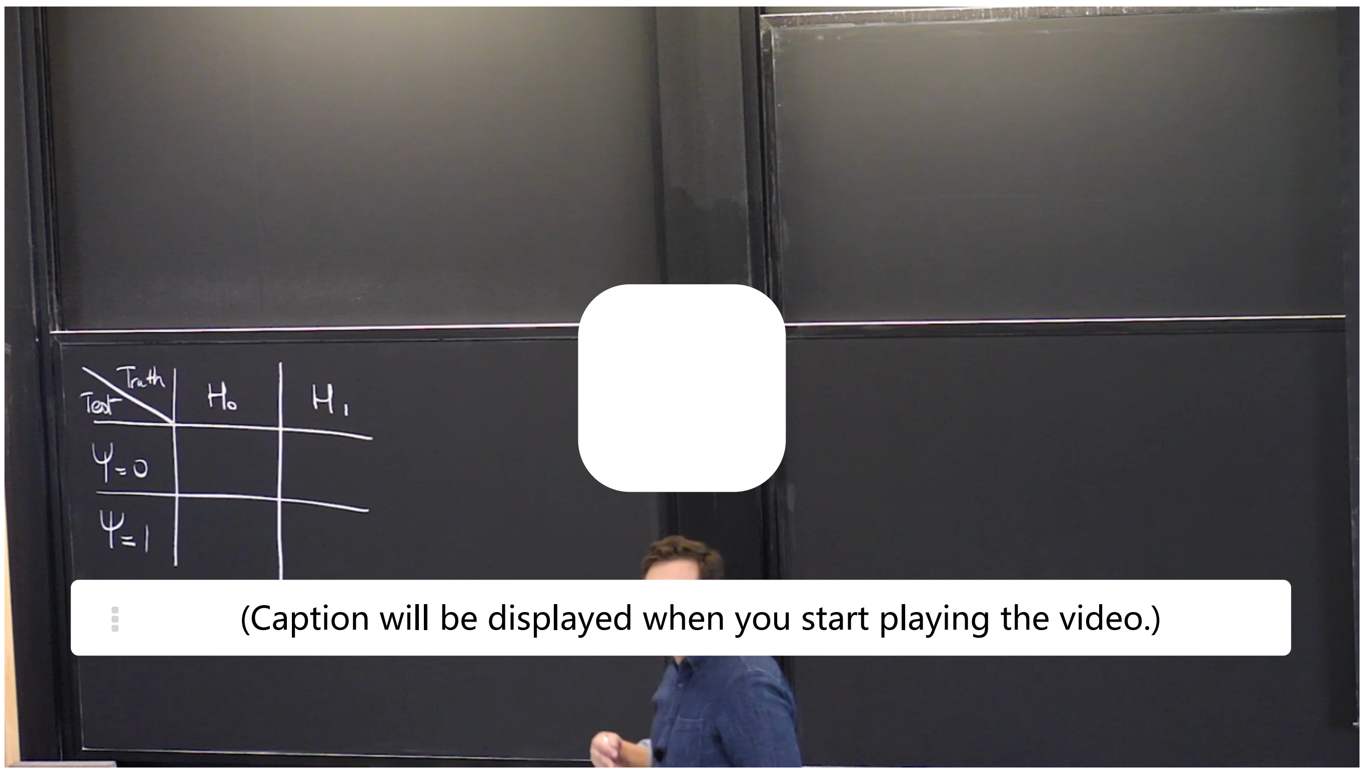
3. Review of Parametric Hypothesis Testing

Worked Example: A Two-Sided Test Associated to a Bernoulli Experiment



All right.
So the type I error.
So the probability of type I error
was denoted by alpha theta.
So it was, by definition, the probability under theta
that psi is equal to 1.
If we really want this to be an error,
we need to actually force theta to be in the set associated
to the null hypothesis, otherwise, as we said,
we're actually in this range where H1 is true.
But we conclude H1.
So it's certainly not a mistake.
So this is our probability of type I error.
And then we said, well, if for all of thetas,
we have that this thing is less than alpha--
so we said, for all theta in theta
equals 0, which is the same as saying

视频
下载视频文件

字幕
下载 SubRip (.srt) file
下载 Text (.txt) file

Review: Interpreting the Level

1/1 point (graded)
Which of the following is a correct interpretation of the (smallest) **level** of a test? (Choose all that apply.)

- [x] The level of a test is an upper bound on the type 1 error. ✔
- [ ] The level of a test is an upper bound on the type 2 error.
- [ ] The level of a test is a random variable that depends on the sample.
- [x] The level of a test gives an upper bound on the worst-case probability of making an error under the null hypothesis. ✔

✔

Solution:

We recall the definition of the **level** of a test ψ. We have a statistical model given by $(E, \{P_\theta\}_{\theta\in\Theta})$ and null and alternative hypotheses H_0 and H_1, respectively. Let Θ_0 denote the region corresponding to the null hypothesis. Let

$$\begin{aligned} \alpha_\psi : \Theta &\to [0,1] \\ \theta &\mapsto P_\theta(\psi = 1) \end{aligned}$$

denote the type 1 error. Then the **level** of ψ is any real number α such that

Generating Speech Output

$$\alpha_\psi(\theta) \le \alpha, \quad \text{for all } \theta \in \Theta_0.$$

Having reviewed this definition, we now examine the choices in order.

- "The level of a test is an upper bound on the type 1 error." is correct by definition.
- "The level of a test is an upper bound on the type 2 error." is incorrect. The definition of level is given in terms of the *type 1* error.
- "The level of a test is a random variable that depends on the sample." is incorrect. The level of a test is given by $P_\theta(\psi = 1)$, which is a probability with respect to $X_1, \ldots, X_n \overset{iid}{\sim} P_\theta$. Hence, the level is a number that does not depend on the data.
- "The level of a test gives an upper bound on the worst-case probability of making an error under the null hypothesis." is correct. This is a restatement of the formal definition given at the start of this solution.

Remark: The final choice gives a convenient description of the type 1 error that is useful to keep in mind.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Concept Check: Test Statistics

1/1 point (graded)

Setup:

Recall the **statistical experiment** in which you flip a coin n times to decide the coin is fair.

You model the coin flips as $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$ where p is an unknown parameter, and formulate the hypothesis:

$$H_0 : p = 0.5$$
$$H_1 : p \neq 0.5,$$

and design the test ψ using the statistic T_n:

$$\begin{aligned} \psi_n &= \mathbf{1}(T_n > C) \\ \text{where} \quad T_n &= \sqrt{n}\frac{\left|\overline{X}_n - 0.5\right|}{\sqrt{0.5(1-0.5)}} \end{aligned}$$

where the number C is the threshold. Note the absolute value in T_n for this two sided test.

Question:

If it is true that $p = 1/2$, which of the following are true about T_n?

(Choose all that apply.)

- [ ] T_n is a consistent estimator of the true parameter $p = 1/2$.
- [x] $\lim_{n\to\infty} T_n \xrightarrow[n\to\infty]{(d)} |Z|$ where $Z \sim N(0,1)$ is a standard Gaussian. ✔
- [x] T_n involves a shift and rescaling of the sample average so that as $n \to \infty$, this random variable will converge in distribution. ✔
- [x] The limiting distribution of T_n can be understood using computational software or tables. ✔

✔

Solution:

We examine the choices in order.

- The first choice is incorrect. The statistic T_n does **not** converge to a real number as $n \to \infty$. By the CLT, T_n converges in *distribution*, Generating Speech Output ymptotically, it is a random variable.

- The remaining choices are correct. To construct T_n we have shifted the sample mean $\overline{X}_n$ by $1/2$, rescaled by $\sqrt{\frac{n}{0.5(1-0.5)}}$. The CLT guarantees that T_n converges in distribution to a random variable $|Z|$ where $Z \sim N(0,1)$. Since the density of Z is given explicitly, we can work with the limiting distribution using computational software. Alternatively, there are also tables available containing the quantiles of a standard Gaussian.

Remark: This example illustrates one of the main strategies involved in hypothesis testing. Namely, we want to work with a test statistic, that, asymptotically, tends to a distribution that we can easily work with. In many cases, this will involve shifting and rescaling the sample mean so that the CLT applies and we can just work with a standard Gaussian $\mathcal{N}(0,1)$.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Designing a Test to have a Given Asymptotic Level

4/4 points (graded)
In this problem, we will see the condition for a threshold of a hypothesis test graphically.

Setup as above:

You observe $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Ber}(p^*)$ (each X_i models a coin flip) and want to decide if $p^* = 1/2$. Let the null and alternative hypotheses be

- $H_0 : p^* = 0.5$
- $H_1 : p^* \neq 0.5$.

You construct the statistical test:

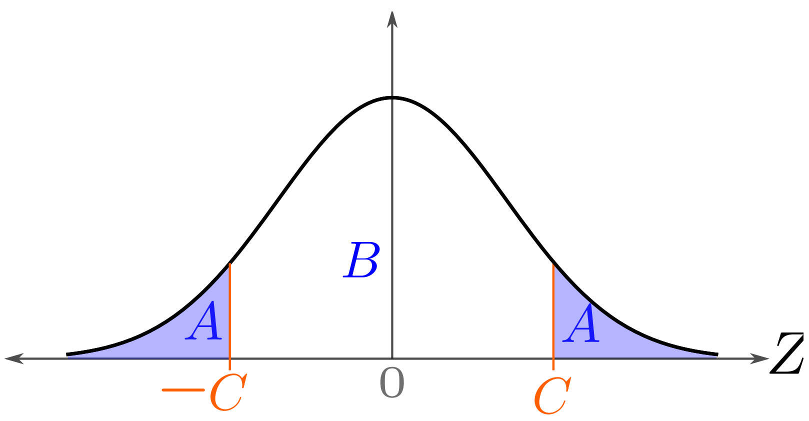
$$\begin{aligned} \psi_n &= \mathbf{1}(T_n > C) \\ \text{where} \quad T_n &= \sqrt{n}\frac{\left|\overline{X}_n - 0.5\right|}{\sqrt{0.5(1-0.5)}} \end{aligned}$$

where the number C is the threshold to be determined. Note the absolute value in T_n; this is a two-sided test.

Recall that the test ψ has **asymptotic level** α if

$$\lim_{n\to\infty} P_\theta(\psi = 1) \leq \alpha.$$

Questions:



The graph of the standard normal distribution $\mathcal{N}(0,1)$, along with the lines $Z = \pm C$. The letters A, B denote the areas of the corresponding shaded regions:

$$\begin{aligned} A &= \mathbf{P}(Z < -C) = \mathbf{P}(Z > C) \qquad (\mathbf{P}(Z < -C) = \mathbf{P}(Z > C) \text{ by symmetry}), \\ B &= \mathbf{P}(-C \leq Z \leq C) \end{aligned}$$

where $\mathbf{P}$ is the probability distribution of $\mathcal{N}(0,1)$.

Generating Speech Output

What is the smallest C such that the test $\psi(T_n > C)$ has asymptotic level α? (The level is often given as a specification for the test.)

Answer not by giving the value of C, but by **giving the condition** that C must satisfy, i.e.refer to the figure above, the smallest C such that $\psi(T_n > C)$ has asymptotic level α must be chosen such that, in terms of A and B in the figure above, α equals...

$\alpha =$ 2*A ✔ **Answer:** 2*A

Hence, as a function of α, what is C_α? (To enter the quantiles of the standard Gaussian, for instance q_α, type **q(alpha)**. Recall q_α denotes the $1-\alpha$-quantile of a standard Gaussian, i.e. the value such that $P(Z \geq q_\alpha) = \alpha$ for $Z \sim \mathcal{N}(0,1)$.)
Denote by C_α the smallest C such that the test $\psi(T_n > C)$ has asymptotic level α.

$C_\alpha =$ q(alpha/2) ✔ **Answer:** q(alpha/2)

Let the rejection region for the test $\psi(T_n > C_\alpha)$ be

$$R_\alpha = \left\{(X_1, \ldots, X_n) \in \{0,1\}^n : \overline{X}_n < L \cup \overline{X}_n > R\right\}.$$

What are L and R?

(Your answers will depend on α and n.)
(To enter quantiles, for instance q_α, type **q(alpha)**.)

$L =$ 0.5-q(alpha/2)*0.5/sqrt(r ✔ **Answer:** 0.5-q(alpha/2)*sqrt(0.5*(1 - 0.5))/(sqrt(n))

$R =$ 0.5+q(alpha/2)*0.5/sqrt(✔ **Answer:** 0.5+q(alpha/2)*sqrt(0.5*(1 - 0.5))/(sqrt(n))

STANDARD NOTATION

Solution:

- By the central limit theorem, if $\mathbb{E}[X] = p^* = 0.5$, then

$$\sqrt{n}\frac{\overline{X}_n - 0.5}{\sqrt{0.5(1-0.5)}} \xrightarrow[n\to\infty]{(d)} N(0,1).$$

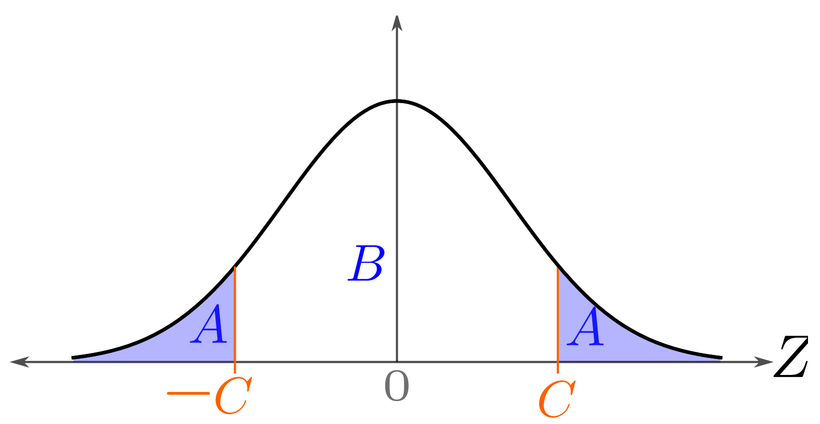
Let $\mathbf{P}_{1/2} = \mathrm{Ber}(1/2)$ for notational convenience. Then for the test statistics

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

we have

$$\mathbf{P}_{1/2}(T_n > C) \xrightarrow[n\to\infty]{} A + A = 2A$$

where $2A$ are the total area of the shaded regions under the graph of the normal distribution:



Generating Speech Output

The graph of the standard normal distribution $\mathcal{N}(0,1)$, along with the lines $Z = \pm C$. The letters A, B denote the areas of the corresponding shaded regions; hence:

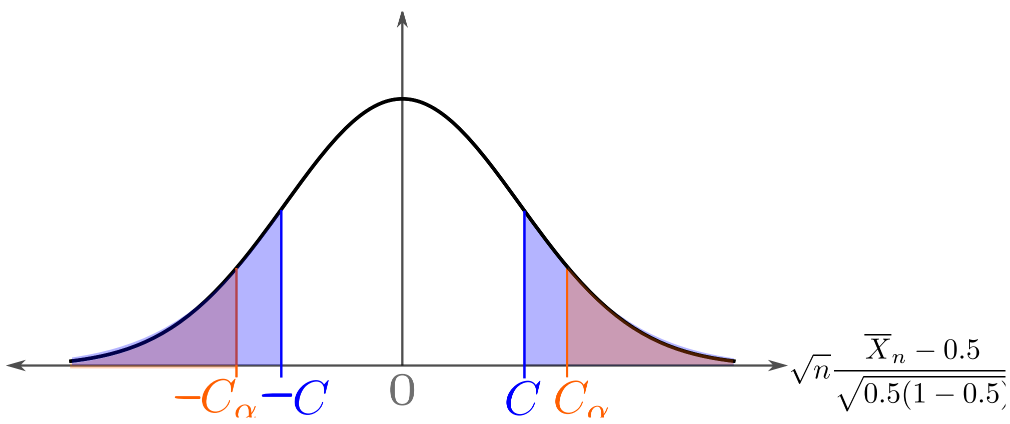
$$\begin{aligned} A &= P(Z < -C) \\ B &= P(Z \le C) \\ A &= P(Z > C) \end{aligned}$$

where P is the probability distribution of $\mathcal{N}(0,1)$

Since H_0 is defined by a single value $p = 1/2$, the asymptotic level is equal to the asymptotical type 1 error at $p = 1/2$, which is $\mathbf{P}_{1/2}(T_n > C)$. Therefore, given a desired asymptotic level α, choosing a threshold C_α such that

$$\alpha \;=\; P(Z < -C_\alpha) + P(Z > C_\alpha) \;=\; A + A = 2A \qquad Z \sim \mathcal{N}(0,1)$$

will result in a test $\psi = \mathbf{1}(T_n > C_\alpha)$ that has asymptotic level α. Furthermore, for any threshold $C < C_\alpha$ will yield a larger asymptotic type 1 error, as shown in the figure below



The graph of the standard normal distribution $\mathcal{N}(0,1)$;
For $C < C_\alpha$, the **type 1 error for** $\psi = \mathbf{1}(T_n > C)$ (shaded **blue**) is larger than the **type 1 error for** $\psi = \mathbf{1}(T_n > C_\alpha)$ (shaded **orange**).

This means that C_α is the smallest choice of threshold C such that the test $\psi(T_n > C)$ has asymptotic level α.

- Since $\alpha = P(Z < -C_\alpha) + P(Z > C_\alpha) = 2P(Z > C_\alpha)$ by symmetry, we have $C_\alpha = q_{\alpha/2}$.
- The rejection region of $\psi = \mathbf{1}(T_n > q_{\alpha/2})$ is defined by

$$\begin{aligned} T_n \;&=\; \left|\sqrt{n}\frac{\overline{X}_n - 0.5}{\sqrt{0.5(1-0.5)}}\right| \quad > \; q_{\alpha/2} \\ \Longrightarrow \; \overline{X}_n \; &< \; 0.5 - q_{\alpha/2}\frac{\sqrt{0.5(1-0.5)}}{\sqrt{n}} \;\cup\; \overline{X}_n \; > \; 0.5 + q_{\alpha/2}\frac{\sqrt{0.5(1-0.5)}}{\sqrt{n}}. \end{aligned}$$

Remark: We have done similar manipulations when looking for two-sided confidence interval of level $1 = \alpha$. But here, we look for a range of $\overline{X}_n$ in terms of the assumed value of the parameter p under the null hypothesis.

Remark: Since the limiting distribution of our test statistic is well-known (the absolute value of a standard Gaussian), it is straightforward to specify the asymptotic level of our test using computational tools or tables. Later in this course, we will also encounter tests where for fixed n we can compute the (non-asymptotic) level of ψ_n using computational tools or tables.

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

Concept Check: Rejection Region

1/2 points (graded)

You observe $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathbf{P}_{\theta^*}$ and design a test ψ to test between a null hypothesis and an alternative hypothesis.

True or False: The rejection region of ψ depends on the value of the true unknown parameter θ^*.

Generating Speech Output

True ✖

False ✔

你不知道theta的时候，你才要通过一个假设的测试来推断。

True or False: To define a statistical test ψ, it is enough to define the rejection region R_ψ.

True ✔

False

Solution:

- The rejection region does not depend on the true parameter. It is fixed when a test is designed, as in the example in the problem above.
- As pointed out above, a test is by definition an indicator function of its rejection region:

$$\psi = \mathbf{1}\left((X_1, \ldots, X_n) \in R_\psi\right)$$

Hence, yes, to define a test, all that is needed is to define its rejection region.

提交 你已经尝试了1次（总共可以尝试1次）

Answers are displayed within the problem

讨论

显示讨论

主题： Unit 2 Foundation of Inference:Lecture 7: Hypothesis Testing (Continued): Levels and P-values / 3. Review of Parametric Hypothesis Testing

认证证书是什么?

© 保留所有权利

Generating Speech Output