Lecture 6: Statistics

Robert Garrard

Estimation

Statistical questions typically begin in the same place:

$$\text{Let } X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} F$$

be an independent and identically distributed (iid) sample from a distribution F.

F is called the **data generating process**.

We usually want to estimate a property of F based on the sample.

Often we may assume F belongs to a particular family, such as $F = N(\theta, \sigma^2)$, for which we need to estimate a parameter. This is called **parametric** statistics.

Other times we may make no assumptions about the distribution F, this is called **non-parametric** statistics.

Estimation

Let F be a distribution with parameter θ. An **estimator** is a function of a sample from F.

$$\hat{\theta} = T(X_1, \dots, X_n; F)$$

Since it is a function of a random sample, the estimator is itself a random variable.

Ideally, we want to use estimators whose distributions have nice properties.

Estimation

An estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$.

Let $X_1, \ldots, X_n \sim F$ be an iid sample from a distribution with unknown mean $\mathbb{E}[X] = \mu$.

Let the **sample mean** be the estimator:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Exercise 6.1 Show that $\hat{\theta} = \bar{X}$ is unbiased for μ. Is $\hat{\theta} = X_1$ unbiased?

Estimation

Let $X_1, \ldots, X_n \sim F$ be an iid sample from a distribution with unknown variance $\mathbb{E}[(X-\mu)^2] = \sigma^2$.

Let the sample variance be the estimator

$$s^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

Exercise 6.2

1. Is the sample variance unbiased for σ^2?
2. What about $\frac{1}{n} \sum_i (X_i - \bar{X})^2$?
3. Is s unbiased for σ?

Convergence of Random Variables

Consider a sequence of tosses of a fair coin with the random variable

$$X_i = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$$

After each toss, compute the mean $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. This generates a sequence

$$\{\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots\}$$

Question: Does this sequence converge? (in the ε-ball sense)

Convergence of Random Variables

We say that a sequence of random variables $\{X_n\}$ **converges in probability** to X, written $X_n \xrightarrow{p} X$, if $\forall \varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

We say that a sequence of random variables $\{X_n\}$ **converges almost surely** to X, written $X_n \xrightarrow{a.s.} X$, if

$$\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$$

We say that a sequence of random variables $\{X_n\}$ **converges in distribution** to X, written $X_n \xrightarrow{d} X$, if

$$F_n(x) \to F(x) \quad \forall x$$

Convergence almost surely $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in distribution

Convergence of Random Variables

Exercise 6.3 Show that the mean in the coin tossing example above converges in probability to $1/2$.

Convergence of Random Variables

When trying to prove certain convergence properties, two theorems will be particularly useful:

Theorem (Continuous Mapping Theorem)

Let $g(\cdot)$ be a continuous function. Then

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$
$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$
$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

Consequently, the usual arithmetic operations preserve convergence. And

$$\mathbf{A}_n \to \mathbf{A} \Rightarrow \mathbf{A}_n^{-1} \to \mathbf{A}^{-1}$$

Convergence of Random Variables

Theorem (Slutsky's Theorem)

Let $X_n \xrightarrow{d} X$ *and* $Y_n \rightarrow c$.

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} Xc$$

$$X_n / Y_n \xrightarrow{d} X/c, \textit{provided } c \not 0 \ .$$

In addition, if $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$ and $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$, then $\mathbf{A}_n \mathbf{x}_n \xrightarrow{d} \mathbf{A}\mathbf{x}$.

Consistency

An estimator, $\hat{\theta}$, is **consistent** for a parameter θ if $\hat{\theta} \xrightarrow{p} \theta$.

Exercise 6.4 Let $X_1, \ldots, X_n \sim F$ be an iid sample drawn from a distribution with finite mean $\mu = \mathbb{E}[X]$ and variance σ^2. Show that the sample mean, $\bar{X}$ is a consistent estimator of the population mean, μ.

Laws of Large Numbers

Theorem (Khinchin's weak law of large numbers)

Let $X_1, \ldots, X_n \sim F$ be an iid sample from a distribution with mean μ.

$$\bar{X}_n = \frac{1}{n} \sum X_i \xrightarrow{p} \mu$$

Theorem (Kolmogorov's strong law of large numbers)

Let $X_1, \ldots, X_n \sim F$ be an iid sample from a distribution with finite mean μ and variance σ^2.

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

Laws of Large Numbers

Exercise 6.5 Let $X_1, \ldots, X_n \sim F$, $\mathrm{Var}(X) = \sigma^2 < \infty$. Show that $s = \sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2} \xrightarrow{p} \sigma$.

Sampling Distribution

Recall that a statistic is itself a random variable and has a corresponding probability distribution.

However, the statistic may have a different distribution depending on the sample size.

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from distribution F. Let $\hat{\theta} = \hat{\theta}(\mathbf{X})$ be a statistic. The **sampling distribution** of $\hat{\theta}$ is

$$F_{\theta_n}(\tau) = \mathbb{P}_n(\hat{\theta}(\mathbf{X}) \leq \tau)$$

Sampling Distribution

Proposition

Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. The sample mean, $\bar{X}$, has the following sampling distribution.

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$$

or equivalently

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

Proof is a tutorial exercise.

Central Limit Theorems

Usually we won't be able to characterize the finite sample distribution of a statistic exactly.

Instead, we may be able to approximate it with an **asymptotic** distribution.

Theorem (Lindeberg-Lévy CLT)

Let $X_1, \ldots X_n \overset{iid}{\sim} F$, *with* $\mathbb{E}[X] = \mu$ *and* $\mathrm{Var}(X) = \sigma^2$, *both finite. Let* $\bar{X}_n = \frac{1}{n} \sum X_i$.

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

or equivalently

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

Central Limit Theorems

Theorem (Multivariate CLT)

Suppose $\mathbf{X}$ *is a random vector with finite mean* μ *and variance* $\mathbf{\Sigma}$*. Then*

$$\sqrt{n}\left(\bar{\mathbf{X}}_n - \mu\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$$

Central Limit Theorems

We'll sketch out a proof of the CLT using moment generating functions.

Let $Y_i = \frac{X_i - \mu}{\sigma}$. Then $\mathbb{E}[Y_i] = 0$, $\mathbb{E}[Y_i^2] = 1$.

$\bar{X} = n^{-1/2} \sum Y_i$. Therefore, $M_{\bar{X}}(t) = \left[M_{Y_i}(\frac{t}{\sqrt{n}}) \right]^n$.

Performing a Taylor expansion about 0 gives: $M_{Y_i}(t) = 1 + \frac{t^2}{2} + \mathcal{O}\left(t^3\right)$

$M_{\bar{X}}(t) = \left[M_{Y_i}(\frac{t}{\sqrt{n}}) \right]^n = \left[1 + \frac{t^2/2}{n} + \mathcal{O}\left(\frac{t^3}{n^{3/2}} \right) \right]^n \to e^{t^2/2}$

Inference - Hypothesis Tests

We care about some unknown parameter θ.

Suppose we have a consistent estimator, $\hat{\theta}_n \xrightarrow{p} \theta$.

Suppose we know the distribution of this estimator $\mathbb{P}(\hat{\theta}_n \leq x)$.

Can we make some claims about θ with confidence?

E.g., "the average height of an Australian adult is above 1m", or "smoking a pack of cigarettes a day increases your risk of lung cancer".

Inference - Hypothesis Tests

Hypothesis testing is always about trying to reject some sort of hypothesis as false. If you were to hypothesise what you were trying to prove, then the hypothesis might be retained for one of two reasons:

1. The hypothesis is true.
2. You don't have a sufficient sample size to resolve the difference between your hypothesis and the true value. It may be the case that you would retain pretty much any hypothesis you started with!

Typically our hypothesis will be one of $\theta = 0$, $\theta \leq 0$, or $\theta \geq 0$.

Inference - Hypothesis Tests

So we construct a **null hypothesis** which we wish to disprove.

This is usually a statement about one or more parameters of a distribtuion and is typically written

$$H_0 : \theta = \theta_0$$

In the event that we reject the null hypothesis, we retain an **alternative hypothesis**, which is usually the converse of the null.

$$H_1 : \theta \neq \theta_0$$

Inference - Hypothesis Tests

Once we know what we're trying to reject, we need to set a tolerance for which we're willing to make mistakes.

We pick an α, the **(nominal) level** of the test. Often $\alpha = 0.05$.

We then need to pick a **test statistic**, which is usually a function of an estimator (which is itself a function of the sample). E.g.

$$T = T(\mathbf{X}) = \sqrt{n}\frac{(\hat{\theta} - \theta_0)}{\hat{\sigma}}$$

Inference - Hypothesis Tests

A typical hypothesis test proceeds as follows.

1. Have a parameter of interest and an estimator for it.
2. State the null and alternative hypotheses in terms of your parameter of interest.
3. Choose a test statistic.
4. If the null were *true*, what would the sampling distribution of the test statistic be?
5. Find the appropriate quantiles of the sampling distribution. These are your **critical values**.
6. Calculate the value of your test statistic from your sample. Compare to the critical value.
7. If the value of the statistic is greater than the critical value, reject the null and retain the alternative. Else retain the null.

Inference - Hypothesis Tests

Example

Let $X_1, \ldots, X_{10} \sim N(\mu, 4)$. We wish to test

$$H_0 : \mu = 1 \text{ vs } H_A : \mu \neq 1$$

at the 5% level. We find a sample mean of $\bar{X} = 2.3$. Under the null hypothesis, we have the following sampling distribution.

$$\sqrt{10}\frac{\bar{X} - 1}{2} \sim N(0, 1)$$

We have a two sided test, so the $\alpha/2$ and $1 - \alpha/2$ quantiles are -1.96 and 1.96 respectively. Let's calculate our test statistic given our observed sample mean.

$$\sqrt{10}\frac{2.3 - 1}{2} = 2.05$$

$2.05 > 1.96$, so we may reject the null and retain the alternative.

Inference - Hypothesis Tests

A test might incorrectly reject a null hypothesis that is in fact true. This is called a **type-I error** (or **false discovery**).

The probability of a type-I error, α, is called the **size** of the test.

A test m ight incorrectly retain a null hypothesis that is in fact false. This is called a **type-II error**.

The probability of a type-II error, β, is called the **power** of the test.

Inference - Hypothesis Tests

If we have correctly specified the distribution of the test statistic, and used the correct critical values, then the size of the test will be equal to the nominal level.

Note that the distribution will not always be correctly specified, since we may be using asymptotic approximations.

If the size of a test is less its nominal level, it is called **conservative**.

If the size of a test is greater than its nominal level, it is called **anti-conservative**.

Inference - Hypothesis Tests

Finding a test statistic with a known distribution can be challanging.

In particular, to extract critical values, we'd prefer that the distribution not depend on any unknown parameters (**nuisance parameters**).

Such distributions are called **pivotal**.

Exercise 6.6

1. Let $X_1, \ldots, X_n \sim N(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_n \sim N(\mu_Y, \sigma^2)$ be samples drawn from two distributions with a known common variance. We wish to test the hypothesis $H_0 : \mu_X \leq \mu_Y$ against the alternative $H_1 : \mu_X > \mu_Y$. What might make a good test statistic?

Inference - Hypothesis Tests

Recall our unbiased estimator of the population variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

We'll need the following fact.

Proposition

Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Pivotal Quantities

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, where both μ and σ^2 are unknown.

If we appeal to the central limit theorem, we get

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$$

But we don't know σ^2!

We can make this statistic pivotal by **studentizing** it.

Pivotal Quantities

Proposition

Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$$\sqrt{n}\frac{\bar{X} - \mu}{s} \sim t(n-1)$$

Proof:

$$\sqrt{n}\frac{\bar{X} - \mu}{s} = \frac{\sqrt{n}\frac{\bar{X}-\mu}{\sigma}}{\frac{s}{\sigma}} = \frac{\sqrt{n}\frac{\bar{X}-\mu}{\sigma}}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\sqrt{n}\frac{\bar{X}-\mu}{\sigma}}{\frac{(n-1)s^2}{\sigma^2}/(n-1)}$$

Recall the numerator is $N(0,1)$. The denomitator is $\chi^2(n-1)/(n-1)$. And by definition

$$\frac{N(0,1)}{\sqrt{\chi^2(n-1)/(n-1)}} \sim t(n-1)$$

Pivotal Quantities

Quantities that converge to a pivotal distribution are called **asymptotically pivotal**.

Proposition

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$, $\text{Var}(X) = \sigma^2 < \infty$.

$$\sqrt{n}\frac{\bar{X} - \mu}{s} \xrightarrow{d} N(0,1)$$

Proof:

$X_n = \sqrt{n}(\bar{X} - \mu)$, $Y_n = s_n$.

$X_n \xrightarrow{d} N(0, \sigma^2)$. $Y_n \xrightarrow{p} \sigma$.

$X_n / Y_n \xrightarrow{d} \frac{1}{\sigma} N(0, \sigma^2) = N(0,1)$.

Confidence Intervals

A $1 - \alpha$ **confidence interval** for a parameter θ is an interval $C_n = (a, b)$ such that

$$\mathbb{P}_\theta(\theta \in C_n) \geq 1 - \alpha$$

That is, it's an interval that traps θ with probability $1 - \alpha$. We call $1 - \alpha$ the **coverage** of the interval.

θ is fixed, it's C_n that's random!

Confidence Intervals

For estimators that are asymptotically normal, such that $\sqrt{n}\frac{(\hat{\theta}-\theta)}{\hat{\sigma}} \xrightarrow{d} N(0,1)$, we'll typically construct the interval as

$$C_n = \left(\hat{\theta} - z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\theta} + z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the normal distribution.

Exercise 6.7 Show that the above interval has asymptotic coverage $1-\alpha$.

Regression

Regression models attempt to explain an outcome variable in terms of a set of predictors.

We'll use be given a set of n obervations on p predictors: $(y_i, \mathbf{x}_i')_{i=1}^n$.

We'll assume independence across samples, i, but y_i and $\mathbf{x}_i'$ are jointly distributed.

Our objective is to find a function $f(\cdot)$ that explains the outcome in terms of the predictors: $y_i = f(\mathbf{x}_i') + \varepsilon_i$.

For a few reasons, we'll restrict $f(\cdot)$ to be a linear combination of the predictors: $y_i = \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon_i$.

Note: you can allow a constant by setting $x_1 = 1$.

Regression Assumptions

We're given a sample of n iid draws from a joint distribution $(y_i, \mathbf{x}_i') \sim F$.

Let's make some highly spurious assumptions, and see what traction we can get:

Gauss-Markov Assumptions

1. The true DGP is linear: $y = X\beta + \varepsilon$. (linearity)
2. X is full column rank. (no multicollinearity)
3. $\mathbb{E}[\varepsilon_i \mid X] = 0$. (strict exogeneity)
4. $\mathbb{E}[\varepsilon_i^2 \mid X] = \sigma^2$. (spherical errors)

Least Squares

Now we have a linear model $y = X\beta + \varepsilon$, and a parameter β we need to estimate. How?

$$\hat{\beta}_{OLS} = \underset{\beta}{\operatorname{argmin}} \; ||y - X\beta||_2^2$$

We've seen previously that the solution to this minimization problem is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

Least Squares

Theorem (Gauss-Markov Theorem)

Under the Gauss-Markov assumptions, the OLS estimator is the ***best linear unbiased estimator****.*

Exercise 6.8

1. Show that the OLS estimator is unbiased.
2. Show that the OLS estimator is consistent.
3. Show that the OLS estimator is asymptotically normal.

Learning Outcomes

That's a lot already, but there's so much more good stuff! We didn't even touch on: maximum likelihood, method of moments, delta method, shrinkage estimation, kernel density estimation, etc.

You should be able to:

1. Determine the bias of a given estimator.
2. Apply the CMT and Slutsky's theorem.
3. Determine consistency of an estimator using LLN.
4. Apply the CLT to show asymptotic normality.
5. Conduct a hypothesis test.
6. Construct a confidence interval.
7. Estimate regression coefficients with OLS.