Introduction to asymptotic theory

James Scott (UT-Austin)

Reference: Bertsekas Chapter 7

Outline

- Some important inequalities
- Convergence in probability
- Weak law of large numbers
- Convergence in distribution
- Central limit theorem and asymptotic normality
- Normal-theory confidence intervals
- Normal-theory hypothesis tests

Some important inequalities

We'll study three important inequalities in probability theory:

- Markov's inequality
- Chebyshev's inequality
- Hoeffding's inequality

These are useful for bounding quantities that might otherwise be hard to compute. They are also used in the study of *convergence of random variables*.

Markov's inequality

Markov's inequality bounds the upper tail probability for a nonnegative random variable.

Theorem: Let X be a nonnegative random variable and suppose that $E(X) < \infty$. Then for any $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Markov's inequality: proof

Let's fix any positive number t and consider the random variable Y_t defined as

$$Y_t = \begin{cases} 0, & \text{if } X < t \\ t, & \text{if } X \geq t \end{cases}$$

Since it is always the case that $Y_t \leq X$, then $E(Y_t) \leq E(X)$. Therefore

$$E(X) \geq E(Y_t) = t \cdot P(Y_t = t) = t \cdot P(X_t \geq t)$$

Chebyshev's inequality

Chebyshev's inequality bounds the probability that a random variable will deviate very far from its mean.

Theorem: Let X be a random variable and suppose that $E(X) = \mu < \infty$ and $\mathrm{var}(X) = \sigma^2 < \infty$. Let $t > 0$. Then

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

In words: if the variance of a random variable is small, then the probability that it takes a value far from its mean is also small. In particular, the probability that X is more than t units away from its mean falls at least as fast as $1/t^2$.

Chebyshev's inequality: proof

Clearly $P(|X - \mu| \geq t) = P(|X - \mu|^2 \geq t^2)$.

Now since $|X - \mu|^2$ is nonnegative random variable, we can use Markov's inequality to conclude that

$$P(|X - \mu|^2 \geq t^2) \leq \frac{E(|X - \mu|^2)}{t^2} = \frac{\sigma^2}{t^2}$$

Corollary

An alternative for Chebyshev's inequality is obtained by setting $t = k\sigma$. Then

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Chebyshev's inequality: example

Suppose we randomly poll N voters in order to estimate p, the true fraction of voters in the population who intend to vote for the incumbent.

- Suppose of our N voters, X support the incumbent.
- Then $\hat{p} = X/N$ is a natural estimate: intuitively, we should expect that $\hat{p}$ is close to p.
- But how likely is $\hat{p}$ to fall more than ϵ away from the true p? E.g. if $\epsilon = 0.05$, how likely is a 5% polling error or more?

Chebyshev's inequality: example

Note that X is a binomial random variable with $E(X) = Np$ and X $= Np(1-p)$. So:

- $E(\hat{p}) = E(X/N) = Np/N = p.$
- $\mathrm{var}(\hat{p}) = \mathrm{var}(X/N) = \mathrm{var}(X)/N^2 = p(1-p)/N$

Now using Chebyshev's inequality,

$$\begin{aligned} P(|\hat{p} - p| > \epsilon) \leq \frac{\mathrm{var}(\hat{p})}{\epsilon^2} &= \frac{p(1-p)}{N\epsilon^2} \\ &\leq \frac{1}{4N\epsilon^2} \end{aligned}$$

since $p(1-p) \leq 1/4$ for all p.

Chebyshev's inequality: example

So if we poll $N = 1000$ people, then

$$P(|\hat{p} - p| > 0.05) \leq \frac{1}{4 \cdot 1000 \cdot 0.05^2} = 0.1$$

Less than a 10% chance of a polling error of 5% or more. So if I report the interval $\hat{p} \pm 0.05$, I can be at least 90% confidence that my interval contains the truth.

This is a very conservative bound: the real probability is much smaller. (Markov and Chebyshev usually give conservative bounds.)

Hoeffding's inequality

Suppose that $X_1, \ldots, X_N \sim$ Bernoulli(p) be independent random variables. Define

$$\bar{X}_N = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$$

Then for any $\epsilon > 0$,

$$P(|\bar{X}_N - p| \geq \epsilon) \leq 2e^{-2N\epsilon^2}$$

Hoeffding's inequality is similar to Markov's inequality, but it is sharper. (This is actually a special case of a more general version of the inequality, which we don't really need.)

Hoeffding's inequality

Exercise: go back to our political poll example and again suppose $N = 1000$.

Use Hoeffding's inequality to bound the probability that $P(|\hat{p} - p| \geq \epsilon)$ for $\epsilon = 0.05$ and $\epsilon = 0.01$ (that is, polling errors of 5% and 1%, respectively.)

Compare these to the bounds you get from the Chebyshev inequality.

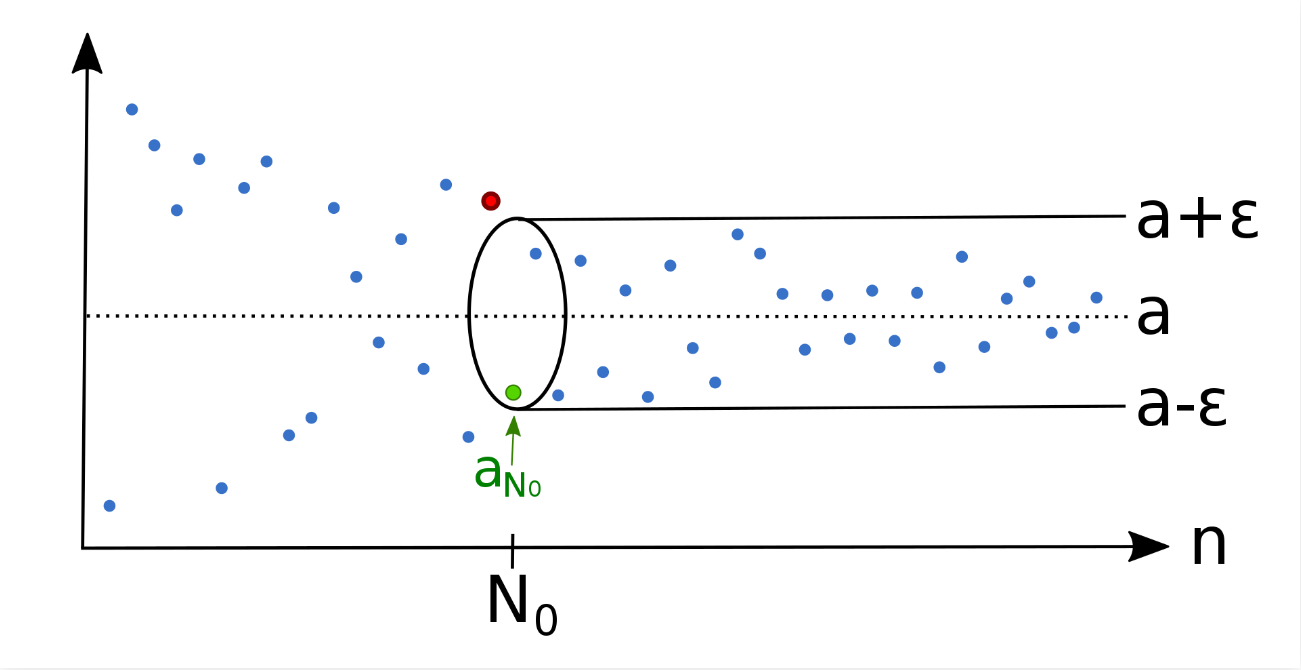
Convergence of random variables

We now turn to a topic called “large sample theory” or “limit theory” or “asymptotic theory.”

- The basic question is: what can we say about the limiting behavior about a sequence of random variables $X_1, X_2, X_3, \ldots$?
- Since statistics is all about gathering data, we will be interested in what happens when we gather more and more data.

Convergence of random variables

In calculus, we say that a sequence x_N converges to a limit a if, for every $\epsilon > 0$, then exists an n_0 such that $|x_N - a| \leq \epsilon$ for every $n \geq n_0$.



But in probability theory, convergence is more subtle.

Convergence of random variables

For example, consider the trivial calculus sequence of $x_N = 0$ for all n. Then clearly x_N converges to 0.

But now consider the sequence of random variables $X_1, X_2, \ldots,$ where each $X_N \sim N(0, 1)$.

- Since these all have the same distribution, we'd like to say that X_N "converges to" a standard normal distribution Z.
- But how can we make sense of this idea, when $P(X_N = Z) = 0$ for all N?

Convergence of random variables

Or consider the sequence of random variables $X_1, X_2, \ldots$, where each $X_N \sim N(0, 1/n^2)$.

- Intuitively, X_N becomes very spiky and concentrated around 0.
- So we'd like to say that X_N "converges to" 0.
- Yet for every n, $P(X_N = 0) = 0$.

Convergence of random variables

We need to develop some appropriate tools for describing the convergence of random variables.

- Convergence in probability: typically used to describe when a sequence of random variables "settles down" to a specific number. Key result: law of large numbers.
- Convergence in distribution: typically used to describe when a sequence of random variables never settles down to a specific number, but begins to look more and more like some specific *distribution*. Key result: central limit theorem.

Convergence in probability

Definition. Let $X_1, X_2, \ldots$ be a sequence of random variables. We say that the sequence X_N converges in probability to some random variable X if, for every $\epsilon > 0$,

$$P(|X_N - X| \geq \epsilon) \to 0$$

as $n \to \infty$. We write this as $X_N \xrightarrow{p} X$.

Key point: we have turned a claim about random variables into a claim about real numbers (i.e. probabilities) so that we can describe those real numbers using ordinary calculus tools.

Convergence in probability: example

Consider the sequence $X_N \sim N(0, 1/n^2)$. Show that $X_N \xrightarrow{P} 0$ (i.e. the trivial random variable that always takes the value 0).

Proof: We need to show that $P(|X_N| \geq \epsilon) \to 0$ for all ϵ. From Chebyshev's inequality, we know that, since $E(X_N) = 0$ and $\mathrm{var}(X_N) = 1/n^2$,

$$P(|X_N| \geq \epsilon) \leq \frac{\mathrm{var}(X_N)}{\epsilon^2} = \frac{1}{n^2 \epsilon^2}$$

And for fixed ϵ, $1/(n^2\epsilon^2) \to 0$ as $n \to \infty$.

The law of large numbers

The law of large numbers is one of the most important results in all of probability theory.

It says says that the mean of a large sample is close to the mean of the distribution.

We now make this idea precise.

The law of large numbers

Theorem. Suppose that $X_1, X_2, \ldots, X_N$ are independent, identically distributed random variables. (Recall that we refer to this as an IID sample.) Suppose each X_i has mean $\mu = E(X_i) < \infty$ and variance $\sigma^2 = \mathrm{var}(X_i) < \infty$.

Let $\bar{X}_N = n^{-1} \sum_{i=1}^{n} X_i$ be the sample mean. Then $\bar{X}_N \xrightarrow{P} \mu$.

The law of large numbers

As an example, consider flipping a biased coin where the probability of heads is p.

- Let X_i denote the outcome of a single toss (0 or 1). Hence, $p = P(X_i = 1) = E(X_i)$.
- The fraction of heads after N tosses is $\bar{X}_N = \hat{p}$.
- According to the LLN, as you toss the coin more and more times, $\bar{X}_N$ converges in probability to the true probability p.

This doesn't mean that $\bar{X}_N = p$ for any N: you never get the answer *exactly* right. It does mean that for large N, the distribution of $\bar{X}_N$ is very tightly concentrated around p.

The law of large numbers

Some comments:

1. This is the fundamental math result that connects the formal definition of an expected value with the idea of an expected value as a long-run average.
2. The requirement that $\sigma^2 < \infty$ isn't essential, but it does simplify the proof.
3. Technically I've shown you the Weak Law of Large Numbers.
    - There is a stronger version of the theorem, called the Strong Law of Large Numbers.
    - It replaces “convergence in probability” with “almost sure convergence,” a different and stronger type of convergence.
    - The differences aren't super important for our purposes. The “Weak Law” is strong enough for us!

The law of large numbers: proof

With Chebyshev's inequality, it's a two liner! First, note that

$$\mathrm{var}(\bar{X}_N) = \mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\cdot N \cdot \mathrm{var}(X_i) = \frac{\sigma^2}{N}$$

Therefore,

$$P(|\bar{X}_N - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}\,,$$

which converges to 0 for fixed ϵ as $N \to \infty$. **QED!**

The law of large numbers

A whole city built by the Law of Large Numbers…

The law of large numbers

Like a good neighbor, State Farm is there.®

...because of the Law of Large Numbers!

A whole industry, too!

Convergence in distribution

Our second type of convergence is *convergence in distribution*. Let $X_1, X_2, \ldots$ be a sequence of random variables, and suppose that X_N has CDF F_N. Let X be some other random variable with CDF F. We say that the sequence X_N converges in distribution to X, written $X_n \rightsquigarrow X$, if

$$\lim_{N \to \infty} F_N(t) = F(t)$$

at all t where $F(t)$ is continuous.

Intuitively: probability statements about X_N can be approximated using probability statements about X, and the approximation gets arbitrarily good as N diverges.

Convergence in distribution

Best math symbol ever!

$$X_n \rightsquigarrow X$$

OK, you can also write this as $X_N \stackrel{D}{\to} X$.

Central limit theorem

Suppose that $X_1, \ldots, X_n$ are IID random variables with mean μ and variance σ^2. Let $\bar{X}_n$ be the sample mean:

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

Now define the z-score of the sample mean as

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - \mu}{\mathrm{se}(\bar{X}_n)} \\ &= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \end{aligned}$$

Note: the second line is from de Moivre's equation.

Central limit theorem

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

The central limit theorem says that $Z_n \rightsquigarrow Z$, where Z is a standard normal random variable. That is,

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z)$$

In words: we can approximate statements about Z_n using statements about a standard normal random variable. Informally, we write this as $Z_n \approx N(0, 1)$.

Central limit theorem

Here's an equivalent way of expressing the central limit theorem. Since

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \approx N(0, 1)$$

then

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Informally, this means that the CDF of $\bar{X}_n$ is close to that of a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Central limit theorem: comments

1. The central limit theorem holds basically no matter what the distribution of the X_i's are. The only conditions are that the mean and variance are finite!

2. Some terminology: the CLT implies that $\bar{X}_n$ is **asymptotically normal.**

3. The bigger the sample size, the closer the approximation gets. By $n = 30$, the normal approximation is pretty good, unless the distribution of X_i is super heavy tailed and/or skewed.

Central limit theorem: comments

Let's see some examples in `CLT.R`.

An example: FedEx logistics

- A FedEx 737 cargo plane has a max cargo capacity of 11,422 pounds.
- From a long run of experience, FedEx knows that the average weight of a package is 6.1 pounds, with a standard deviation of 5.6 pounds (the distribution is skewed right: most packages are pretty light, a few packages are really heavy).
- Today's logistics problem: 1810 packages are in the system and scheduled to be shipped to Austin. But these packages have yet to be weighed at the sorting facility.

Your turn: What is the probability that FedEx will need more than one flight to Austin to get all 1810 packages there?

An example: FedEx logistics

Let X_i be the weight of package $i = 1, \ldots, 1810$. FedEx will need only one flight to Austin if

$$\sum_{i=1}^{1810} X_i \leq 11422$$

or equivalently, if

$$\frac{1}{1810} \sum_{i=1}^{1810} X_i \leq \frac{11422}{1810} = 6.31$$

That is, average package weight $\bar{X}_n$ cannot exceed 6.31 pounds (recall $\mu = 6.1$ and $\sigma = 5.6$).

An example: FedEx logistics

So what is $P(\bar{X}_n \leq 6.31)$?

The sample size is $n = 1810$: pretty large! So use the central limit theorem: **make approximate statements about $\bar{X}$ using statements about the normal distribution.**

An example: FedEx logistics

This gives us:

$$\begin{aligned}
P(\bar{X}_n \leq 6.31) &= P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n}(6.31 - \mu)}{\sigma}\right) \\
&\approx P\left(Z \leq \frac{\sqrt{n}(6.31 - \mu)}{\sigma}\right) \quad \text{where } Z \sim N(0, 1) \\
&\approx P\left(Z \leq \frac{\sqrt{1810}(6.31 - 6.1)}{5.6}\right) \\
&= P\left(Z \leq 1.595\right) \\
&= \Phi(1.595) \approx 0.945
\end{aligned}$$

Central limit theorem: plug-in version

The central limit theorem assumes that we know σ, the true standard deviation of the data points, so that we can z-score using the true standard error, $\sigma/\sqrt{n}$.

What if we don't know σ? A natural thing to do is to plug-in the *sample* standard deviation $\hat{\sigma}$ to give us an approximation:

$$\hat{se}(\bar{X}_n) = \frac{\hat{\sigma}}{\sqrt{n}} \quad \text{where} \quad \hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{m}(X_i - \bar{X}_n)^2}$$

Remember: we call this the *plug-in estimate* of the standard error.

Central limit theorem: plug-in version

It turns out that the central limit theorem still holds even if we use the plug-in estimate of the standard error.

That is, let

$$T_n = \frac{\bar{X}_n - \mu}{\hat{se}(\bar{X}_n)} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}}$$

Then $T_n \rightsquigarrow Z$, where Z is a standard normal random variable. (We call T_n a t-statistic rather than a z-score.)

As aside

If you've taken a statistics class before, you may remember spending a lot of time on t statistics and the t distribution.

As aside

Your experience might have involved a horrible, soul-crushing table that looked something like this:

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

As aside

The Z statistic and T statistic are really similar:

- Z statistic: divide an estimator by its true standard error, $\mathrm{se}(\hat{\theta})$.
- T statistic: divide an estimator by **an estimate of** its true standard error, $\hat{se}(\hat{\theta})$.

In particular, both Z and T are asymptotically normal!

- We can make approximate probability statements about either of them using the normal distribution.
- The difference between $\mathrm{se}(\hat{\theta})$ and $\hat{se}(\hat{\theta})$ is usually much smaller than the uncertainty you have about θ in the first place.

So what gives?

The whole point of that awful table was to correct for the small errors in the normal approximation that emerge when n is modest (say, less than 30).

- This was common in 1908, when the t distribution was invented, and when most statisticians were concerned with analyzing data from small agricultural experiments.
- This is pretty rare today. Data sets are bigger than they were a century ago. (And even for small samples, the differences are pretty small.)

So what gives?

The whole point of that awful table was to correct for the small errors in the normal approximation that emerge when n is modest (say, less than 30).

- This was common in 1908, when the t distribution was invented, and when most statisticians were concerned with analyzing data from small agricultural experiments.
- This is pretty rare today. Data sets are bigger than they were a century ago. (And even for small samples, the differences are pretty small.)
- I have **never, in my entire career,** encountered a real-life example where the difference between using the normal and t distributions made a substantive scientific difference. (I asked three other statisticians on my hallway and they all said the same thing.)

As aside

So we'll basically ignore this horrible table and treat the T statistic as if it were normal (which it is, asymptotically!)

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

But...

I cannot promise that everyone you encounter will be so sensible.

- journal editors…
- peer reviewers…
- econometrics teachers…
- PhD advisors…
- the FDA…

Some of these people may make you do calculations with the t distribution, in which case you will have to learn/remind yourself about it.

But...

I cannot promise that everyone you encounter will be so sensible.

- journal editors…
- peer reviewers…
- econometrics teachers…
- PhD advisors…
- the FDA…

Some of these people may make you do calculations with the t distribution, in which case you will have to learn/remind yourself about it. **Sorry :-(**

CLT: summary

Suppose we're trying to estimate θ using an estimator $\hat{\theta}_n$. Let $\hat{se}(\hat{\theta}_n)$ be the (estimated) standard error of $\hat{\theta}_n$. Define

$$Z_n = \frac{\hat{\theta}_n - \theta}{\hat{se}(\hat{\theta}_n)}$$

If $Z_n \rightsquigarrow Z$ where $Z \sim N(0, 1)$, we say that $\hat{\theta}$ is *asymptotically normal.*

CLT: summary

There are variations on the central limit theorem showing that lots of common estimators are asymptotically normal:

- sample means and proportions
- differences of sample means and proportions
- sample standard deviations and correlations
- OLS estimators of the intercept and slope
- basically anything that looks like an average of some sample quantity! **Averages eventually look normal.**

Confidence intervals

One consequence of this fact is that we can use the normal distribution to produce approximate confidence intervals for lots of common statistics problems.

How does that work?

Confidence intervals

Suppose that we have some $\hat{\theta}_n$ that we know to be asymptotically normal. Then we have the following approximation:

$$\frac{\hat{\theta}_n - \theta}{\widehat{se}(\hat{\theta}_n)} \sim N(0, 1)$$

That is, our estimation error is (approximately) normally distributed:

$$\hat{\theta}_n - \theta \sim N\left(0, \widehat{se}(\hat{\theta}_n)\right)$$

Confidence intervals

So, for example, if we go out one standard error:

$$P\left[\hat{\theta}_n - \theta \in \left(-\hat{se}, \hat{se}\right)\right] \approx 0.68$$

And if we go out two standard errors:

$$P\left[\hat{\theta}_n - \theta \in \left(-2\hat{se}, 2\hat{se}\right)\right] \approx 0.95$$

Probalistic bounds on estimation error = confidence intervals!

Confidence intervals

In general, if $\hat{\theta}_n$ is asymptotically normal, and if we let z_α denote the $(1 - \alpha/2)$ quantile of the normal distribution, then the interval estimate

$$I_n = [\hat{L}_N, \hat{U}_N] = \hat{\theta}_n \pm z_\alpha \cdot \widehat{se}(\hat{\theta}_n)$$

is an approximate confidence interval at level $1 - \alpha$. That is, it covers the true value θ with probability approximately equal to $1 - \alpha$.

Confidence intervals

Remember the frequentist coverage principle?

$$P_\theta \left(\theta \in [\hat{L}_N, \hat{U}_N] \right) \geq 1 - \alpha \,,$$

Remember our three questions? (What is fixed? What is random? What is the source of this randomness?)

It's the same statement again. Here, as with bootstrapped confidence intervals, the probability claim is *approximately* true, and the approximation gets better with larger n.

Confidence intervals

Let's see a couple of examples in `normalCI_examples.R`.

Testing

A really similar line of thinking allows us to construct hypothesis tests.
Suppose we have:

- an unknown parameter θ
- an asymptotically normal test statistic T_n
- a null hypothesis that $\theta = \theta_0$

Testing

Because of asymptotic normality, we know that if the null hypothesis is true (i.e. $\theta = \theta_0$), then

$$T_n \sim N(\bar{T}_0, \mathrm{se}_0(T_n))$$

where $\bar{T}_0 = E(T \mid \theta = \theta_0)$ and $\mathrm{se}_0(T_n)$ are the mean and standard error of T_n, assuming that $\theta = \theta_0$.

We can then do any kind of test we want:

- Fisher: calculate a p-value
- Neyman-Pearson: specify an alternative hypothesis, choose a rejection region, calculate α and power

Example

Ikea produces 148.9 million of those tiny little Allen wrenches each year (5 wrenches per second, all day every day). The wrenches should be 5.0 mm in diameter, on average. If they're not, something has gone awry in the manufacturing process.

The last 50 wrenches that were sampled off the assembly line have measured 5.03 mm in average diameter, with a standard deviation of 0.07 mm.

Should we stop making wrenches and figure out what's wrong? Or is this just a chance statistical fluctuation?

Example

Let X_i be the diameter of wrench i, and let $\mu = E(X_i)$. The null hypothesis is that $\mu = 5.0$. Let's compute a p-value under this null hypothesis.

- Our test statistic is $T = \bar{X}_n$. From the CLT, we know this is asymptotically normal.
- From de Moivre's equation, our (estimated) standard error is $\hat{se}(\bar{X}_N) = 0.07/\sqrt{30} = 0.0128$.

Example

So under the null hypothesis, we can approximate the sampling distribution of $\bar{X}_n$ as

$$\bar{X}_n \sim N(5.0, 0.0128)$$

Under this null hypothesis, we have $p = P(\bar{X}_n \geq 5.03) = 0.0095$.

Example

Note: in this case we might want to calculate a two-sided p-value as:

$$p = P(\bar{X}_n \leq 4.97) + P(\bar{X}_n \geq 5.03) = 0.019$$

The thinking here is that deviations of ≥ 0.03 up and down are both equally important, and they should both count as "more extreme than" a deviation of 5.03.

Remember: in general the p value is $P(T \in \Gamma(t_{ob}) \mid H_0)$, where $\Gamma(t)$ is the set of test statistics that are judged "more extreme than" some specific value t.

Standardized test statistics

Any time we're doing a test under the assumption of asymptotic normality, we can always *standardize* our test statistic to compare it to a standard $N(0, 1)$ distribution. If $T_n \sim N(\bar{T}_0, \mathrm{se}_0(T_n))$ under the null, then

$$Z_n = \frac{(T_n - \bar{T}_0)}{\mathrm{se}_0(T_n)} \sim N(0, 1)$$

Standardized test statistics

So the general recipe for testing under asymptotic normality is:

1. Calculate $\bar{T}_0$ and $\mathrm{se}_0(T_n)$, the mean and standard error of your test statistic under the null hypothesis.
2. Standardize your observe test statistic as
$$z_{ob} = \frac{(t_{ob} - \bar{T}_0)}{\mathrm{se}_0(T_n)}$$
3. Compare z_{ob} to a standard normal distribution (using either a Fisher or a Neyman-Pearson test).

A few notes

- z_{ob} can be interpreted as a signal-to-noise ratio: "how many standard deviations away is z_{ob} from what we'd expect based on the null hypothesis?"
- Remember, some fastidious people will call z_{ob} a "t statistic" and want you to compare it to a t distribution rather than a normal distribution. (This is called a "t test.") But the difference between these two is usually small, and virtually always dominated by other sources of error.
- If you want to sound fancy, you can call the test based on asymptotic normality a **Wald test.**

Testing vs. confidence interval

Note: compare our test with the normal-theory confidence interval for μ in light of the data.

$$\begin{aligned} \mu &\in \bar{X}_n \pm 2 \cdot \hat{se}(\bar{X}_m) \\ &= 5.03 \pm 2 \cdot 0.0128 \\ &= (5.0044, 5.0556) \end{aligned}$$

Isn't that way more informative than quoting $p = 0.019$?

Some examples

I'll show you a few examples of normal-based testing and confidence intervals in action:

- Requests for Abortion in Latin America Related to Concern about Zika Virus Exposure, Figure S1, S3, and Table 1.
- The influence of hours worked prior to delivery on maternal and neonatal outcomes: a retrospective cohort study, Table 3 and Figure 1.
- Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity, Table 3.

In each case, the confidence interval is more scientifically meaningful and interpretable, compared to the p-value.