GOV 2001/ 1002/ E-2001 Section 3
Theories of Inference

Solé Prillaman

Harvard University

February 12, 2014

Logistics

Reading Assignment- Unifying Political Methodology chs 2 and 4.

Problem Set 3- Due by 6pm next Wednesday on Canvas.

Assessment Question- Due by 6pm next Wednesday on Canvas. You must work alone and only one attempt.

Study Groups- New groups posted on Canvas discussion.

Replication Paper

1. Read *Publication, Publication*

2. Find a coauthor. See the Canvas discussion board to help with this.

3. Choose a paper based on the crieria in *Publication, Publication*.

4. Have a classmate sign-off on your paper choice.

5. Upload the paper or papers to Canvas quiz by **2/26**.

Discussion Forum

Everyone should change their notification preferences in Canvas! This makes sure that when updates are made on the discussion board, you know about them.

- Go to 'Settings' in the top right, then under 'Ways to Contact' you make sure your correct email address is there.
- Go to 'Settings' in the top right, and then 'Notifications', and choose 'Notify me right away' for 'Discussion' and 'Discussion Post'.

Model Set-up

$$Y_i \sim f_N(\mu_i, \sigma^2)$$
$$\mu_i = \beta x_i$$

$$Y_i \sim f_N(\mu, \sigma_i^2)$$
$$\sigma_i^2 = \beta x_i$$

$$Y_i \sim f_N(\mu_i, \sigma_i^2)$$
$$\mu_i = \beta x_{1i}$$
$$\sigma_i^2 = \alpha x_{2i}$$

Quiz Break!

If $P(Y_i = y_i) = \binom{n}{y_i} \pi^{y_i} (1 - \pi_i)^{n - y_i}$ and π_i is a function of x_i and β, how do we write out the systematic and stochastic components of this model?

Bootstrapped Standard Errors

We can estimate the sampling distribution of our parameters using bootstrapping.

1. Start with observed sample and model.
2. Determine the quantity of interest (the estimand). Ex: β.
3. Choose an estimator. Ex: OLS, MLE, etc.
4. Take a re-sample of size n (with replacement) from the sample and calculate the estimate of the parameter using the estimator.
5. Repeat step 4 many times, say 1000, each time storing the estimate of the parameter.
6. Calculate the standard deviation of your 1000 estimates of the parameter.

Outline

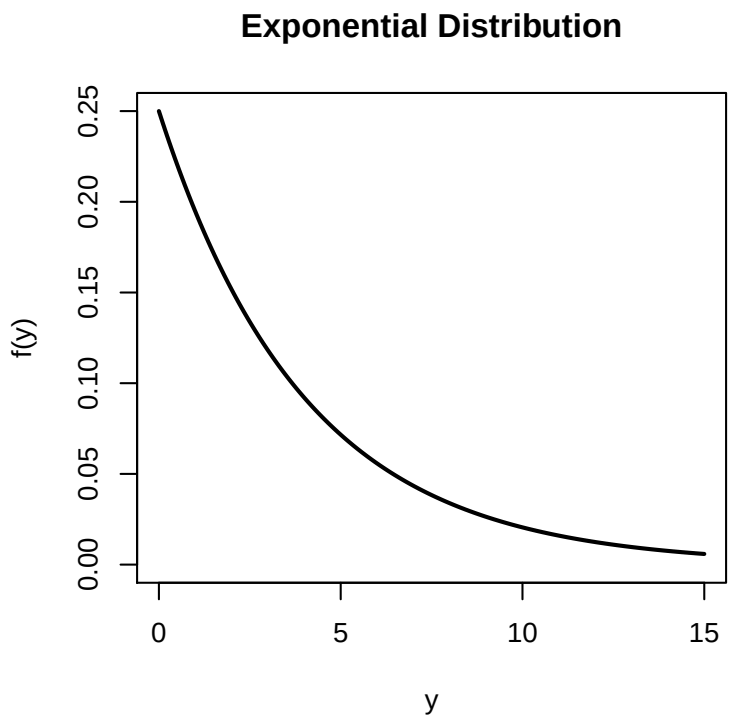
Likelihood: An Example

Ex. Waiting for the Redline – How long will it take for the next T to get here?

Likelihood: An Example

Y is a Exponential random variable with parameter λ.

$$f(y) = \lambda e^{-\lambda y}$$



What is Likelihood?

Last week we assumed $\lambda = .25$ to get the probability of waiting for the redline for Y mins, i.e. we set λ to look at the probability of drawing some data.

$$p(y|\lambda) = .25e^{-.25y}$$

- $\lambda = .25 \rightarrow data.$
- $p(y|\lambda = .25) = .25e^{-.25y} \rightarrow p(2 < y < 10|\lambda) = .525$

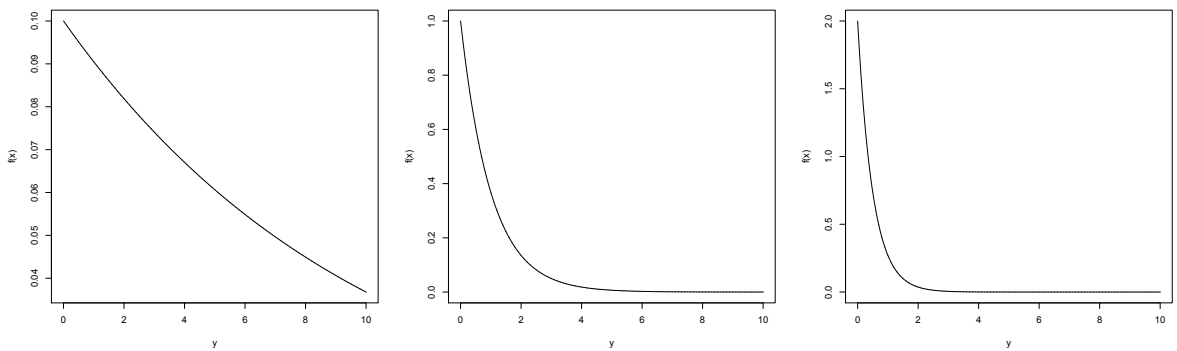
So we can calculate $p(y|\lambda)$.

What is Likelihood?

BUT, what we care about estimating is λ!

Given our observed data, what is the probability of getting a particular λ?

- $data \rightarrow \lambda$.
- $p(\lambda|y) = ?$



Likelihood and Bayes' Rule

Why is Bayes' rule important?

- Often we want to know $P(\theta|\text{data})$.
- But what we *do* know is $P(\text{data}|\theta)$.
- We'll be able to *infer* θ by using a variant of Bayes rule.

Bayes' Rule:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

LIKELIHOOD

The whole point of likelihood is to leverage information about the data generating process into our inferences.

Here are the basic steps:

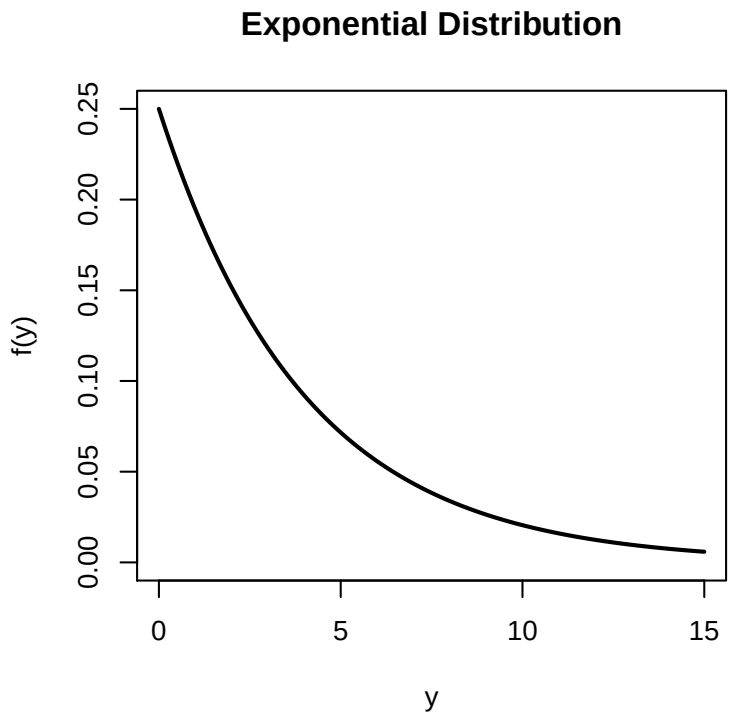
1. Think about your data generating process. (What do the data look like? Use your substantive knowledge.)
2. Find a distribution that you think explains the data. (Poisson, Binomial, Normal? Something else?)
3. Derive the likelihood.
4. Plot the likelihood.
5. Maximize the likelihood to get the MLE.

Note: This is the case in the univariate context. We'll be introducing covariates later on in the term.

Data Generating Process and Distribution of Y

Y is a Exponential random variable with parameter λ.

$$f(y|\lambda) = \lambda e^{-\lambda y}$$



Derive the Likelihood

- From Bayes' Rule:

$$p(\lambda|y) = \frac{p(y|\lambda)p(\lambda)}{p(y)}$$

- Let

$$k(y) = \frac{p(\lambda)}{p(y)}$$

(Note that the λ in $k(y)$ is the true λ, a constant that doesn't vary. So $k(y)$ is just a function of y and therefore a constant since we know y.)

Derive the Likelihood

$$
\begin{aligned}
p(\lambda|y) &= \frac{p(y|\lambda)p(\lambda)}{p(y)} \\
L(\lambda|y) &= p(y|\lambda)k(y) \\
&\propto p(y|\lambda)
\end{aligned}
$$

$$
\begin{aligned}
L(\lambda|y_1) &\propto p(y_1|\lambda) \\
&= \lambda e^{-\lambda * y_1} \\
&= \lambda e^{-\lambda * 12}
\end{aligned}
$$

On Monday it took 12 minutes for the train to come.

Quiz Break!

Why is the Likelihood only a measure of relative uncertainty and *not* absolute uncertainty?

Plot the likelihood

First, note that we can take advantage of a lot of pre-packaged R functions

- `rbinom`, `rpoisson`, `rnorm`, `runif` → gives random values from that distribution
- `pbinom`, `ppoisson`, `pnorm`, `punif`→ gives the cumulative distribution (the probability of that value or less)
- `dbinom`, `dpoisson`, `dnorm`, `dunif`→ gives the density (i.e., height of the PDF – useful for drawing)
- `qbinom`, `qpoisson`, `qnorm`, `qunif`→ gives the quantile function (given quantile, tells you the value)

Plot the Likelihood

We want to plot $L(\lambda|y) \propto \lambda e^{-\lambda * 12}$

1. We can do this easily since we know that y is exponentially distributed.

```
# dexp provides the PDF of the exponential distribution
# If we set x to be 12 and lambda to be .25,
# then p(y | lambda) =
dexp(x = 12, rate = .25, log=FALSE)
[1] 0.01244677

# To plot the likelihood, we can plot the PDF for different
# values of lambda
curve(dexp(12, rate = x),
  xlim = c(0,1),
  xlab = "lambda",
  ylab = "likelihood")
```

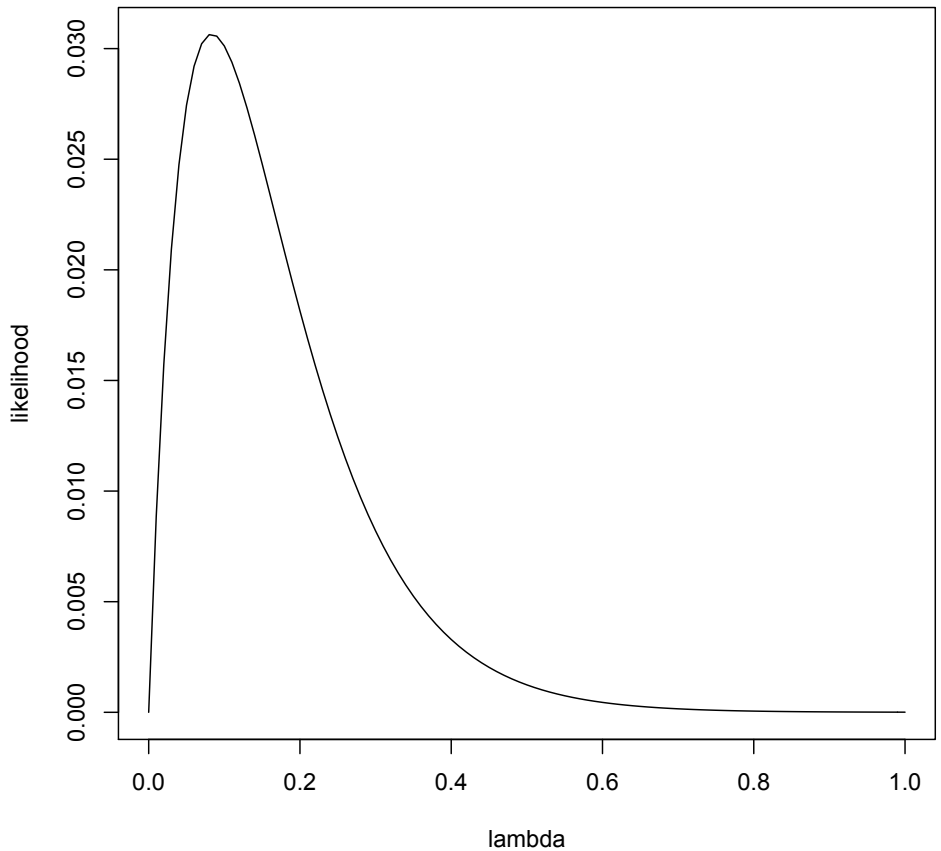
Plot the Likelihood

2. Or we can write a function for our likelihood and plot that function.

```
# Write a function for L( lambda | y)
# The inputs are lambda and y (our data)
likelihood <- function(lambda, y){
  lambda * exp(-lambda * y)
}

# We can use curve to plot this likelihood function
curve(likelihood(lambda = x, y = 12),
  xlim = c(0,1),
      xlab = "lambda",
      ylab = "likelihood")
```

Maximize the Likelihood



What do you think the maximum likelihood estimate will be?

Outline

Likelihoods and Bayesian Posteriors

Likelihood:

$$\begin{aligned} p(\lambda|y) &= \frac{p(\lambda)p(y|\lambda)}{p(y)} \\ L(\lambda|y) &= k(y)p(y|\lambda) \\ &\propto p(y|\lambda) \end{aligned}$$

There is a fixed, true value of λ. We use the likelihood to estimate λ with the MLE.

Bayesian Posterior Density:

$$\begin{aligned} p(\lambda|y) &= \frac{p(\lambda)p(y|\lambda)}{p(y)} \\ &= \frac{p(\lambda)p(y|\lambda)}{\int_\lambda p(\lambda)p(y|\lambda)d\lambda} \\ &\propto p(\lambda)p(y|\lambda) \end{aligned}$$

λ is a random variable and therefore has fundamental uncertainty. We use the posterior density to make probability statements about λ.

Understanding the Posterior Density

In Bayesian inference, we have a prior *subjective* belief about λ, which we update with the data to form posterior beliefs about λ.

$$p(\lambda|y) \propto p(\lambda)p(y|\lambda)$$

- $p(\lambda|y)$ is the posterior density
- $p(\lambda)$ is the prior density
- $p(y|\lambda)$ is proportional to the likelihood

Bayesian Inference

The whole point of bayesian inference is to leverage information about the data generating process along with subjective beliefs about our parameters into our inference.

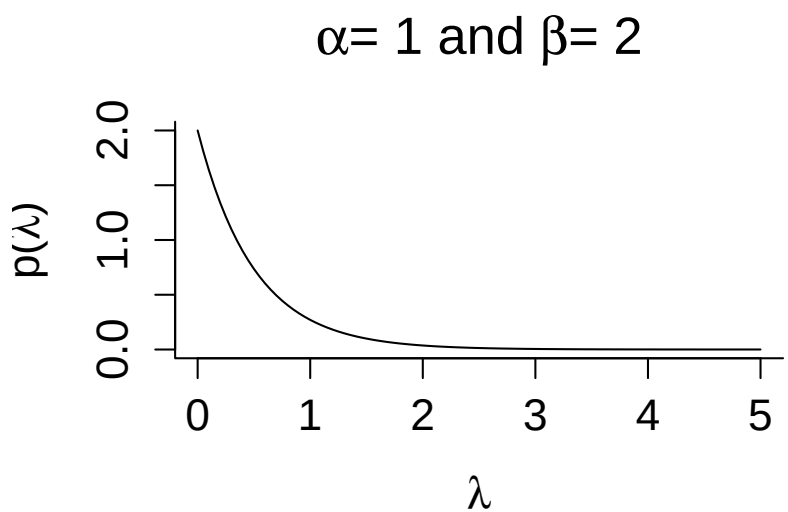
Here are the basic steps:

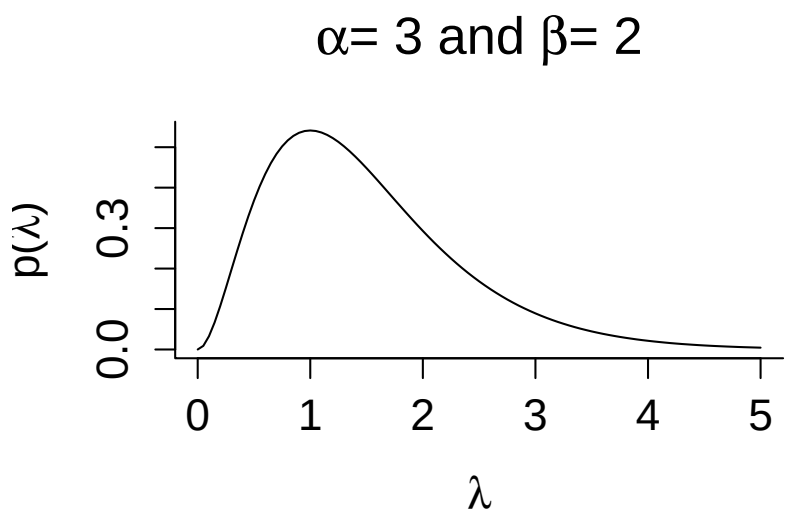
1. Think about your subjective beliefs about the parameters you want to estimate.
2. Find a distribution that you think explains your prior beliefs of the parameter.
3. Think about your data generating process.
4. Find a distribution that you think explains the data.
5. Derive the posterior distribution.
6. Plot the posterior distribution.
7. Summarize the posterior distribution. (posterior mean, posterior standard deviation, posterior probabilities)

Setting the Prior

Continuing from our example about the redline, we believe that the rate at which trains arrive is distributed according to the Gamma distribution with **known** parameters α and β. That is:

$$p(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$





Derive the Posterior

$$
\begin{aligned}
p(\lambda|y) &\propto p(\lambda)p(y|\lambda) \\
&\propto \frac{\beta^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\lambda e^{-\lambda*12} \\
&\propto \lambda^{\alpha-1}e^{-\beta\lambda}\lambda e^{-\lambda*12} && \text{dropping constants} \\
&\propto \lambda^{\alpha}e^{-\beta\lambda-\lambda*12} && \text{combining terms} \\
&\propto \lambda^{\alpha}e^{-\lambda(\beta+12)}
\end{aligned}
$$

So $p(\lambda|y) \sim Gamma(\alpha+1, \beta+12)$

Plot the Posterior

We want to plot $p(\lambda|y)$

1. We can do this easily since we know that $\lambda|y$ is distributed gamma.

```
# dgamma provides the PDF of the gamma distribution
# To plot the posterior, we can plot the PDF for different
# values of lambda
# Let's set alpha = 1 and beta = 2
curve(dgamma(x, shape = 2, rate = 14),
  xlim = c(0,1),
  xlab = "lambda",
  ylab = "posterior")
```

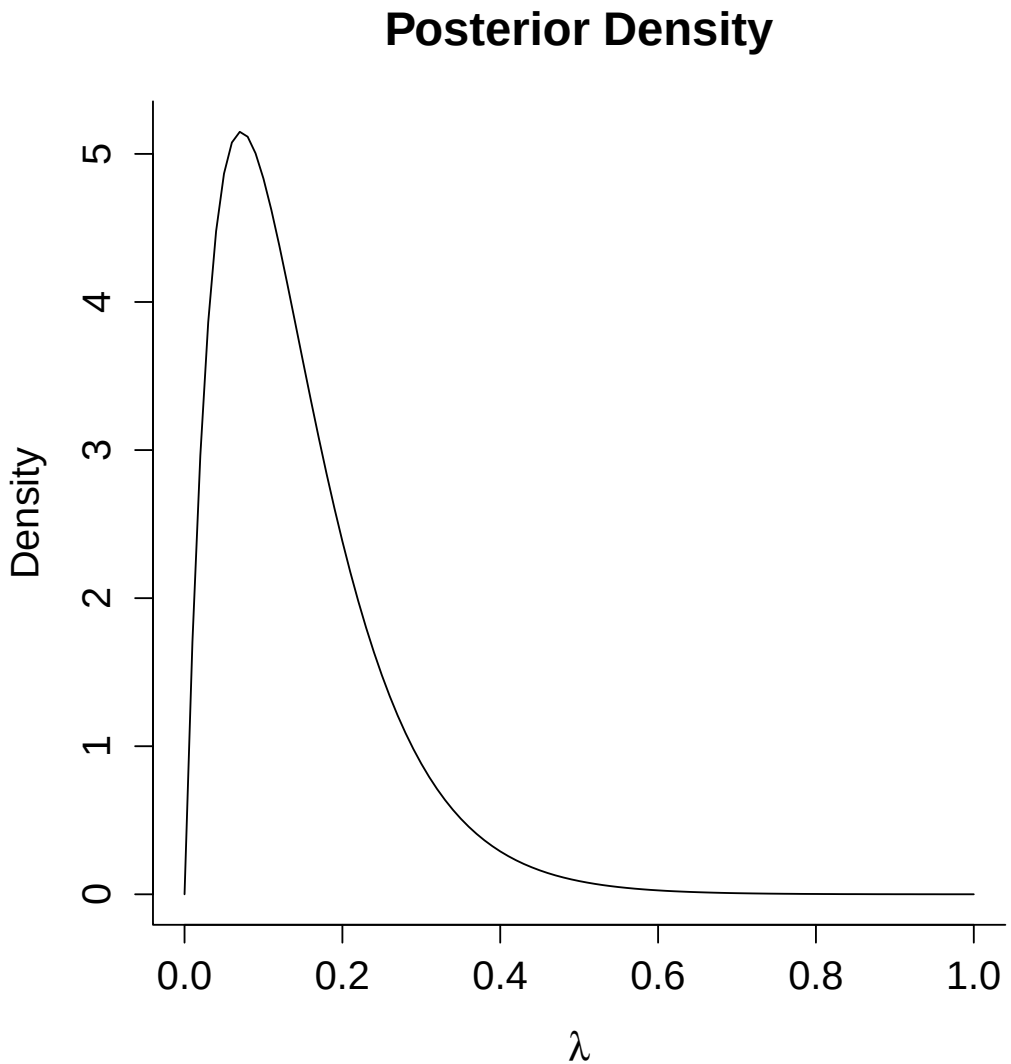
Plot the Posterior

2. Or we can do this without knowing the exact distribution of our posterior and plot a proportional distribution.

```
# Write a function for p( lambda | y)
# The inputs are lambda, y (our data), beta, and alpha
posterior.propto <- function(lambda, y, beta, alpha){
  lambda^alpha * exp(-lambda * (beta + y))
}

# We can use curve to plot this likelihood function
# Let's set alpha = 1 and beta = 2
curve(posterior(lambda = x, y = 12, alpha = 1, beta = 2),
  xlim = c(0,1),
      xlab = "lambda",
      ylab = "posterior")
```

Plot the Posterior



Summarize the Posterior Distribution

What is the mean of the posterior distribution?

1. We can simulate to get the mean of the posterior

```
# Again let's assume that alpha is 1 and beta is 2
sim.posterior <- rgamma(10000, shape = 2, rate = 14)
mean(sim.posterior)
```

2. Or we can calculate this analytically because the mean of a gamma is $\frac{\alpha}{\beta}$

```
mean <- (1 + 1)/(2 + 12)
mean
```

Summarize the Posterior Distribution

What is the probability that λ is between 0.2 and 0.4?

1. We can integrate our proportional function and divide this by the total area from our proportional function

```
density1 <- integrate(posterior.propto, alpha = 1,
beta = 2, y = 12, lower = .2, upper = .4)
total.density <- integrate(posterior.propto, alpha = 1,
beta = 2, y = 12, lower = 0, upper = 100)
density1$value / total.density$value
```

2. Or we can calculate this directly because we know the functional form of the posterior:

```
pgamma(.4, shape = 2, rate = 14) -
pgamma(.2, shape = 2, rate = 14)
```

Outline

Hypothesis Testing

1. Set your **null hypothesis-** the assumed null state of the world.

   ex. $H_0 : \beta = c$

2. Set your **alternative hypothesis-** the claim to be tested.

   ex. $H_A : \beta \neq c$

Hypothesis Testing

3. Choose your **test statistic (estimator)-** function of the sample and the null hypothesis used to test your hypothesis.

   ex. $\frac{\hat{\beta}-c}{\hat{SE}[\hat{\beta}]}$

4. Determine the **null distribution-** the sampling distribution of the test statistic assuming that the null hypothesis is true.

   ex. $\frac{\hat{\beta}-c}{\hat{SE}[\hat{\beta}]} \sim t_{n-k-1}$

Hypothesis Testing

How do we choose our estimator?

By the CLT, if n is large, then $\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2})$.

If we standardize, we get that $Z = \frac{\hat{\beta} - \beta}{SE[\hat{\beta}]} \sim N(0, 1)$

If we estimate σ^2, we get that $Z = \frac{\hat{\beta} - \beta}{\hat{SE}[\hat{\beta}]} \sim t_{n-k-1}$

We can plug in c for β since we are *assuming* the null hypothesis is true.

Hypothesis Testing

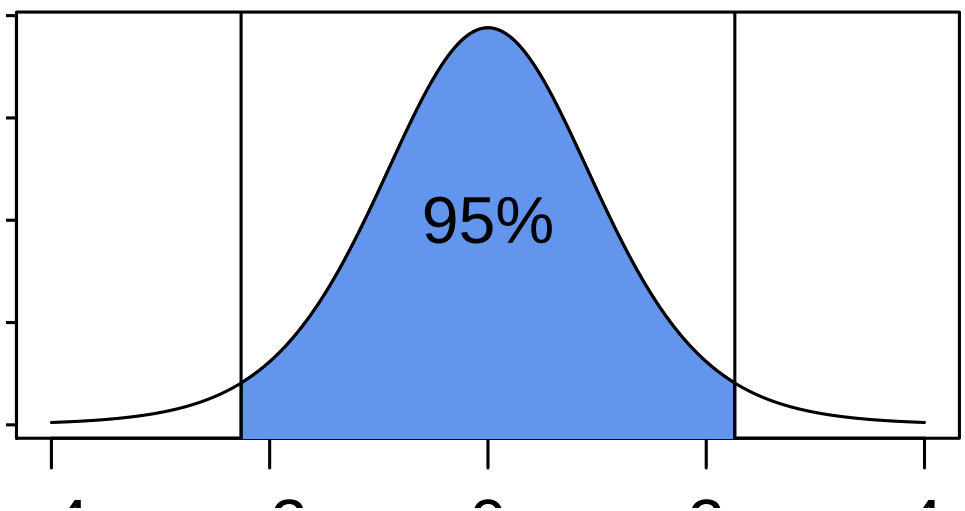
5. Choose your **significance level**, α. This is the accepted amount of Type 1 Error.
6. Set your **decision rule** and **crititcal value-** the function that specifies where we accept or reject the null hypothesis for a given value of the test statistic; a function of α.

ex. reject null if $\left| \frac{\hat{\beta} - c}{\hat{SE}[\hat{\beta}]} \right| \geq t_{1-\alpha/2, n-k-1}$

Hypothesis Testing

How do we get $t_{1-\alpha/2,n-k-1}$?

```
## alpha = .05
## 9 degrees of freedom
qt(1-.025, df=9)
```



Hypothesis Testing

6. Calculate the **rejection region-** the set of values for which we will reject the null hypothesis.

   ex. $[-\infty, -t_{1-\alpha/2,n-k-1}]$ and $[t_{1-\alpha/2,n-k-1}, \infty]$

7. Using your sample, calculate your observed estimate.
8. Determine if your estimate is within the rejection region and draw conclusion about hypothesis.

   ex.

Reject null	if	$\frac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}} \in [-\infty, -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}, \infty]$
Fail to reject null	if	$\frac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}} \notin [-\infty, -z_{1-\alpha/2}] \cup [z_{1-\alpha/2}, \infty]$

P-VALUES

- **p-value-** Assuming that the null hypothesis is true, the probability of getting something at least as extreme as our observed test statistic, where extreme is defined in terms of the alternative hypothesis.

How do we find p-values?

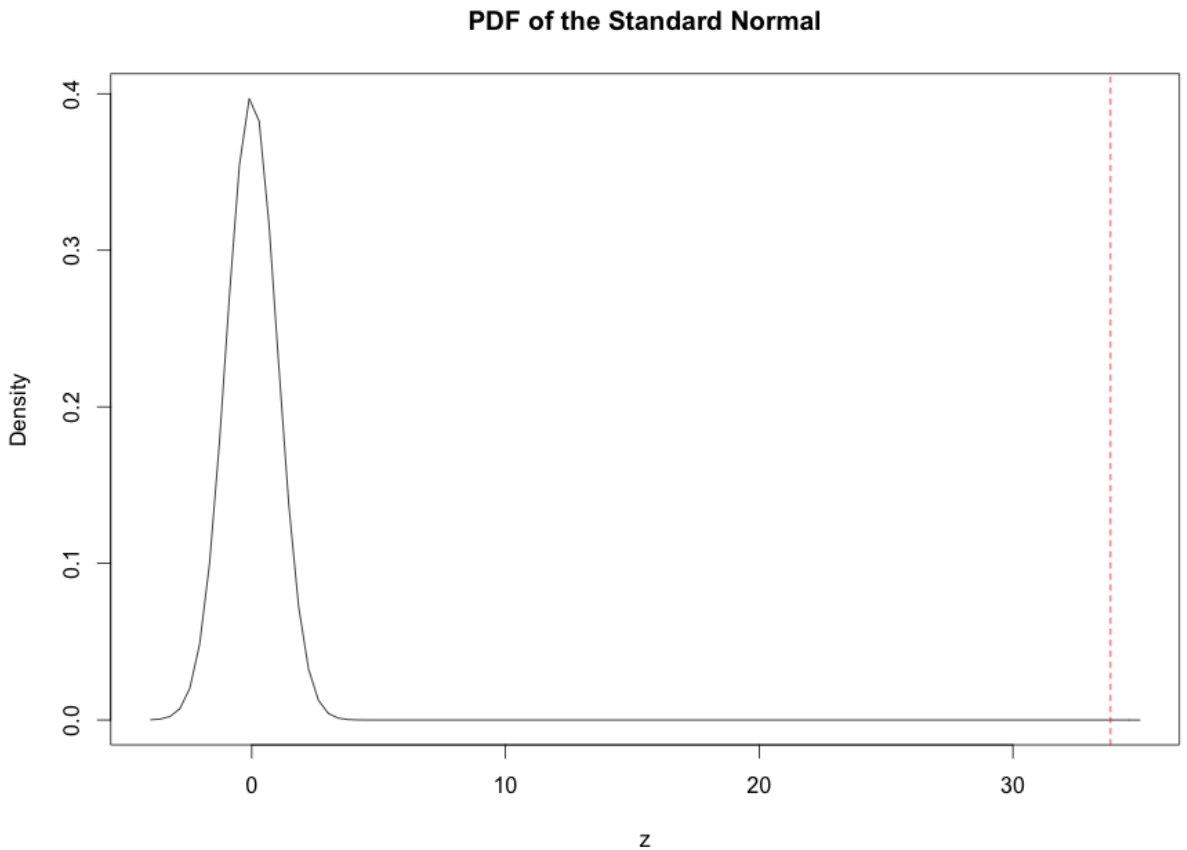
```
# create a population with mean 12, sd 4
pop <- rnorm(n=1e6, mean=12, sd=4)

# draw a single sample of size 100 from population
my.sample <- sample(pop, size=100, replace=T)

# calculate our test statistic
# this is a test statistic for the sample mean
# as an estimator of the true mean (which we set to be 12)
test.statistic <- mean(my.sample) /(sd(my.sample)/10)

# find the p-value
p.value <- 2*(1-pnorm(test.statistic))
```

P-VALUES



Questions

Questions?