STA261: Lecture 5

Sampling Distributions

Alex Stringer

July 18th, 2018

Disclaimer

The materials in these slides are intended to be a companion to the course textbook, *Mathematical Statistics and Data Analysis, Third Edition*, by John A Rice. Material in the slides may or may not be taken directly from this source. These slides were organized and typeset by Alex Stringer.

A big thanks to Jerry Brunner as well for providing inspiration for assignment questions.

License

Sampling Distribution

We defined the sampling distribution of an estimator as its probability distribution.

Estimators are random variables, because they are functions of the sample, which is itself random.

So they have probability distributions.

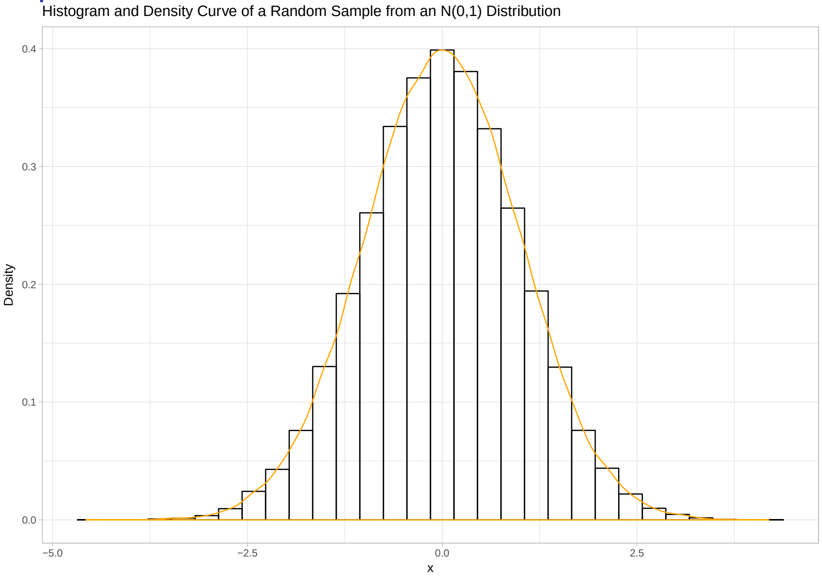
But what does this mean?

Sampling Distribution

When we talk of a random variable having a probability distribution, $X \sim F_\theta$, we usually mean to describe the set of plausible outcomes if we sampled many values of it.

For example if $X \sim N(0, 1)$, we picture a bell curve (the density). If we sampled many values of X and made a histogram, the density curve would touch the tops of the bars.

Example


Histogram and Density Curve of a Random Sample from an N(0,1) Distribution
Density
0.0
0.1
0.2
0.3
0.4
−5.0
−2.5
0.0
2.5
x

Sampling Distribution

For an estimator, we only observe one dataset, and calculate one value.

Key point: the dataset we observed was one of may possible datasets we could have observed.

The data is *random*. If we repeated our experiment, we would get a different dataset, and a different estimate of θ.

$\hat{\theta}$ is a realization of a random variable.

Example: normal sample mean

We say that if $X_i \sim N(\mu, \sigma^2)$, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Proof: ...

Simulate

To illustrate this computationally, fix $\mu = 0$ and $\sigma^2 = 1$ and consider the following procedure:

- Randomly sample $B = 10,000$ datasets of size $n = 5$ from a $N(0, 1)$ distribution
- Calculate $\bar{X}$ for each, so we get $B = 10,000 \bar{X}$'s
- Those $B = 10,000$ $\bar{X}$'s are a random sample from the sampling distribution of $\bar{X}$. They should follow a $N\left(0, \frac{1}{\sqrt{5}}\right)$ distribution

Example


Histogram and Density Curve of Xbar
For X sampled from a N(0,1) distribution
Density
0.75
0.50
0.25
0.00
−1
0
1
2
x

Example

```
## Mean of Xbar =  0.007
## SD of Xbar =  0.449
```

```
## Theoretical mean of Xbar = 0,
## Theoretical SD of Xbar =  0.447
```

Example: normal variance with known mean

If $X_i \sim N(\mu, \sigma^2)$ and we know μ, and

$$s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

then

$$\frac{n s_n^2}{\sigma^2} \sim \chi_n^2$$

where the so-called "Chi-Square" distribution is a $Gamma(n/2, 2)$.

Proof: ...

Example: normal variance with unknown mean

Now let $X_i \sim N(\mu, \sigma^2)$, but we don't know μ, so we need to plug in $\bar{X}$. Defining

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

we have a similar result,

$$\frac{(n-1)s_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$$

Proof: ...

Example: standardized normal with known variance

Recall if $X \sim N(\mu, \sigma^2)$, then

$$Z \equiv \frac{X - \mu}{\sigma} \sim N(0, 1)$$

(proof: ...)

This is useful because it gives us

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

which we will use later in the course when we learn about *interval estimation*.

This allows us to make probabilistic statements about how far away we might expect $\bar{X}$ to be from μ in a given sample.

We don't know σ

...but not really, because we don't know σ.

Consider these two results side-by-side: for $X \sim N(\mu, \sigma^2)$,

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

$$\frac{(n-1)s_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$$

Define

$$T = \frac{\bar{X} - \mu}{s_{n-1}/\sqrt{n}}$$

which looks like the simple replacement of σ by s_{n-1}.

Studentizing

Notice that

$$
\begin{aligned}
T &= \frac{\bar{X} - \mu}{s_{n-1}/\sqrt{n}} \\
&= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \times \left(\frac{(n-1)s_{n-1}^2}{\sigma^2} / (n-1) \right)^{-1/2} \\
&\overset{d}{=} \frac{Z}{\sqrt{W_k/k}}
\end{aligned}
$$

with $Z \sim N(0,1)$ and $W_k \sim \chi_k^2$

T-Distribution

Defintion: t-distribution: a random variable T follows a *t-distribution* with k degrees of freedom, $T \sim t_k$, if

$$T \stackrel{d}{=} \frac{Z}{\sqrt{W_k/k}}$$

with $Z \sim N(0,1)$, $W_k \sim \chi^2_k$, and $Z \perp W_k$

We can find the density of T: ...

Studentizing

This is relevant because now we have the distribution of

$$T = \frac{\bar{X} - \mu}{s_{n-1}/\sqrt{n}}$$

in which the only unknown for a given sample is μ. Hence we can look at the value of T for a given sample by plugging in candidate values of μ, and seeing how likely it was that we observed the sample we observed.

Take that in, because that's the basis of *frequentist inference*.

. . . or do we? We never said that $Z \perp W_k$ in our original definition of T.

Studentizing

Theorem: independence of normal sample mean and sample variance. Let $X \sim N(\mu, \sigma^2)$ and let

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

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

Then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, $\frac{(n-1)s_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$, and $\bar{X} \perp s_{n-1}^2$.

Proof 1 (moment-generating functions): ...

Proof 2 (geometrical): ...

Approximate sampling distributions

We don't always know the sampling distribution of our estimator exactly; the normal distribution is very special.

But, recall the (slightly modified) version of the Central Limit Theorem: as long as $E(X) < \infty$ and $Var(X) < \infty$,

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

More practically, for any fixed $n \in \mathbb{N}$, we can say

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

where $\dot{\sim}$ means "approximately distributed as".

Approximate sampling distributions

So we can use the results derived here as long as the sample size is “large”.

We will also prove later a different Central Limit Theorem for the Maximum Likelihood Estimator (and you'll see why we gave names to the score and observed/expected information functions).