Gov 2001: Section 4

February 20, 2013

Outline

1. The Likelihood Model with Covariates
2. Likelihood Ratio Test
3. The Central Limit Theorem and the MLE
4. What We Can Do with a Normal MLE
5. Efficiency

Outline

1. The Likelihood Model with Covariates
2. Likelihood Ratio Test
3. The Central Limit Theorem and the MLE
4. What We Can Do with a Normal MLE
5. Efficiency

A roadmap

- We started by introducing the concept of likelihood in the simplest univariate context – one observation, one variable.
- Then we moved forward with more than one observation – and multiplied likelihoods together.
- Now, we are introducing covariates.

$$\begin{aligned} Stochastic : Y_i &\sim f(y_i|\gamma) \\ Systematic : \gamma &= g(X_i, \theta) \end{aligned}$$

- This allows us to find estimated coefficients for our covariates – ultimately what we are really interested in!

A roadmap (ctd.)

Key to all of this is the distinction between stochastic and systematic components:

- *Stochastic* - the probability distribution of the data; key to identifying what model (Poisson, binomial, etc.) you should use., E.g., $Y_i \sim f(y_i|\gamma)$.
- *Systematic* - how the parameters of the probability distribution vary over your covariates; key to incorporating covariates into your model. E.g., $\gamma = g(X_i, \theta)$.

You'll need both parts to model the likelihood.

Back to our Running Example

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

Y is a Exponential random variable with parameter λ.

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

Back to our Running Example

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

But this time we want to add covariates. What do you think affects the wait for the Redline?

How would we model this?

- We know the stochastic component:

$$\begin{aligned} Y_i &\sim Exponential(\lambda_i) \\ Y_i &\sim \lambda_i e^{-\lambda_i y_i} \end{aligned}$$

- Remember, for an Exponential

$$\mu_i = \frac{1}{\lambda_i}$$

- So we're going to set the systematic component

$$\begin{aligned} \mu_i &= \exp(X_i\beta) \\ \lambda_i &= \frac{1}{\exp(X_i\beta)} \end{aligned}$$

Why do we use the exp?
What are the parameters here?

Solve for the log-likelihood

First, write the log-likelihood in terms of λ_i:

$$\begin{aligned} L(\lambda_i|y_i) &\propto Exponential(y_i|\lambda_i) \\ &\propto \lambda_i e^{-\lambda_i y_i} \\ L(\lambda|y) &\propto \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i} \\ \ln L(\lambda|y) &\propto \sum_{i=1}^{n} (\ln \lambda_i - \lambda_i y_i) \end{aligned}$$

Solve for the log-likelihood

Next, plug in the systematic component

$$\begin{aligned}
\ln L(\beta|y) &\propto \sum_{i=1}^{n}\left(\ln\left(\frac{1}{\exp(X_i\beta)}\right) - \frac{1}{\exp(X_i\beta)}y_i\right) \\
\ln L(\beta|y) &\propto \sum_{i=1}^{n}\left(\ln 1 - \ln\exp(X_i\beta) - \frac{1}{\exp(X_i\beta)}y_i\right) \\
\ln L(\beta|y) &\propto \sum_{i=1}^{n}\left(-(X_i\beta) - \frac{1}{\exp(X_i\beta)}y_i\right)
\end{aligned}$$

Solve Using R

I'm going to say whether or not it is Friday and the minutes behind schedule are important covariates.
I'm going to create some fake data:

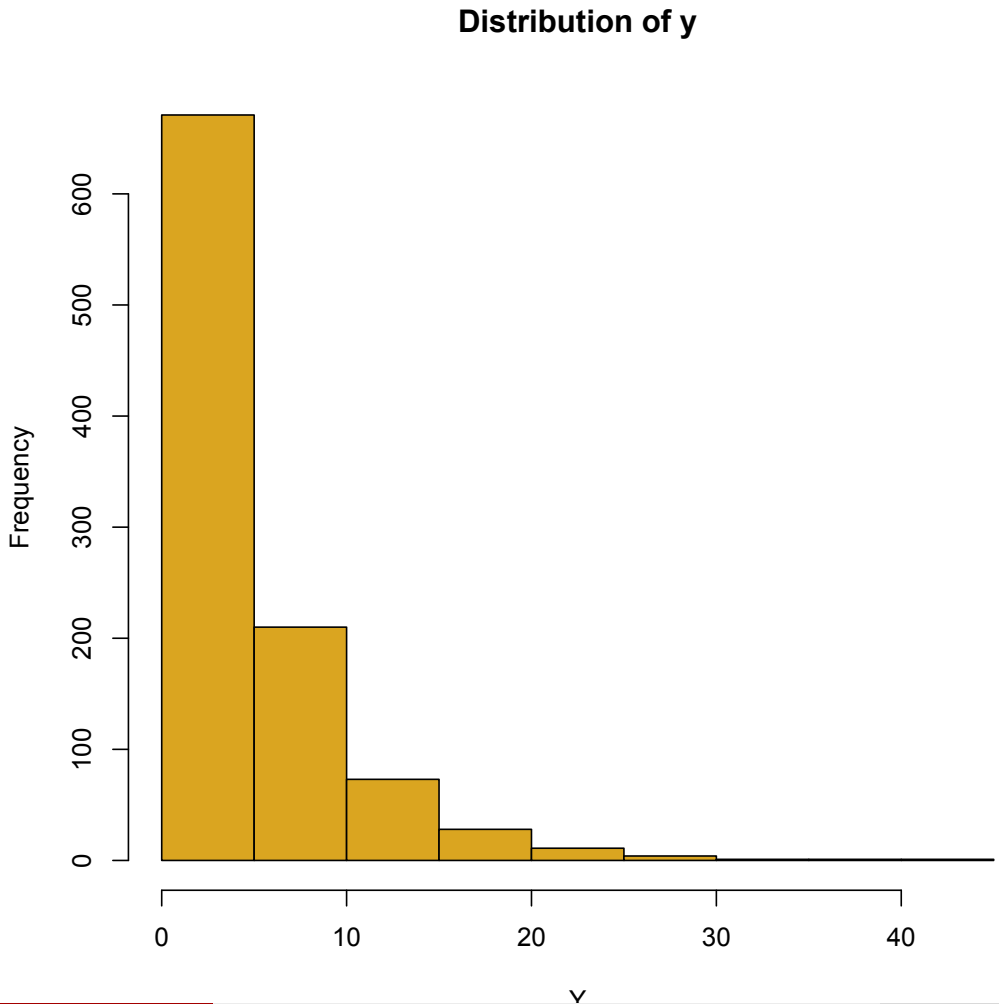
```
set.seed(02139)
n = 1000
Friday <- sample(c(0,1), n, replace=T)

minsSch <- rnorm(n, 3, .5)

Y <- rexp(n, rate = 1/exp(1.25 - .5*Friday +.2*minsSch))
data <- as.data.frame(cbind(Y, Friday, minsSch))
```

Let's look at Y

```
hist(Y, col = "goldenrod", main = "Distribution of y")
```



Solve with Zelig

We could solve with Zelig.

```
library(Zelig)

#First, create an indicator that indicates 100% death rate.
data$ind <- rep(1,n)

#Next, specify the model
z.out <- zelig(Surv(Y, ind) ~ Friday + minsSch,
       data=data, model="exp")
```

Solve with Zelig

```
summary(z.out)

Call:
zelig(formula = Surv(Y, ind) ~ Friday + minsSch,
  model = "exp", data = data)
             Value Std. Error     z        p
(Intercept)  1.089     0.1928  5.65 1.64e-08
Friday      -0.463     0.0635 -7.29 3.02e-13
minsSch      0.212     0.0613  3.46 5.41e-04

Scale fixed at 1

Exponential distribution
Loglik(model)= -2503.4   Loglik(intercept only)= -2538.4
Chisq= 69.99 on 2 degrees of freedom, p= 6.7e-16
Number of Newton-Raphson Iterations: 4
```

Solving Manually

Remember the log-likelihood we solved for before:

$$\ln L(\beta|y) \propto \sum_{i=1}^{n} \left(-(X_i\beta) - \frac{1}{\exp(X_i\beta)} y_i \right)$$

We can program the log-likelihood in two ways.

```
llexp <- function(param, y, x){
rate <- 1/exp(x%*%param)
sum(dexp(y, rate=rate, log=T))
}

llexp2 <- function(param, y,x){
cov <- x%*%param
sum(-cov - 1/exp(cov)*y)
}
```

Solving Manually: Optimize

```
#Create X with an intercept
X <- cbind(1, Friday, minsSch)

#Specify starting values
param <- c(1,1,1)

#Solve using optim
out <- optim(param, fn=llexp, y=Y, x=X, method="BFGS",
   hessian=T, control=list(fnscale=-1))
```

Solving Manually: Output

```
> out$par
[1]  1.0885871 -0.4634621  0.2120591
```

Does this check with the Zelig output?

Outline

1 The Likelihood Model with Covariates

2 Likelihood Ratio Test

3 The Central Limit Theorem and the MLE

4 What We Can Do with a Normal MLE

5 Efficiency

Likelihood Ratio Tests

- Useful for when you are comparing two models.
- We'll call these restricted and unrestricted:

$$\textit{Unrestricted} : \beta_0 + \beta_1 X_1 + \beta 2 X_2$$
$$\textit{Restricted} : \beta_0 + \beta_2 X_2$$

- We want to test the usefulness of the parameters in the unrestricted model but omitted in the restricted model.

Likelihood Ratio Tests (ctd.)

Here's how to operationalize this:

- Let L^* be the maximum of the unrestricted likelihood, and let L_r^* the maximum of the restricted likelihood.
- But adding more variables can only increase the likelihood.
- Thus, $L^* \geq L_r^*$, or $\frac{L_r^*}{L} \leq 1$.
- If the likelihood ratio is exactly 1, then there's no effect of the extra parameters at all.

Likelihood Ratio Tests (ctd.)

Now, let's define a test statistic:

$$\begin{aligned} define: \Re &= -2ln\frac{L_r^*}{L^*} \\ &= 2(lnL^* - lnL_r^*) \end{aligned}$$

- $\Re$ will always be greater than zero.
- It follows a χ^2 distribution with m degrees of freedom, where m is the number of restrictions.
- Key question: how much greater than zero does $\Re$ have to be in order to convince us that the difference is due to systematic differences between the two models?

Back to Our Example

What if we wanted to test whether the minutes behind schedule should be in our model at all?

```
> unrestricted <- optim(param, fn=llexp, y=Y, x=X,
      method="BFGS", hessian=T, control=list(fnscale=-1))
> unrestricted$value
[1] -2503.445
```

versus

```
>restricted <- optim(c(1,1), fn=llexp, y=Y,
      x=cbind(1, Friday), method="BFGS",
      hessian=T, control=list(fnscale=-1))
> restricted$value
[1] -2509.471
```

Back to Our Example (ctd.)

Under the null that the restrictions are valid, the test statistic would be distributed χ^2 with one degree of freedom:

```
r <- 2*(unrestricted$value - restricted$value)
```

```
> 1-pchisq(r,df=1)
[1] 0.0005176814
```

So the probability of getting this test statistic under the null is extremely small. We reject.

Outline

1 The Likelihood Model with Covariates

2 Likelihood Ratio Test

3 The Central Limit Theorem and the MLE

4 What We Can Do with a Normal MLE

5 Efficiency

- Once an ML estimates are calculated, we'll want to know how good they are.
- How much information does the MLE contain about the underlying parameter?
- The MLE alone isn't satisfying – we need a way to quantify uncertainty.

Convince Yourself of a Normal Distribution for the MLE

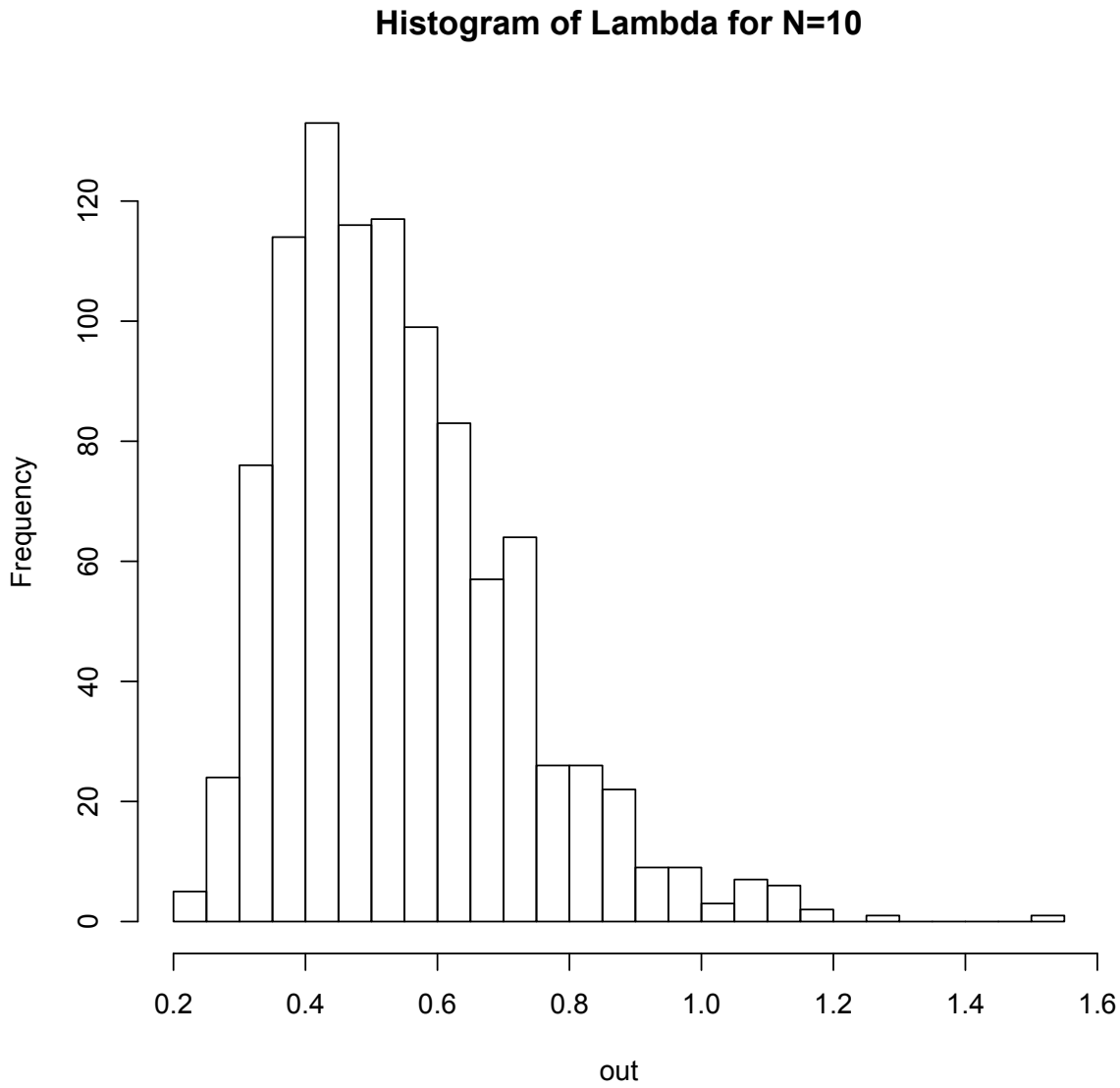
- I'm going to generate 1000 datasets, of 10 observations each from an Exponential with $\lambda = .5$.

```
> n <- 10
> data <- sapply(seq(1,1000)function(x)
        rexp(n, rate=.5))
> dim(data)
[1]   10 1000
```

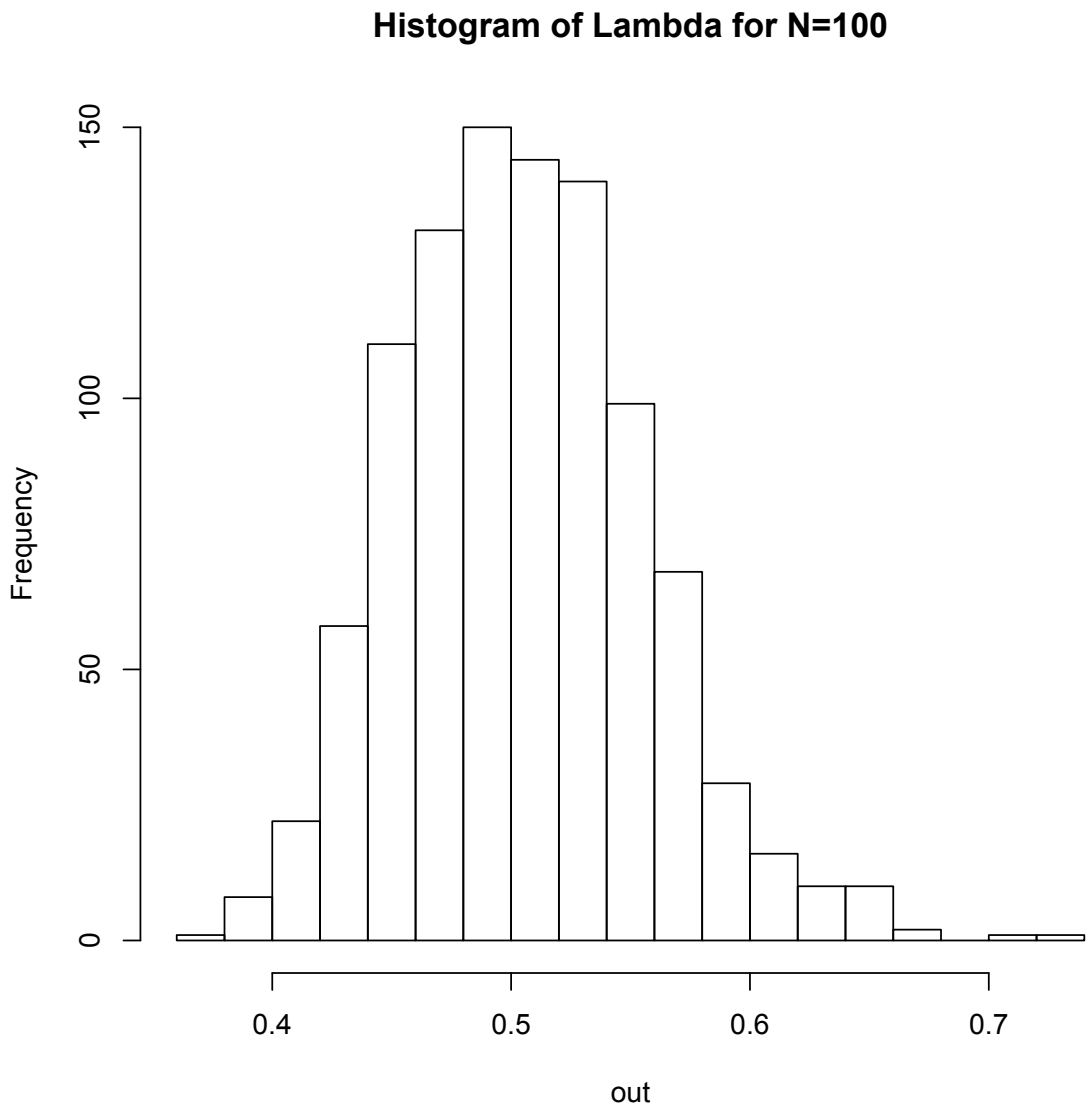
- For each of these datasets, I'm going to find the maximum likelihood estimate for λ.

```
llexp <- function(param, y){
sum(dexp(y, rate=param, log=T))
}
out <- NULL
for(i in 1:1000){
out[i] <- optim(c(1), fn=llexp, y=data[,i],
  method="BFGS", control=list(fnscale=-1))$par
```

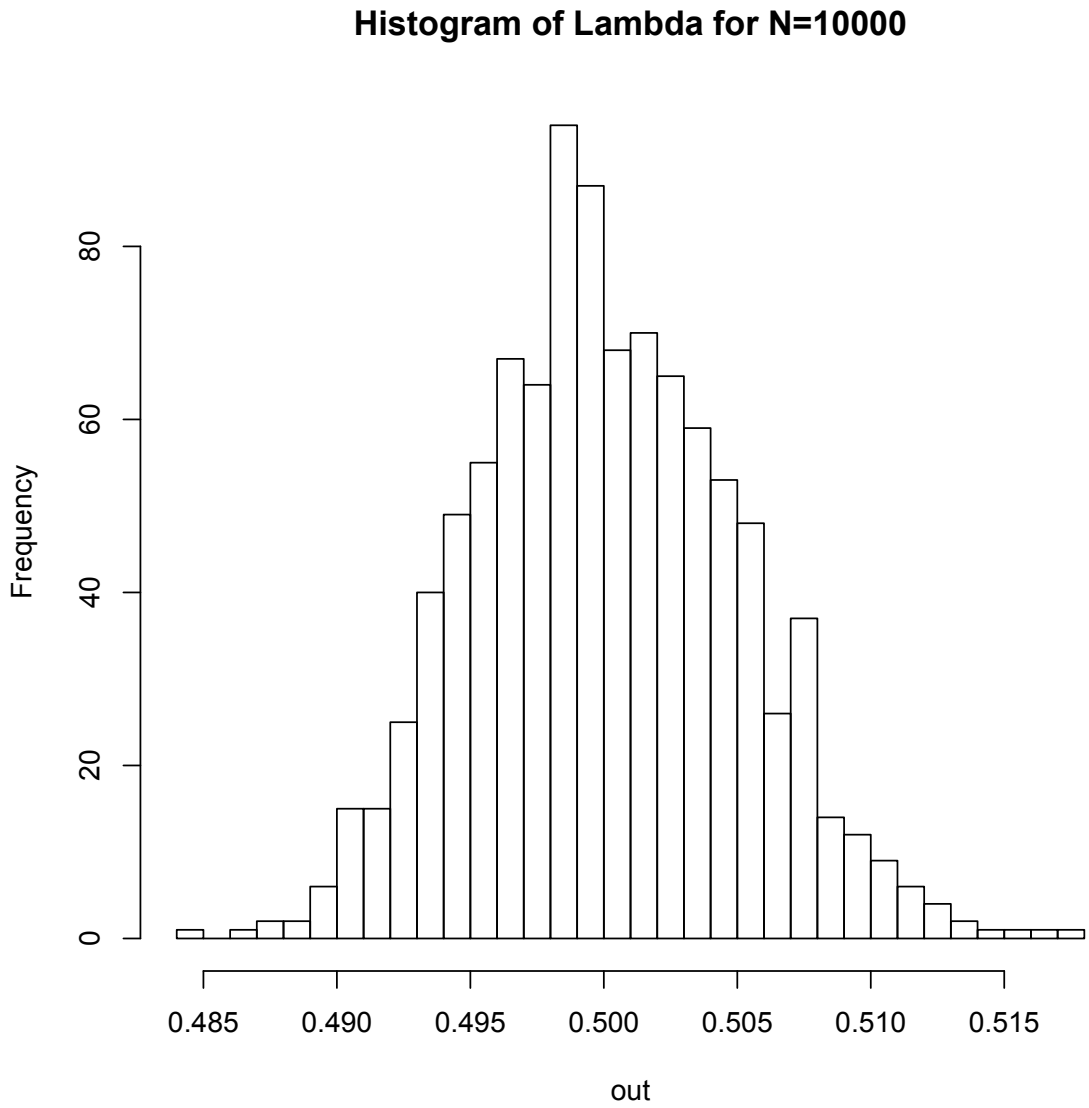
Plot of the Distribution of Lambda when N is 10



Plot of the Distribution of Lambda when N is 100



Plot of the Distribution of Lambda when N is 10000



How do we think about this intuitively?

- The Central Limit Theorem states that the mean of independent random variables will become approximately Normal as n goes to infinity.
- But we're not talking about the mean!
- Yes, but the maximum of the log-likelihood is essentially the mean of a lot of likelihoods. And we use this maximum to estimate our parameter.
- Therefore, as n gets larger, and the more likelihoods we have to conglomerate, the more normal the distribution of the parameter becomes.

Our Normal Variable

So for large n, our parameter θ is distributed Normally with the mean as the true value of θ and the variance $[I(\hat{\theta})]^{-1}$

- Measure of curvature: Fisher Information Matrix

$$I(\hat{\theta}) = -\frac{\partial^2 lnL(\theta)}{\partial^2 \theta}(\hat{\theta})$$

- Inverse of the Fisher Information gives us $Var(\hat{\theta})$

$$[I(\hat{\theta})]^{-1} \approx Var(\hat{\theta})$$

- Square root of $Var(\hat{\theta})$ gives us $SE(\hat{\theta})$

$$SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$$

Outline

1 The Likelihood Model with Covariates

2 Likelihood Ratio Test

3 The Central Limit Theorem and the MLE

4 What We Can Do with a Normal MLE

5 Efficiency

We Can Compute Confidence Intervals

- Under H_0: $\theta = \theta_0$

$$Z_{\theta_0} = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \sim N(0,1)$$

- Given α, the Confidence Interval:

$$[\hat{\theta} \pm N(0,1)_{\frac{\alpha}{2}} * SE(\hat{\theta})]$$

- Why? Find all θ_0, s.t.

$$P(-N(0,1)_{\frac{\alpha}{2}} \leq Z_{\theta_0} \leq N(0,1)_{\frac{\alpha}{2}}) = 1 - \alpha$$

We Can Make Predictions

- First, we find the variance-covariance matrix of our parameters we had before.

```
out <- optim(param, fn=llexp, y=Y, x=X, method="BFGS",
       hessian=T, control=list(fnscale=-1))
varcv <- -solve(out$hessian)
```

- Using our assumption that for large n, our β's are distributed Normally, we simulate β's.

```
simbetas <- rmvnorm(1000, mean=out$par, sigma=varcv)
```

We Can Make Predictions

Say I wanted to know how much longer I will have to wait for the redline on Friday.

- First, I would have to create covariates that will let me predict the wait on Friday

```
predCovs <- c(1,1, mean(minsSch))
```

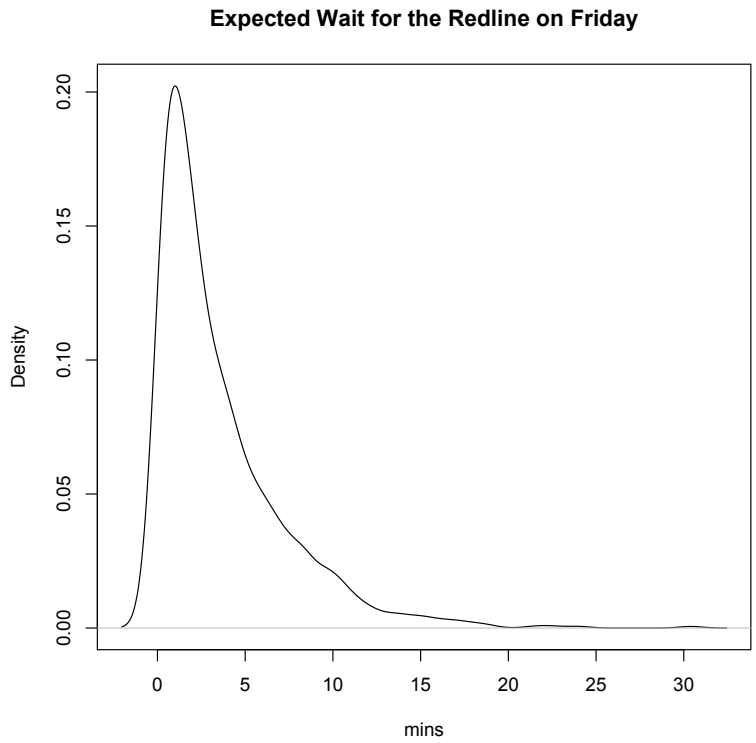
- Then I would have to simulate y's from our model and simulated betas

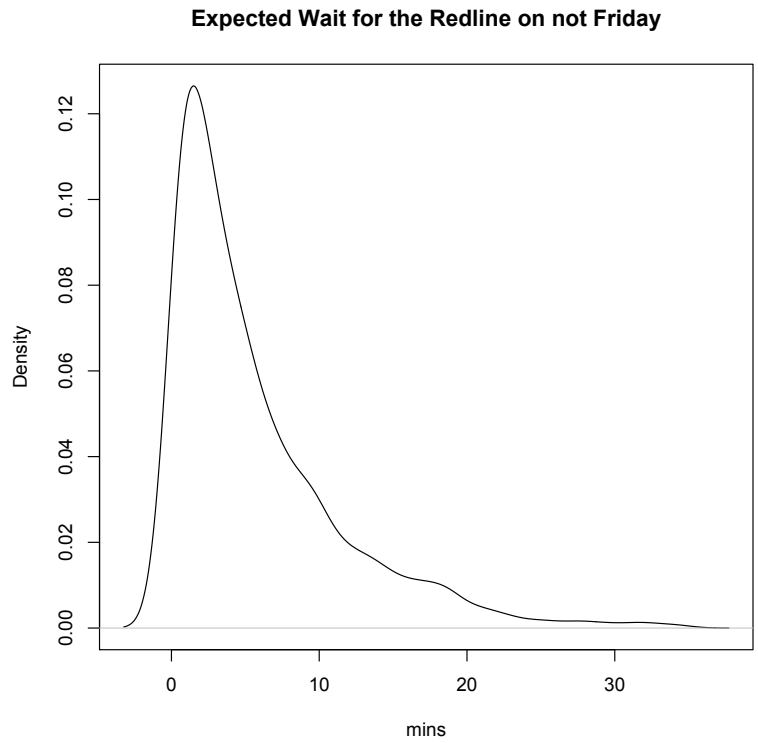
```
simyFriday <- apply(simbetas,1,function(x)
      rexp(n=1, rate=1/exp(predCovs%*%x)))
plot(density(simyFriday), main= "Expected Wait for the
      Redline on Friday", xlab="mins")
```

- I would do the same for not Friday.

```
predCovs <- c(1,0, mean(minsSch))
simy <- apply(simbetas,1,function(x)
       rexp(n=1, rate=1/exp(predCovs%*%x)))
plot(density(simy), main= "Expected Wait for
        the Redline not Friday", xlab="mins")
```

We Can Make Predictions





Outline

1 The Likelihood Model with Covariates

2 Likelihood Ratio Test

3 The Central Limit Theorem and the MLE

4 What We Can Do with a Normal MLE

5 Efficiency

Mean Squared Error and Efficiency

For estimator $\hat{\theta}$ and true parameter θ_0,

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - \theta_0)^2] \\ &= Var(\hat{\theta}) + (Bias(\hat{\theta}, \theta_0)^2) \end{aligned}$$

This is the *bias-variance tradeoff*

When two estimates $\hat{\theta}$ and $\tilde{\theta}$ are unbiased, we can compare their *efficiency* by using

$$eff(\hat{\theta}, \tilde{\theta}) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$$

Efficiency and the Cramer-Rao Inequality

So basically, for unbiased estimators, we want parameters with low variance.

Cramer-Rao Inequality Let $X_1 \dots X_n$ be i.i.d. with density function $f(x|\theta)$. Let $T = t(X_1 \dots X_n)$ be an unbiased estimate of θ. Then, under smoothness assumptions of $f(x|\theta)$.

$$Var(T) \geq \frac{1}{nI_n(\theta)} = \frac{1}{I(\theta)}$$

$\Rightarrow$Among unbiased estimators, the MLE has the smallest asymptotic variance The MLE is thus said to be *asymptotically efficient*.