# Precept 2: Likelihood inference

## Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

February 16, 2017

# Outline

# Replication Paper

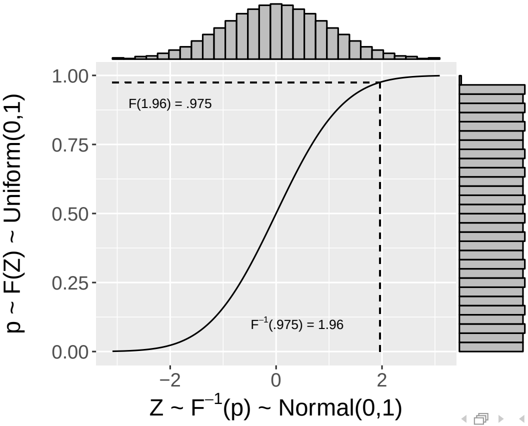
Any thoughts or issues to discuss?

# Universality of the Uniform

aka Probability Integral Transform, PIT

## Theorem

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



# Integral of PDF

All PDFs integrate to 1. Why?

# Integral of PDF

All PDFs integrate to 1. Why?

Because the probability of observing a value somewhere in the support of the random variable is 1!

# Expected value as an integral

$$E(X) =$$

# Expected value as an integral

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

# Expected value as an integral

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Suppose

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{for } y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

This is called the **beta distribution**.

# Expected value as an integral

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Suppose

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{for } y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

This is called the **beta distribution**.

Can we write a formula for $E(Y)$?

$$E(Y) =$$

# Expected value as an integral

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Suppose

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{for } y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

This is called the **beta distribution**.

Can we write a formula for $E(Y)$?

$$E(Y) = \int_0^1 y f_Y(y) dy =$$

# Expected value as an integral

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Suppose

$$f_Y(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} & \text{for } y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

This is called the **beta distribution**.

Can we write a formula for $E(Y)$?

$$E(Y) = \int_0^1 y f_Y(y) dy = \int_0^1 y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} dy$$

# Calculating $E(Y)$ analytically

NOTE: We will not ask you to do this. It's cool though!

$$E(Y) = \int_0^1 y f_Y(y) dy$$

# Calculating $E(Y)$ analytically

NOTE: We will not ask you to do this. It's cool though!

$$
\begin{aligned}
E(Y) &= \int_0^1 y f_Y(y) dy \\
&= \int_0^1 y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} dy
\end{aligned}
$$

# Calculating $E(Y)$ analytically

NOTE: We will not ask you to do this. It's cool though!

$$
\begin{aligned}
E(Y) &= \int_0^1 y f_Y(y) dy \\
&= \int_0^1 y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} dy \\
&= \underbrace{\left(\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}\right)\left(\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\right)}_{\text{Multiplying by 1}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} dy
\end{aligned}
$$

# Calculating $E(Y)$ analytically

NOTE: We will not ask you to do this. It's cool though!

$$
\begin{aligned}
E(Y) &= \int_0^1 y f_Y(y) dy \\
&= \int_0^1 y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} dy \\
&= \underbrace{\left(\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}\right)\left(\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\right)}_{\text{Multiplying by 1}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} dy \\
&= \left(\frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)}\right) \underbrace{\int_0^1 \left(\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}\right) y^{(\alpha+1)-1}(1-y)^{\beta-1} dy}_{\text{Beta pdf integrates to 1}}
\end{aligned}
$$

# Calculating $E(Y)$ analytically

NOTE: We will not ask you to do this. It's cool though!

$$
\begin{aligned}
E(Y) &= \int_0^1 y f_Y(y) dy \\
&= \int_0^1 y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} dy \\
&= \underbrace{\left(\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}\right)\left(\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\right)}_{\text{Multiplying by 1}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{(\alpha+1)-1}(1-y)^{\beta-1} dy \\
&= \left(\frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)}\right) \underbrace{\int_0^1 \left(\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}\right) y^{(\alpha+1)-1}(1-y)^{\beta-1} dy}_{\text{Beta pdf integrates to 1}} \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

# Numeric integration

On the homework, we'll have you do integrals with the `integrate` function in R.

# Numeric integration

On the homework, we'll have you do integrals with the `integrate` function in R.

# Numeric integration

On the homework, we'll have you do integrals with the `integrate` function in R.

```
beta.pdf <- function(alpha,beta,x) {
  gamma(alpha + beta) / (gamma(alpha) * gamma(beta)) *
    x^(alpha - 1) * (1 - x)^(beta - 1)
}
```

# Numeric integration

On the homework, we'll have you do integrals with the `integrate` function in R.

```
beta.pdf <- function(alpha,beta,x) {
  gamma(alpha + beta) / (gamma(alpha) * gamma(beta)) *
    x^(alpha - 1) * (1 - x)^(beta - 1)
}

integrate(f = function(x) x * beta.pdf(1,2,x),
          lower = 0, upper = 1)
```

# Numeric integration

On the homework, we'll have you do integrals with the `integrate` function in R.

```
beta.pdf <- function(alpha,beta,x) {
  gamma(alpha + beta) / (gamma(alpha) * gamma(beta)) *
    x^(alpha - 1) * (1 - x)^(beta - 1)
}

integrate(f = function(x) x * beta.pdf(1,2,x),
          lower = 0, upper = 1)

0.3333333 with absolute error < 3.7e-15
```

# Variance as an integral

$$V(X) =$$

# Variance as an integral

$$V(X) = E(X - E[X])^2$$
$$=$$

# Variance as an integral

$$\begin{aligned} V(X) &= E(X - E[X])^2 \\ &= E(X^2) - (E[X])^2 \\ &= \end{aligned}$$

# Variance as an integral

$$
\begin{aligned}
V(X) &= E(X - E[X])^2 \\
&= E(X^2) - (E[X])^2 \\
&= \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left( \int_{-\infty}^{\infty} x f_X(x) dx \right)^2
\end{aligned}
$$

# 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

Suppose we would like to know the probability that a Princeton Ph.D. student in sociology who submits a paper to a major journal is offered the chance to revise and resubmit the paper. We have data on several students who each submit 5 papers over the course of the program. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

# Binomial example

Suppose we would like to know the probability that a Princeton Ph.D. student in sociology who submits a paper to a major journal is offered the chance to revise and resubmit the paper. We have data on several students who each submit 5 papers over the course of the program. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Can we translate this into a data generating process?

# Binomial example

Suppose we would like to know the probability that a Princeton Ph.D. student in sociology who submits a paper to a major journal is offered the chance to revise and resubmit the paper. We have data on several students who each submit 5 papers over the course of the program. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Can we translate this into a data generating process?

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

# Binomial example

Suppose we would like to know the probability that a Princeton Ph.D. student in sociology who submits a paper to a major journal is offered the chance to revise and resubmit the paper. We have data on several students who each submit 5 papers over the course of the program. For each student, we observe the number of these papers that receive a revise and resubmit on the first submission.

Can we translate this into a data generating process?

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

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

We **assume** that the response is binomial, with each paper independent, and with all submissions from all students having the same probability $p$ of success.

We **assume** that the response is binomial, with each paper independent, and with all submissions from all students having the same probability $p$ of success.

What does the model help us to learn?

We **assume** that the response is binomial, with each paper independent, and with all submissions from all students having the same probability $p$ of success.

What does the model help us to learn?

The model helps us to learn the value of the parameter $\hat{p}$ that makes the observed data the most likely.

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

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

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

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

What is the systematic component?
What is the stochastic component

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

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

What is the systematic component?
What is the stochastic component

**Systematic component:** The probability $p$

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

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

What is the systematic component?
What is the stochastic component

**Systematic component:** The probability $p$
**Stochastic component:** The outcome $Y_i$

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

$$p(y_i \mid p) = p^{y_i}(1-p)^{5-y_i}$$

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

$$p(y_i \mid p) = p^{y_i}(1-p)^{5-y_i}$$

What is the probability of all $n$ observations given $p$?

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

$$p(y_i \mid p) = p^{y_i}(1-p)^{5-y_i}$$

What is the probability of all $n$ observations given $p$?

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

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

$$p(y_i \mid p) = p^{y_i}(1-p)^{5-y_i}$$

What is the probability of all $n$ observations given $p$?

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

$$=$$

# $P(y \mid p)$

What is the probability of one $y_i$ given $p$?

$$p(y_i \mid p) = p^{y_i}(1-p)^{5-y_i}$$

What is the probability of all $n$ observations given $p$?

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

# $L(p \mid y_1, \ldots, y_n)$

What is the likelihood?

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

# $L(p \mid y_1, \ldots, y_n)$

What is the likelihood?

$$L(p \mid y_1, \ldots, y_n) \propto p(y_1, \ldots, y_n \mid p)$$

$$=$$

# $L(p \mid y_1, \ldots, y_n)$

What is the likelihood?

$$L(p \mid y_1, \ldots, y_n) \propto p(y_1, \ldots, y_n \mid p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{5-y_i}$$

# Review of log rules

# Review of log rules

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

# Review of log rules

$$
\begin{aligned}
\log(ab) &= \log(a) + \log(b) \\
\log(e^a) &= a
\end{aligned}
$$

# $\ell(p \mid y_1, \ldots, y_n)$

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

# $\ell(p \mid y_1, \ldots, y_n)$

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

$$\doteq$$

# $\ell(p \mid y_1, \ldots, y_n)$

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

# $\ell(p \mid y_1, \ldots, y_n)$

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

# $\ell(p \mid y_1, \ldots, y_n)$

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

# $\ell(p \mid y_1, \ldots, y_n)$

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

# $\ell(p \mid y_1, \ldots, y_n)$ (continued)

(we can go further)

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

$$=$$

# $\ell(p \mid y_1, \dots, y_n)$ (continued)

(we can go further)

$$\begin{aligned}
\ell(p \mid y_1, \dots, y_n) &= \sum_{i=1}^{n} \Big( y_i \log p + (5 - y_i) \log(1-p) \Big) \\
&= \log p \sum_{i=1}^{n} y_i + 5n \log(1-p) - \log(1-p) \sum_{i=1}^{n} y_i \\
&=
\end{aligned}$$

# $\ell(p \mid y_1, \ldots, y_n)$ (continued)

(we can go further)

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

So...the data $y_1, \ldots, y_n$ only enter the likelihood through their sum.

# $\ell(p \mid y_1, \ldots, y_n)$ (continued)

(we can go further)

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

So...the data $y_1, \ldots, y_n$ only enter the likelihood through their sum. We call $\sum_{i=1}^{n} y_i$ a **sufficient statistic** since it's all you need to compute the likelihood.

# Sufficient statistic discussion

Does it seem reasonable that we could compute the likelihood only knowing the number of graduate student submissions that are given R&Rs?

# Sufficient statistic discussion

Does it seem reasonable that we could compute the likelihood only knowing the number of graduate student submissions that are given R&Rs?

This makes sense because we assumed every submission had the same probability of success, so it's like we had 5n Bernoulli trials. There is no need to distinguish who submitted them!

# Sufficient statistic discussion

Does it seem reasonable that we could compute the likelihood only knowing the number of graduate student submissions that are given R&Rs?

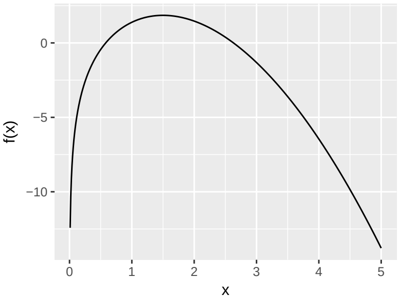
This makes sense because we assumed every submission had the same probability of success, so it's like we had 5n Bernoulli trials. There is no need to distinguish who submitted them!

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

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

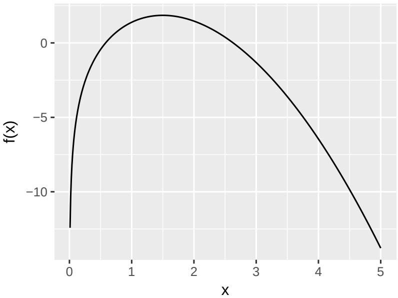



What is the derivative?

# Calculus review: Derivatives

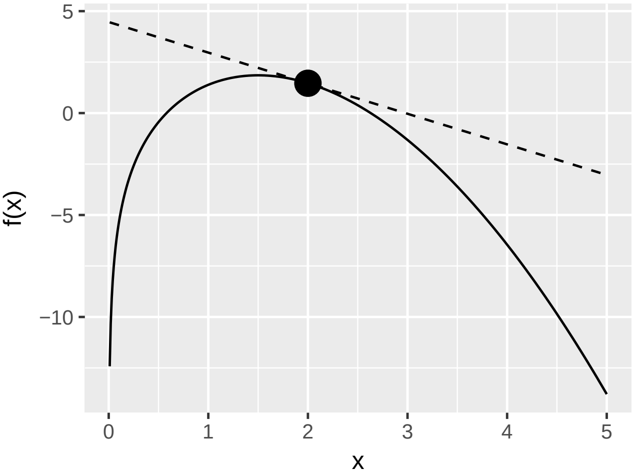
Suppose we have a function

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




What is the derivative? **It is just 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

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

The derivative is

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

# Calculus review: Derivatives

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

The derivative is

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

$$=$$

# Calculus review: Derivatives

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

The derivative is

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

# Calculus review: Derivatives

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

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

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

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 $x = 2$

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

# 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?

# 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

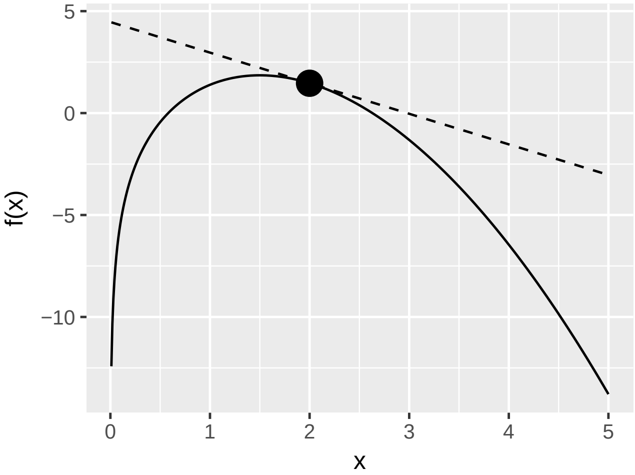
$$f(x) = x - x^2 + \log(x) + \log(3x^2)$$
$$f'(x) = 1 - 2x + \frac{3}{x}$$

How do we maximize this?

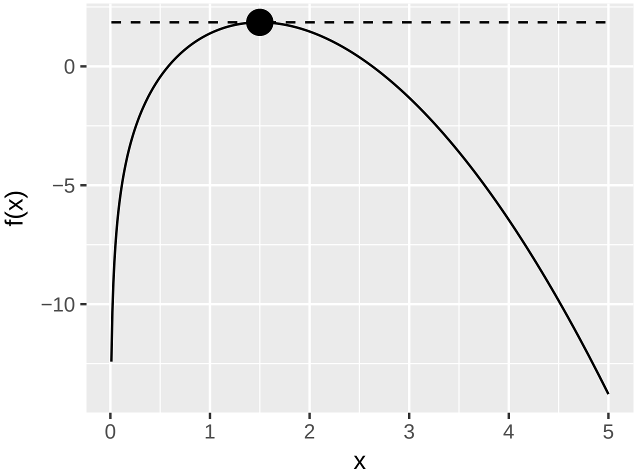
Set the derivative equal to 0 and solve!

(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

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

# Calculus review: Maximizing a function

Set the derivative equal to 0

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

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.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2} \\
&= -2 - \frac{3}{x^2} \\
f''(-1) &= -2 - \frac{3}{(-1)^2} = -5 \rightarrow
\end{aligned}
$$

# Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2} \\
&= -2 - \frac{3}{x^2} \\
f''(-1) &= -2 - \frac{3}{(-1)^2} = -5 \rightarrow \text{Maximum}
\end{aligned}
$$

# Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2} \\
&= -2 - \frac{3}{x^2} \\
f''(-1) &= -2 - \frac{3}{(-1)^2} = -5 \rightarrow \text{Maximum} \\
f''(1.5) &= -2 - \frac{3}{1.5^2} = -3.333 \rightarrow
\end{aligned}
$$

# Calculus review: Second derivative

The second derivative captures the curvature of the function.

$$
\begin{aligned}
f''(x) &= \frac{\partial}{\partial x} 1 - 2x + \frac{3}{x} \\
&= \frac{\partial}{\partial x} 1 - 2x + 3x^{-1} \\
&= -2 - 3x^{-2} \\
&= -2 - \frac{3}{x^2} \\
f''(-1) &= -2 - \frac{3}{(-1)^2} = -5 \rightarrow \text{Maximum} \\
f''(1.5) &= -2 - \frac{3}{1.5^2} = -3.333 \rightarrow \text{Maximum}
\end{aligned}
$$

# Back to our example: Maximizing the log likelihood

We had

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

How do we maximize this?

# Back to our example: Maximizing the log likelihood

We had

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

How do we maximize this? Take derivative and set equal to 0!

# Back to our example: Maximizing the log likelihood

We had

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

How do we maximize this? Take derivative and set equal to 0!

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

$$=$$

# Back to our example: Maximizing the log likelihood

We had

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

How do we maximize this? Take derivative and set equal to 0!

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

# Back to our example: Maximizing the log likelihood

We had

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

How do we maximize this? Take derivative and set equal to 0!

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

# Back to our example: Maximizing the log likelihood

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

# Back to our example: Maximizing the log likelihood

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

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

# Back to our example: Maximizing the log likelihood

$$
\begin{aligned}
0 &= \frac{1}{p^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-p^*} \\
\frac{5n - \sum_{i=1}^{n} y_i}{1-p^*} &= \frac{1}{p^*}\sum_{i=1}^{n} y_i \\
(5n - \sum_{i=1}^{n} y_i)p^* &= (1-p^*)\sum_{i=1}^{n} y_i
\end{aligned}
$$

# Back to our example: Maximizing the log likelihood

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

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

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

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

# Back to our example: Maximizing the log likelihood

$$
\begin{aligned}
0 &= \frac{1}{p^*}\sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} y_i - 5n}{1-p^*} \\
\frac{5n - \sum_{i=1}^{n} y_i}{1-p^*} &= \frac{1}{p^*}\sum_{i=1}^{n} y_i \\
(5n - \sum_{i=1}^{n} y_i)p^* &= (1-p^*)\sum_{i=1}^{n} y_i \\
5np^* &= \sum_{i=1}^{n} y_i \\
p^* &= \frac{\sum_{i=1}^{n} y_i}{5n}
\end{aligned}
$$

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

# Back to our example: Maximizing the log likelihood

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

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

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

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

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

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

# Back to our example: Maximizing the log likelihood

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

# Back to our example: Maximizing the log likelihood

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

# Back to our example: Maximizing the log likelihood

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

# Back to our example: Maximizing the log likelihood

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

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

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

# Plotting the log likelihood

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

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

# Plotting the log likelihood

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

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

```
log.lik <- function(p) {
```

# Plotting the log likelihood

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

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

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

# 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(p = 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(p = seq(0.1,0.3,0.01)) %>%
  mutate(`Log likelihood` = log.lik(p)) %>%
```

# Plotting the log likelihood

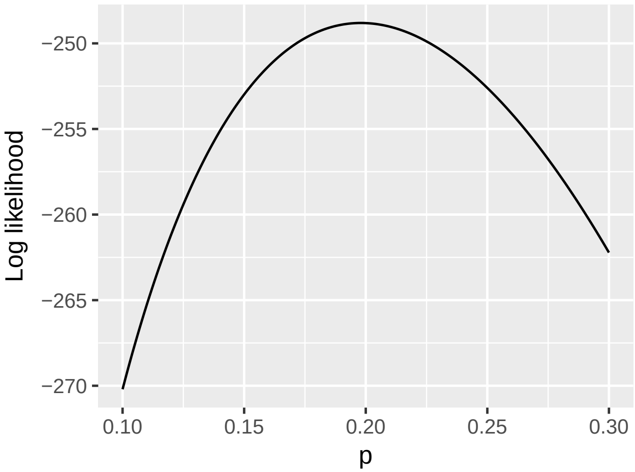
Define some data and make a plot.

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

# Plotting the log likelihood

Define some data and make a plot.

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


−250
−255
−260
−265
−270
Log likelihood
0.10
0.15
0.20
0.25
0.30
p

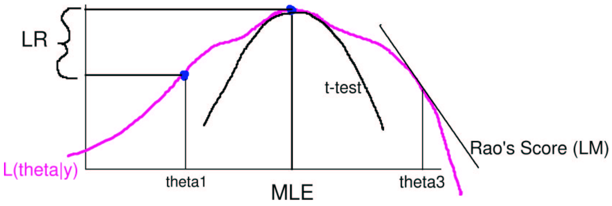
# Finding the maximum numerically

# Finding the maximum numerically

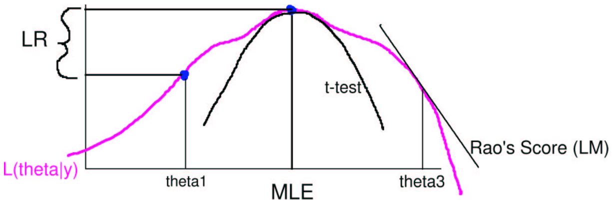
```
> y <- rbinom(100,5,.2)
> optimize(f = log.lik,
+          interval = c(0,1),
+          maximum = T)
$maximum
[1] 0.2020157

$objective
[1] -251.5813
```

The next 3 slides are exactly copied from lecture so we can discuss uncertainty.

# Uncertainty: Likelihood Ratios for nested models

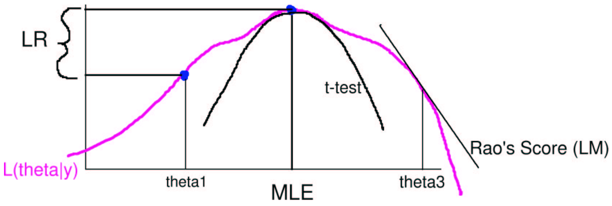


# Uncertainty: Likelihood Ratios for nested models



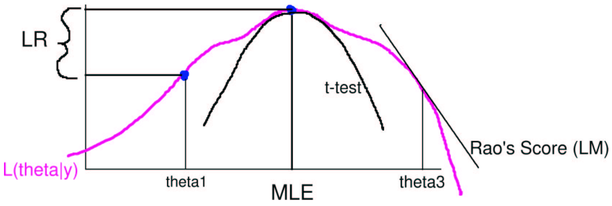

- $L^*$ is the likelihood value for the **unrestricted** model

# Uncertainty: Likelihood Ratios for nested models




- $L^*$ is the likelihood value for the **unrestricted** model
- $L^*_R$ is the likelihood value for the (nested) **restricted** model

# Uncertainty: Likelihood Ratios for nested models




- $L^*$ is the likelihood value for the **unrestricted** model
- $L_R^*$ is the likelihood value for the (nested) **restricted** model
- $\implies L^* \geq L_R^* \implies \frac{L_R^*}{L^*} \leq 1$

# Meaning of the likelihood ratio

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$L(\theta_1|y) \propto k(y)\P(y|\theta_1)$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$L(\theta_1|y) \propto k(y)\P(y|\theta_1)$$
$$L(\theta_2|y) \propto k(y)\P(y|\theta_2)$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$L(\theta_1|y) \propto k(y)\P(y|\theta_1)$$
$$L(\theta_2|y) \propto k(y)\P(y|\theta_2)$$
$$\frac{L(\theta_1|y)}{L(\theta_2|y)} = \frac{k(y)}{k(y)}\frac{\P(y|\theta_1)}{\P(y|\theta_2)}$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$
\begin{aligned}
L(\theta_1|y) &\propto k(y)\P(y|\theta_1) \\
L(\theta_2|y) &\propto k(y)\P(y|\theta_2) \\
\frac{L(\theta_1|y)}{L(\theta_2|y)} &= \frac{k(y)}{k(y)}\frac{\P(y|\theta_1)}{\P(y|\theta_2)} \\
&= \frac{\P(y|\theta_1)}{\P(y|\theta_2)}
\end{aligned}
$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$
\begin{aligned}
L(\theta_1|y) &\propto k(y)\P(y|\theta_1) \\
L(\theta_2|y) &\propto k(y)\P(y|\theta_2) \\
\frac{L(\theta_1|y)}{L(\theta_2|y)} &= \frac{k(y)}{k(y)} \frac{\P(y|\theta_1)}{\P(y|\theta_2)} \\
&= \frac{\P(y|\theta_1)}{\P(y|\theta_2)}
\end{aligned}
$$

- **Statistically** (from the Neyman-Pearson Hypothesis Testing viewpoint), let

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$L(\theta_1|y) \propto k(y)¶(y|\theta_1)$$
$$L(\theta_2|y) \propto k(y)¶(y|\theta_2)$$
$$\frac{L(\theta_1|y)}{L(\theta_2|y)} = \frac{k(y)}{k(y)}\frac{¶(y|\theta_1)}{¶(y|\theta_2)}$$
$$= \frac{¶(y|\theta_1)}{¶(y|\theta_2)}$$

- **Statistically** (from the Neyman-Pearson Hypothesis Testing viewpoint), let

$$R = -2\ln\left(\frac{L_R^*}{L^*}\right) = 2(\ln L^* - \ln L_R^*)$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$
\begin{aligned}
L(\theta_1|y) &\propto k(y)\P(y|\theta_1) \\
L(\theta_2|y) &\propto k(y)\P(y|\theta_2) \\
\frac{L(\theta_1|y)}{L(\theta_2|y)} &= \frac{k(y)}{k(y)} \frac{\P(y|\theta_1)}{\P(y|\theta_2)} \\
&= \frac{\P(y|\theta_1)}{\P(y|\theta_2)}
\end{aligned}
$$

- **Statistically** (from the Neyman-Pearson Hypothesis Testing viewpoint), let

$$R = -2 \ln \left( \frac{L_R^*}{L^*} \right) = 2(\ln L^* - \ln L_R^*)$$

Then, under the null of no difference between the 2 models,

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$
\begin{aligned}
L(\theta_1|y) &\propto k(y)\P(y|\theta_1) \\
L(\theta_2|y) &\propto k(y)\P(y|\theta_2) \\
\frac{L(\theta_1|y)}{L(\theta_2|y)} &= \frac{k(y)}{k(y)}\frac{\P(y|\theta_1)}{\P(y|\theta_2)} \\
&= \frac{\P(y|\theta_1)}{\P(y|\theta_2)}
\end{aligned}
$$

- **Statistically** (from the Neyman-Pearson Hypothesis Testing viewpoint), let

$$R = -2\ln\left(\frac{L_R^*}{L^*}\right) = 2(\ln L^* - \ln L_R^*)$$

Then, under the null of no difference between the 2 models,

$$R \sim f_{\chi^2}(r|m)$$

# Meaning of the likelihood ratio

- **Substantively**, its the ratio of 2 traditional probabilities:

$$\begin{aligned} L(\theta_1|y) &\propto k(y)¶(y|\theta_1) \\ L(\theta_2|y) &\propto k(y)¶(y|\theta_2) \\ \frac{L(\theta_1|y)}{L(\theta_2|y)} &= \frac{k(y)}{k(y)}\frac{¶(y|\theta_1)}{¶(y|\theta_2)} \\ &= \frac{¶(y|\theta_1)}{¶(y|\theta_2)} \end{aligned}$$

- **Statistically** (from the Neyman-Pearson Hypothesis Testing viewpoint), let

$$R = -2\ln\left(\frac{L_R^*}{L^*}\right) = 2(\ln L^* - \ln L_R^*)$$

Then, under the null of no difference between the 2 models,

$$R \sim f_{\chi^2}(r|m)$$

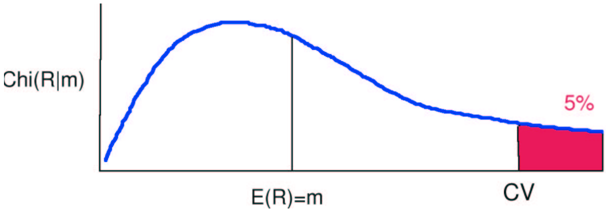
where $r$ is the observed value of $R$ and $m$ is the number of restricted parameters.

# Meaning of the likelihood ratio

From lecture slides

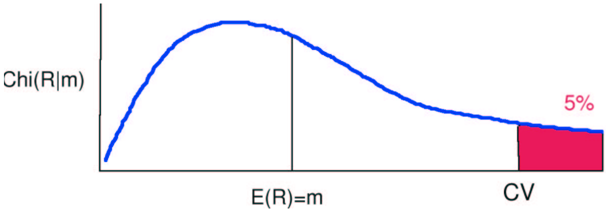
# Meaning of the likelihood ratio

From lecture slides



# Meaning of the likelihood ratio

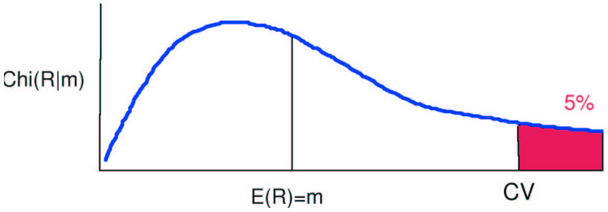
From lecture slides




- If restrictions have no effect, $E(R) = m$.

# Meaning of the likelihood ratio

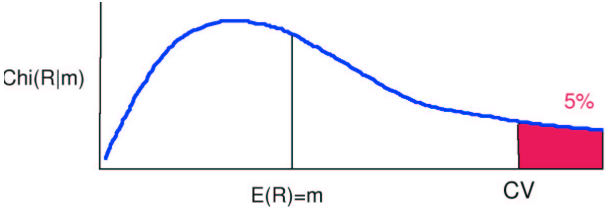
From lecture slides




- If restrictions have no effect, $E(R) = m$.
- So only if $r >> m$ will the test parameters be clearly different from zero.

# Meaning of the likelihood ratio

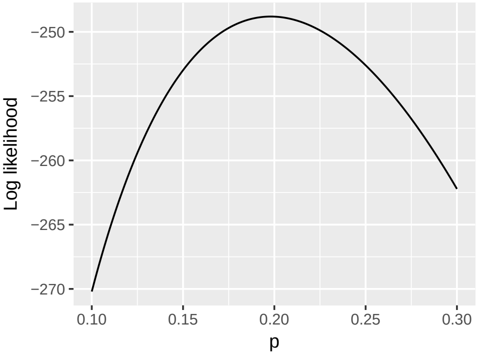
From lecture slides




- If restrictions have no effect, $E(R) = m$.
- So only if $r >> m$ will the test parameters be clearly different from zero.
- Disadvantage: Too many likelihood ratio tests may be required to test all points of interest

# Uncertainty: Curvature at the maximum

Because of the logic of likelihood ratio tests, we can think of uncertainty as curvature 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**.

$$\mathcal{I}_n(\theta) = -E\left( \frac{\partial^2}{\partial \theta^2} \log f(X \mid \theta) \mid \theta \right)$$

# Uncertainty: Curvature at the maximum

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

$$\mathcal{I}_n(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta) \mid \theta\right)$$

The expectation is taken over the distribution of possible samples $X$. In practice, we often use the **observed fisher information** from our one sample as an estimate.

# Uncertainty: Curvature at the maximum

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

$$\mathcal{I}_n(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta) \mid \theta\right)$$

The expectation is taken over the distribution of possible samples $X$. In practice, we often use the **observed fisher information** from our one sample as an estimate.

The variance is the inverse of the Fisher information:

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

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!

# 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?

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

$$=$$

What is the likelihood for $n$ observations?

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

What is the likelihood for $n$ observations?

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

What is the log likelihood?

What is the log likelihood?

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

What is the log likelihood?

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

$$=$$

What is the log likelihood?

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

What is the log likelihood?

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

What is the log likelihood?

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

What is the log likelihood?

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

## What is the derivative of the log likelihood?

What is the derivative of the log likelihood?

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

What is the derivative of the log likelihood?

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

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

What is the derivative of the log likelihood?

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

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?

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

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

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

What is the derivative of the log likelihood?

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

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?

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

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

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.

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

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

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

$$=$$

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!