PLSC 502: "Statistical Methods for Political Research"

Confidence Intervals
October 13, 2016

Introduction

The topic is confidence intervals, which will serve as a (hopefully gentle) introduction to statistical inference more generally. We'll discuss *hypothesis testing* – a related concept – next Tuesday.

We'll introduce a bit of terminology first, then go on to some specific examples. For terminology, define θ as a population parameter of interest, one which could (in theory) be calculated from a population of $\mathfrak{N}$ units. The associated sample statistic we'll call $\hat{\theta}$, which itself is based upon N units drawn in a simple random sample from the population. What we're interested in doing in general is knowing something about θ from $\hat{\theta}$, and in particular having some idea about how "close" $\hat{\theta}$ is to θ.

Confidence Intervals

A *confidence interval* (often abbreviated "c.i.") is a range of estimates $\hat{\theta}$ for θ. We'll occasionally refer to this range as $[\hat{\theta}_L, \hat{\theta}_U]$, and we'll call the former the *lower bound* of the confidence interval and the latter its *upper bound* (always remembering that $\hat{\theta}_L \leq \hat{\theta}_U$).[1].

To be useful, that range should have two traits:

1. It should contain θ, and
2. It should be as narrow as possible.

The challenge, then, is to come up with a way of getting a confidence interval that is both narrow and has a high probability of including θ. If we can do so, then we can be reasonably sure ("confident") that the range we construct in that way is a good guess of our population parameter. In the limit, if we know that $\theta \in [\hat{\theta}_L, \hat{\theta}_U]$ and $\hat{\theta}_L = \hat{\theta}_U$, then we have a "perfect" estimator of θ.[2]

Of course, in practice there is a tradeoff between the two desiderata above: we can be arbitrarily confident that the c.i. we construct contains θ if it is sufficiently wide. In the limit, and infinitely wide confidence interval is assured of containing θ; but such an interval is

[1] I should note that this is what is called a *two-sided* confidence interval. It is also possible to construct *one-sided* c.i.s, such as $(-\infty, \hat{\theta}_U]$ and, alternatively, $[\hat{\theta}_L, \infty)$. More on those a bit later.

[2] Note that this never actually happens in practice, unless $N = \mathfrak{N}$.

completely uninformative.

To begin, then, it is useful to talk about the likelihood that $\theta \in [\hat{\theta}_L, \hat{\theta}_U]$. If we define:

$$\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha, \tag{1}$$

then we say that $1 - \alpha$ is the *confidence coefficient* (or *level of confidence*) in which we are interested in making inferences about θ. The value of $1 - \alpha$ is something that is determined by the researcher, and is usually set with an eye to whether s/he is more concerned with the parameter θ being in the c.i., or with the relative precision of the interval estimate.[3]

One way of getting an interval estimate is through what is termed the *pivotal method* (Wackerly et al. 2008, p. 407). That approach requires that we find an estimator $\hat{\theta}$ of θ that

1. is a function *only* of the sample data and the population parameter θ, and

2. whose sampling distribution *does not* depend on θ.

If an estimator has these characteristics, then (as we'll discuss below) we can use simple linear transformations to construct confidence intervals.

Large-Sample Theory, the Mean, and the Normal Distribution

As we discussed and demonstrated last week, the sampling distribution of a mean (and, for that matter, any sum of a sufficiently large number of independent random variables) follows a normal distribution. This provides the nicest/simplest place to start the discussion. Note that our typical estimator of μ, denoted $\bar{X}$, can be though of as being normally distributed:

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2_{\bar{X}}) \tag{2}$$

where we defined $\sigma^2_{\bar{X}} = \frac{\sigma^2}{N}$ and σ^2 is (as usual) just the variance of X. Comparing this to our list of two criteria discussed above, we can note that

1. $\bar{X}$ itself depends only on the values of X in the sample, and on the value of μ, and

2. the *shape* of its sampling distribution does not depend on μ, but only on other things (like the size of the sample).

Now, to construct a confidence interval, we can start with the the sample statistic $\bar{X}$. Since we know that, in expectation, $\mathrm{E}(\bar{X}) = \mu$, it makes sense to use the sample value $\bar{X}$ as the "center" of our confidence interval.

[3]These are somewhat analogous to committing *Type I* or *Type II* errors, respectively.

Next, we choose a level of confidence; tradition suggests that we set $1 - \alpha = 0.95$, that is, that we select a "95 percent level of confidence," though there's absolutely nothing sacred about this number. This means that we want to create a confidence interval such that

$$\Pr(\bar{X}_L \leq \mu \leq \bar{X}_U) = 0.95 \tag{3}$$

One way of calculating the bounds of the confidence interval is to choose $\bar{X}_L$ and $\bar{X}_U$ so that

$$\Pr(\mu < \bar{X}_L) = \int_{-\infty}^{\bar{X}_L} \phi_{\bar{X}}(u)\, du = 0.025$$

and

$$\Pr(\mu > \bar{X}_H) = \int_{\bar{X}_H}^{\infty} \phi_{\bar{X}}(u)\, du = 0.025.$$

Since we know the parameters of $\phi_{\bar{X}}$ – that is, the distribution is $\mathcal{N}(\mu, \sigma^2_{\bar{X}})$ – calculating values for the upper and lower limits of a confidence interval is straightforward.

More Generally...

More generally, for any sample statistic $\hat{\theta}$ which is an estimator of θ and whose sampling distribution is Normal, the statistic

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \tag{4}$$

is distributed according to a standard normal distribution. As a result, Z is a pivotal quantity, and we can consider two values in the tails of that standard normal distribution $-z_{\alpha/2}$ and $z_{\alpha/2}$ such that

$$\Pr(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha. \tag{5}$$

We can rewrite this as

$$\begin{aligned}
1 - \alpha &= \Pr\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) \\
&= \Pr\left(-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \hat{\theta} - \theta \leq z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\
&= \Pr\left(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq -\theta \leq -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) \\
&= \Pr\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right)
\end{aligned}$$

This means that a $(1 - \alpha) \times 100$-percent confidence interval for θ is given by

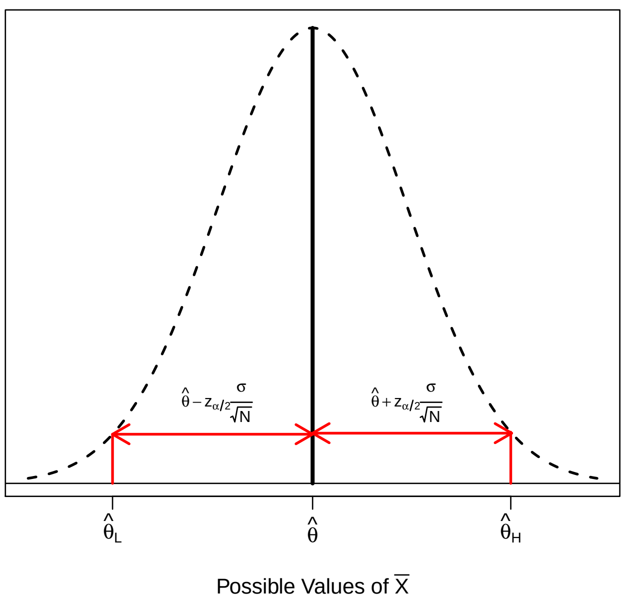
$$[\hat{\theta}_L, \hat{\theta}_U] = \left[\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}},\ \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right] \tag{6}$$

As a practical matter, this means that constructing a confidence interval for a variable whose (asymptotic) sampling distribution is normal consists of five steps:

1. Select your level of confidence $1 - \alpha$,
2. Calculate the sample statistic $\hat{\theta}$,
3. Calculate the z-value associated with the $1 - \alpha$ level of confidence,
4. Divide that z-value by $\sigma_{\hat{\theta}}$, the standard error of the sampling statistic, and
5. Construct the confidence interval according to (6).

This is illustrated in Figure 1.

Figure 1: Normal Confidence Intervals via the Pivotal Method



Another Example: Estimating Proportions

For a proportion, we know that $\theta = \pi$, and that $\sigma^2 = \pi(1 - \pi)$. In our sample of N observations, we have $\hat{\theta} = \hat{\pi} = \frac{1}{N}\sum_{i=1}^{N} X_i$, and we showed before that $\sigma^2_{\hat{\pi}} = \frac{\pi(1-\pi)}{N}$, so that $\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{N}}$.

For π sufficiently different from either zero or one, and N sufficiently large,[4] we discussed previously how the sampling distribution of $\hat{\pi}$ is $\mathcal{N}(\pi, \sigma^2_{\hat{\pi}})$. That means that we can calculate confidence intervals for an estimated proportion as

$$\hat{\pi}_L = \hat{\pi} - z_{\alpha/2}\left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}}\right] \tag{7}$$

and

$$\hat{\pi}_U = \hat{\pi} + z_{\alpha/2}\left[\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}}\right]. \tag{8}$$

As an example, consider taking a sample of size 20, and calculating $\hat{\pi} = 0.390$. The lower bound of the associated 95% confidence interval is

$$\begin{aligned}
\hat{\pi}_L &= 0.390 - 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\
&= 0.390 - 0.214 \\
&= \mathbf{0.176}
\end{aligned}$$

while the upper bound is

$$\begin{aligned}
\hat{\pi}_U &= 0.390 + 1.96\left[\sqrt{\frac{0.39(0.61)}{20}}\right] \\
&= 0.390 + 0.214 \\
&= \mathbf{0.604}
\end{aligned}$$

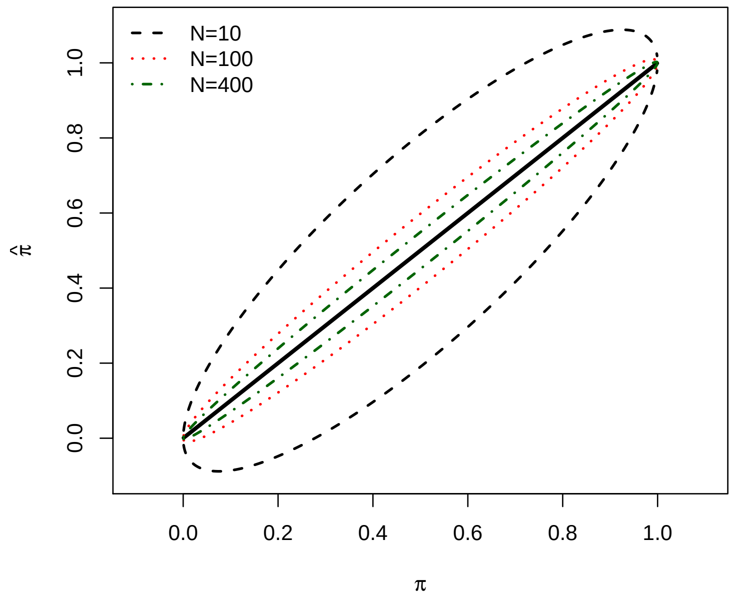
More generally, (7) and (8) indicate that the confidence interval for a proportion is a straightforward function of two quantities: the estimated proportion $\hat{\pi}$, and the sample size N. This makes it easy to illustrate how the width of the confidence interval changes with each, as in Figure 2.

Small Samples and Estimates: The t Distribution

For "small" samples, the asymptotic normal properties of the estimator of the mean $\bar{X}$ don't hold. We can see why if we unpack the sampling distribution of $\bar{X}$ a bit. Using (4), we might write:

[4] A good rule of thumb is that we need $\pi N \geq 5$ for $\pi \leq 0.5$ and $(1-\pi)N > 5$ otherwise; in other words, for $\pi = 0.25$, we can get away with as few as 20 observations, but for $\pi = 0.01$ (or, equivalently, $\pi = 0.99$) we need closer to 500 observations for the normal distribution to hold.

Figure 2: Confidence Intervals for $\hat{\pi}$ for $N = 20$ (black dashes), $N = 100$ (red dashes), and $N = 400$ (green dashes)



$$T = \frac{\bar{X} - \mu}{s/\sqrt{N}} \tag{9}$$

Here, $\bar{X}$ is the sample mean, and s^2 is the sample variance (i.e., the sample analogue to σ^2, the variance of X). Note that we are using s rather than σ because we do not know σ directly; you can think of s as an estimate of it.

In large samples, s converges in distribution to σ; that, in turn, means that (9) takes on a $\mathcal{N}(0, 1)$ distribution. In small samples, however, the two diverge; in such circumstances, the quantity in (9) follows a t distribution with $N - 1$ degrees of freedom. That, in turn, means that we need to use "critical values" from a t distribution (rather than a standard normal) when calculating confidence intervals for means in small samples. That is, we now construct our c.i.s according to

$$[\bar{X}_L, \bar{X}_U] = \bar{X} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{N}}\right) \tag{10}$$

Talking About Confidence Intervals

The value of a confidence interval lies in our ability to use it to say something about the population from which the sample is drawn. We do this by integrating our knowledge of the sample statistic value itself, its sampling distribution, and the sample size to consider the "long run" distribution of sample statistics $\hat{\theta}$.

Over a large number of repeated samples, confidence intervals constructed by the methods described here will contain the true value of θ $[(1-\alpha)\times 100]\%$ of the time. Put differently, $[(1-\alpha)\times 100]\%$ of all confidence intervals constructed from independent simple random samples will contain the population parameter θ, and $(\alpha \times 100)\%$ of them will not. In this way, a given confidence interval can tell us something about the population, though it can never tell us for *certain* whether θ is in $[\hat{\theta}_L, \hat{\theta}_U]$.

It is critically important to remember that *any given confidence interval, based on any single sample, either contains the "true" population value θ or it does not.* It makes no sense to – and so, one should never – say (e.g.) "There is a 95% chance that our confidence interval contains the true population value θ." It is more accurate to say simply that "The estimated 95% confidence interval for θ is $[\hat{\theta}_L, \hat{\theta}_U]$."

An Example with Data: Back to the Court...

Let's return to the Warren & Burger Court data from last week. One variable – `constit` – was coded one if the cases was decided on constitutional grounds, and zero otherwise. The true "population" proportion is $\pi = 0.2536$; we'll use this as an example of how we can learn about that parameter through the use of confidence intervals.

```
> summary(WB)
        us              id             amrev          amaff
 394/0310:  15   Min.   :   1   Min.   : 0   Min.   : 0
 390/0747:  14   1st Qu.:1791   1st Qu.: 0   1st Qu.: 0
 389/0486:  12   Median :3581   Median : 0   Median : 0
 375/0002:  10   Mean   :3581   Mean   : 0   Mean   : 0
 375/0032:   9   3rd Qu.:5371   3rd Qu.: 0   3rd Qu.: 0
 391/0009:   9   Max.   :7161   Max.   :33   Max.   :37
 (Other) :7092
     sumam         fedpet          constit          sgam
 Min.   : 0   Min.   :0.00   Min.   :0.00   Min.   :0.00
 1st Qu.: 0   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
 Median : 0   Median :0.00   Median :0.00   Median :0.00
 Mean   : 1   Mean   :0.17   Mean   :0.25   Mean   :0.08
 3rd Qu.: 1   3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:0.00
 Max.   :39   Max.   :1.00   Max.   :1.00   Max.   :1.00
```

We can begin by considering a rather small random sample of cases ($N = 20$), and calculating the confidence interval for π based on that sample:

```
> set.seed(7222009)
```

```
> WBsample <- with(WB, sample(constit,20,replace=F))
> summary(WBsample)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     0.0     0.0     0.1     0.0     1.0
> LB <- mean(WBsample) - 1.96*sqrt((mean(WBsample)*(1-mean(WBsample)))/(20))
> UB <- mean(WBsample) + 1.96*sqrt((mean(WBsample)*(1-mean(WBsample)))/(20))
> print(c(LB,UB))
[1] -0.031  0.231
```

The c.i. for this sample is $[-0.031, 0.231]$, which means that in repeated random samples from this population, we would expect the "true" population parameter to be contained in an interval constructed in this way 95% of the time.[5]

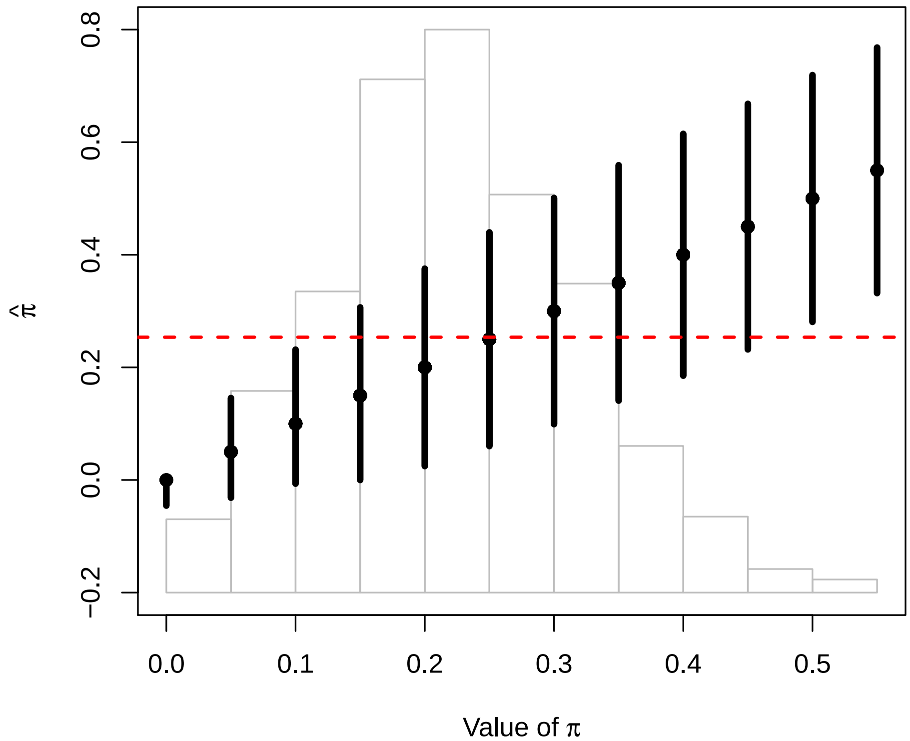
Of course, here we are 100% sure that π is in the c.i. given, because we (unrealistically) know what π is. We can illustrate the idea of a confidence interval by doing what we did above, say, 100 times, and then seeing how many of the resulting c.i.s contain the value 0.2536. To do that, we can modify (e.g.) the R code we used last week. First, a little code:

```
N <- 20
reps <- 1000
PI20 <- numeric(reps)
UB20<-numeric(reps)
LB20<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI20[i] <- bar$estimate
  LB20[i] <- PI20[i] - 1.96 * sqrt((PI20[i] * (1-PI20[i]))/(N))
  UB20[i] <- PI20[i] + 1.96 * sqrt((PI20[i] * (1-PI20[i]))/(N))
  }
```

Here, we drew 1000 samples with $N = 20$, calculated each sample's 95% c.i., and then saved them. The results look like this:

[5] Note again: this does *not* mean that "we are 95% sure that *this particular confidence interval* contains π; as we noted above, it either does or it doesn't...

Figure 3: 1000 Confidence Intervals for $\hat{\pi}_{\texttt{constit}}$ for $N = 20$



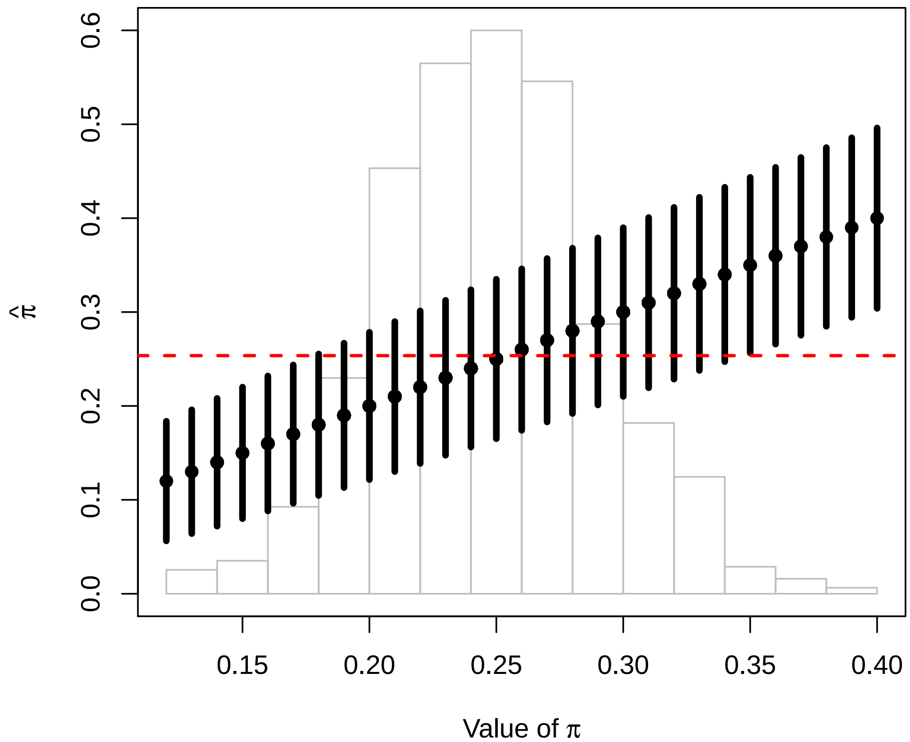
Note that we have three observations with $\hat{\pi} = 0$, 25 with $\hat{\pi} = 0.05$, 77 with $\hat{\pi} = 0.10$, and 14 with $\hat{\pi} \geq 0.50$, all of which have calculated c.i.s that do not include the "true" value $\pi = 0.2536$. That's 139/1000, or $\alpha = 0.14$, which is quite different from $\alpha = 0.05$. We'll discuss why that is in a minute.

We can modify the code slightly to do the same thing with 1000 samples of $N = 100$:

```
N <- 100
reps <- 1000
PI100 <- numeric(reps)
UB100<-numeric(reps)
LB100<-numeric(reps)
set.seed(7222009)
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI100[i] <- bar$estimate
  LB100[i] <- PI100[i] - 1.96 * sqrt((PI100[i] * (1-PI100[i]))/(N))
  UB100[i] <- PI100[i] + 1.96 * sqrt((PI100[i] * (1-PI100[i]))/(N))
```

```
}
```

Figure 4: 1000 Confidence Intervals for $\hat{\pi}_{\texttt{constit}}$ for $N = 100$



Here, there are only far fewer samples whose confidence intervals do not include π. That gets us much closer to $\alpha = 0.05$, as we expect it to be. What we say here is that the "coverage probabilities" are getting better as the size of the sample increases.[6]

If we do the same for 1000 samples each with $N = 400$, the coverage is also very good:

```
N <- 400
reps <- 1000
PI400 <- numeric(reps)
UB400<-numeric(reps)
LB400<-numeric(reps)
set.seed(7222009)
```

[6] What's actually happening is that, as the sample size increases, the Normal distribution that we've been using implicitly to calculate our confidence intervals becomes a more and more accurate representation of the shape of the sampling distribution for $\hat{\pi}$.

```
for (i in 1:reps) {
  foo <- with(WB, sample(constit,N,replace=F))
  bar <- prop.test(sum(foo),length(foo),correct=FALSE)
  PI400[i] <- bar$estimate
  LB400[i] <- PI400[i] - 1.96 * sqrt((PI400[i] * (1-PI400[i]))/(N))
  UB400[i] <- PI400[i] + 1.96 * sqrt((PI400[i] * (1-PI400[i]))/(N))
}
```

While I haven't shown them here the coverage probability is more-or-less perfect (955/1000). In each figure, the density plot overlaid shows the distribution of estimated means ($\hat{\pi}$s). Note that they look both increasingly Normal, and that their range / standard deviation declines, as the sample sizes increase.

Figure 5: 1000 Confidence Intervals for $\hat{\pi}_{\texttt{constit}}$ for $N = 400$