ECMA31000: Introduction to Empirical Analysis
Hypothesis Testing

Joe Hardwick

University of Chicago

Autumn 2021

Statistical Model

F_{θ_0} is generating sample data.

- Suppose there exists a population distribution with CDF $F_{\theta_0} \in \mathcal{F} = \{F_s : s \in \Theta\}$, where Θ is an index set.
  - $\mathcal{F}$ is called an <u>indexed family of distributions</u>.
- Let $\{X_i\}_{i=1}^n$ be an iid sample of random variables with marginal distribution F_{θ_0}, for some $\theta_0 \in \Theta$.
- If we assume $\Theta \subset \mathbb{R}^d$, our statistical model comprises a parametric class of distributions.

Definitions

- Given a sample $\{X_i\}_{i=1}^n$ of draws from a distribution F, a <u>statistic</u> is a function T_n mapping observed data to some set V:

$$T_n : (X_1, \ldots, X_n) \rightarrow V,$$

  where V could be e.g. a subset of $\mathbb{R}^d$, or a function space.
- An <u>estimator</u> is a statistic used to learn about some feature of F, $\theta(F)$.
- If the class of distributions $\mathcal{F}$ is indexed by a k-dimensional parameter θ, and we estimate $d \leq k$ of these parameters, then

$$\hat{\theta}_n : (X_1, \ldots, X_n) \rightarrow \mathbb{R}^d.$$

- Today we will use estimators to construct confidence sets and test hypotheses about the distribution generating the data.

Confidence Sets

- Fix a desired *coverage probability* $1-\alpha$, for some $\alpha \in (0,1)$.
- Our goal is to construct a subset of Θ such that the true parameter θ_0 lies in this subset with probability at least $1-\alpha$, either in a finite sample, or asymptotically.
- This subset must be observable: It must be determined by the data and any modelling assumptions.
- A random set $C_n(X_1, \ldots, X_n) \subset \Theta$ is called a $1-\alpha$ confidence set if it is observable and for any $\theta \in \Theta$:

$$\mathrm{P}_\theta(\theta \in C_n(X_1, \ldots, X_n)) \geq 1-\alpha.$$

- Confidence sets which are intervals are called confidence intervals.

Confidence Sets

- We have seen in PSETS 2+3 that constructing intervals with these properties is hard.
- One one hand, the definition requires that the true coverage probability exceeds $1 - \alpha$ for <u>any</u> value of the unknown parameter.
- On the other hand, we don't want to exceed $1 - \alpha$ too much, since then the confidence set may be much larger than necessary for the desired coverage probability.
- Recall that confidence sets based on Chebyshev's inequality were wider than necessary, often much wider!

Example I: Exact finite sample coverage

- Let $\{X_i\}_{\cancel{i \geq 1}\; i=1}^{n}$ be iid draws from $\mathcal{N}\left(\mu, \sigma^2\right)$, where μ is unknown and σ^2 is known.
- We know that

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

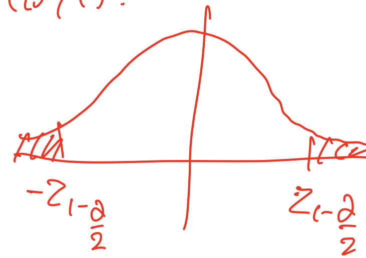
from which it follows that

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \sim \mathcal{N}\left(0, 1\right).$$

- The quantity $\sqrt{n}\left(\bar{X}_n - \mu\right)/\sigma$ is called a <u>pivot</u>, because its distribution does not depend on the actual values of the unknown parameters $\left(\mu, \sigma^2\right)$.
- Pivots are not necessarily statistics, because they may depend on unknown parameters, though their distribution cannot.

Example I: Exact finite sample coverage

$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$.



- It follows that for every n,

$$\mathrm{P}_\mu \left(\left| \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \right| \leq z_{1-\alpha/2} \right) = 1 - \alpha,$$

$z_{1-\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of $N(0,1)$.

where $z_{1-\alpha/2} = \Phi\left(1 - \frac{\alpha}{2}\right)$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

$P_\mu(\mu \in C_n) = 1-\alpha$.

- Rearranging gives

$$\mathrm{P}_\mu \left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right) = 1 - \alpha.$$

$\left|\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right| \le z_{1-\frac{\alpha}{2}} \iff \mu \in \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$.

Example I: Exact finite sample coverage

- Alternatively, $\mathrm{P}_{\mu}\left(\mu \in \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right]\right) = 1 - \alpha$, so since σ is known, the set

$$C_n\left(X_1, \ldots, X_n\right) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right]$$

is observable and is an (exact) $1 - \alpha$ confidence interval for μ.

Example II: Exact finite sample coverage

- Let $\{X_i\}_{i\geq 1}$ be iid draws from $\mathcal{N}\left(\mu, \sigma^2\right)$, where μ and σ^2 are unknown. It can be shown that

Pivot

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}} \sim t_{n-1},$$

Estimated σ^2 with

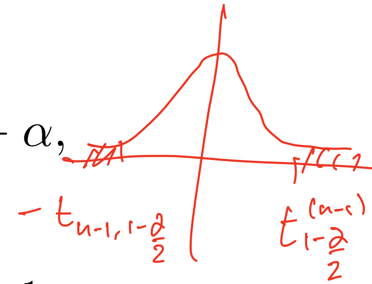
where t_{n-1} denotes the t distribution with $n-1$ degrees of freedom.

- This quantity is also a pivot, and now only depends on the data and unknown parameter μ.

Example II: Exact finite sample coverage

- Let $\tilde{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$. It follows that for every n,

$$\mathrm{P}_\mu\left(\left|\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\tilde{S}_n}\right| \leq t_{n-1,1-\alpha/2}\right) = 1-\alpha,$$



  where $t_{n-1,1-\alpha/2}$ is the $1-\alpha/2$ quantile of the t_{n-1} distribution.

- Rearranging yields

$$\mathrm{P}_\mu\left(\bar{X}_n - \frac{\tilde{S}_n}{\sqrt{n}}t_{n-1,1-\alpha/2} \leq \mu \leq \bar{X}_n + \frac{\tilde{S}_n}{\sqrt{n}}t_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

Example II: Exact finite sample coverage

$$P_\mu\left(\mu \in \left[\bar{X} - \frac{\tilde{S}_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \bar{X} + \frac{\tilde{S}_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right]\right)$$

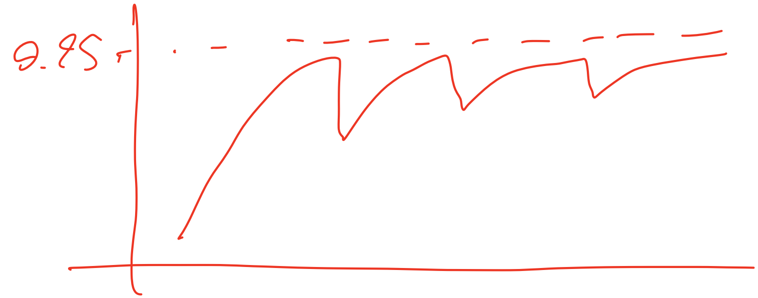
$$= 1-\alpha.$$

- In summary, the set

$$C_n(X_1, \ldots, X_n) = \left[\bar{X}_n - \frac{\tilde{S}_n}{\sqrt{n}} t_{n-1,1-\alpha/2}, \bar{X}_n + \frac{\tilde{S}_n}{\sqrt{n}} t_{n-1,1-\alpha/2}\right]$$

is observable and is an (exact) $1-\alpha$ confidence interval for μ.

Asymptotic Confidence Sets



- It is rarely the case that we can easily describe the exact distribution of an estimator and construct a confidence set based on it with exact coverage probability.
- Let $\{X_i\}_{i \geq 1}$ be iid draws from *Bernoulli* (p).
- In PSET 2, we noted that a $1 - \alpha$ confidence interval valid in finite samples may have true coverage probability much higher than $1 - \alpha$. We call such a confidence interval conservative.
- We may weaken the requirement on our confidence set to require that it is an *asymptotic* $1 - \alpha$ confidence interval.

Asymptotic Confidence Sets

- Fix a desired *coverage probability* $1-\alpha$, for some $\alpha \in (0,1)$.
- A random set $C_n(X_1, \ldots, X_n) \subset \Theta$ is called an asymptotic $\underline{1-\alpha \text{ confidence set}}$ if it is observable and for any $\theta \in \Theta$:

$$\lim_{n \to \infty} \mathrm{P}_\theta \left(\theta \in C_n(X_1, \ldots, X_n)\right) \geq 1-\alpha.$$

- In PSET 3, we saw in simulations that, for a fixed sample size (e.g. $n = 100$), an asymptotic confidence interval based on a normal approximation can have true coverage probability close to $1-\alpha$ for some parameter values e.g. $p = 0.4$ and very far from $1-\alpha$ for others e.g. $p = 0.99$.

Asymptotic Confidence Sets

- Suppose that for any $\theta \in \Theta$, $\hat{\theta}_n$ is an asymptotically normal estimator of θ:

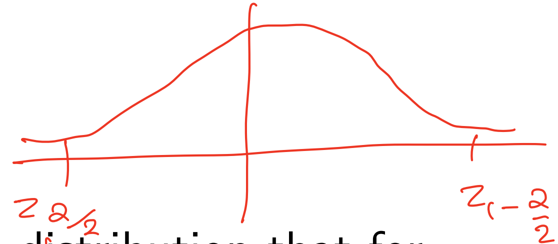
$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\theta\right)\right).$$

- Suppose we have a consistent estimator $\hat{V}\left(\theta\right) \xrightarrow{p} V\left(\theta\right)$. Then by Slutsky's theorem

$$\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{V}\left(\theta\right)}} \xrightarrow{d} \mathcal{N}\left(0, 1\right).$$

- This quantity is *asymptotically pivotal,* since its limiting distribution does not depend on θ.
- In PSET 3: $X_i \sim \mathit{Bernoulli}\left(p\right)$, $\hat{\theta}_n = \bar{X}_n$, $\theta = p$, $V\left(p\right) = \mathit{Var}\left(X_i\right) = p\left(1-p\right)$, and $\hat{V}\left(p\right) = \bar{X}_n\left(1 - \bar{X}_n\right)$.

Asymptotic Confidence Sets



- It follows by definition of convergence in distribution that for any $\theta \in \Theta$:

$-z_{1-\frac{\alpha}{2}}$ by symmetry.

$$\lim_{n\to\infty} \mathrm{P}_\theta\left(z_{\alpha/2} \le \frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{V}\left(\theta\right)}} \le z_{1-\alpha/2} \right) = 1 - \alpha.$$

- Since $z_{\alpha/2} = -z_{1-\alpha/2}$ by symmetry of the standard normal about 0, rearranging yields

$$\lim_{n\to\infty} \mathrm{P}_\theta\left(\hat{\theta}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{V}\left(\theta\right)}{n}} \le \theta \le \hat{\theta}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{V}\left(\theta\right)}{n}} \right) = 1-\alpha.$$

$$C_n = \left[\hat{\theta}_n \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{V}(\theta)}{n}}\right].$$

Example: Sample Mean

- Suppose $\{X_i\}_{i\geq 1}$ is an iid sequence with $\mathrm{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$. Then by the CLT:

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right),$$

  and

$$\tilde{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \xrightarrow{p} \sigma^2.$$

- By Slutsky's theorem

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\tilde{S}_n} \xrightarrow{d} \mathcal{N}(0, 1),$$

  so $C_n = \left[\bar{X}_n - z_{1-\alpha/2}\frac{\tilde{S}_n}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2}\frac{\tilde{S}_n}{\sqrt{n}}\right]$ is an asymptotic $1-\alpha$ confidence interval for μ.

Questions?

Hypothesis Testing: Definitions

$X_i \sim N(\mu, \sigma^2)$.

$\Theta = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$.

$H_0 : \mu \leq 0$.

$\Theta_0 = \{(\mu, \sigma^2) : \mu \leq 0, \sigma^2 > 0\}$

- The null hypothesis is a subset $\Theta_0 \subset \Theta$ of hypothesised values for θ_0, written as:
$$H_0 : \theta_0 \in \Theta_0.$$

$X_i \sim U[0, \theta]$ $\quad H_0 : \theta = 1$ Simple Hypothesis

- If Θ_0 is a singleton, the null hypothesis is called a simple hypothesis. If not, the hypothesis is composite.
- e.g. If $X \sim \mathcal{N}(\mu, \sigma^2)$, $H_0 : \mu = 0$ is a simple null, $H_0 : \mu \leq 0$ is a composite null.

$H_0 : \theta \in [1, 2]$ Composite hypothesis

- A test of H_0 is therefore a test of whether the data were generated by some F_θ such that $\theta \in \Theta_0$.
- The alternative hypothesis is $\Theta \backslash \Theta_0$.

$H_0 : \mu \leq 0 \quad H_1 : \mu > 0$.

$\Theta_1 = \Theta \backslash \Theta_0$

Hypothesis Testing: Definitions

- A test of H_0 is a function $\phi_n(X_1, \ldots, X_n) \to \{0, 1\}$.
- We reject H_0 at sample size n iff $\phi_n = 1$.
- e.g. If $T_n(X_1, \ldots, X_n)$ is a sequence of statistics, and c_n a sequence of real numbers,

$$\phi_n(X_1, \ldots, X_n) = \mathbf{1}(T_n > c_n)$$

is a test which rejects H_0 iff $T_n > c_n$.

Two undesirable outcomes

- Suppose $\theta_0 \in \Theta_0$ but $\phi_n = 1$. This is called a Type I Error.
- Suppose $\theta_0 \notin \Theta_0$ but $\phi_n = 0$. This is called a Type II Error.
- It is customary to control the probability of a Type I Error first, and then minimize the probability of a Type II Error subject to this constraint.
- The power function associated with ϕ_n is the function

$$\beta_n(\theta) = \mathrm{P}_\theta\left(\phi_n\left(X_1, \ldots, X_n\right) = 1\right),$$

which is the probability that ϕ_n rejects H_0 if the true parameter is θ.

Properties of tests

- It is trivial to construct a test such that $\beta_n(\theta) = 0$ for all $\theta \in \Theta_0$ (never reject, no matter the draws).
- Also trivial to construct a test such that $\beta_n(\theta) = 1$ for all $\theta \notin \Theta_0$ (always reject, no matter the draws).
- The problem: A test which never rejects does not alert us that our hypothesis is false. A test which always rejects always does so even when H_0 is true.
- Our task is to construct a test with specified (low) null rejection probability, that also rejects as often as possible when H_0 is false.

Properties of tests

- Given a significance level α, we select a test ϕ_n such that

$$\beta_n(\theta) \leq \alpha \quad \text{for all } \theta \in \Theta_0,$$

  or, alternatively

$$\sup_{\theta \in \Theta_0} \beta_n(\theta) \leq \alpha.$$

- Similarly, given a test ϕ_n, the size of ϕ_n with power function β_n is

$$\alpha := \sup_{\theta \in \Theta_0} \beta_n(\theta).$$

- The probability that ϕ_n rejects when H_0 is true cannot exceed α.

Example: Normal Distribution

- Suppose $\{X_i\}_{i=1}^n$ is an iid sample from $\mathcal{N}(\mu, 1)$. We wish to test $H_0 : \mu \leq 0$, so $\Theta_0 = (-\infty, 0]$.
- Let $c > 0$ and consider the test

$$\phi_n(X_1, \ldots, X_n) = \mathbf{1}\left(\bar{X}_n > c\right).$$

(Cutoff rule).

- The power function is

$P(Z > \sqrt{n}(c-\mu))$

$$\begin{aligned}\beta_n(\mu) &= \mathrm{P}_\mu\left(\bar{X}_n > c\right) \\ &= \mathrm{P}_\mu\left(\sqrt{n}\left(\bar{X}_n - \mu\right) > \sqrt{n}(c-\mu)\right) \\ &= 1 - \Phi\left(\sqrt{n}(c-\mu)\right).\end{aligned}$$

Φ is the CDF of $N(0,1)$.

Example: Normal Distribution

$\sqrt{n}c - \sqrt{n}\mu$ ≤ 0.

$\Phi(x)$ is strictly increasing in x.

- Note that if $\mu \leq 0$, $\sqrt{n}\,(c-\mu) \geq \sqrt{n}c$, so

$$\beta\,(\mu) = 1 - \Phi\left(\sqrt{n}\,(c-\mu)\right) \leq 1 - \Phi\left(\sqrt{n}c\right),$$

with equality attained iff $\mu = 0$. Alternatively:

$$\max_{\mu\leq 0} \beta_n\,(\mu) = 1 - \Phi\left(\sqrt{n}c\right). \quad = \alpha.$$

- Suppose we fix a desired size $\alpha \in (0,1)$. When the null is true, the rejection probability may equal but should not exceed α.
- We must choose c such that

$$1 - \Phi\left(\sqrt{n}c\right) \leq \alpha.$$

Example: Normal Distribution

$1-\alpha = \Phi(\sqrt{n}c)$.

$\sqrt{n}c = \Phi^{-1}(1-\alpha)$

$c(\alpha) = \dfrac{\Phi^{-1}(1-\alpha)}{\sqrt{n}}$

- Solving $1 - \Phi\left(\sqrt{n}c\right) = \alpha$ for c yields:

$$c^*(\alpha) = \frac{\Phi^{-1}(1-\alpha)}{\sqrt{n}} = \frac{z_{1-\alpha}}{\sqrt{n}}.$$

- Notice that any value of $c \geq c^*$ would also keep size below α, so why pick $c = c^*$?
- Remember that subject to controlling size, we want our test to maximize power when H_0 is false.
- If $\max_{\mu \leq 0} \beta_n(\mu) < \alpha$, we say the test is <u>conservative</u>.

Example: Normal Distribution

$\sqrt{n}c^* < \sqrt{n}c$.

- If $c > c^*$,

$$\beta_n(\mu) = 1 - \Phi\left(\sqrt{n}(c-\mu)\right) < 1 - \Phi\left(\sqrt{n}(c^*-\mu)\right).$$

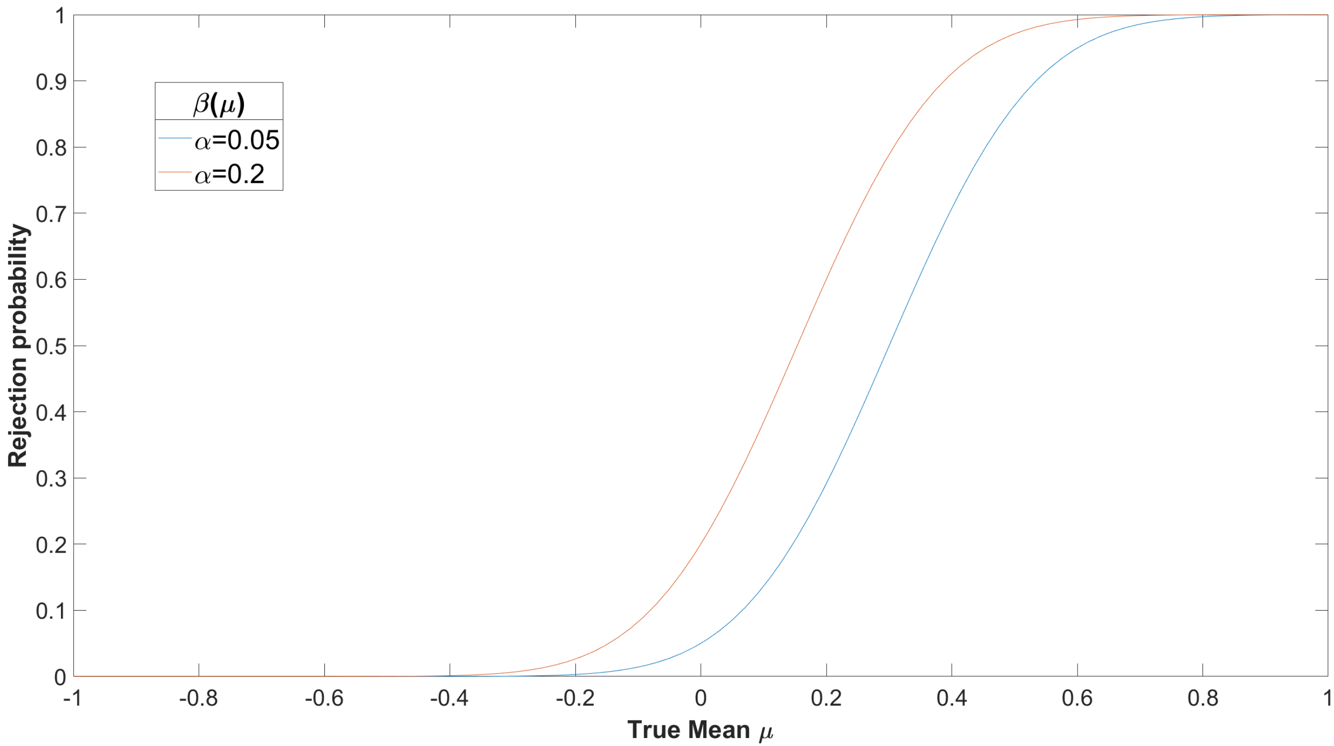
- Therefore, if H_0 is false, the test which rejects when $\bar{X}_n > c^*(\alpha)$ has greater power (at any value of μ) than the test which rejects when $\bar{X}_n > c$, while both tests have acceptable size.
- In fact, $c = c^*$ yields the uniformly (in μ) most powerful test of $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ subject to the constraint that

$$\max_{\mu \leq 0} \beta_n(\mu) \leq \alpha.$$

- This shows that if $\mu > 0$ but $\mu \approx 0$, *any* test which has size α will have power not much greater than α.

Example: Normal Distribution

- Probability of rejecting H_0 is known for any true value of μ if, e.g. $X \sim \mathcal{N}(\mu, 1)$. This is the power curve:



Θ_0 Θ_1

Questions?

Asymptotic Tests

- Constructing tests of a given size that have good power properties in finite samples is difficult in general since the distribution of the test statistic is unknown, whether or not the null hypothesis is true.
- We may appeal to asymptotic approximations to the distribution of the test statistic to yield a test with *asymptotically* correct null rejection probability.
- A test is called <u>asymptotically of size α</u> if

$$\lim_{n\to\infty} \beta_n(\theta) \leq \alpha \quad \text{for all } \theta \in \Theta_0.$$

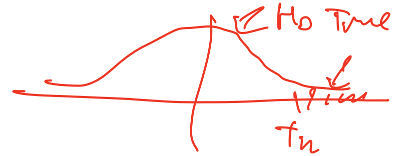
Asymptotic Tests

- If $\{X_i\}_{i=1}^{n}$ is an iid sample with $\mathrm{E}\left(X_i\right) = \mu$ and $Var\left(X_i\right) = \sigma^2 > 0$, and we wish to test $\mu \leq \mu_0$, it follows that if
$$t_n\left(X_1, \ldots, X_n\right) = \frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{\tilde{S}_n}$$
and
$$\phi_n\left(X_1, \ldots, X_n\right) = \mathbf{1}\left(t_n > z_{1-\alpha}\right),$$
then the test ϕ_n is asymptotically of size α, where
$$c\left(\alpha\right) = \Phi^{-1}\left(1 - \alpha\right).$$

$= z_{1-\alpha}$.

p-values

eg:



Prob of observing something more extreme than T_n?

- Given $\{X_i\}_{i=1}^{n}$ and a test $\phi_n = \mathbf{1}\left(T_n\left(X_1, \ldots, X_n\right) > c\left(\alpha\right)\right)$ of size α, the p-value is defined as

$$\hat{p}_n = \inf\left\{\alpha \in (0,1) : T_n\left(X_1, \ldots, X_n\right) > c\left(\alpha\right)\right\}.$$

$\hat{p}_n$ is the smallest value of α for which H_0 is rejected.

- In the preceding example,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{\tilde{S}_n} \geq c\left(\alpha\right) \iff \frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{\tilde{S}_n} \geq \Phi^{-1}\left(1-\alpha\right),$$

$$\Phi\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\tilde{S}_n}\right) \geq 1-\alpha.$$

which yields

$$\hat{p}_n = 1 - \Phi\left(\frac{\sqrt{n}\left(\bar{X}_n - \mu_0\right)}{\tilde{S}_n}\right).$$

$$\alpha \geq 1 - \Phi\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\tilde{S}_n}\right)$$

Consistency

- While power in finite samples is (hopefully) larger than size when H_0 is false, in our examples, any departure from the null will be detected with probability approaching 1.
- Let $\alpha \in (0, 1)$. We say a sequence of tests ϕ_n is <u>consistent</u> if for any $\theta \notin \Theta_0$,

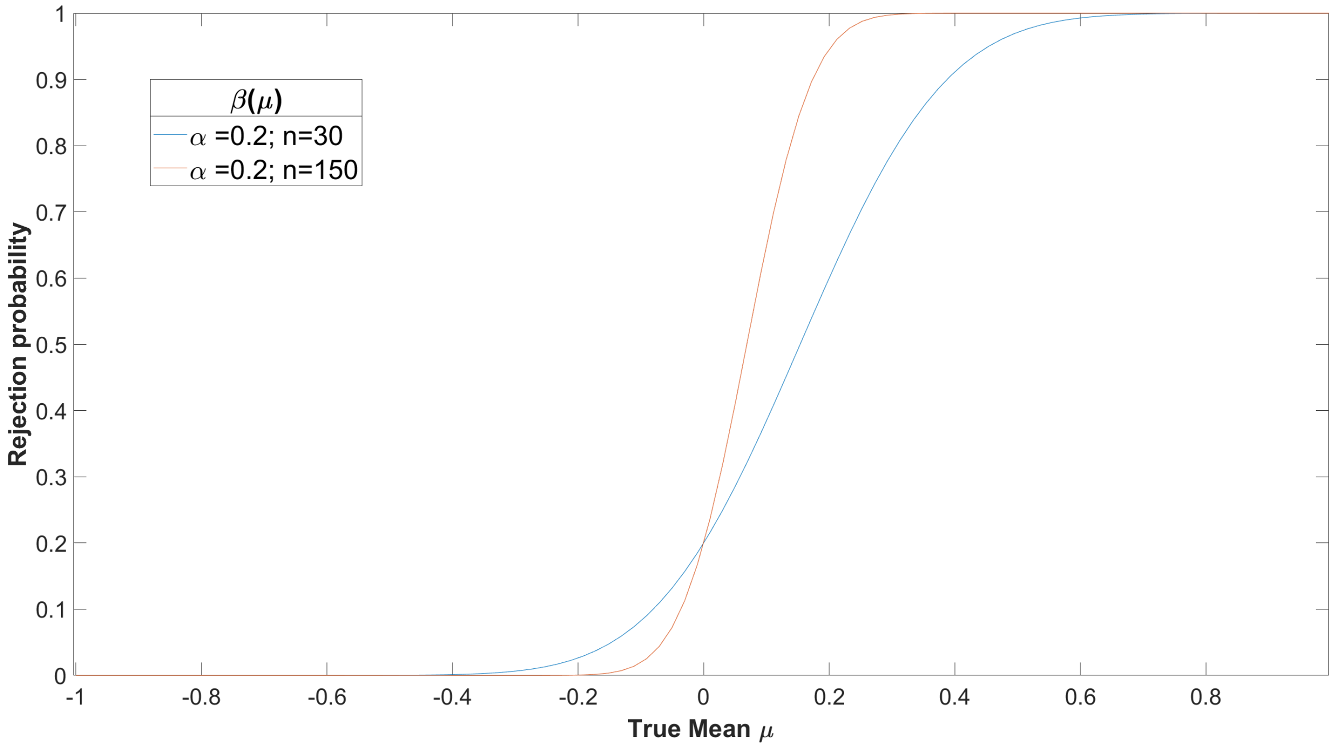
$$\beta_n(\theta) \to 1 \quad \text{as } n \to \infty.$$

- Suppose $\{X_i\}_{i=1}^n$ is an iid sample from $\mathcal{N}(\mu, 1)$. We wish to test $H_0 : \mu \leq 0$, so $\Theta_0 = (-\infty, 0]$.
- For a test of size α, we derived $c^*(\alpha) = \underbrace{\Phi^{-1}(1-\alpha)}_{z_{1-\alpha}} / \sqrt{n}$, so

$$\begin{aligned} \beta_n(\mu) &= 1 - \Phi\left(\sqrt{n}(c^* - \mu)\right). \\ &= 1 - \Phi\left(z_{1-\alpha} - \mu\sqrt{n}\right) \end{aligned}$$

Consistency

- As the sample size increases, test is better able to detect small departures from H_0 since test has higher power for a given μ:



Questions?