Guilty in reality, but will walk away free.

✔

If the jury commits a type 2 error, the defendant is...

- ◯ Innocent in reality, and will walk away free.
- ◯ Guilty in reality, and will go to jail.
- ◯ Innocent in reality, but still will go to jail.
- ◉ Guilty in reality, but will walk away free.

✔

What strategy could the jurors follow if they wanted to never commit a type 2 error?

- ◯ Always acquit– *i.e.*, always decide that the defendant is innocent.
- ◉ Always convict– *i.e.*, always decide that the defendant is guilty.

✔

Solution:

Let's examine the questions in order.

1. Since the null hypothesis is that X is innocent, in a type 1 error, the jury will convict X even though the defendant is innocent. Hence, the correct choice is that the defendant is "Innocent in reality, but still will go to jail."

2. Similary, since the alternative hypothesis is that X is guilty, in a type 2 error, the jury deems that X is innocent even though the defendant committed the crime. Hence the correct choice is that the defendant is "Guilty in reality, but will walk away free."

3. The correct response is "Always convict." If the jury always convicts, then there will never be a case where a guilty defendant walks away free: this strategy minimizes the type 2 error. However, it also maximizes the type 1 error. Every defendant who is innocent will be convicted, so practically speaking, this is a very questionable strategy.

The Threshold for a Statistical Test

1/1 point (graded)

Continuing from problem on the previous page, let $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\mu, 1)$ where μ is an unknown parameter. You are interested in answering the **question of interest**: "**Does** $\mu = 0$ **?**".

To do so, you construct

- the **null hypothesis** $H_0 : \mu = 0$;
- the **alternative hypothesis** $H_1 : \mu \neq 0$.

Motivated by the central limit theorem, you decide to use a test of the form

$$\psi_C = \mathbf{1}\left(\sqrt{n}\,|\overline{X}_n| > C\right)$$

where $C > 0$ is a constant known as the **threshold** that you will choose in designing the test. (In the previous problem, C was chosen to be 0.25.) On observing the data set, if $\psi = 1,$ you will **reject** H_0. If $\psi = 0,$ then you will **fail to reject** H_0.

Suppose that indeed $\mu = 0$. Then $\mathbf{P}(\psi_C = 1)$, the probability of rejecting H_0, quantifies how likely we are to make the error of rejecting H_0 even though H_0 holds.

Under the assumption that $H_0 : \mu = 0$, for which value of C is $\mathbf{P}(\psi_C = 1)$ likely the largest?

- (selected) $C = 0.01$
- $C = 0.1$
- $C = 0.5$
- $C = 1.0$

Solution:

The probability $\mathbf{P}(\mathbf{1}(|\overline{X}_n| > 0.01)$ is the largest.

Consider the events A_1, A_2, A_3, A_4 defined by

$$\begin{aligned} A_1 &: |\overline{X}_n| > 0.01, \quad A_2 : |\overline{X}_n| > 0.1 \\ A_3 &: |\overline{X}_n| > 0.5, \quad A_4 : |\overline{X}_n| > 1. \end{aligned}$$

Observe that $A_4 \subset A_3 \subset A_2 \subset A_1$, hence, by basic probability, $P(A_4) \leq P(A_3) \leq P(A_2) \leq P(A_1)$. Indeed, A_1 has the highest probability, so $P(\psi_1 = 1) = P(\mathbf{1}(|\overline{X}_n| > 0.01) = 1)$ is the largest out of $\psi_1, \ldots, \psi_4$. Thus, the test where $C = 0.01$ has the highest probability of rejection.
Remark: We did not need to know the shape of the distribution of X_n to make this conclusion; so in particular, we did not rely on the CLT.

Compute the Type 1 Error

0/1 point (graded)

As above, let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, 1)$ where μ is an unknown parameter. You are interested in answering the **question of interest**: "**Does $\mu = 0$?**".

To do so, you construct

- the **null hypothesis** $H_0 : \mu = 0$;
- the **alternative hypothesis** $H_1 : \mu \neq 0$.

Motivated by the central limit theorem, you decide to use a test of the form

$$\psi_C = \mathbf{1}\left(\sqrt{n}\,|\overline{X}_n| > C\right).$$

Recall from lecture that the **type 1 error** (also known as **type 1 error rate**) of a test ψ is the **function**

$$\begin{aligned} \alpha_\psi : \Theta_0 &\to [0,1] \\ \theta &\mapsto \mathbf{P}_\theta(\psi = 1) \end{aligned}$$

If you choose the threshold $C = q_{0.05}$, what is the type 1 error α_ψ?
(In this case, since H_0 only consists of one point, the function α_ψ is defined only at one point, and we loosely use the termininology "type 1 error" to mean the value of α_ψ at that point.)

Type 1 Error α_ψ : 1 ✖ **Answer:** 0.1

Solution:

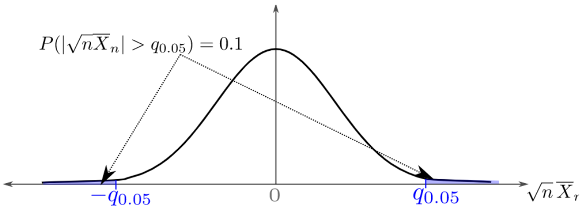
If we assume the null hypothesis $H_0 : \mu = 0$, and since the variance is known to be 1, the CLT gives

$$\sqrt{n}\,\overline{X}_n \sim \mathcal{N}(0,1) \qquad \text{for large } n.$$

The probability of a type 1 error is

$$\alpha_\psi(0) = \mathbf{P}_0(\psi_C = 1) = \mathbf{P}_0\left(\sqrt{n}\,|\overline{X}_n| > q_{0.05}\right) = 0.1.$$

as depicted in the figure below:



If H_0 is true, i.e. $\mu = 0$, then $\sqrt{n}\,\overline{X}_n$ is asymptotically normal. Hence, the total area of the two shaded regions is $\mathbf{P}_0(\psi_C = 1) = \mathbf{P}_0\left(\sqrt{n}\,|\overline{X}_n| > q_{0.05}\right)$, the probability that H_0 is rejected even though it is true.

14. Example: a Non-Asymptotic Test for the Support of a Uniform Variable

Bookmark this page

Testing the Support of a Uniform Variable: Designing a Test

4/4 points (graded)

The next few problems cover a test that is not motivated by the CLT.

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Unif}\,[0, \theta]$ where θ is an unknown parameter. Let $\left(\mathbb{R}_{\geq 0}, \{\text{Unif}\,[0, \theta]\}_{\theta > 0}\right)$ denote the associated statistical model. (Here, $\mathbb{R}_{\geq 0}$ denotes the nonnegative real numbers.)

You want to answer the **question of interest**: **"Is $\theta \leq 1/2$?"**. To do so you formulate a hypothesis test with

$$\begin{aligned} H_0 &: \theta \leq 1/2 \qquad \text{(null hypothesis)} \\ H_1 &: \theta > 1/2 \qquad \text{(alternative hypothesis) .} \end{aligned}$$

You also design the **test**

$$\psi_n = \mathbf{1}(\max_{1 \leq i \leq n} X_i > 1/2).$$

(If $\psi_n = 1$, then we will **reject** the null hypothesis. Note the dependence of ψ_n on the sample size.)

We use Θ_0 to denote the region of Θ defined by the null hypothesis. In this example, Θ_0 can be written as an interval $(A, B]$. What are the numbers A and B?

$A =$ 0 ✔

$B =$ .5 ✔

Similarly, we let Θ_1 denote the region of Θ defined by the alternative hypothesis. In this example, Θ_1 can be written as an interval (C, ∞). What is the number C?

$C =$ 0.5 ✔ **Answer:** 1/2

Suppose you observe the sample

$$0.1, 0.53, 0.002, 0.1234, 0.24, 0.48.$$

Should you **reject** or **fail to reject** the null hypothesis with the above test for this data?

- (●) Reject
- ( ) Fail to reject

✔

Solution:

The parameter space is $\Theta = \{\theta : \theta > 0\}$. Since the null hypothesis is $H_0 : \theta \leq 1/2$, then $\Theta_0 = (0, 1/2]$. Similarly, $\Theta_1 = (1/2, \infty)$.

On observing the sample

$$0.1, 0.53, 0.002, 0.1234, 0.24, 0.48,$$

the null hypothesis $H_0 : \theta \leq 1/2$ should be rejected. Recall that the test is $\psi_n = \mathbf{1}(\max_{1 \leq i \leq n} X_i > 1/2)$ which evaluates to 1 on the given sample. (Here $n = 6$.)

Testing the Support of a Uniform Variable: Complement of the Rejection Region of a Test

3/3 points (graded)

As above, let $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Unif}\,[0, \theta]$ for an unknown parameter θ, and recall that we designed the statistical test

$$\psi_n = \mathbf{1}(\max_{1 \le i \le n} X_i > 1/2)$$

to decide between the null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta \ \le 1/2 \\ H_1 &: \theta \ > 1/2. \end{aligned}$$

(Going forward we will simply write the null and alternative hypotheses and omit the motivating yes/no question.)

Recall from lecture that the **rejection region** for a test ψ_n is

$$R_{\psi_n} := \{(x_1, \ldots, x_n) \in E^n : \ \psi_n(x_1, \ldots, x_n) = 1\}$$

where E is the sample space of the i.i.d. variables X_i, which is $\mathbb{R}_{\ge 0}$ in this example since X_i are uniform random variables.

Consider the complement C_n of the rejection region: this is all the points in $(\mathbb{R}_{\ge 0})^n$ that do not lie in R_{ψ_n}. Note that the dimension of C_n is determined by the sample size n.

What is the length of C_1?

1/2 ✔ **Answer:** 1/2

What is the area of C_2?

1/4 ✔ **Answer:** 1/4

What is the volume of C_3?

1/8 ✔ **Answer:** 1/8

Solution:

The complement C_n of the rejection region is the set of all $(x_1, \ldots, x_n) \in \mathbb{R}^n_{\ge 0}$ such that $\max_{1 \le i \le n} x_i \le 1/2$. (Equivalently, it is the set of all $(x_1, \ldots, x_n)$ such that $\psi_n = \mathbf{1}(\max_{1 \le i \le n} x_i > 1/2) = 0$). The region defined by the constraint $x_i \le 1/2$ for all $1 \le i \le n$ is the set $[0, 1/2]^n$.

In one dimension, this is the interval $[0, 1/2]$ which has length $1/2$. In two dimensions, this is the square $[0, 1/2] \times [0, 1/2]$, which has area $(1/2)^2 = 1/4$. Finally in three dimensions, C_3 is a cube $[0, 1/2] \times [0, 1/2] \times [0, 1/2]$, which has volume $(1/2)^3 = 1/8$.

Testing the Support of a Uniform Variable: Type 1 Error of a Test

0/1 point (graded)

As above, let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Unif}[0, \theta]$ for an unknown parameter θ, and recall that we designed the statistical test

$$\psi_n = \mathbf{1}(\max_{1 \le i \le n} X_i > 1/2)$$

to decide between the null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta \le 1/2 \\ H_1 &: \theta > 1/2. \end{aligned}$$

The region defined by the null hypothesis is $\Theta_0 = (0, 1/2]$. Therefore, the **type 1 error (or error rate)** of the test ψ_n is the **function**

$$\begin{aligned} \alpha_{\psi_n} : (0, 1/2] &\to \mathbb{R} \\ \theta &\mapsto P_\theta(\psi_n = 1) \end{aligned}$$

where $P_\theta = \mathrm{Unif}[0, \theta]$, and $P_\theta(\psi_n = 1)$ is the probability of the event $\{\psi_n = 1\}$ under the probability distribution P_θ when $\theta \in \Theta_0$, i.e. the probability of rejecting H_0 when H_0 is true.

What is $\alpha_{\psi_n}(\theta)$?

$\alpha_{\psi_n}(\theta) =$ theta/4 ✖ **Answer:** 0

$\frac{\theta}{4}$

Solution:

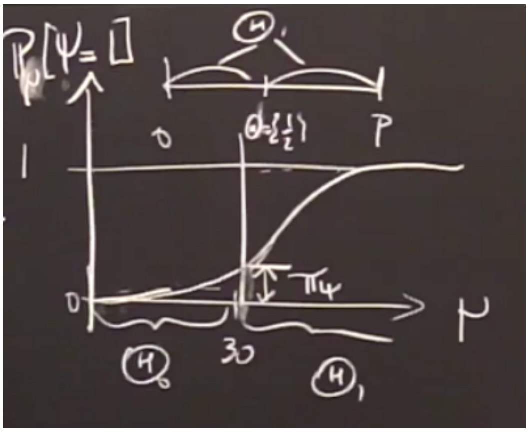
By definition,

$$\alpha_{\psi_n}(\theta) = P_\theta(\max_{1 \le i \le n} X_i > 1/2)$$

where $P_\theta = \mathrm{Unif}[0, \theta]$ and we restrict $\theta \in \Theta_0 = \{\theta : \theta \le 1/2\}$. Observe that if $\theta \le 1/2$, then there is a 0% chance of generating an observation which is larger than $1/2$. Hence, the type 1 error $\alpha_{\psi_n}(\theta)$ is 0 for all $\theta \in \Theta_0$.

Remark: In general, the type 1 error will be a function of θ, but in this special case it is constant.

slide 52



As on the previous page, let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,[0, \theta]$ for an unknown parameter θ and we designed the statistical test

$$\psi_n = \mathbf{1}(\max_{1 \leq i \leq n} X_i > 1/2)$$

to decide between the null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta \leq 1/2 \\ H_1 &: \theta > 1/2. \end{aligned}$$

Recall from lecture that the **type 2 error (rate)** of a test ψ_n is the **function**

$$\begin{aligned} \beta_{\psi_n} : \Theta_1 &\to \mathbb{R} \\ \theta &\mapsto \mathbf{P}_\theta\,(\psi_n = 0) \end{aligned}$$

where $\mathbf{P}_\theta\,(\psi_n = 0)$ is the probability of the event $\psi_n = 0$ under the probability distribution $\mathbf{P}_\theta$ when $\theta \in \Theta_1$, i.e. the probability of not rejecting H_0 when H_1 is true. In this example, the region Θ_1 defining the alternative hypothesis is $(1/2, \infty)$, and $\mathbf{P}_\theta = \text{Unif}\,[0, \theta]$.

Evaluate $\mathbf{P}_\theta\,(\psi_n = 0) = \mathbf{P}_\theta\left(\max_{1 \leq i \leq n} X_i \leq 1/2\right)$ at $\theta = 1/2$, the boundary between Θ_0 and Θ_1.

$\mathbf{P}_{\theta=1/2}\left(\max_{1 \leq i \leq n} X_i \leq 1/2\right) =$ 1 ✔ **Answer:** 1

Solution:

$$\begin{aligned} \beta_{\psi_n}\,(1/2) &= \mathbf{P}_{1/2}(\max_{1 \leq i \leq n} X_i < 1/2) \\ &= \mathbf{P}_{1/2}\,(X_1 < 1/2) \ldots \mathbf{P}_{1/2}\,(X_n < 1/2) \\ &= 1 \times 1 \ldots \times 1 \; = \; 1 \end{aligned}$$

where we applied independence of the X_i's in the second line.

Testing the Support of a Uniform Variable: Type 2 Error of a Test Continued

2/3 points (graded)

As above, let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,[0, \theta]$ for an unknown parameter θ and we designed the statistical test

$$\psi_n = \mathbf{1}(\max_{1 \leq i \leq n} X_i > 1/2)$$

to decide between the null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta \leq 1/2 \\ H_1 &: \theta > 1/2. \end{aligned}$$

Recall from lecture that the **type 2 error** of a test ψ_n is the **function**

$$\begin{aligned} \beta_{\psi_n} : \Theta_1 &\to [0, 1] \\ \theta &\mapsto \mathbf{P}_\theta\,(\psi_n = 0) \end{aligned}$$

where $\mathbf{P}_\theta\,(\psi_n = 0)$ is the probability of the event $\psi_n = 0$ under the probability distribution $\mathbf{P}_\theta$ when $\theta \in \Theta_1$, i.e. the probability of not rejecting H_0 when H_1 is true.

In this example, $\Theta_1 = (1/2, \infty)$, and $\mathbf{P}_\theta = \text{Unif}\,[0, \theta]$.

What is $\beta_{\psi_n}(\theta)$?

$\beta_{\psi_n}(\theta) =$ theta - 1/2 ✖

$\theta - \frac{1}{2}$

Find $\lim_{\theta \to 1/2} \beta_{\psi_n}(\theta)$.

$\lim_{\theta \to 1/2} \beta_{\psi_n}(\theta) =$ 1 ✔

1

Find $\lim_{\theta \to \infty} \beta_{\psi_n}(\theta)$.

$\lim_{\theta \to \infty} \beta_{\psi_n}(\theta) =$ 0 ✔

0

Solution:

For any $\theta \in \Theta_1 = [1/2, \infty)$,

$$\begin{aligned}\beta_{\psi_n}(\theta) = \mathbf{P}_\theta(\psi_n = 0) &= \mathbf{P}_\theta\left(\max_{1 \le i \le n} X_i < 1/2\right) \\ &= \mathbf{P}_\theta(X_1 < 1/2) \dots \mathbf{P}_\theta(X_n < 1/2) = \left(\frac{1/2}{\theta}\right)^n.\end{aligned}$$

As $\theta \to 1/2$,

$$\beta_{\psi_n}(\theta) \to \left(\frac{1/2}{1/2}\right)^n = 1.$$

As $\theta \to \infty$,

$$\beta_{\psi_n}(\theta) = \left(\frac{1/2}{\theta}\right)^n \to 0.$$

Remark: This test is rather extreme example in that it minimizes type-1 error while maximizing the type-2 error. In general, we want to design tests so that the type-1 and type-2 error are both controlled. These types of trade-offs are crucial to consider in the context of hypothesis testing.

Testing the Support of a Uniform Variable: : Power of a Test

1/1 point (graded)

The **power** of the test ψ_n is defined to be

$$\pi_{\psi_n} = \inf_{\theta \in \Theta_1} \left(1 - \beta_{\psi_n}(\theta)\right).$$

Continuing from the problem above, what is the power π_{ψ_n}?

$\pi_{\psi_n} =$ 0 ✔ **Answer:** 0

Solution:

A priori we have that

$$\pi_{\psi_n} = \inf_{\theta \in [1/2, \infty)} \left(1 - P_\theta\left(\psi_n = 0\right)\right) = \inf_{\theta \in [1/2, \infty)} P_\theta\left(\psi_n = 1\right) \geq 0].$$

Moreover, we computed above that $\beta_{\psi_n}(1/2) = P_{0.5}\left[\psi_n = 0\right] = 1$. Thus,

$$\pi_{\psi_n} = 0.$$

Remark: The power of a test is the largest lower bound on the probability that if H_1 is true, that indeed H_0 is rejected in favor of H_1. In this example, as $\theta \in \Theta_1$ approaches the boundary $1/2$, the probability of rejecting H_0 decreases and approaches 0.

Testing the Support of a Uniform Variable: Graphing the errors

0/1 point (graded)

As above, let $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{Unif}\,[0, \theta]$ for an unknown parameter θ and we designed the statistical test

$$\psi_n = \mathbf{1}\left(\max_{1 \leq i \leq n} X_i > 1/2\right)$$

to decide between the null and alternative hypotheses

$$H_0 : \theta \leq 1/2$$
$$H_1 : \theta > 1/2.$$

Let $\alpha_{\psi_n}(\theta)$ and $\beta_{\psi_n}(\theta)$ denote the type 1 and type 2 errors respectively.

On the graph below, do the following:

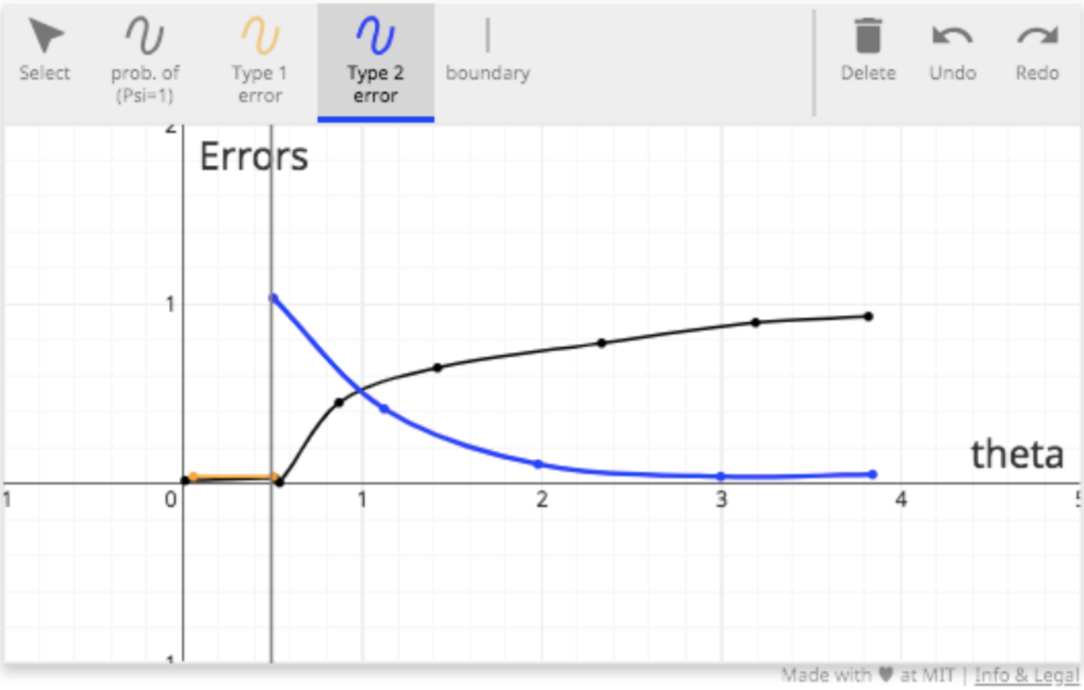
- Place a vertical line at the boundary of Θ_0 and Θ_1 using the **boundary tool**.
- Sketch the graph of $\mathbf{P}_\theta\left(\psi_n = 1\right)$ as a function of θ using the **probabilty of rejecting null** tool.
- Sketch the graph of the type 1 error $\alpha_{\psi_n}(\theta)$ on the **correct domain** using the **type 1 error** tool.
- Sketch the graph of the type 2 error $\beta_{\psi_n}(\theta)$ on the **correct domain** using the **type 2 error** tool.

Note: To use the spline tool for sketching the graphs, click on point on the graph, and the tool will connect these points with a smooth curve.

For each curve, you will be graded on its domain, its limiting values, its value on the boundary between Θ_0 and Θ_1, and its shape and continuity.

https://courses.edx.org/assets/courseware/v1/
b852eceb69416795f389f94a3b03df0f/asset-

v1:MITx+18.6501x+2T2020+type@asset+block/images_u2s4_errorsketch.svg



Testing the Support of a Uniform Variable: Level and Threshold

2/2 points (graded)

As in the problems on the previous page, let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}\,[0, \theta]$ for an unknown parameter θ and we designed the statistical test

$$\psi_n = \mathbf{1}(\max_{1 \leq i \leq n} X_i > 1/2)$$

to decide between the null and alternative hypotheses

$$\begin{aligned} H_0 &: \theta \leq 1/2 \\ H_1 &: \theta > 1/2. \end{aligned}$$

Let $\alpha_{\psi_n}(\theta)$ and $\beta_{\psi_n}(\theta)$ be the type 1 and type 2 errors respectively.

Recall from lecture that a test ψ has **level α** if

$$\alpha \geq \alpha_\psi(\theta) \qquad \text{for all } \theta \in \Theta_0,$$

where $\alpha_\psi = \mathbf{P}_\theta(\psi = 1)$ is the type 1 error. We will often use the word "level" to mean the "smallest" such level, i.e. the least upper bound of the type 1 error, defined as follows:

$$\alpha = \sup_{\theta \in \Theta_0} \alpha_\psi(\theta)$$

Here, $\sup_{\theta \in \Theta_0}$ stands for the supremum over all values of θ within Θ_0. If Θ_0 is a closed (*resp.* closed half-interval), and if $\alpha_\psi(\theta)$ is continuous (*resp.* continuous and decreasing as it approaches infinity), then its supremum equals the maximum.

Using the graph of the errors on the previous page, what is the smallest level α of the test ψ_n?

$\alpha =$ 0 ✔

How should the threshold of the test be changed to increase the smallest level α? In other words, consider tests of the form

$$\psi_{n,C} = \mathbf{1}(\max_{1 \le i \le n} X_i > C)$$

where C is the threshold. In the original test above, $C = 1/2$. What should the value of C be so that the level of $\psi_{n,C}$ is greater than the level of the $\psi_{n,1/2}$?
(Think of how the graph of $\mathbf{P}_\theta\,(\psi_C)$ changes with the threshold C.)

- ○ $C > 1/2$
- ◉ $C < 1/2$

✔

Solution:

Since the type 1 error $\alpha_{\psi_n}\,(\theta)$ is constantly zero over Θ_0, the smallest level of this test ψ is $\alpha = 0$.

To increase the smallest level α from 0, note that $\mathbf{P}_\theta\left(\max_{1 \le i \le n} X_i > C\right) = 0$ if and only if $\theta \le C$. This means the constant zero region of graph of $\mathbf{P}_\theta\,(\psi_C) = 0$ shifts to the right as C increases from $1/2$, and to the left as C decreases from $1/2$. Since the maximum of type 1 error occurs at the boundary $\theta = 1/2$, this means $C < 1/2$ is required for the level to be positive.
Remark: The reason behind increasing the level in this example is to increase the power of the test from 0. In general, one of the first requirements of a test is to have a small-enough level so that the probability of concluding a false positive, (i.e. rejecting the null while the null is true) is controlled.

Testing the Support of a Uniform Variable: Determine the Threshold

0/1 point (graded)

As above, let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Unif}\,[0, \theta]$ for an unknown parameter θ and consider tests of the form

$$\psi_{n,C} = \mathbf{1}(\max_{1 \le i \le n} X_i > C)$$

to decide between the null and alternative hypotheses

$$H_0 : \theta \le 1/2$$
$$H_1 : \theta > 1/2.$$

Let $\alpha_{\psi_{n,C}}\,(\theta)$ and $\beta_{\psi_{n,C}}\,(\theta)$ be the type 1 and type 2 errors respectively.

Determine the smallest threshold C such that the test $\psi_{n,C}$ has level α.

(Enter the roots of x as a power of x, e.g. enter **x^(1/3)** for $\sqrt[3]{x} = x^{1/3}$.)

$C =$ (alpha)^(n) ✖ **Answer:** 1/2*(1-alpha)^(1/n)

$(\alpha)^n$

Solution:

Following similar computation as in a previous problem where $C = 1/2$, we have $\mathbf{P}_\theta\left(\psi_{n,C} = 1\right) = 1 - \left(\frac{C}{\theta}\right)^n$. Since the smallest level is

$$\begin{aligned} \alpha &= \max_{\theta \in \Theta_0} p_\theta\left(\psi_{n,C} = 1\right) \\ &= p_{1/2}\left(\psi_{n,C} = 1\right) = 1 - \left(\frac{C}{1/2}\right)^n, \end{aligned}$$

a test with threshold $C = \frac{1}{2}\sqrt[n]{1-\alpha}$ or smaller will have level α.
Remark: Notice the threshold C depends on n, α, as well as the value of θ at the boundary of Θ_0 and Θ_1.