A Primer in Econometric Theory

Lecture 9: Confidence Intervals and Tests

John Stachurski
Lectures by Akshay Shanker

March 26, 2017

Focus till now has been on point estimation

Data are not infinite — point estimation involves uncertainty

In this lecture:

- quantify estimation uncertainty using the language of confidence sets
- cover hypothesis tests, which are a second key method of inference in classical statistics

We'll work with parametric language

Similar notation as previously in Ch. 8, except:

1. the marginal distribution P of each observation $\mathbf{z}_n$ is written as P_θ
2. the universe $\mathscr{P}$ of possible distributions is $\{P_\theta : \theta \in \Theta\}$
3. the joint distribution of the data $\mathbf{z}_{\mathcal{D}} = (\mathbf{z}_1, \ldots, \mathbf{z}_N)$ is written as $P_\theta^{\mathcal{D}}$

Not restricting ourselves to the case of finitely many parameters

- think of θ as an index over the set of distributions

Confidence Sets

Confidence sets are random subsets of model space that contain the true data generating process with high probability

Finite Sample Confidence Sets

Fix $\alpha \in (0, 1)$

A random set $C(\mathbf{z}_\mathcal{D}) \subset \Theta$ is called a $1 - \alpha$ **confidence set** for θ if $C(\mathbf{z}_\mathcal{D})$ is observable given the data and, for any $\theta \in \Theta$,

$$\mathbb{P}\{\theta \in C(\mathbf{z}_\mathcal{D})\} \geq 1 - \alpha \quad \text{whenever} \quad \mathcal{L}(\mathbf{z}_\mathcal{D}) = P_\theta^\mathcal{D}$$

Note that it's the set that's random here, not θ

Suppose we want to estimate the mean of an unknown distribution based on a set of observations $\mathbf{z}_{\mathcal{D}}$

A 95% confidence set $C(\mathbf{z}_{\mathcal{D}})$ will contain the mean in about 95% of our experiments, *regardless of the underlying distribution*

We don't know what θ is, and so the set *must be designed* such that the probability of containing θ exceeds $1 - \alpha$ regardless of which θ is generating the data

The difficulty is to actually construct a confidence set with this property for a reasonably large class of distributions

A traditional approach that uses strong parametric assumptions:

Example. Let $\mathbf{z}_{\mathcal{D}} = \mathbf{x} = (x_1, \dots, x_N)$ where each x_n is an independent draw from $\mathrm{N}(\mu, \sigma^2)$

Suppose for now that only $\mu \in \mathbb{R}$ is unknown

We wish to form a confidence interval for μ. Since $\mathcal{L}(\bar{x}_N) = \mathrm{N}\left(\mu, \frac{\sigma^2}{N}\right)$, we have

$$\mathcal{L}\left[\sqrt{N}\frac{(\bar{x}_N - \mu)}{\sigma}\right] = \mathrm{N}(0,1) \tag{1}$$

Example. (cont.) Above statement true regardless of the values of μ and σ

Recall that if $\mathcal{L}(x) = F$, x has a symmetric density and F is strictly increasing, then

$$c = F^{-1}(1 - \alpha/2) \quad \implies \quad \mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha$$

As such,

$$\mathbb{P}\left\{ \frac{\sqrt{N}}{\sigma} |\bar{x}_N - \mu| \leq z_{\alpha/2} \right\} = 1 - \alpha$$

when $\quad z_{\alpha/2} := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

Here Φ is the standard normal CDF

Example. (cont.) Rearranging the above gives

$$\mathbb{P}\left\{\bar{x}_N - \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \leq \mu \leq \bar{x}_N + \frac{\sigma}{\sqrt{N}} z_{\alpha/2}\right\} = 1 - \alpha$$

The above argument is true regardless of the value of μ. We then conclude

$$C(\mathbf{x}) := (\bar{x}_N - e_n, \bar{x}_N + e_n) \quad \text{with} \quad e_n := \frac{\sigma}{\sqrt{N}} z_{\alpha/2}$$

is a $1 - \alpha$ confidence interval for μ. (Note that $C(\mathbf{x})$ is indeed observable under our assumption that σ is known.)

In the above example, assumption that σ is known is clearly unrealistic

If σ is unknown, replace the term with the sample standard deviation s_N. In doing so, we will use the following fact

Fact. (10.1.1) If $x_1, \ldots, x_N$ are IID draws from $\mathrm{N}(\mu, \sigma^2)$, then

$$\sqrt{N-1}\frac{(\bar{x}_N - \mu)}{s_N} \tag{2}$$

has a Student's t-distribution with $N-1$ degrees of freedom

The random variable (2) is called a **pivot**, its distribution does not depend on the unknown parameters

Example. Continue with setting of the previous example

Recall $\mathbf{z}_{\mathcal{D}} = \mathbf{x} = (x_1, \ldots, x_N)$ where each x_n is an independent draw from $\mathrm{N}(\mu, \sigma^2)$

Now let both σ and μ be unknown

Let F_{N-1} be the CDF of the t-distribution with $N-1$ degrees of freedom. Using the same reasoning as in the above example, with $z_{\alpha/2}$ replaced by $t_{\alpha/2} := F_{N-1}^{-1}(1 - \alpha/2)$, we obtain

$$\mathbb{P}\left\{\bar{x}_N - \frac{s_N}{\sqrt{N-1}} t_{\alpha/2} \leq \mu \leq \bar{x}_N + \frac{s_N}{\sqrt{N-1}} t_{\alpha/2}\right\} = 1 - \alpha$$

Recall that the standard deviation of $\bar{x}_N$ is $\sigma/\sqrt{N}$

The term $s_N/\sqrt{N-1}$ is a sample estimate of $\sigma/\sqrt{N}$

Call $s_N/\sqrt{N-1}$ the standard error of $\bar{x}_N$ and write it as $\mathrm{se}(\bar{x}_N)$

Now write the confidence interval for μ in the previous example as

$$C(\mathbf{x}) := (\bar{x}_N - \mathrm{se}(\bar{x}_N) t_{\alpha/2},\ \bar{x}_N + \mathrm{se}(\bar{x}_N) t_{\alpha/2}) \tag{3}$$

Asymptotic Methods

Given $\alpha \in (0,1)$, a set $C_N(\mathbf{z}_{\mathcal{D}}) \subset \Theta$ is called an **asymptotic $1-\alpha$ confidence set** for θ if $C_N(\mathbf{z}_{\mathcal{D}})$ is observable given the data and

$$\lim_{N \to \infty} \mathbb{P}\{\theta \in C_N(\mathbf{z}_{\mathcal{D}})\} \geq 1 - \alpha$$

for all $\theta \in \Theta$

Note $C_N(\mathbf{z}_{\mathcal{D}})$ is a sequence of sets and the definition concerns this sequence

Consider first the case of scalar θ

Suppose we have an estimator $\hat{\theta}_N$ of θ that is asymptotically normal for all $\theta \in \Theta$

For all $\theta \in \Theta$, there exists a positive constant $v(\theta)$ such that

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} \mathrm{N}(0, v(\theta)) \quad \text{as} \quad N \to \infty \tag{4}$$

(Recall results in §9.2.4.1 where we discuss asymptotic properties of estimators and specialise to the scalar case)

The constant $v(\theta)$ is called the asymptotic variance of $\hat{\theta}_N$

Suppose we have a sequence of statistics $\mathrm{se}(\hat{\theta}_N)$ such that

$$\sqrt{N}\,\mathrm{se}(\hat{\theta}_N) \xrightarrow{p} \sqrt{v(\theta)} \quad \text{as} \quad N \to \infty \tag{5}$$

Next, (4) and (5) imply

$$\frac{\hat{\theta}_N - \theta}{\mathrm{se}(\hat{\theta}_N)} \xrightarrow{d} \mathrm{N}(0, 1) \quad \text{as} \quad N \to \infty \tag{6}$$

(Exercise 10.4.2)

We can now create our asymptotic confidence intervals as

$$C_N := (\hat{\theta}_N - \mathrm{se}(\hat{\theta}_N) z_{\alpha/2}, \hat{\theta}_N + \mathrm{se}(\hat{\theta}_N) z_{\alpha/2})$$

To see this, take the limit and apply (6) to get

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}\{\theta \in C_N(\mathbf{z}_{\mathcal{D}})\} &= \lim_{N\to\infty} \mathbb{P}\{\hat{\theta}_N - \mathrm{se}(\hat{\theta}_N) z_{\alpha/2} \leq \theta \leq \hat{\theta}_N \\
&\quad + \mathrm{se}(\hat{\theta}_N) z_{\alpha/2}\} \\
&= \lim_{N\to\infty} \mathbb{P}\left\{-z_{\alpha/2} \leq \frac{\hat{\theta}_N - \theta}{\mathrm{se}(\hat{\theta}_N)} \leq z_{\alpha/2}\right\} \\
&= 1 - \alpha
\end{aligned}$$

Example. Let $\bar{x}_N$ be the sample mean of IID data $\{x_n\}$

Suppose that these data come from some distribution with finite second moment

Let μ denote the common mean and let σ^2 denote the variance

Let s_N be the sample standard deviation

Example. (cont.) Combine the central limit theorem and consistency of s_N for σ (see page 263):

$$\sqrt{N}(\bar{x}_N - \mu) \xrightarrow{d} \mathrm{N}(0, \sigma^2) \quad \text{and} \quad \sqrt{N}\,\mathrm{se}(\bar{x}_N) \xrightarrow{p} \sigma$$

$$\text{for} \quad \mathrm{se}(\bar{x}_N) := \frac{s_N}{\sqrt{N}}$$

With this definition of $\mathrm{se}(\bar{x}_N)$, the set

$$(\bar{x}_N - \mathrm{se}(\bar{x}_N) z_{\alpha/2}, \, \bar{x}_N + \mathrm{se}(\bar{x}_N) z_{\alpha/2})$$

is an asymptotic $1 - \alpha$ confidence interval for $\bar{x}_N$

Non-Parametric Example

In §8.1.3 of ET we learned that if $\mathbf{x} = (x_1, \ldots, x_N)$ is a vector of IID draws from some CDF F and $\hat{F}_N$ is the corresponding ECDF, then $\|F - \hat{F}_N\|_\infty$ converges in probability to 0

In 1933, A. N. Kolmogorov used an extension of the central limit theorem to obtain an asymptotic distribution for this term

In particular, he showed that when F is continuous,

$$\sqrt{N} \sup_{s \in \mathbb{R}} \left| \hat{F}_N(s) - F(s) \right| \xrightarrow{d} K \tag{7}$$

where K is the **Kolmogorov** CDF

$$K(s) := \frac{\sqrt{2\pi}}{s} \sum_{i=1}^{\infty} \exp\left[- \frac{(2i-1)^2 \pi^2}{8s^2} \right] \qquad (s \geq 0)$$

As in the CLT, the limiting distribution K is independent of the CDF F that generates the data

We can use Kolmogorov's result, Equation (7), to produce an asymptotic $1-\alpha$ confidence set for F

Let $\mathfrak{F}$ be the set of all CDFs on $\mathbb{R}$, let $x_1, \dots, x_N$ be IID draws from $F \in \mathfrak{F}$, let $k_{1-\alpha} := K^{-1}(1-\alpha)$, and let

$$C_N(\mathbf{x}) := \left\{ G \in \mathfrak{F} \, : \, \hat{F}_N(s) - \frac{k_{1-\alpha}}{\sqrt{N}} \leq G(s) \leq \hat{F}_N(s) + \frac{k_{1-\alpha}}{\sqrt{N}} \text{ for all } s \in \mathbb{R} \right\}$$

The set $C_N(\mathbf{x}) \subset \mathfrak{F}$ is an asymptotic $1-\alpha$ confidence set for F

Indeed, after rearranging the expression, we get

$$F \in C_N(\mathbf{x}) \iff -k_{1-\alpha} \leq \sqrt{N}(\hat{F}_N(s) - F(s)) \leq k_{1-\alpha} \text{ for all } s$$

$$\iff \sqrt{N}|\hat{F}_N(s) - F(s)| \leq k_{1-\alpha} \text{ for all } s$$

$$\iff \sup_s \sqrt{N}|\hat{F}_N(s) - F(s)| \leq k_{1-\alpha}$$

Apply (7) to confirm our claim:

$$\lim_{N \to \infty} \mathbb{P}\{F \in C_N(\mathbf{x})\}$$

$$= \lim_{N \to \infty} \mathbb{P}\left\{\sup_s \sqrt{N}|\hat{F}_N(s) - F(s)| \leq k_{1-\alpha}\right\}$$

$$= 1 - \alpha$$

Present the confidence set $C_N(\mathbf{x})$ visually:

- highlighting the area between the lower bound $\hat{F}_N(s) - k_{1-\alpha}/\sqrt{N}$ and the upper bound $\hat{F}_N(s) + k_{1-\alpha}/\sqrt{N}$ over s

Consider data drawn from a t-distribution F with 10 degrees of freedom and α is set to 0.05...

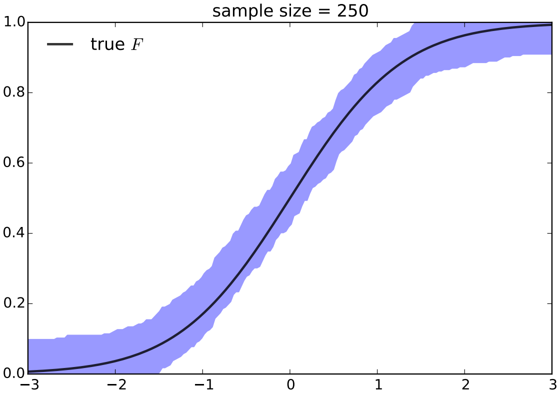

Figure: Confidence set for the ECDF with 250 observations

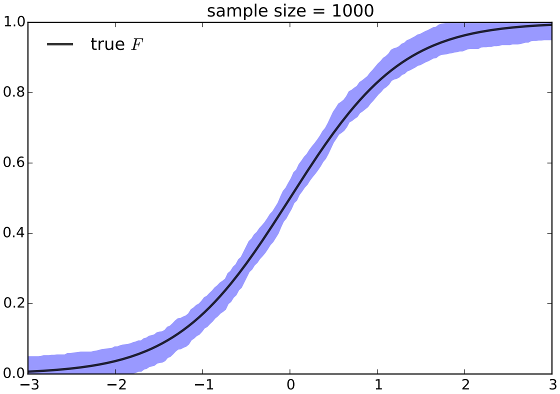

Figure: Confidence set for the ECDF with 1,000 observations

Hypothesis Testing

This section considers problems of inference where we

1. hold a belief or theory concerning the probabilities generating the data and
2. consider whether the data provides evidence for or against that theory

Example. Asset price returns data tend to have heavier tails than the normal distribution

For certain assets, normality might still be a reasonable and convenient approximation

Let's look at whether or not normality provides a reasonable fit in the case of daily returns on the Nikkei 225

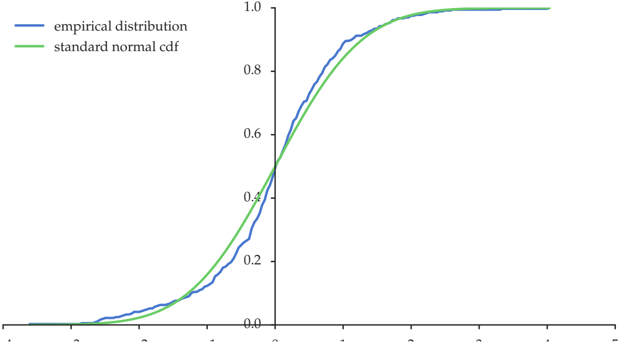

Figure: ECDF for standardized returns versus the standard normal CDF

While the ECDF of the data and the CDF are not identical, we wouldn't expect this even if the hypothesis of normality is completely true

Question: is the sample "unlikely" given the hypothesis of normality?

Null Hypothesis

Hypothesis testing begins with a **null hypothesis**:

- a statement that the observed data are being generated by a certain model, or one of a certain class of models

A hypothesis test is an attempt to reject the null

Black Swan Karl Popper (1902–1994):

- claim: *all* swans are white
- no quantity of white swan sightings can confirm the claim
- a single black swan can show the claim to be false

Attempting to falsify a theory on the basis of observation is more constructive than attempting to confirm it

Reject the null hypothesis only if strong evidence against it is observed – we don't want to mistakenly discard a good theory

Constructing Tests

We will use notation we used to discuss confidence intervals

A null hypothesis is a specification of a set of models that we believe the data-generating process belongs to

This amounts to specifying a subset Θ_0 of Θ

The null hypothesis often written as

$$H_0 : \theta \in \Theta_0$$

If Θ_0 is a singleton, then the null hypothesis called a **simple hypothesis**

If not, then the null hypothesis called a **composite hypothesis**

A test of the null hypothesis is a test of whether or not observed data generated by P_θ for some $\theta \in \Theta_0$

Example. The hypothesis of purchasing power parity is tested by considering models such as

$$p = a + \beta e p^* + \sigma u$$

where p is a price or index of domestic prices, e is an exchange rate, p^* is a corresponding foreign price, u is a disturbance term and a, β, σ are parameters

The null hypothesis for absolute purchasing power parity is that (35) holds with $a = 0$ and $\beta = 1$

Since σ is not pinned down by the null, this is a composite null hypothesis

Rejection Region

Formally, a **test** of H_0 is a binary function ϕ mapping the observed data $\mathbf{z}_\mathcal{D}$ into $\{0, 1\}$

The decision rule:

$$\text{if } \phi(\mathbf{z}_\mathcal{D}) = 1, \text{ then reject } H_0$$

$$\text{if } \phi(\mathbf{z}_\mathcal{D}) = 0, \text{ then do not reject } H_0$$

(Failing to reject H_0 should not be confused with accepting H_0. More on this below.)

Our aim will be to design ϕ such that it takes the value 1 when the data present strong evidence against H_0

Example. Let $\mathbf{z}_{\mathcal{D}}$ be a set of pairs $\mathbf{z}_n = (x_n, y_n)$ from some unknown bivariate distribution P_θ

If H_0 is the hypothesis that the correlation between x and y is negative, then a large positive sample correlation would constitute evidence against H_0

Hence our test might take the form $\phi(\mathbf{z}_{\mathcal{D}}) = \mathbb{1}\{\hat{\varrho} > c\}$, where $\hat{\varrho}$ is the sample correlation

An appropriate value of c remains to be determined

Type I and Type II error

Two different ways in which the realization of *random* $\mathbf{z}_{\mathcal{D}}$ can mislead us

1. mistakenly reject the null hypothesis when it is in fact true — **type I error**
2. fail to reject the null hypothesis when it is false — **type II error**

The **power function** associated with the test ϕ is the function

$$\beta(\theta) := \mathbb{P}_\theta\{\phi(\mathbf{z}_{\mathcal{D}}) = 1\} \qquad (\theta \in \Theta)$$

The symbol $\mathbb{P}_\theta$ indicates we are computing probabilities under the assumption that $\mathcal{L}(\mathbf{z}_{\mathcal{D}}) = P_\theta^{\mathcal{D}}$

- $\beta(\theta)$ is the probability the test rejects when the data are generated by the probabilistic model identified by θ

Ideally, $\beta(\theta) = 0$ when $\theta \in \Theta_0$, and $\beta(\theta) = 1$ when $\theta \notin \Theta_0$ — in practice, difficult to achieve

We want to be conservative in rejecting the null

- traditional to keep the probability of type I error small

Standard procedure is to choose a small number α such as 0.05 or 0.01, and then adjust the test such that

$$\beta(\theta) \leq \alpha \quad \text{for all } \theta \in \Theta_0 \tag{8}$$

If (8) holds, then the test is said to be of **size** α

The **size** of any test with power function β is

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

This is the maximal rejection probability when the null hypothesis is true

Critical Values and Tests

In constructing tests, common to define a real-valued **test statistic** T and a **critical value** c, and then set

$$\phi(\mathbf{z}_{\mathcal{D}}) := \mathbb{1}\{T(\mathbf{z}_{\mathcal{D}}) > c\} \tag{9}$$

The pair (T, c) then defines the test, and the rule becomes:

$$\text{reject } H_0 \text{ if and only if } T(\mathbf{z}_{\mathcal{D}}) > c$$

Example. Suppose $x_1, \ldots, x_N$ are independent draws from $\mathrm{N}(\mu, 1)$ where the value of μ is unknown

The null hypothesis is $\mu \leq 0$, or $\Theta_0 = (-\infty, 0]$

Since we want to make inference about the mean, a natural choice for our test statistic is the sample mean. Thus

$$T(\mathbf{z}_{\mathcal{D}}) =: T(x_1, \ldots, x_N) := \bar{x}_N$$

Each $c \in \mathbb{R}$ gives a test via (9) on the previous slide, with power function $\beta(\mu) = \mathbb{P}_{\mu}\{\bar{x}_N > c\}$

Example. (cont.) To evaluate β, recall $\mathcal{L}(\bar{x}_N) = \mathrm{N}(\mu, 1/N)$

As a result, if Φ is the CDF of the standard normal distribution and $\mathcal{L}(z) = \Phi$, then

$$\mathbb{P}_\mu\{\bar{x}_N \leq c\} = \mathbb{P}\{\mu + N^{-1/2} z \leq c\}$$

$$= \mathbb{P}\{z \leq N^{1/2}(c - \mu)\} = \Phi[N^{1/2}(c - \mu)]$$

$$\therefore \quad \beta(\mu) = 1 - \Phi[N^{1/2}(c - \mu)] \tag{10}$$

Given c, the power function is increasing in μ because higher μ pushes up the mean of $\bar{x}_N$, making the event $\{\bar{x}_N > c\}$ more likely

Given μ, the function is decreasing in c, because higher c makes the event $\{\bar{x}_N > c\}$ less likely

Plots of the power function β in (10)

- two different values of c
- N is fixed at 10.
- since $\Theta_0 = (-\infty, 0]$, the size of the test in each case is $\sup_{\mu \leq 0} \beta(\mu)$. Since the power curves are increasing, this is just $\beta(0)$

Typical trade-off between type I and type II error. If we increase c,

- we make rejection less likely for all values of μ
- push down type I error, but also increase the probability we fail to reject a false null

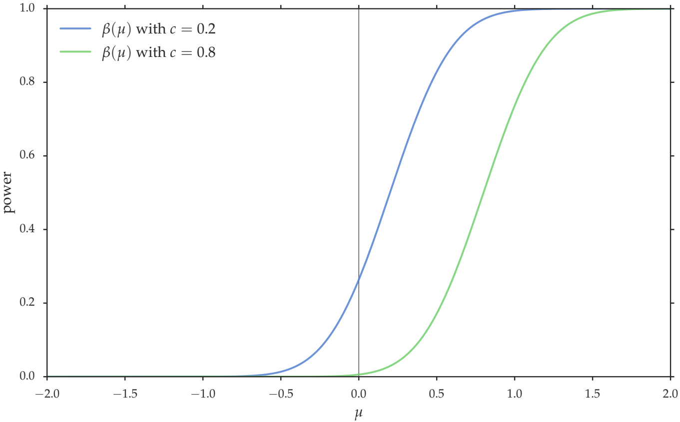

Figure: Power function β

Choosing Critical Values

Typically, the choice of test statistic T suggested by the problem

- example: if our hypothesis is a statement about the second moment of a random variable, then take T to be the sample second moment

Once T is fixed, we need to adjust the critical region c such that (T, c) attains the appropriate size

Standard procedure:

1. choose a desired size α according to our tolerance for type I error
2. identify a suitable test statistic T
3. choose a critical value c so that (T, c) is of size α

In performing the last step, balance our desire to minimize type II error while maintaining a size α — choose c to solve

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta\{T(\mathbf{z}_\mathcal{D}) > c\} \tag{11}$$

Figure on next slide:

- $\Theta_0 = \{\theta_a, \theta_b\}$
- blue line is distribution of $T(\mathbf{z}_\mathcal{D})$ (density) when $\mathbf{z}_\mathcal{D}$ generated by θ_a
- black line is the same for θ_b
- choose c such that the largest of the two shaded areas is equal to α

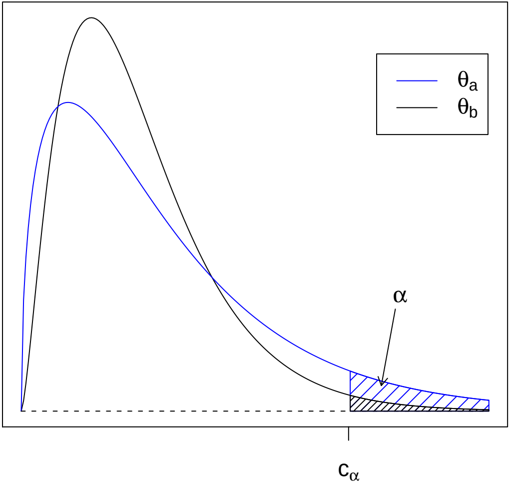

Figure: Determining the critical value

Example. Consider again the previous example. Recall $x_1, \ldots, x_N \overset{\text{IID}}{\sim} \mathrm{N}(\mu, 1)$ for μ unknown, and our null hypothesis is $\mu \leq 0$

Given α, our task is to find the appropriate critical value c so that the test (T, c) is of size α

To solve for c given α, we choose c to solve

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta\{T(\mathbf{z}_\mathcal{D}) > c\} \tag{12}$$

Recall the expression for the power function for this setting that we calculated in the previous example

$$\beta(\mu) = 1 - \Phi[N^{1/2}(c - \mu)] \tag{13}$$

Example. (cont.) We then choose c to solve

$$\alpha = \sup_{\mu \leq 0} \{1 - \Phi[N^{1/2}(c - \mu)]\}$$

The right-hand side is increasing in μ, so the supremum is obtained by setting $\mu = 0$

Setting $\mu = 0$ and solving for c, we obtain

$$c(\alpha) := N^{-1/2} \Phi^{-1}(1 - \alpha)$$

where Φ^{-1} is the quantile function of the standard normal distribution

Example. (cont.)

Since Φ^{-1} is increasing:

- smaller α corresponds to larger $c(\alpha)$ — we reduce the probability of type I error by increasing the critical value the mean $\bar{x}_N$ must obtain for rejection to occur
- higher N brings down $c(\alpha)$, without increasing the probability of rejecting a true null

Asymptotic Tests

In many cases, we know little about the distribution of the test statistic

One approach is to determine the asymptotic distribution of the test statistic

- CLT: pin down distributions without actually assuming a parametric structure

Notation β_N emphasizes that the power function depends on sample size

A test is called **asymptotically of size α** if

$$\lim_{N\to\infty} \beta_N(\theta) \leq \alpha \quad \text{for all } \theta \in \Theta_0$$

Test for the Empirical Distribution

Return to the data on standardized daily returns on the Nikkei 225

Let Φ be the standard normal CDF as before

Let H_0 be that standardized returns are IID draws from Φ

Let α be given, and let $k_{1-\alpha} = K^{-1}(1-\alpha)$ be the $1-\alpha$ quantile of the Kolmogorov distribution K:

$$K(s) := \frac{\sqrt{2\pi}}{s} \sum_{i=1}^{\infty} \exp\left[-\frac{(2i-1)^2 \pi^2}{8s^2} \right] \qquad (s \geq 0)$$

Finally, let $\hat{F}_N$ be the ECDF of the data

If the null hypothesis is true, then

$$\sqrt{N} \sup_{s \in \mathbb{R}} |\hat{F}_N(s) - \Phi(s)| \xrightarrow{d} K \tag{14}$$

For the test

$$\phi_N(\mathbf{x}) := \mathbb{1} \left\{ \sqrt{N} \sup_{s \in \mathbb{R}} |\hat{F}_N(s) - \Phi(s)| > k_{1-\alpha} \right\}$$

let $\beta_N(\Phi)$ be the value of the power function when the null hypothesis is true. By (14),

$$\lim_{N \to \infty} \beta_N(\Phi) = \lim_{N \to \infty} \mathbb{P} \left\{ \sqrt{N} \sup_{s \in \mathbb{R}} |\hat{F}_N(s) - \Phi(s)| > k_{1-\alpha} \right\} = \alpha$$

And the test is asymptotically of size α

The Nikkei data set used here has 364 observations

The value of the test statistic $\sqrt{N} \sup_{s \in \mathbb{R}} |\hat{F}_N(s) - \Phi(s)|$ is 1.548

If $\alpha = 0.05$, then the critical value $k_{1-\alpha}$ is 1.36 (recall our discussion on the Kolmogorov CDF above or in §10.1.3 of ET)

Hence the test statistic exceeds the critical value, and the null hypothesis is rejected

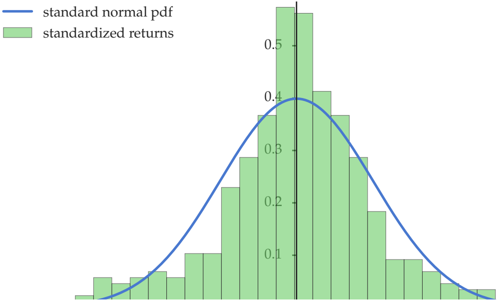

Figure: Standardized daily returns for the Nikkei 225

P-values

Typically, a test that rejects at size 0.05 will also reject at size 0.1, but may not reject at size 0.01

Lower α means less tolerance for type I error, and forces the critical value to become larger

The p-value is the smallest value of α at which we can still reject given a test statistic

The null hypothesis is $H_0 : \theta \in \Theta_0$ and, for each $\alpha \in (0,1)$, a test $(T, c(\alpha))$ of size α

Assume $c(\alpha)$ determined to solve

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta\{T(\mathbf{z}_\mathcal{D}) > c\}$$

In this setting, the p-value of the test defined as

$$p(\mathbf{z}_\mathcal{D}) := \inf\{\alpha \in (0,1) \, : \, c(\alpha) < T(\mathbf{z}_\mathcal{D})\}$$

When $\alpha \mapsto c(\alpha)$ is continuous, the expression for $p(\mathbf{z}_\mathcal{D})$ reduces to

$$p(\mathbf{z}_\mathcal{D}) := \text{ the } \alpha \text{ such that } c(\alpha) = T(\mathbf{z}_\mathcal{D})$$

Example. Let $x_1, \ldots, x_N$ be an IID sample with mean θ and variance σ^2

Assume both θ and σ are unknown

We wish to test the hypothesis $H_0 \colon \theta = \theta_0$

Consider the statistic

$$t_N := \sqrt{N} \left\{ \frac{\bar{x}_N - \theta_0}{s_N} \right\} = \frac{\bar{x}_N - \theta_0}{\mathrm{se}(\bar{x}_N)}$$

where $\mathrm{se}(\bar{x}_N) := \frac{s_N}{\sqrt{N}}$

Expression converges in distribution to a standard normal

Example. (cont.) Hence

$$\phi_N(\mathbf{x}) := \mathbb{1}\{|t_N| > z_{\alpha/2}\}$$

is asymptotically of size α (see ex. 10.4.3 in ET)

To evaluate the p-value, $c(\alpha) := \Phi^{-1}(1 - \alpha/2)$, and $c(\alpha)$ is continuous in α, so we can apply the following definition of $p(\mathbf{z}_{\mathcal{D}})$

$$p(\mathbf{z}_{\mathcal{D}}) := \text{ the } \alpha \text{ such that } c(\alpha) = T(\mathbf{z}_{\mathcal{D}})$$

To solve for $p(\mathbf{z}_{\mathcal{D}})$, solve for α in the expression

$$\Phi^{-1}(1 - \alpha/2) = |t_N(\mathbf{z}_{\mathcal{D}})|$$

Rearrange to obtain

$$p(\mathbf{z}_{\mathcal{D}}) = 2\Phi(-|t_N(\mathbf{z}_{\mathcal{D}})|)$$

Accepting the Null?

Failure to reject the null not necessarily evidence in favour of the null

Economists discovered that a newly developed test due to Dickey and Fuller (1979) did not reject the null hypothesis of a unit root for a variety of economic time series

- called into question earlier time series methods that assumed trend stationarity
- The idea that innovations to major economic time series contain a permanent component became something like a stylized fact

Specialisze to unemployment rates

- unit root hypothesis was associated with the concept of hysteresis

Unit root null can be expressed as the hypothesis $a = 1$ in the AR(1) process

$$u_{t+1} = au_t + b + \epsilon_{t+1} \tag{15}$$

Here $\{u_t\}$ is unemployment, a and b are parameters and $\{\epsilon_t\}$ is a zero-mean innovation

Standard one-sided Dickey–Fuller test of the unit root null:

1. estimate the parameters as $\hat{a}$ and $\hat{b}$ by least squares
2. reject the null if $T(\hat{a} - 1) < c$, where c is a critical value and T is length of the sample

Dickey and Fuller (1979) showed the test statistic converges in distribution under the null and tabulated critical values for different test sizes

A number of studies have been unable to reject a unit root null for unemployment rate data

Interpretation?

- (15) is a good model for the data with $a = 1$
- Another possibility: data generated by another process against which this test has little power

The second possibility seems more plausible than the first — employment rates don't diverge

Adopt a nonlinear model

$$u_{t+1} = h(u_t) + \epsilon_{t+1} \tag{16}$$

where h is the generalized logistic function

- for unemployment rates in a band between about 5 and 15, the process exhibits strong persistence
- for values above and below this number the shape of h causes drift back to the band — "equilibrium forces"

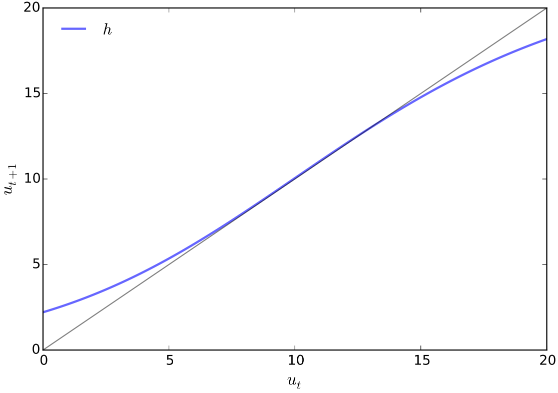

Figure: A nonlinear process for unemployment dynamics

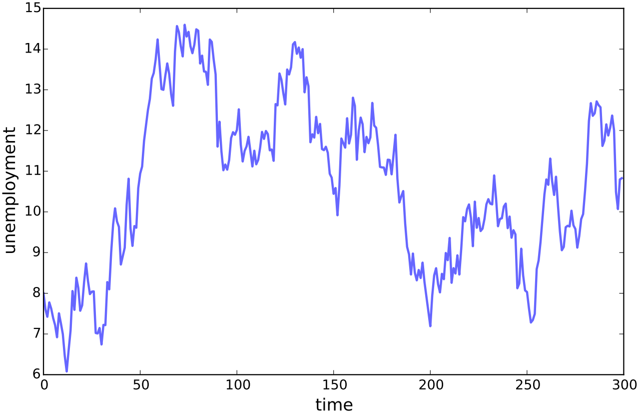

Figure: Simulated unemployment dynamics

Listing 7 in ET gives code for a simulation

- shocks $\{\epsilon_t\}$ are independent
- time series of length 100 is drawn and the Dickey–Fuller test statistic is calculated
- size of the test set to 0.05
- critical value for the test becomes -13.7

Rejection frequency over 5,000 repetitions is around 0.05 – same rejection frequency as when null is true

Yet, nonlinear data-generating process has very different properties from the null: stationary and uniformly ergodic