Precept 2: Likelihood inference

Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

February 16, 2018

Replication Paper

An example of how to write to authors requesting data:

> Dear Professor [name],
>
> I am a graduate student in sociology at Princeton, and I really enjoyed your paper, "[name of paper]." I would like to replicate it and think about ways to extend it to new research questions. Before e-mailing, I found the [data source] online and downloaded it, but it seems difficult to re-create what you had from scratch. Would you be willing to share your data and code files with me?
>
> Thanks so much,
>
> [Your name]

Any other replication issues to discuss?

Likelihood

Steps of likelihood inference:

Likelihood

Steps of likelihood inference:

1. Assume a **data generating process**.

Likelihood

Steps of likelihood inference:

1. Assume a **data generating process**.
2. Derive the **likelihood**.

Likelihood

Steps of likelihood inference:

1. Assume a **data generating process**.
2. Derive the **likelihood**.
3. Maximize the likelihood to get the **MLE**.

Likelihood

Steps of likelihood inference:

1. Assume a **data generating process**.
2. Derive the **likelihood**.
3. Maximize the likelihood to get the **MLE**.
4. Derive standard errors from the inverse of the **Fisher information**

Binomial example

What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis?
2. What is the outcome?
3. What is its support?
4. What distribution might it follow?

Binomial example

What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis? **Students indicated by i**
2. What is the outcome?
3. What is its support?
4. What distribution might it follow?

Binomial example

What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis? **Students indicated by i**
2. What is the outcome? **$Y_i = \#$ R&Rs out of 5 papers**
3. What is its support?
4. What distribution might it follow?

Binomial example

What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis? **Students indicated by i**
2. What is the outcome? **$Y_i = \#$ R&Rs out of 5 papers**
3. What is its support? $\{\mathbf{0,1,2,3,4,5}\}$
4. What distribution might it follow?

Binomial example

What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis? **Students indicated by i**
2. What is the outcome? **Y_i = # R&Rs out of 5 papers**
3. What is its support? **$\{0,1,2,3,4,5\}$**
4. What distribution might it follow? **Binomial**

Binomial example

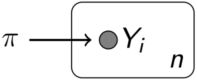
What is the probability that a Princeton Ph.D. student in sociology who submits a paper to a journal is invited to revise and resubmit? We have data on $n = 20$ students who each submit 5 papers. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Think, pair, share: Translate this into a data generating process.

1. What is the unit of analysis? **Students indicated by i**
2. What is the outcome? **Y_i = # R&Rs out of 5 papers**
3. What is its support? **$\{0,1,2,3,4,5\}$**
4. What distribution might it follow? **Binomial**

For $i = 1, \ldots, n$:

$$Y_i \sim \text{Binomial}(5, \pi)$$



For more on these graphs, see Airoldi 2007 [link]

We **assume:**

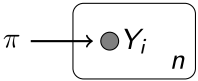
- Response is binomial, with each paper independent
- All submissions from all students having the same probability π of success.

From the data, we **learn**:

- The value of the parameter $\hat{\pi}$ under which the observed data would be most likely.

Binomial likelihood

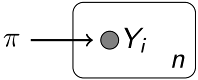
For $i = 1, \ldots, n$:
$Y_i \sim \text{Binomial}(5, \pi)$



What is the likelihood of π given one y_i?

Binomial likelihood

For $i = 1, \ldots, n$:
$$Y_i \sim \text{Binomial}(5, \pi)$$

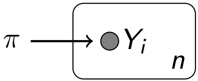


What is the likelihood of π given one y_i?

$$L(\pi \mid y_i) = \mathrm{P}(y_i \mid \pi) = \binom{5}{y_i} \pi^{y_i}(1-\pi)^{5-y_i}$$

Binomial likelihood

For $i = 1, \ldots, n$:
$Y_i \sim \text{Binomial}(5, \pi)$



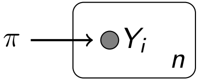
What is the likelihood of π given one y_i?

$$L(\pi \mid y_i) = \text{P}(y_i \mid \pi) = \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

What is the likelihood of π given all n observations?

Binomial likelihood

For $i = 1, \ldots, n$:
$$Y_i \sim \text{Binomial}(5, \pi)$$



What is the likelihood of π given one y_i?

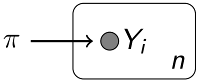
$$L(\pi \mid y_i) = \mathrm{P}(y_i \mid \pi) = \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

What is the likelihood of π given all n observations?

$$L(\pi \mid \vec{y}) = \mathrm{P}(\vec{y} \mid \pi) = \mathrm{P}(y_1, \ldots, y_n \mid \pi)$$

Binomial likelihood

For $i = 1, \dots, n$:
$Y_i \sim \text{Binomial}(5, \pi)$



What is the likelihood of π given one y_i?

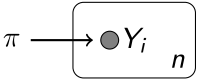
$$L(\pi \mid y_i) = \text{P}(y_i \mid \pi) = \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

What is the likelihood of π given all n observations?

$$L(\pi \mid \vec{y}) = \text{P}(\vec{y} \mid \pi) = \text{P}(y_1, \dots, y_n \mid \pi) \overset{\substack{\text{Assumes independence} \\ \downarrow}}{=} \prod_{i=1}^{n} \text{P}(y_i \mid \pi)$$

Binomial likelihood

For $i = 1, \ldots, n$:
$Y_i \sim \text{Binomial}(5, \pi)$



What is the likelihood of π given one y_i?

$$L(\pi \mid y_i) = \mathrm{P}(y_i \mid \pi) = \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

What is the likelihood of π given all n observations?

Assumes independence

$$L(\pi \mid \vec{y}) = \mathrm{P}(\vec{y} \mid \pi) = \mathrm{P}(y_1, \ldots, y_n \mid \pi) = \prod_{i=1}^{n} \mathrm{P}(y_i \mid \pi)$$

$$= \prod_{i=1}^{n} \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

Review of log rules

$$\log(ab) = \log(a) + \log(b)$$
$$\log(e^a) = a$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Products into sums

$$\ell(\pi \mid y_1, \ldots, y_n) =$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Products into sums

$$\ell(\pi \mid y_1, \ldots, y_n) = \log L(\pi \mid y_1, \ldots, y_n)$$

$$=$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Products into sums

$$
\begin{aligned}
\ell(\pi \mid y_1, \ldots, y_n) &= \log L(\pi \mid y_1, \ldots, y_n) \\
&= \log \prod_{i=1}^{n} \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}
\end{aligned}
$$

Apply rules of the log to turn products into sums

$$
=
$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Products into sums

$$\ell(\pi \mid y_1, \ldots, y_n) = \log L(\pi \mid y_1, \ldots, y_n)$$

$$= \log \prod_{i=1}^{n} \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}$$

Apply rules of the log to turn products into sums

$$= \sum_{i=1}^{n} \log \left(\binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i} \right)$$

$$=$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Products into sums

$$
\begin{aligned}
\ell(\pi \mid y_1, \ldots, y_n) &= \log L(\pi \mid y_1, \ldots, y_n) \\
&= \log \prod_{i=1}^{n} \binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i}
\end{aligned}
$$

Apply rules of the log to turn products into sums

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log \left(\binom{5}{y_i} \pi^{y_i} (1-\pi)^{5-y_i} \right) \\
&= \sum_{i=1}^{n} \left(\log \binom{5}{y_i} + y_i \log(\pi) + (5-y_i) \log(1-\pi) \right)
\end{aligned}
$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Remove constants

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(\log \binom{5}{y_i} + y_i \log(\pi) + (5 - y_i) \log(1 - \pi) \right)$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Remove constants

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(\log \binom{5}{y_i} + y_i \log(\pi) + (5 - y_i) \log(1 - \pi) \right)$$

Pull the first term out of the sum

$$=$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Remove constants

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(\log \binom{5}{y_i} + y_i \log(\pi) + (5 - y_i) \log(1 - \pi) \right)$$

Pull the first term out of the sum

$$= \sum_{i=1}^{n} \log \binom{5}{y_i} + \sum_{i=1}^{n} (y_i \log(\pi) + (5 - y_i) \log(1 - \pi))$$

Drop the constant which does not involve π

$$=$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Remove constants

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(\log \binom{5}{y_i} + y_i \log(\pi) + (5 - y_i) \log(1 - \pi) \right)$$

Pull the first term out of the sum

$$= \sum_{i=1}^{n} \log \binom{5}{y_i} + \sum_{i=1}^{n} (y_i \log(\pi) + (5 - y_i) \log(1 - \pi))$$

Drop the constant which does not involve π

$$= \sum_{i=1}^{n} (y_i \log(\pi) + (5 - y_i) \log(1 - \pi)) + \text{constant}$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} (y_i \log(\pi) + (5 - y_i) \log(1 - \pi))$$

$$=$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$\begin{aligned}
\ell(\pi \mid y_1, \ldots, y_n) &= \sum_{i=1}^{n} \left(y_i \log(\pi) + (5 - y_i) \log(1 - \pi)\right) \\
&= \log \pi \sum_{i=1}^{n} y_i + 5n \log(1 - \pi) - \log(1 - \pi) \sum_{i=1}^{n} y_i \\
&=
\end{aligned}$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$
\begin{aligned}
\ell(\pi \mid y_1, \ldots, y_n) &= \sum_{i=1}^{n} \left(y_i \log(\pi) + (5 - y_i) \log(1 - \pi) \right) \\
&= \log \pi \sum_{i=1}^{n} y_i + 5n \log(1 - \pi) - \log(1 - \pi) \sum_{i=1}^{n} y_i \\
&= (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)
\end{aligned}
$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$\begin{aligned}
\ell(\pi \mid y_1, \ldots, y_n) &= \sum_{i=1}^{n} (y_i \log(\pi) + (5 - y_i) \log(1 - \pi)) \\
&= \log \pi \sum_{i=1}^{n} y_i + 5n \log(1 - \pi) - \log(1 - \pi) \sum_{i=1}^{n} y_i \\
&= (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)
\end{aligned}$$

Simplify by rules of logs

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(y_i \log(\pi) + (5 - y_i) \log(1 - \pi)\right)$$

$$= \log \pi \sum_{i=1}^{n} y_i + 5n \log(1 - \pi) - \log(1 - \pi) \sum_{i=1}^{n} y_i$$

$$= (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

Simplify by rules of logs

$$= \log\left(\frac{\pi}{1 - \pi}\right) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

$\ell(\pi \mid y_1, \ldots, y_n)$: Collect terms involving data

$$\ell(\pi \mid y_1, \ldots, y_n) = \sum_{i=1}^{n} \left(y_i \log(\pi) + (5 - y_i) \log(1 - \pi)\right)$$

$$= \log \pi \sum_{i=1}^{n} y_i + 5n \log(1 - \pi) - \log(1 - \pi) \sum_{i=1}^{n} y_i$$

$$= (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

Simplify by rules of logs

$$= \log \left(\frac{\pi}{1 - \pi}\right) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

We're finished!

Sufficient statistics

The data $y_1, \ldots, y_n$ only enter the likelihood through their **sum**!

$$\ell(\pi \mid y_1, \ldots, y_n) = \log\left(\frac{\pi}{1-\pi}\right)\sum_{i=1}^{n} y_i + 5n\log(1-\pi)$$

We call $\sum_{i=1}^{n} y_i$ a **sufficient statistic**: it provides sufficient information to compute the likelihood.

Think, Pair, Share:

Sufficient statistics

The data $y_1, \ldots, y_n$ only enter the likelihood through their **sum**!

$$\ell(\pi \mid y_1, \ldots, y_n) = \log\left(\frac{\pi}{1-\pi}\right)\sum_{i=1}^{n} y_i + 5n\log(1-\pi)$$

We call $\sum_{i=1}^{n} y_i$ a **sufficient statistic**: it provides sufficient information to compute the likelihood.

Think, Pair, Share:

1. We can compute the likelihood only knowing the total number of graduate student submissions that are given R&Rs. Can you explain this?
2. Why might we want to work with sufficient statistics rather than the full data?

Sufficient statistics

The data $y_1, \ldots, y_n$ only enter the likelihood through their **sum**!

$$\ell(\pi \mid y_1, \ldots, y_n) = \log\left(\frac{\pi}{1-\pi}\right)\sum_{i=1}^{n} y_i + 5n\log(1-\pi)$$

We call $\sum_{i=1}^{n} y_i$ a **sufficient statistic**: it provides sufficient information to compute the likelihood.

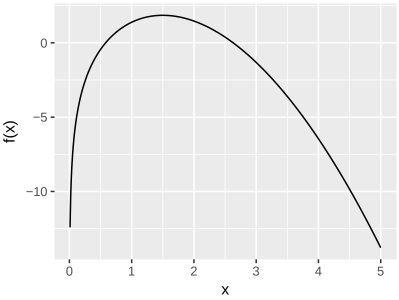
Think, Pair, Share:

1. We can compute the likelihood only knowing the total number of graduate student submissions that are given R&Rs. Can you explain this?
    - Since we assumed every submission had the same probability of success, it's like we had 5n Bernoulli trials. There is no need to distinguish who submitted them!
2. Why might we want to work with sufficient statistics rather than the full data?
    - Sufficient statistics can save disk space in more complex problems - no need to store all the data!

Calculus review: Derivatives

Suppose we have a function

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

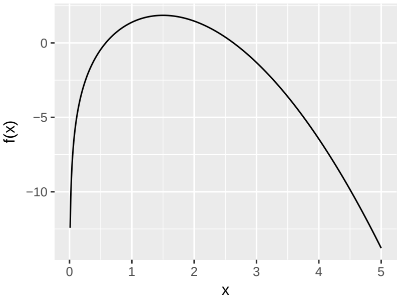


What is the derivative?

Calculus review: Derivatives

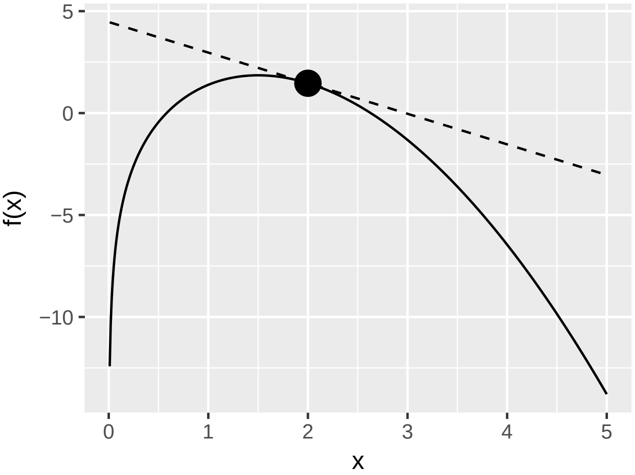
Suppose we have a function

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$



What is the derivative? **It is the slope.**

Calculus review: Derivatives



Calculus review: A few derivative rules

$$\frac{\partial}{\partial x} x^a = a x^{a-1}$$

$$\frac{\partial}{\partial x} \log x = \frac{1}{x}$$

$$\frac{\partial}{\partial x} f(x) \text{ is often denoted } f'(x)$$

$$\frac{\partial}{\partial x} f(g[x]) = f'(g[x]) g'(x) \text{ (often called the chain rule)}$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\frac{\partial}{\partial x} f(x) =$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\frac{\partial}{\partial x} f(x) = 1 -$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\frac{\partial}{\partial x} f(x) = 1 - 2x +$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\frac{\partial}{\partial x} f(x) = 1 - 2x + \frac{1}{x} +$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\frac{\partial}{\partial x} f(x) = 1 - 2x + \frac{1}{x} + \frac{6x}{3x^2}$$

$$=$$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\begin{aligned} \frac{\partial}{\partial x} f(x) &= 1 - 2x + \frac{1}{x} + \frac{6x}{3x^2} \\ &= 1 - 2x + \frac{1}{x} + \frac{2}{x} \\ &= 1 - 2x + \frac{3}{x} \end{aligned}$$

Let's evaluate the derivative at $x = 2$

Calculus review: Derivatives

Check for understanding:

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$

The derivative is

$$\begin{aligned}\frac{\partial}{\partial x} f(x) &= 1 - 2x + \frac{1}{x} + \frac{6x}{3x^2} \\ &= 1 - 2x + \frac{1}{x} + \frac{2}{x} \\ &= 1 - 2x + \frac{3}{x}\end{aligned}$$

Let's evaluate the derivative at $x = 2$

$$f'(2) = 1 - 2(2) + \frac{3}{2} = -1.5$$

Calculus review: Maximizing a function

$$
\begin{aligned}
f(x) &= x - x^2 + \log(x) + \log(3x^2) \\
f'(x) &= 1 - 2x + \frac{3}{x}
\end{aligned}
$$

How do we maximize this?

Calculus review: Maximizing a function

$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$
$$f'(x) = 1 - 2x + \frac{3}{x}$$

How do we maximize this?

Set the derivative equal to 0 and solve!

Calculus review: Maximizing a function

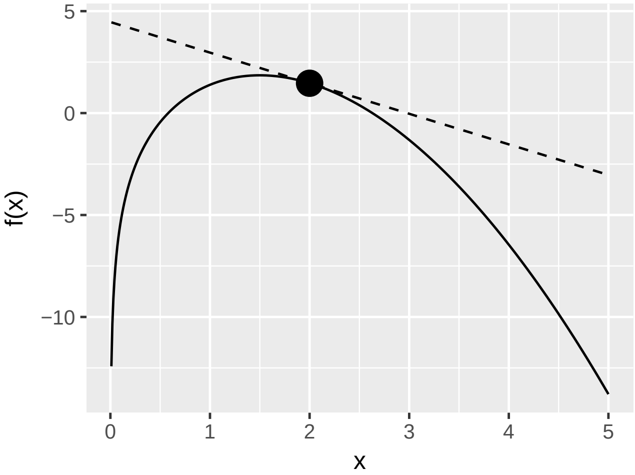
$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$
$$f'(x) = 1 - 2x + \frac{3}{x}$$

How do we maximize this?

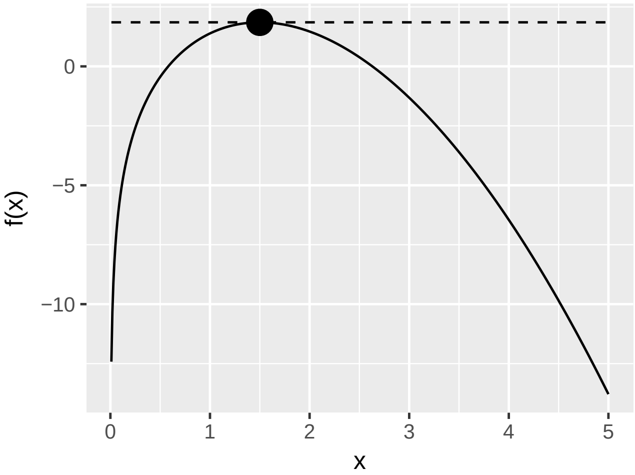
Set the derivative equal to 0 and solve!

(Then check that you find a maximum)

Calculus review: Maximizing a function



Calculus review: Maximizing a function



Calculus review: Maximizing a function

Set the derivative equal to 0

$$
\begin{aligned}
f'(x^*) &= 0 \\
1 - 2x^* + \frac{3}{x^*} &= 0
\end{aligned}
$$

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$f'(x^*) = 0$$

$$1 - 2x^* + \frac{3}{x^*} = 0$$

Useful skill:
Set derivative equal to 0

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$f'(x^*) = 0$$

$$1 - 2x^* + \frac{3}{x^*} = 0$$

$$\frac{3}{x^*} = 2x^* - 1$$

Useful skill:
Set derivative equal to 0

Algebra: Not important

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$f'(x^*) = 0$$

$$1 - 2x^* + \frac{3}{x^*} = 0$$

$$\frac{3}{x^*} = 2x^* - 1$$

$$3 = 2x^{*2} - x^*$$

Useful skill:
Set derivative equal to 0

Algebra: Not important

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$\boxed{\begin{aligned} f'(x^*) &= 0 \\ 1 - 2x^* + \frac{3}{x^*} &= 0 \end{aligned}}$$

Useful skill:
Set derivative equal to 0

$$\begin{aligned} \frac{3}{x^*} &= 2x^* - 1 \\ 3 &= 2x^{*2} - x^* \\ 0 &= 2x^{*2} - x^* - 3 \end{aligned}$$

Algebra: Not important

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$f'(x^*) = 0$$

$$1 - 2x^* + \frac{3}{x^*} = 0$$

Useful skill:
Set derivative equal to 0

$$\frac{3}{x^*} = 2x^* - 1$$

$$3 = 2x^{*2} - x^*$$

$$0 = 2x^{*2} - x^* - 3$$

$$0 = 2x^{*2} - x^* - 3$$

Algebra: Not important

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

$$\begin{aligned}
f'(x^*) &= 0 \\
1 - 2x^* + \frac{3}{x^*} &= 0 \\
\frac{3}{x^*} &= 2x^* - 1 \\
3 &= 2x^{*2} - x^* \\
0 &= 2x^{*2} - x^* - 3 \\
0 &= 2x^{*2} - x^* - 3 \\
0 &= (2x^* - 3)(x^* + 1)
\end{aligned}$$

Useful skill:
Set derivative equal to 0

Algebra: Not important

These are our **critical values**.

Calculus review: Maximizing a function

Set the derivative equal to 0

Useful skill:
Set derivative equal to 0

$$
\begin{aligned}
f'(x^*) &= 0 \\
1 - 2x^* + \frac{3}{x^*} &= 0 \\
\frac{3}{x^*} &= 2x^* - 1 \\
3 &= 2x^{*2} - x^* \\
0 &= 2x^{*2} - x^* - 3 \\
0 &= 2x^{*2} - x^* - 3 \\
0 &= (2x^* - 3)(x^* + 1) \\
x^* &= \{-1, 1.5\}
\end{aligned}
$$

Algebra:
Not important

These are our **critical values**.

Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$f''(x) = \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x}$$

Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1}
\end{aligned}
$$

Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2}
\end{aligned}
$$

Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2} \\
&= -2 - \frac{3}{x^2}
\end{aligned}$$

Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$\begin{aligned} f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\ &= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\ &= -2 - 3x^{-2} \\ &= -2 - \frac{3}{x^2} \end{aligned}$$

$f''(-1) = -2 - \frac{3}{(-1)^2} = -5 < 0$ $\downarrow$ Maximum

$f''(1.5) = -2 - \frac{3}{1.5^2} = -3.333 < 0$ $\downarrow$ Maximum

Back to our example: Maximizing the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

To maximize,

Back to our example: Maximizing the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

To maximize, take the derivative and set it equal to 0!

Back to our example: Maximizing the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

To maximize, take the derivative and set it equal to 0!

$$\frac{\partial}{\partial \pi} \ell(\pi \mid y_1, \ldots, y_n) = \frac{\partial}{\partial \pi} (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

$$=$$

Back to our example: Maximizing the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

To maximize, take the derivative and set it equal to 0!

$$\begin{aligned}
\frac{\partial}{\partial \pi}\ell(\pi \mid y_1, \ldots, y_n) &= \frac{\partial}{\partial \pi}(\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi) \\
&= \frac{\partial}{\partial \pi}(\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi) \\
&=
\end{aligned}$$

Back to our example: Maximizing the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

To maximize, take the derivative and set it equal to 0!

$$\begin{aligned}
\frac{\partial}{\partial \pi}\ell(\pi \mid y_1, \ldots, y_n) &= \frac{\partial}{\partial \pi}(\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi) \\
&= \frac{\partial}{\partial \pi}(\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi) \\
&= \frac{1}{\pi} \sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi}
\end{aligned}$$

$$0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}$$

$$0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}$$

Useful skill:
Set derivative
equal to 0

$$\boxed{0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}}$$

Useful skill:
Set derivative
equal to 0

$$\frac{5n - \sum_{i=1}^{n} y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i$$

Algebra: Not important

$$\boxed{0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}}$$

Useful skill:
Set derivative
equal to 0

$$\frac{5n - \sum_{i=1}^{n} y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i$$

$$(5n - \sum_{i=1}^{n} y_i)\pi^* = (1-\pi^*)\sum_{i=1}^{n} y_i$$

Algebra:
Not important

$$\boxed{0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}}$$

Useful skill: Set derivative equal to 0

$$\frac{5n - \sum_{i=1}^{n} y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i$$

$$(5n - \sum_{i=1}^{n} y_i)\pi^* = (1-\pi^*)\sum_{i=1}^{n} y_i$$

$$5n\pi^* = \sum_{i=1}^{n} y_i$$

Algebra: Not important

$$0 = \frac{1}{\pi^*}\sum_{i=1}^n y_i + \frac{\sum_{i=1}^n y_i - 5n}{1-\pi^*}$$

Useful skill: Set derivative equal to 0

$$\frac{5n - \sum_{i=1}^n y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^n y_i$$

$$(5n - \sum_{i=1}^n y_i)\pi^* = (1-\pi^*)\sum_{i=1}^n y_i$$

$$5n\pi^* = \sum_{i=1}^n y_i$$

$$\pi^* = \frac{\sum_{i=1}^n y_i}{5n}$$

Algebra: Not important

$$0 = \frac{1}{\pi^*}\sum_{i=1}^n y_i + \frac{\sum_{i=1}^n y_i - 5n}{1-\pi^*}$$

Useful skill:
Set derivative equal to 0

$$\frac{5n - \sum_{i=1}^n y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^n y_i$$

Algebra: Not important

$$(5n - \sum_{i=1}^n y_i)\pi^* = (1-\pi^*)\sum_{i=1}^n y_i$$

$$5n\pi^* = \sum_{i=1}^n y_i$$

$$\pi^* = \frac{\sum_{i=1}^n y_i}{5n}$$

This π^* such that $\ell'(\pi \mid y) = 0$ is the **critical value**.

$$\boxed{0 = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-\pi^*}}$$

Useful skill:
Set derivative
equal to 0

$$\frac{5n - \sum_{i=1}^{n} y_i}{1-\pi^*} = \frac{1}{\pi^*}\sum_{i=1}^{n} y_i$$

$$(5n - \sum_{i=1}^{n} y_i)\pi^* = (1-\pi^*)\sum_{i=1}^{n} y_i$$

$$5n\pi^* = \sum_{i=1}^{n} y_i$$

$$\pi^* = \frac{\sum_{i=1}^{n} y_i}{5n}$$

**Algebra:
Not important**

This π^* such that $\ell'(\pi \mid y) = 0$ is the **critical value**. Is it a max?

Back to our example: Maximizing the log likelihood

$$\frac{\partial^2}{\partial \pi^2}\ell(\pi \mid y) = \frac{\partial}{\partial \pi}\left(\frac{\sum_{i=1}^n y_i}{\pi} - \frac{5n - \sum_{i=1}^n y_i}{1-\pi}\right)$$

Back to our example: Maximizing the log likelihood

$$\begin{aligned}
\frac{\partial^2}{\partial \pi^2} \ell(\pi \mid y) &= \frac{\partial}{\partial \pi} \left(\frac{\sum_{i=1}^n y_i}{\pi} - \frac{5n - \sum_{i=1}^n y_i}{1-\pi} \right) \\
&= -\frac{\sum_{i=1}^n y_i}{\pi^2} - \frac{5n - \sum_{i=1}^n y_i}{(1-\pi)^2}
\end{aligned}$$

Back to our example: Maximizing the log likelihood

$$
\begin{aligned}
\frac{\partial^2}{\partial \pi^2}\ell(\pi \mid y) &= \frac{\partial}{\partial \pi}\left(\frac{\sum_{i=1}^n y_i}{\pi} - \frac{5n - \sum_{i=1}^n y_i}{1-\pi}\right) \\
&= -\frac{\sum_{i=1}^n y_i}{\pi^2} - \frac{5n - \sum_{i=1}^n y_i}{(1-\pi)^2} \\
&< 0 \quad \forall \quad \pi
\end{aligned}
$$

The $\forall$ symbol just means "for all"

Back to our example: Maximizing the log likelihood

$$\begin{aligned}\frac{\partial^2}{\partial \pi^2}\ell(\pi \mid y) &= \frac{\partial}{\partial \pi}\left(\frac{\sum_{i=1}^n y_i}{\pi} - \frac{5n - \sum_{i=1}^n y_i}{1-\pi}\right) \\ &= -\frac{\sum_{i=1}^n y_i}{\pi^2} - \frac{5n - \sum_{i=1}^n y_i}{(1-\pi)^2} \\ &< 0 \quad \forall \quad \pi\end{aligned}$$

The $\forall$ symbol just means "for all"

Since the first derivative is 0 and the second derivative is negative, the critical value $\pi^* = \frac{\sum_{i=1}^n y_i}{5n}$ is a maximum.

$$\hat{\pi}_{\text{MLE}} = \frac{\sum_{i=1}^n y_i}{5n}$$

Plotting the log likelihood

$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1 - \pi]) \sum_{i=1}^{n} y_i + 5n \log(1 - \pi)$$

Let's define a function in R that returns the log likelihood given a vector `y`

Plotting the log likelihood

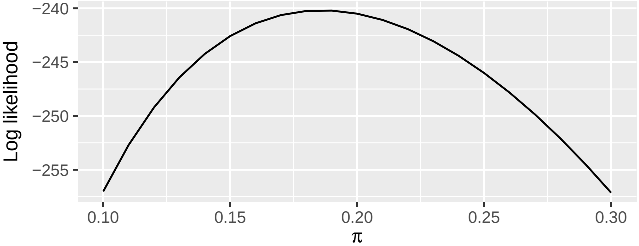
$$\ell(\pi \mid y_1, \ldots, y_n) = (\log \pi - \log[1-\pi]) \sum_{i=1}^{n} y_i + 5n \log(1-\pi)$$

Let's define a function in R that returns the log likelihood given a vector y

```
log.lik <- function(pi,y) {
     (log(pi) - log(1 - pi)) * sum(y) + 5 * length(y) * log(1 - pi)
}
```

Plotting the log likelihood

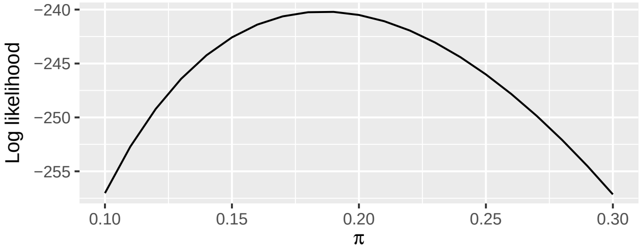
Define some data and make a plot.




```
y <- rbinom(100,5,.2)
```

Plotting the log likelihood

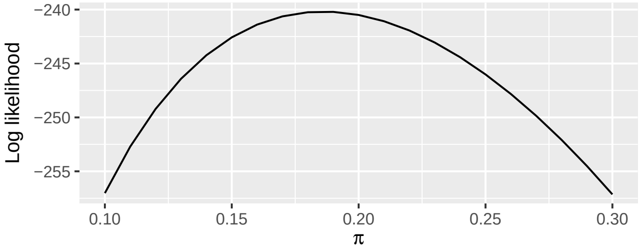
Define some data and make a plot.




```
y <- rbinom(100,5,.2)
data.frame(pi = seq(0.1,0.3,0.01)) %>%
```

Plotting the log likelihood

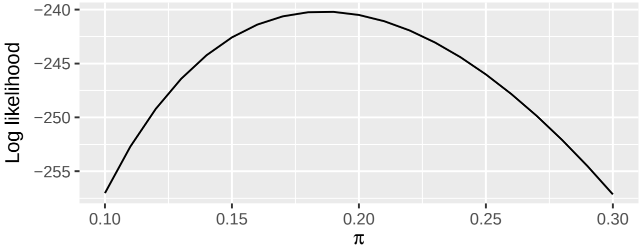
Define some data and make a plot.




```
y <- rbinom(100,5,.2)
data.frame(pi = seq(0.1,0.3,0.01)) %>%
  mutate(`Log likelihood` = log.lik(pi,y)) %>%
```

Plotting the log likelihood

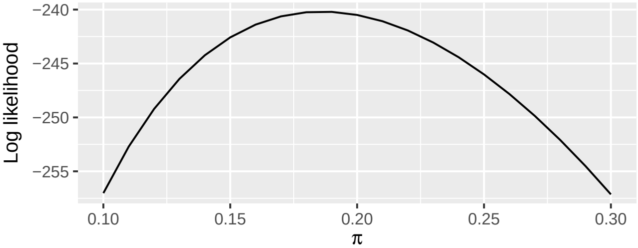
Define some data and make a plot.




```
y <- rbinom(100,5,.2)
data.frame(pi = seq(0.1,0.3,0.01)) %>%
  mutate(`Log likelihood` = log.lik(pi,y)) %>%
  ggplot(aes(x = pi, y = `Log likelihood`)) +
```

Plotting the log likelihood

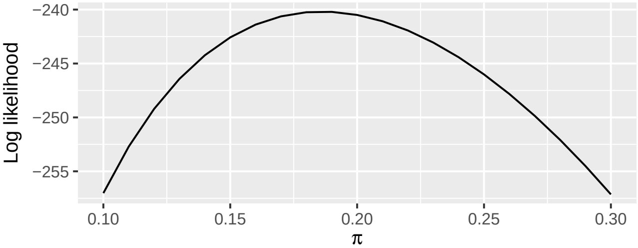
Define some data and make a plot.




```
y <- rbinom(100,5,.2)
data.frame(pi = seq(0.1,0.3,0.01)) %>%
  mutate('Log likelihood' = log.lik(pi,y)) %>%
  ggplot(aes(x = pi, y = 'Log likelihood')) +
  geom_line() +
```

Plotting the log likelihood

Define some data and make a plot.




```
y <- rbinom(100,5,.2)
data.frame(pi = seq(0.1,0.3,0.01)) %>%
  mutate(‘Log likelihood‘ = log.lik(pi,y)) %>%
  ggplot(aes(x = pi, y = ‘Log likelihood‘)) +
  geom_line() +
  scale_x_continuous(name = expression(pi))
```

Finding the maximum numerically

```
> y <- rbinom(100,5,.2)
> optimize(f = log.lik,
+          interval = c(0,1),
+          maximum = T,
+          y = y)
$maximum
[1] 0.1860007

$objective
[1] -240.1853
```

Invariance of the MLE

Theorem (King Sec. 4.4, Casella & Berger Thm 7.2.10)

If $\hat{\theta}_{\text{MLE}}$ is the MLE for θ, then for any function $g(\theta)$ the MLE of $g(\theta)$ is $g(\hat{\theta}_{\text{MLE}})$.

Application: If we knew the true $\pi = \text{P(R\&R)}$, what would be the probability of getting at least 1 R&R out of 5 submissions?

Invariance of the MLE

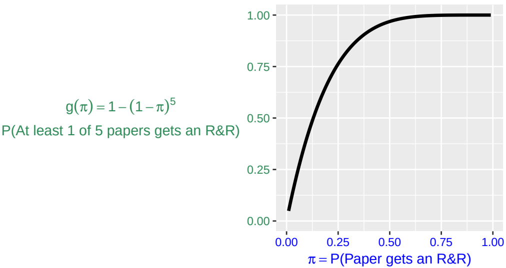
Theorem (King Sec. 4.4, Casella & Berger Thm 7.2.10)

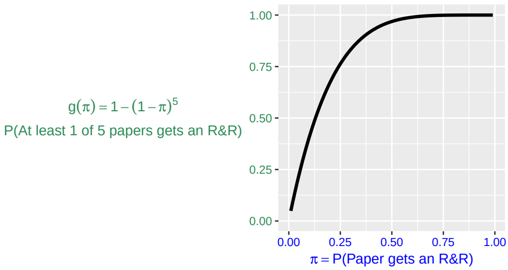
If $\hat{\theta}_{\text{MLE}}$ is the MLE for θ, then for any function $g(\theta)$ the MLE of $g(\theta)$ is $g(\hat{\theta}_{\text{MLE}})$.

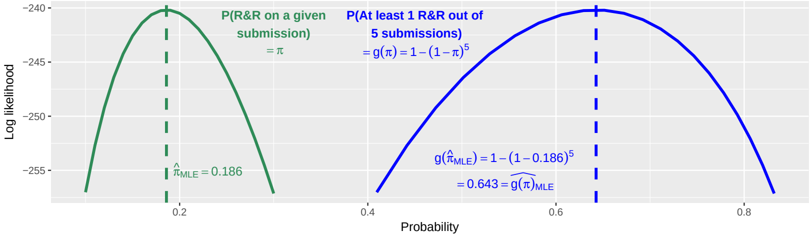
Application: If we knew the true $\pi = \text{P(R\&R)}$, what would be the probability of getting at least 1 R&R out of 5 submissions?

$$\begin{aligned} \text{P(At least 1 R\&R out of 5 submissions)} &= 1 - \text{P(5 rejections)} \\ &= 1 - (1-\pi)^5 = g(\pi) \end{aligned}$$

Question: If we have $\hat{\pi}_{\text{MLE}} = 0.186$, what is $\hat{\tau}_{\text{MLE}} = \widehat{g(\pi)}_{\text{MLE}}$?

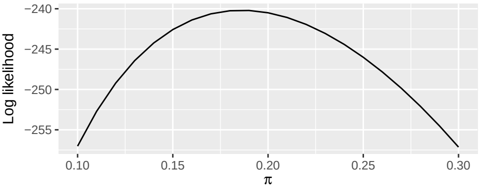

g(π) = 1 − (1 − π)^5
P(At least 1 of 5 papers gets an R&R)
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
π = P(Paper gets an R&R)


$g(\pi) = 1 - (1 - \pi)^5$
P(At least 1 of 5 papers gets an R&R)
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
π = P(Paper gets an R&R)


Log likelihood
−240
−245
−250
−255
P(R&R on a given submission)
$= \pi$
P(At least 1 R&R out of 5 submissions)
$= g(\pi) = 1 - (1 - \pi)^5$
$\hat{\pi}_{MLE} = 0.186$
$g(\hat{\pi}_{MLE}) = 1 - (1 - 0.186)^5$
$= 0.643 = \widehat{g(\pi)}_{MLE}$
0.2
0.4
0.6
0.8
Probability

Uncertainty: Curvature at the maximum

We can estimate uncertainty from the curvature of ℓ around the MLE.



Uncertainty: Curvature at the maximum

The negative of the curvature at the MLE is referred to as the **Fisher information**.

Uncertainty: Curvature at the maximum

The negative of the curvature at the MLE is referred to as the **Fisher information**.

$$\mathcal{I}_n(\theta) = -\mathsf{E}\left(\left. \frac{\partial^2}{\partial \theta^2} \ell(\theta \mid y) \right|_{\theta=\theta_{\text{MLE}}} \right)$$

Uncertainty: Curvature at the maximum

The negative of the curvature at the MLE is referred to as the **Fisher information**.

$$\mathcal{I}_n(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial \theta^2} \ell(\theta \mid y) \bigg|_{\theta = \theta_{\mathrm{MLE}}} \right)$$

The **variance of the MLE** is the inverse of the Fisher information:

$$\mathrm{V}(\hat{\theta}_{\mathrm{MLE}}) = \frac{1}{\mathcal{I}_n(\theta)}$$

Uncertainty: Curvature at the maximum

The negative of the curvature at the MLE is referred to as the **Fisher information**.

$$\mathcal{I}_n(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ell(\theta \mid y)\Big|_{\theta=\theta_{\mathrm{MLE}}} \right)$$

The **variance of the MLE** is the inverse of the Fisher information:

$$\mathrm{V}(\hat{\theta}_{\mathrm{MLE}}) = \frac{1}{\mathcal{I}_n(\theta)}$$

The expectation is taken over the distribution of possible samples y, evaluated at the true θ. We often estimate by using our one sample to calculate the **observed fisher information** at $\hat{\theta}_{\mathrm{MLE}}$.

$$\hat{\mathrm{V}}\left(\hat{\theta}_{\mathrm{MLE}}\right) = \left(\hat{\mathcal{I}}_n(\theta)\right)^{-1}\Big|_{\theta=\hat{\theta}_{\mathrm{MLE}}}$$

Read "evaluated at $\theta = \hat{\theta}_{\mathrm{MLE}}$"

Uncertainty: Curvature at the maximum

What is the uncertainty of our $\hat{\pi}_{\text{MLE}}$? We already calculated:

$$\frac{\partial^2}{\partial \pi^2}\ell(\pi \mid y) = -\frac{\sum_{i=1}^n y_i}{\pi^2} - \frac{5n - \sum_{i=1}^n y_i}{(1-\pi)^2}$$

$$\hat{\pi}_{\text{MLE}} = \frac{\sum_{i=1}^n y_i}{5n} = \frac{\bar{y}}{5}$$

From the prior slide,

$$\hat{\text{V}}\left(\hat{\theta}_{\text{MLE}}\right) = \left(\hat{\mathcal{I}}_n(\theta)\right)^{-1}\Bigg|_{\theta=\hat{\theta}_{\text{MLE}}} = -\left[\frac{\partial^2}{\partial \theta^2}\ell(\theta \mid y)\right]^{-1}\Bigg|_{\theta=\hat{\theta}_{\text{MLE}}}$$

Think, Pair, and Share: What would be the first step to write a formula for $\hat{\text{V}}\left(\hat{\pi}_{\text{MLE}}\right)$?

Uncertainty: Curvature at the maximum

$$\hat{V}\left(\hat{\pi}_{\mathrm{MLE}}\right) = -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\mathrm{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{(1-\hat{\pi}_{\mathrm{MLE}})^2}\right)^{-1}$$

Uncertainty: Curvature at the maximum

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^{n} y_i}{\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1 - \frac{\sum_{i=1}^{n} y_i}{5n}\right)^2}\right)^{-1}$$

Uncertainty: Curvature at the maximum

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^{n} y_i}{\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1 - \frac{\sum_{i=1}^{n} y_i}{5n}\right)^2}\right)^{-1}$$

Useful skill:
Plug a particular estimate into a general MLE formula

Uncertainty: Curvature at the maximum

Useful skill:
Plug a particular estimate into a general MLE formula

$$\begin{aligned}
\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) &= -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1 - \hat{\pi}_{\text{MLE}}\right)^2}\right)^{-1} \\
&= -\left(-\frac{\sum_{i=1}^{n} y_i}{\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1 - \frac{\sum_{i=1}^{n} y_i}{5n}\right)^2}\right)^{-1} \\
&= -\left(-\frac{(5n)^2\left(\sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} y_i\right)^2} - \frac{(5n)^2\left(5n - \sum_{i=1}^{n} y_i\right)}{\left(5n - \sum_{i=1}^{n} y_i\right)^2}\right)^{-1}
\end{aligned}$$

Algebra: Not important

Uncertainty: Curvature at the maximum

Useful skill:
Plug a particular estimate into a general MLE formula

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^{n} y_i}{\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1-\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2}\right)^{-1}$$

$$= -\left(-\frac{(5n)^2\left(\sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} y_i\right)^2} - \frac{(5n)^2\left(5n - \sum_{i=1}^{n} y_i\right)}{\left(5n - \sum_{i=1}^{n} y_i\right)^2}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{1}{\sum_{i=1}^{n} y_i} + \frac{1}{5n - \sum_{i=1}^{n} y_i}\right)^{-1}$$

Algebra: Not important

Uncertainty: Curvature at the maximum

Useful skill: Plug a particular estimate into a general MLE formula

$$\hat{V}(\hat{\pi}_{\text{MLE}}) = -\left(-\frac{\sum_{i=1}^n y_i}{\hat{\pi}^2_{\text{MLE}}} - \frac{5n - \sum_{i=1}^n y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^n y_i}{\left(\frac{\sum_{i=1}^n y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^n y_i}{\left(1-\frac{\sum_{i=1}^n y_i}{5n}\right)^2}\right)^{-1}$$

$$= -\left(-\frac{(5n)^2\left(\sum_{i=1}^n y_i\right)}{\left(\sum_{i=1}^n y_i\right)^2} - \frac{(5n)^2\left(5n - \sum_{i=1}^n y_i\right)}{\left(5n-\sum_{i=1}^n y_i\right)^2}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{1}{\sum_{i=1}^n y_i} + \frac{1}{5n - \sum_{i=1}^n y_i}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{5n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n y_i\right)\left(5n - \sum_{i=1}^n y_i\right)}\right)^{-1}$$

Algebra: Not important

Uncertainty: Curvature at the maximum

Useful skill:
Plug a particular estimate into a general MLE formula

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = -\left(-\frac{\sum_{i=1}^{n} y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^{n} y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^{n} y_i}{\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^{n} y_i}{\left(1-\frac{\sum_{i=1}^{n} y_i}{5n}\right)^2}\right)^{-1}$$

$$= -\left(-\frac{(5n)^2\left(\sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} y_i\right)^2} - \frac{(5n)^2\left(5n - \sum_{i=1}^{n} y_i\right)}{\left(5n - \sum_{i=1}^{n} y_i\right)^2}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{1}{\sum_{i=1}^{n} y_i} + \frac{1}{5n - \sum_{i=1}^{n} y_i}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{5n - \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} y_i}{\left(\sum_{i=1}^{n} y_i\right)\left(5n - \sum_{i=1}^{n} y_i\right)}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{\left(\sum_{i=1}^{n} y_i\right)\left(5n - \sum_{i=1}^{n} y_i\right)}{5n}\right)$$

Algebra: Not important

Uncertainty: Curvature at the maximum

Useful skill:
Plug a particular estimate into a general MLE formula

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = -\left(-\frac{\sum_{i=1}^n y_i}{\hat{\pi}_{\text{MLE}}^2} - \frac{5n - \sum_{i=1}^n y_i}{(1-\hat{\pi}_{\text{MLE}})^2}\right)^{-1}$$

$$= -\left(-\frac{\sum_{i=1}^n y_i}{\left(\frac{\sum_{i=1}^n y_i}{5n}\right)^2} - \frac{5n - \sum_{i=1}^n y_i}{\left(1 - \frac{\sum_{i=1}^n y_i}{5n}\right)^2}\right)^{-1}$$

$$= -\left(-\frac{(5n)^2\left(\sum_{i=1}^n y_i\right)}{\left(\sum_{i=1}^n y_i\right)^2} - \frac{(5n)^2\left(5n - \sum_{i=1}^n y_i\right)}{\left(5n - \sum_{i=1}^n y_i\right)^2}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{1}{\sum_{i=1}^n y_i} + \frac{1}{5n - \sum_{i=1}^n y_i}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{5n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n y_i\right)\left(5n - \sum_{i=1}^n y_i\right)}\right)^{-1}$$

$$= \frac{1}{(5n)^2}\left(\frac{\left(\sum_{i=1}^n y_i\right)\left(5n - \sum_{i=1}^n y_i\right)}{5n}\right)$$

$$= \frac{1}{5n}\left(\frac{\sum_{i=1}^n y_i}{5n}\right)\left(\frac{5n - \sum_{i=1}^n y_i}{5n}\right)$$

Algebra: Not important

Uncertainty result: Building intuition

Question: Does anyone have intuition about our result?

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = \frac{1}{5n}\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)\left(\frac{5n - \sum_{i=1}^{n} y_i}{5n}\right)$$

Uncertainty result: Building intuition

Question: Does anyone have intuition about our result?

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = \frac{1}{5n}\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)\left(\frac{5n-\sum_{i=1}^{n} y_i}{5n}\right)$$

$\frac{1}{\text{Number of trials}}$ Proportion successes   Proportion failures

$$\frac{1}{N}\hat{\pi}(1-\hat{\pi})$$

Uncertainty result: Building intuition

Question: Does anyone have intuition about our result?

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = \frac{1}{5n}\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)\left(\frac{5n-\sum_{i=1}^{n} y_i}{5n}\right)$$

$\frac{1}{\text{Number of trials}}$ Proportion successes Proportion failures

$$\frac{1}{N}\hat{\pi}(1-\hat{\pi})$$

The same as the variance from $5n$ independent Bernoulli trials! As we expected, results are the same as if one student submitted $5n$ times.

Uncertainty result: Building intuition

Question: Does anyone have intuition about our result?

$$\hat{V}\left(\hat{\pi}_{\text{MLE}}\right) = \frac{1}{5n}\left(\frac{\sum_{i=1}^{n} y_i}{5n}\right)\left(\frac{5n-\sum_{i=1}^{n} y_i}{5n}\right)$$

$\frac{1}{\text{Number of trials}}$ Proportion successes Proportion failures

$$\frac{1}{N}\hat{\pi}(1-\hat{\pi})$$

The same as the variance from $5n$ independent Bernoulli trials! As we expected, results are the same as if one student submitted $5n$ times.

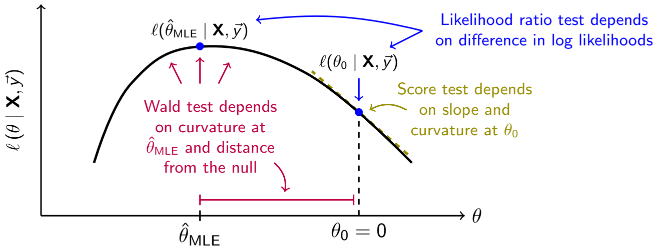
Lesson: A story tying your model to a simpler, known result can help you build confidence in your derivation.

Hypothesis tests[1]

Suppose we want to test the null hypothesis that a subset of the coefficients are zero.

$$\vec{\beta} = \begin{bmatrix} \vec{\beta}_A \\ \vec{\beta}_B \end{bmatrix}, \qquad H_0 : \vec{\beta} = \vec{\beta}_0 \equiv \begin{bmatrix} \vec{0} \\ \vec{\beta}_B \end{bmatrix}$$

There are three main methods for conducting this test.



[1] These tests can be generalized to the null hypothesis $H_0 : h(\vec{\theta}) = \vec{c}$, where h is a function that maps the vector $\vec{\theta} \in \mathbb{R}^p$ to a vector of constraints $c \in \mathbb{R}^k$.

Wald test

The **Wald test** relies on the asymptotic normality of $\hat{\beta}_{\text{MLE}}$:

For $k = 1$

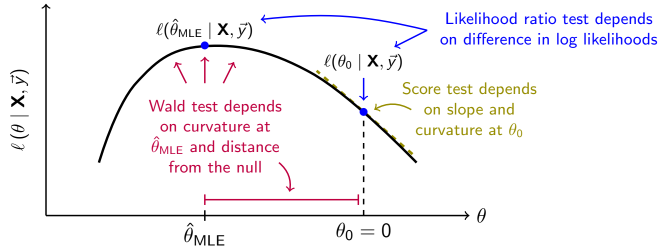
$$\frac{\hat{\beta}_{A,\text{MLE}}}{\widehat{\text{SE}}\left(\hat{\beta}_{A,\text{MLE}}\right)} \xrightarrow{D} N(0,1)$$

Standardized deviation from the null
Squared would be $\sim \chi^2_1$

For any k

$$\left[\hat{\vec{\beta}}_{A,\text{MLE}}\right]^T \left[\hat{\text{V}}\left(\hat{\vec{\beta}}_{A,\text{MLE}}\right)\right]^{-1} \hat{\vec{\beta}}_{A,\text{MLE}} \xrightarrow{D} \chi^2_k$$

Intuition in special case with independent coefficient estimates: Sum of squared standardized deviations from the null



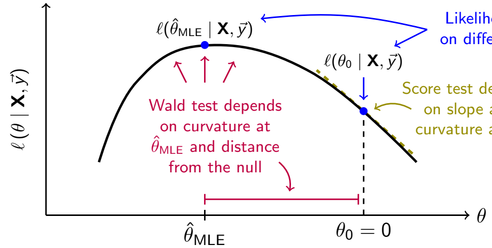
Likelihood ratio test

The **likelihood ratio test** compares nested models:

1. A restricted model estimating $\vec{\beta}_B$ and assuming $\vec{\beta}_A = 0$ with likelihood L_R^*
2. An unrestricted model with likelihood L^*

$$-2\log\left(\frac{L_R^*}{L^*}\right) \xrightarrow{D} \chi_k^2$$

Intuition: We have more evidence in favor of the unrestricted model if the likelihood (L_R^*) under the restricted model is much smaller than the likelihood (L^*) under the unrestricted model. We need more evidence if k is larger.



Score test

The **score test** is based on the score function and the Fisher information evaluated at $\vec{\beta}_0$.

For $k = 1$

$$\left.\frac{\left(\frac{\partial}{\partial \beta_A}\ell(\vec{\beta}|y)\right)^2}{\frac{\partial^2}{\partial {\beta_A}^2}\ell(\vec{\beta}|y)}\right|_{\beta_A=0} \xrightarrow{D} \chi^2_1$$

$\frac{\text{Squared slope}}{\text{Rate of change of slope}}$

Evaluated at the null hypothesis

Intuition: A steep likelihood with minimal curvature is far from the maximum.

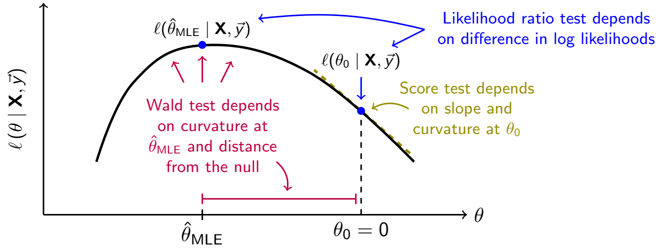
For any k

$$\left.\left[s\left(\tilde{\vec{\beta}}\right)\right]^T\left[\mathcal{I}\left(\tilde{\vec{\beta}}\right)\right]^{-1}\left[s\left(\tilde{\vec{\beta}}\right)\right]\right|_{\tilde{\vec{\beta}}=\vec{\beta}_0} \xrightarrow{D} \chi^2_k$$

Matrix of rate of change of slopes in all directions

Score vector of slopes in all directions

Evaluated at the null hypothesis



Now, we can all practice the whole process on a different distribution: the **Poisson distribution**.

Now, we can all practice the whole process on a different distribution: the **Poisson distribution**.

$$p(y \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Now, we can all practice the whole process on a different distribution: the **Poisson distribution**.

$$p(y \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The Poisson is a discrete distribution for count variables: its support is all nonnegative integers. You can learn more on Wikipedia!

We will use the Poisson to study the **rate of Starbucks stores** in Chicago Census tracts.

Source: Wikipedia

(Motivated by Hwang and Sampson (2014) as a correlate of gentrification.)

There were $\sum y_i = 164$ Starbucks in Chicago as of May 2014 [source]
There were $n = 801$ Census tracts in Chicago in 2010 [source]

Remember the steps for likelihood inference!

1. Assume a data generating process.
2. Derive the likelihood.
3. Maximize the likelihood to get the MLE.
4. Derive standard errors from the inverse of the Fisher information

What is the likelihood for n observations?

What is the likelihood for n observations?

$$L(\lambda \mid y_1, \ldots, y_n) =$$

What is the likelihood for n observations?

$$L(\lambda \mid y_1, \ldots, y_n) = p(y_1, \ldots, y_n \mid \lambda)$$

Assumes independence

$$=$$

What is the likelihood for n observations?

Assumes independence

$$L(\lambda \mid y_1, \ldots, y_n) = p(y_1, \ldots, y_n \mid \lambda)$$
$$= \prod_{i=1}^{n} p(y_i \mid \lambda)$$
$$=$$

What is the likelihood for n observations?

Assumes independence

$$
\begin{aligned}
L(\lambda \mid y_1, \ldots, y_n) &= p(y_1, \ldots, y_n \mid \lambda) \\
&= \prod_{i=1}^{n} p(y_i \mid \lambda) \\
&= \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{(y_i)!}
\end{aligned}
$$

What is the log likelihood?

What is the log likelihood?

$$\ell(\lambda \mid y_1, \ldots, y_n) =$$

What is the log likelihood?

$$\ell(\lambda \mid y_1, \ldots, y_n) = \log L(\lambda \mid y_1, \ldots, y_n)$$

$$=$$

What is the log likelihood?

$$
\begin{aligned}
\ell(\lambda \mid y_1, \ldots, y_n) &= \log L(\lambda \mid y_1, \ldots, y_n) \\
&= \sum_{i=1}^{n} \Big(y_i \log \lambda - \lambda - \log(y_i!) \Big) \\
&=
\end{aligned}
$$

What is the log likelihood?

$$\begin{aligned}
\ell(\lambda \mid y_1, \ldots, y_n) &= \log L(\lambda \mid y_1, \ldots, y_n) \\
&= \sum_{i=1}^{n} \Big(y_i \log \lambda - \lambda - \log(y_i!) \Big) \\
&= \log \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \log(y_i!) \\
&=
\end{aligned}$$

What is the log likelihood?

$$
\begin{aligned}
\ell(\lambda \mid y_1, \ldots, y_n) &= \log L(\lambda \mid y_1, \ldots, y_n) \\
&= \sum_{i=1}^{n} \Big(y_i \log \lambda - \lambda - \log(y_i!) \Big) \\
&= \log \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \log(y_i!) \\
&= \underbrace{\log \lambda \sum_{i=1}^{n} y_i - n\lambda}_{\text{Involves } \lambda} - \underbrace{\sum_{i=1}^{n} \log(y_i!)}_{\text{Does not involve } \lambda} \\
&=
\end{aligned}
$$

What is the log likelihood?

$$
\begin{aligned}
\ell(\lambda \mid y_1, \ldots, y_n) &= \log L(\lambda \mid y_1, \ldots, y_n) \\
&= \sum_{i=1}^{n} \Big(y_i \log \lambda - \lambda - \log(y_i!) \Big) \\
&= \log \lambda \sum_{i=1}^{n} y_i - n\lambda - \sum_{i=1}^{n} \log(y_i!) \\
&= \underbrace{\log \lambda \sum_{i=1}^{n} y_i - n\lambda}_{\text{Involves } \lambda} - \underbrace{\sum_{i=1}^{n} \log(y_i!)}_{\text{Does not involve } \lambda} \\
&= \log \lambda \sum_{i=1}^{n} y_i - n\lambda + \text{constant}
\end{aligned}
$$

What is the derivative of the log likelihood?

What is the derivative of the log likelihood?

$$\frac{\partial}{\partial\lambda}\ell(\lambda \mid y_1, \ldots, y_n) = \frac{\partial}{\partial\lambda}\log\lambda\sum_{i=1}^{n} y_i - n\lambda$$

What is the derivative of the log likelihood?

$$\begin{aligned}\frac{\partial}{\partial \lambda}\ell(\lambda \mid y_1, \ldots, y_n) &= \frac{\partial}{\partial \lambda} \log \lambda \sum_{i=1}^{n} y_i - n\lambda \\ &= \frac{1}{\lambda}\sum_{i=1}^{n} y_i - n\end{aligned}$$

What is the derivative of the log likelihood?

$$\frac{\partial}{\partial\lambda}\ell(\lambda \mid y_1, \ldots, y_n) = \frac{\partial}{\partial\lambda}\log\lambda\sum_{i=1}^{n} y_i - n\lambda$$
$$= \frac{1}{\lambda}\sum_{i=1}^{n} y_i - n$$

Set the derivative equal to 0 to find the critical value.

$$0 = \frac{1}{\lambda^*}\sum_{i=1}^{n} y_i - n$$

What is the derivative of the log likelihood?

$$\frac{\partial}{\partial \lambda} \ell(\lambda \mid y_1, \ldots, y_n) = \frac{\partial}{\partial \lambda} \log \lambda \sum_{i=1}^{n} y_i - n\lambda$$
$$= \frac{1}{\lambda} \sum_{i=1}^{n} y_i - n$$

Set the derivative equal to 0 to find the critical value.

$$0 = \frac{1}{\lambda^*} \sum_{i=1}^{n} y_i - n$$
$$\lambda^* = \frac{\sum_{i=1}^{n} y_i}{n}$$

What is the derivative of the log likelihood?

$$\frac{\partial}{\partial\lambda}\ell(\lambda \mid y_1, \ldots, y_n) = \frac{\partial}{\partial\lambda} \log\lambda \sum_{i=1}^{n} y_i - n\lambda$$

$$= \frac{1}{\lambda}\sum_{i=1}^{n} y_i - n$$

Set the derivative equal to 0 to find the critical value.

$$0 = \frac{1}{\lambda^*}\sum_{i=1}^{n} y_i - n$$

$$\lambda^* = \frac{\sum_{i=1}^{n} y_i}{n}$$

Check second derivative is negative to verify this is a maximum.

What is the derivative of the log likelihood?

$$\frac{\partial}{\partial\lambda}\ell(\lambda \mid y_1, \ldots, y_n) = \frac{\partial}{\partial\lambda}\log\lambda\sum_{i=1}^{n} y_i - n\lambda$$
$$= \frac{1}{\lambda}\sum_{i=1}^{n} y_i - n$$

Set the derivative equal to 0 to find the critical value.

$$0 = \frac{1}{\lambda^*}\sum_{i=1}^{n} y_i - n$$
$$\lambda^* = \frac{\sum_{i=1}^{n} y_i}{n}$$

Check second derivative is negative to verify this is a maximum.

$$\frac{\partial^2}{\partial\lambda^2}\ell(\lambda \mid y_1, \ldots, y_n) = -\frac{\sum_{i=1}^{n} y_i}{\lambda^2} < 0$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$V(\hat{\lambda}_{\text{MLE}}) = \left(\mathcal{I}_n(\lambda)\right)^{-1}$$

$$=$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned} V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\ &= \left(-E\left(\frac{\partial^2}{\partial \lambda^2} \ell(\lambda \mid y)\right)\right)^{-1} \\ &= \end{aligned}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned}
V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\
&= \left(-E\left(\frac{\partial^2}{\partial\lambda^2}\ell(\lambda \mid y)\right)\right)^{-1} \\
&= \left(-E\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right)\right)^{-1} \\
&=
\end{aligned}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned}
V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\
&= \left(-E\left(\frac{\partial^2}{\partial \lambda^2}\ell(\lambda \mid y)\right)\right)^{-1} \\
&= \left(-E\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} E(y_i)}{\lambda^2}\right)\right)^{-1} \\
&=
\end{aligned}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned}
V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\
&= \left(-E\left(\frac{\partial^2}{\partial\lambda^2}\ell(\lambda \mid y)\right)\right)^{-1} \\
&= \left(-E\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} E(y_i)}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} \lambda}{\lambda^2}\right)\right)^{-1} \\
&=
\end{aligned}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned}
V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\
&= \left(-E\left(\frac{\partial^2}{\partial\lambda^2}\ell(\lambda \mid y)\right)\right)^{-1} \\
&= \left(-E\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} E(y_i)}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} \lambda}{\lambda^2}\right)\right)^{-1} \\
&= \left(\frac{n\lambda}{\lambda^2}\right)^{-1} \\
&=
\end{aligned}$$

This is a maximum!

$$\hat{\lambda}_{\text{MLE}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Let's find the variance

$$\begin{aligned}
V(\hat{\lambda}_{\text{MLE}}) &= \left(\mathcal{I}_n(\lambda)\right)^{-1} \\
&= \left(-E\left(\frac{\partial^2}{\partial\lambda^2}\ell(\lambda \mid y)\right)\right)^{-1} \\
&= \left(-E\left(-\frac{\sum_{i=1}^{n} y_i}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} E(y_i)}{\lambda^2}\right)\right)^{-1} \\
&= \left(-\left(-\frac{\sum_{i=1}^{n} \lambda}{\lambda^2}\right)\right)^{-1} \\
&= \left(\frac{n\lambda}{\lambda^2}\right)^{-1} \\
&= \frac{\lambda}{n}
\end{aligned}$$

,

This week is a lot of math. We appreciate the work you're putting in.

This week is a lot of math. We appreciate the work you're putting in.

We'll use this again and again and it will enable you to invent your own models in the future to fit your data generating processes!

This week is a lot of math. We appreciate the work you're putting in.

We'll use this again and again and it will enable you to invent your own models in the future to fit your data generating processes!

Keep it up! Also give us feedback on the cards.

Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

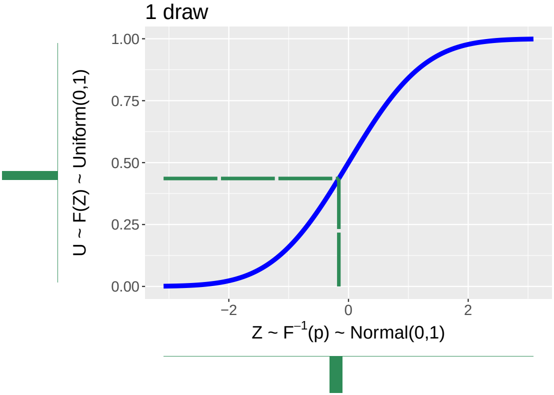
- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

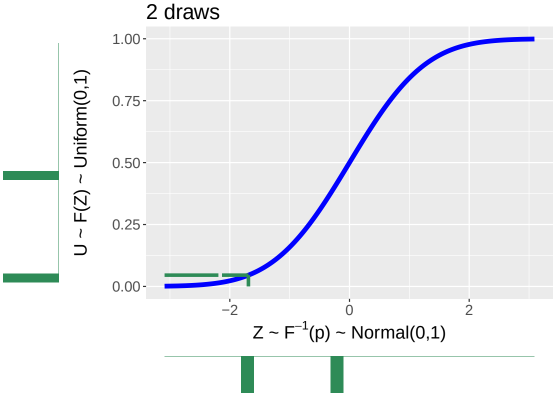


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

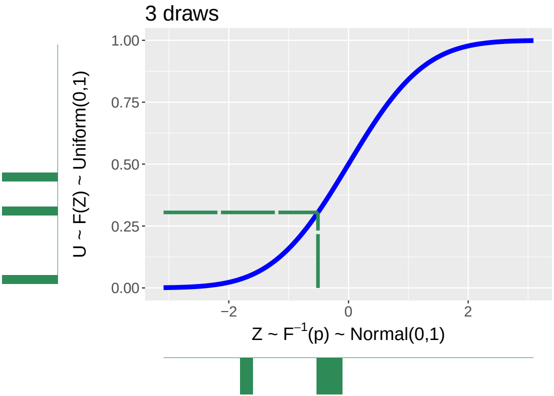


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

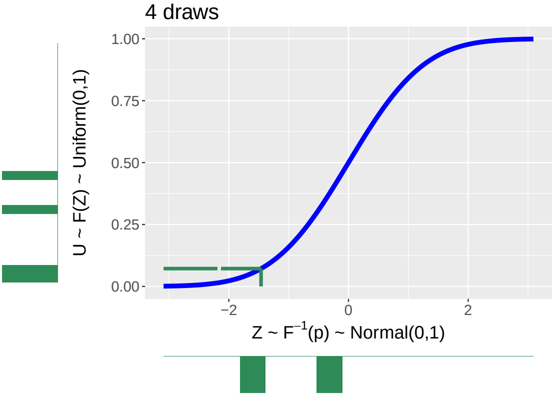


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

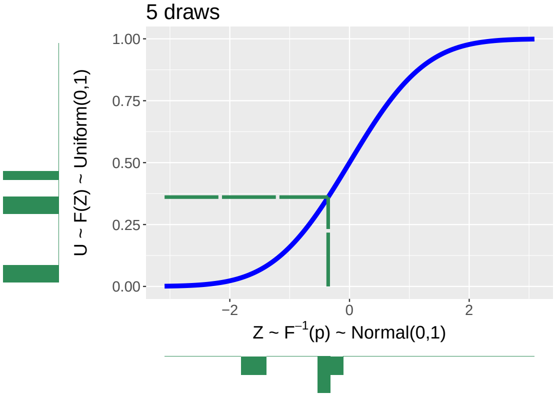


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

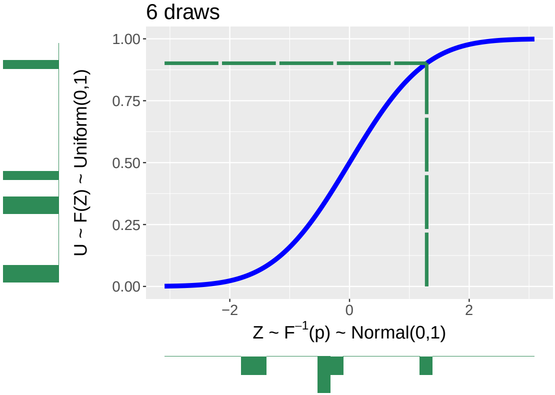


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

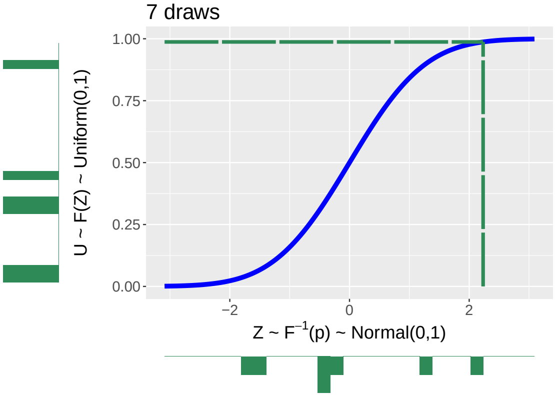


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

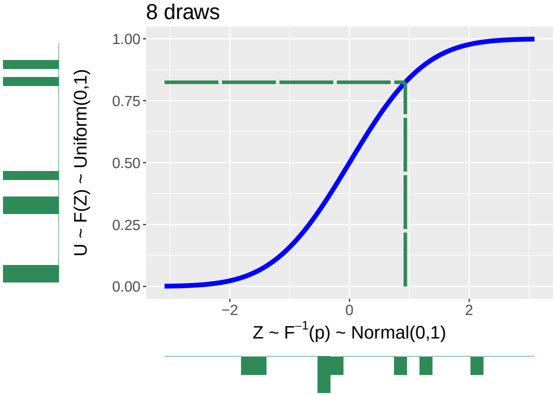


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

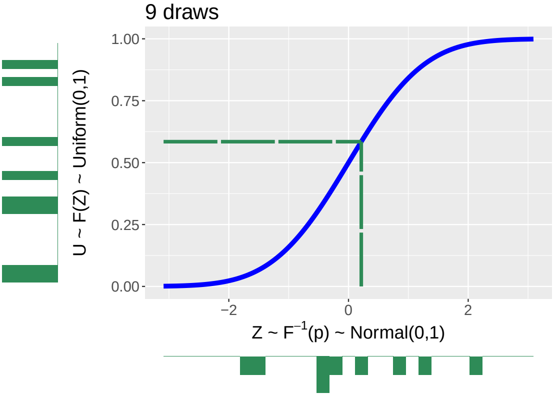


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

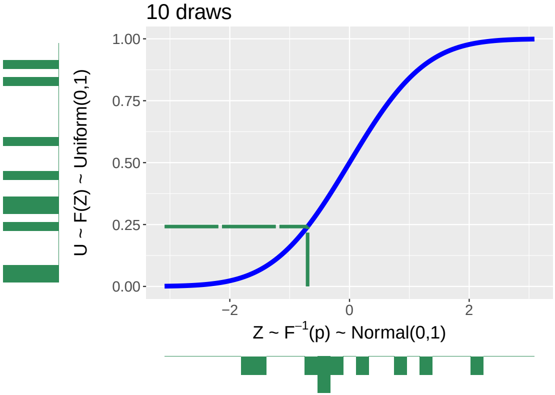


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

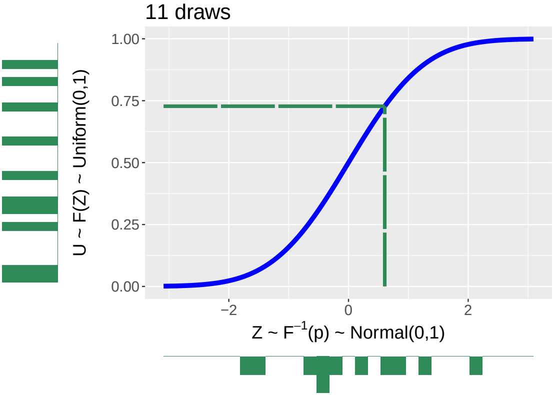


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

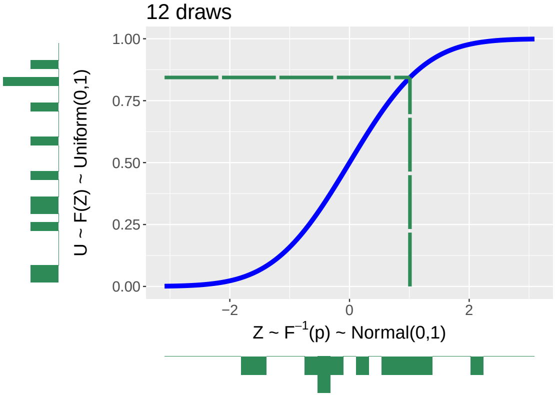


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

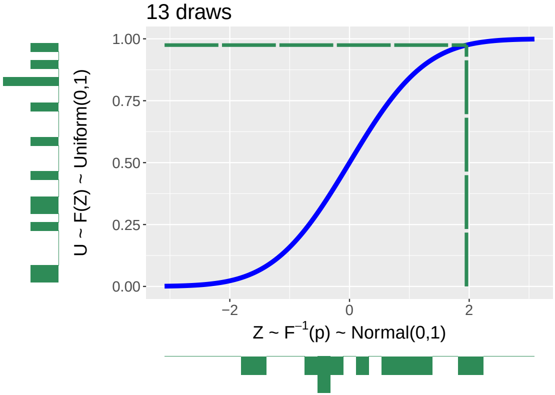


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

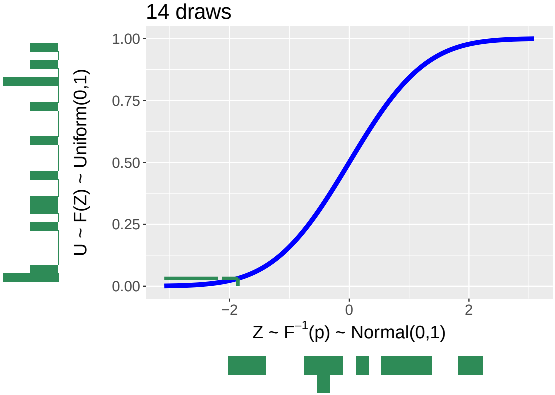


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

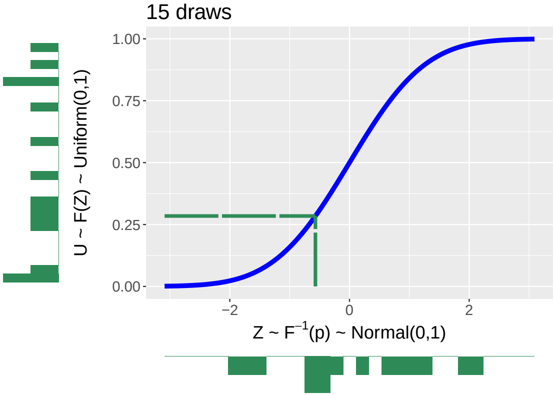


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

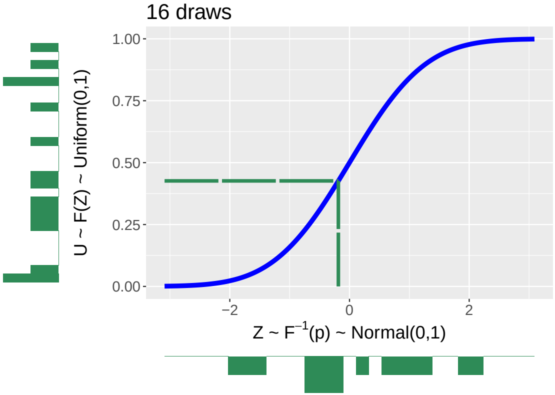


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

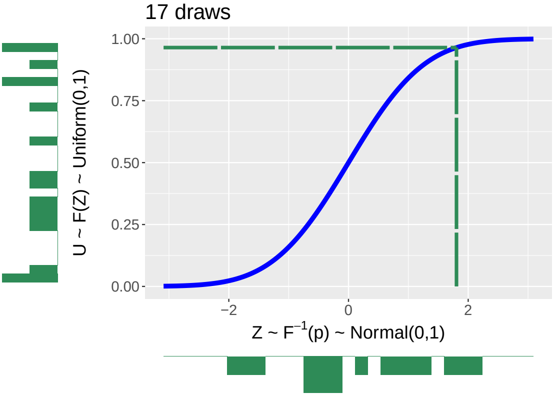


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

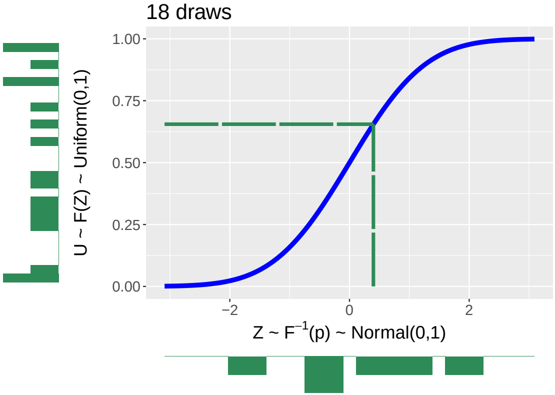


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

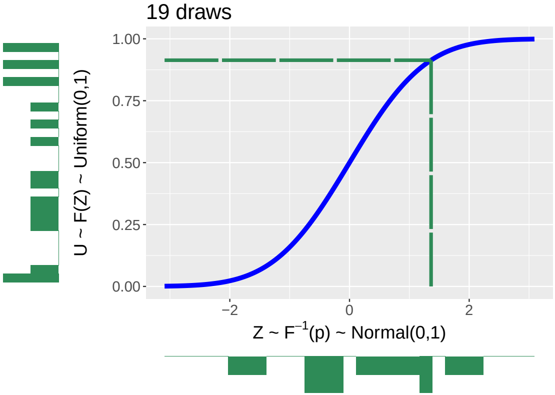


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

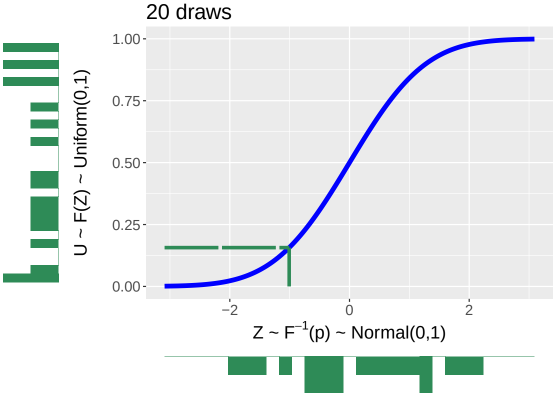
- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$



Universality of the Uniform
aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

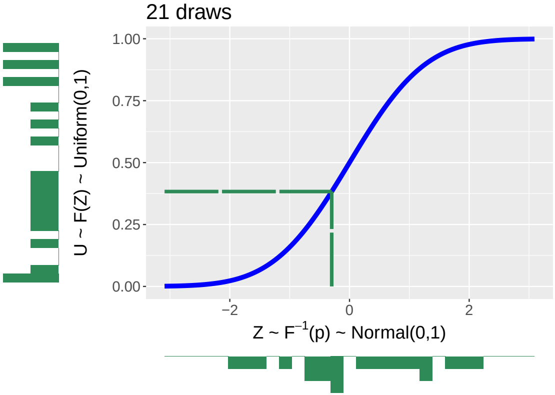


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

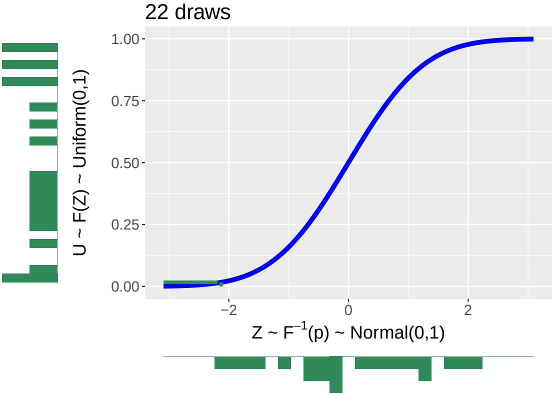


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

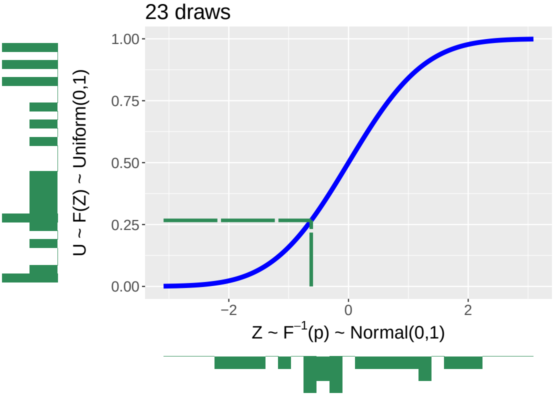


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

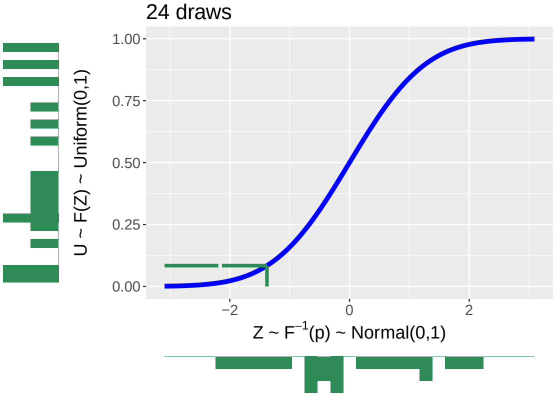


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

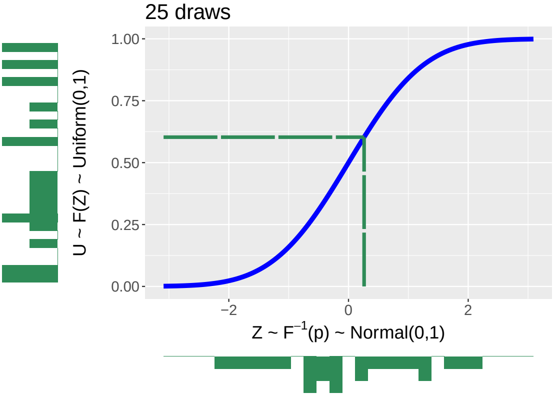


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

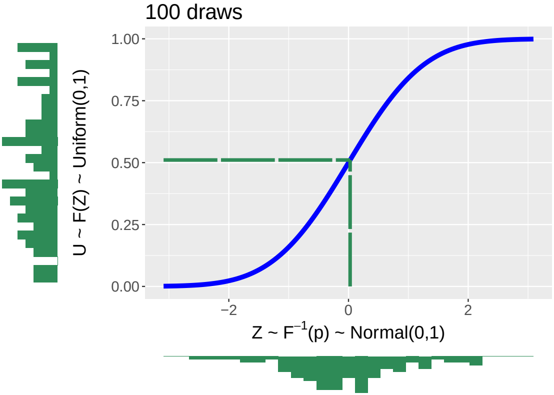


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

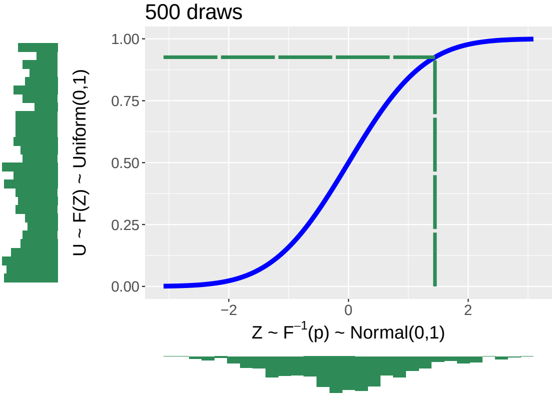


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$

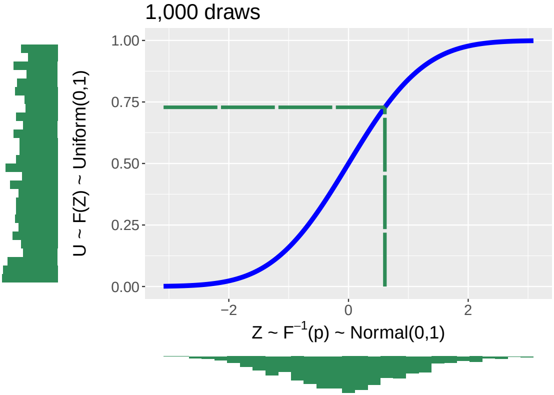


Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0,1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$



Universality of the Uniform

aka Probability Integral Transform, PIT

Theorem

- Regardless of the distribution of X, $F(X) \sim \text{Uniform}(0, 1)$
- For a r.v. X with CDF F and a Uniform r.v. U, $F^{-1}(U) \sim X$