STA286 Lecture 21

Neil Montgomery

Last edited: 2017-03-13 13:06

a quick note—n "large enough" in the Bernoulli(p) special case

The probability p is in some sense a "shape" parameter for Binomial distributions, in the sense that the close p is to 0 (or 1), the more right (left) skewed the distribution is.

a quick note—n "large enough" in the Bernoulli(p) special case

The probability p is in some sense a "shape" parameter for Binomial distributions, in the sense that the close p is to 0 (or 1), the more right (left) skewed the distribution is.

It has been observed through experience and simulation that:

$$np \geqslant 5 \qquad \text{and} \qquad n(1-p) \geqslant 5$$

is sufficient for a good normal approximation.

a quick note—n "large enough" in the Bernoulli(p) special case

The probability p is in some sense a "shape" parameter for Binomial distributions, in the sense that the close p is to 0 (or 1), the more right (left) skewed the distribution is.

It has been observed through experience and simulation that:

$$np \geqslant 5 \qquad \text{and} \qquad n(1-p) \geqslant 5$$

is sufficient for a good normal approximation.

Although this suggestion depends on p, which is usually unknown.

a quick note—n "large enough" in the Bernoulli(p) special case

The probability p is in some sense a "shape" parameter for Binomial distributions, in the sense that the close p is to 0 (or 1), the more right (left) skewed the distribution is.

It has been observed through experience and simulation that:

$$np \geqslant 5 \qquad \text{and} \qquad n(1-p) \geqslant 5$$

is sufficient for a good normal approximation.

Although this suggestion depends on p, which is usually unknown.

We'll revisit this issue when the time comes to discuss ways to evaluate the empirical accuracy of a normal approximation.

more normal-based "sampling distributions"

The main focus will be on $\overline{X}$ and friends, because the most common statistical problem is to make statements about an unknown population mean μ and $\overline{X}$ will be used as the guess.

more normal-based "sampling distributions"

The main focus will be on $\overline{X}$ and friends, because the most common statistical problem is to make statements about an unknown population mean μ and $\overline{X}$ will be used as the guess.

It turns out $\overline{X}$ is BFF with the sample variance:

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

more normal-based "sampling distributions"

The main focus will be on $\overline{X}$ and friends, because the most common statistical problem is to make statements about an unknown population mean μ and $\overline{X}$ will be used as the guess.

It turns out $\overline{X}$ is BFF with the sample variance:

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

Why? Because the analysis of the properties of $\overline{X}$ will depend on:

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

more normal-based "sampling distributions"

The main focus will be on $\overline{X}$ and friends, because the most common statistical problem is to make statements about an unknown population mean μ and $\overline{X}$ will be used as the guess.

It turns out $\overline{X}$ is BFF with the sample variance:

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

Why? Because the analysis of the properties of $\overline{X}$ will depend on:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \text{ (or } \sim^{approx}\text{) } N(0,1)$$

more normal-based "sampling distributions"

The main focus will be on $\overline{X}$ and friends, because the most common statistical problem is to make statements about an unknown population mean μ and $\overline{X}$ will be used as the guess.

It turns out $\overline{X}$ is BFF with the sample variance:

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

Why? Because the analysis of the properties of $\overline{X}$ will depend on:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \text{ (or } \sim^{approx}\text{) } N(0,1)$$

but typically σ is also unknown, and $\sqrt{S^2}$ will be used as *its* guess.

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$M_{Z^2}(t) = E\left(e^{tZ^2}\right) = \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$
\begin{aligned}
M_{Z^2}(t) = E\left(e^{tZ^2}\right) &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2(1-2t)}\, dz
\end{aligned}
$$

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$
\begin{aligned}
M_{Z^2}(t) = E\left(e^{tZ^2}\right) &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2(1-2t)}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{\left(\frac{0.5}{0.5-t}\right)^{0.5}}\right)^2}
\end{aligned}
$$

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$
\begin{aligned}
M_{Z^2}(t) = E\left(e^{tZ^2}\right) &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2 (1-2t)}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\left(\frac{0.5}{0.5-t}\right)^{0.5} \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{\left(\frac{0.5}{0.5-t}\right)^{0.5}}\right)^2}
\end{aligned}
$$

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$
\begin{aligned}
M_{Z^2}(t) = E\left(e^{tZ^2}\right) &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2(1-2t)}\, dz \\
&= \left(\frac{0.5}{0.5-t}\right)^{0.5} \int_{-\infty}^{\infty} \frac{1}{\left(\frac{0.5}{0.5-t}\right)^{0.5}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{\left(\frac{0.5}{0.5-t}\right)^{0.5}}\right)^2}
\end{aligned}
$$

the distribution of the sample variance S^2 - I

First we'll consider the distribution of Z^2 when $Z \sim N(0,1)$.

$$M_{Z^2}(t) = E\left(e^{tZ^2}\right) = \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2(1-2t)}\, dz$$

$$= \left(\frac{0.5}{0.5-t}\right)^{0.5} \int_{-\infty}^{\infty} \frac{1}{\left(\frac{0.5}{0.5-t}\right)^{0.5}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{\left(\frac{0.5}{0.5-t}\right)^{0.5}}\right)^2}$$

The integrand is a $N\left(0, \frac{0.5}{0.5-t}\right)$ density, so the integral equals 1, leaving

$M_{Z^2}(t) = \left(\frac{0.5}{0.5-t}\right)^{0.5}$

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

Synonym: If $X \sim \text{Gamma}\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$ then we give X a special name: "chi-square" distribution is parameter ν.

$$X \sim \chi^2_\nu$$

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

Synonym: If $X \sim \text{Gamma}\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$ then we give X a special name: "chi-square" distribution is parameter ν.

$$X \sim \chi^2_\nu$$

The parameter ν gets a curious name: *degrees of freedom*.

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

Synonym: If $X \sim \text{Gamma}\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$ then we give X a special name: "chi-square" distribution is parameter ν.

$$X \sim \chi^2_\nu$$

The parameter ν gets a curious name: *degrees of freedom*.

Theorem: If $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$, then $\sum_{i=1}^{n} Z^2 \sim \chi^2_n$.

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

Synonym: If $X \sim \text{Gamma}\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$ then we give X a special name: "chi-square" distribution is parameter ν.

$$X \sim \chi^2_\nu$$

The parameter ν gets a curious name: *degrees of freedom*.

Theorem: If $Z_1, \ldots, Z_n$ are i.i.d. $N(0, 1)$, then $\sum_{i=1}^{n} Z^2 \sim \chi^2_n$.

Therefore: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then:

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$$

the distribution of the sample variance S^2 - II

A look at the list of mgfs reveals that $Z^2 \sim \text{Gamma}\left(\alpha = \frac{1}{2}, \lambda = \frac{1}{2}\right)$

Synonym: If $X \sim \text{Gamma}\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$ then we give X a special name: "chi-square" distribution is parameter ν.

$$X \sim \chi^2_\nu$$

The parameter ν gets a curious name: *degrees of freedom*.

Theorem: If $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$, then $\sum_{i=1}^{n} Z^2 \sim \chi^2_n$.

Therefore: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then:

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$$

Theorem: If X and Y are independent with $X \sim \chi^2_m$ and $X + Y \sim \chi^2_{n+m}$, then $Y \sim \chi^2_m$.

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

Finally: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$:

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

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

Finally: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$:

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} + \overline{X} - \mu \right)^2$$

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

Finally: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$:

$$
\begin{aligned}
\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} + \overline{X} - \mu\right)^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\overline{X} - \mu\right)^2 + 0
\end{aligned}
$$

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

Finally: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$:

$$\begin{aligned}
\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} + \overline{X} - \mu \right)^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\overline{X} - \mu \right)^2 + 0 \\
&= \frac{n-1}{\sigma^2} S^2 + \left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2
\end{aligned}$$

the distribution of the sample variance S^2 - III

Magical truth: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$, then $\overline{X}$ and S^2 are independent.

Finally: If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma)$:

$$\begin{aligned}
\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} + \overline{X} - \mu \right)^2 \\
&= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\overline{X} - \mu \right)^2 + 0 \\
&= \frac{n-1}{\sigma^2} S^2 + \left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2
\end{aligned}$$

Conclusion: $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$

the t distributions - I

I mentioned earlier that properties of $\overline{X}$ will depend on:

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

and that σ will have to be guessed by $S = \sqrt{S^2}$.

the t distributions - I

I mentioned earlier that properties of $\overline{X}$ will depend on:

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

and that σ will have to be guessed by $S = \sqrt{S^2}$.

However, when the constant σ is replaced with the random variable S, the result is no longer $N(0, 1)$.

the t distributions - I

I mentioned earlier that properties of $\overline{X}$ will depend on:

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

and that σ will have to be guessed by $S = \sqrt{S^2}$.

However, when the constant σ is replaced with the random variable S, the result is no longer $N(0,1)$.

It turns out it is possible to derive the density of:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

the t distributions - II

The density $f(t)$ of T is is primarily nice to look at:

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi}\nu}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is called the "degrees of freedom" and in this case is $n-1$. You can think of $n-1$ as having been "inherited" from the denominator of T.

the t distributions - II

The density $f(t)$ of T is is primarily nice to look at:

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi}\nu}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is called the "degrees of freedom" and in this case is $n-1$. You can think of $n-1$ as having been "inherited" from the denominator of T.

Important properties:

- Symmetric and bell shaped.

the t distributions - II

The density $f(t)$ of T is is primarily nice to look at:

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi}\nu}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is called the "degrees of freedom" and in this case is $n-1$. You can think of $n-1$ as having been "inherited" from the denominator of T.

Important properties:

- Symmetric and bell shaped.
- As ν gets big, form of $f(t)$ approaches $e^{-t^2/2}\ldots$

the t distributions - II

The density $f(t)$ of T is is primarily nice to look at:

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi}\nu}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is called the "degrees of freedom" and in this case is $n-1$. You can think of $n-1$ as having been "inherited" from the denominator of T.

Important properties:

- Symmetric and bell shaped.
- As ν gets big, form of $f(t)$ approaches $e^{-t^2/2}$...
- ...in other words *as the sample size gets large, T starts to look like $Z \sim N(0,1)$*

the t distributions - II

The density $f(t)$ of T is is primarily nice to look at:

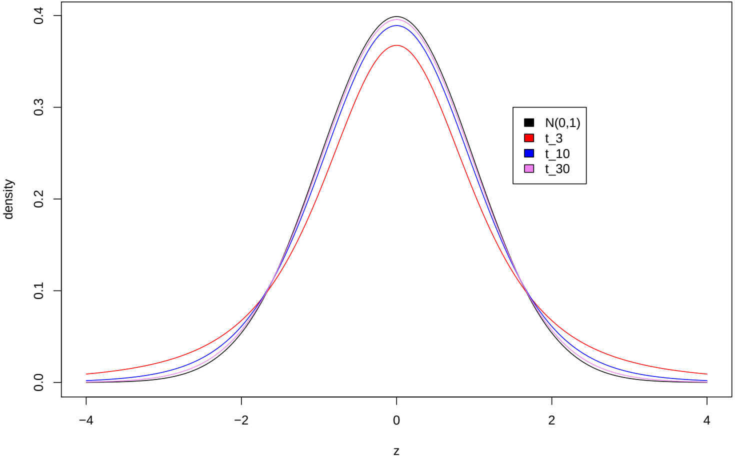
$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]\sqrt{\pi}\nu}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

where ν is called the "degrees of freedom" and in this case is $n-1$. You can think of $n-1$ as having been "inherited" from the denominator of T.

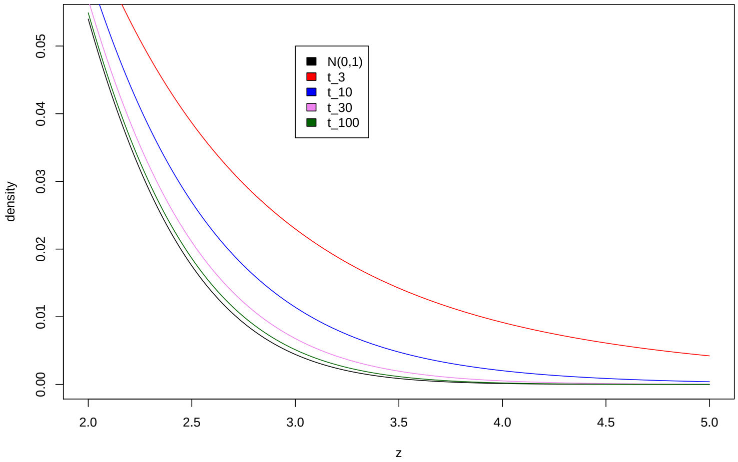
Important properties:

- Symmetric and bell shaped.
- As ν gets big, form of $f(t)$ approaches $e^{-t^2/2}\ldots$
- $\ldots$ in other words *as the sample size gets large, T starts to look like* $Z \sim N(0,1)$
- no anti-derivative, so a table of t probabilities needed on tests.

overall pictures of t_ν



pictures of t_ν in the "tail"



the F distributions

We might need the following technical result as well.

the F distributions

We might need the following technical result as well.

If $U \sim \chi^2_m$ and $V \sim \chi^2_n$ and $U \perp V$ then we say:

$$F = \frac{U/m}{V/n} \sim F_{m,n}$$

or "F has an F distribution with m and n degrees of freedom."

the F distributions

We might need the following technical result as well.

If $U \sim \chi^2_m$ and $V \sim \chi^2_n$ and $U \perp V$ then we say:

$$F = \frac{U/m}{V/n} \sim F_{m,n}$$

or "F has an F distribution with m and n degrees of freedom."

F distributions happen when it makes sense to consider the ratio of sums of squared normals. It is not yet obvious when this might take place.

the F distributions

We might need the following technical result as well.

If $U \sim \chi^2_m$ and $V \sim \chi^2_n$ and $U \perp V$ then we say:

$$F = \frac{U/m}{V/n} \sim F_{m,n}$$

or "F has an F distribution with m and n degrees of freedom."

F distributions happen when it makes sense to consider the ratio of sums of squared normals. It is not yet obvious when this might take place.

The density is nasty, etc.

pictures of some F distributions