

Multiple Linear Regression

Inference for the MLE

Math 392

Mathematical tools

Chain Rule

$$(f(g(x)))' = f'(g(x))g'(x)$$

Product Rule

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

Taylor Expansion

$$f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots$$

Running example

Let

$$X \sim \text{Pois}(\theta)$$

$$f(x|\theta) = \frac{1}{x!} \theta^x e^{-\theta}, \quad x \geq 0, \quad \theta > 0, \quad \hat{\theta}^{MLE} = \bar{x}$$

The likelihood function L :

$$L(\theta) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta)$$

The log-likelihood function l :

$$l(\theta) = \log(f(x_1|\theta)) + \log(f(x_2|\theta)) + \dots + \log(f(x_n|\theta))$$

The $\hat{\theta}^{MLE}$ is the solution to setting $\frac{\partial}{\partial \theta} \log(f(\mathbf{x}|\theta))$ to 0.

Score function

Define

$$U = \frac{\partial}{\partial \theta} \log(f(X|\theta))$$

as the *score function* or *score statistic*.

Note:

- θ is fixed
- X is a single RV
- therefore U is a RV

Understanding U

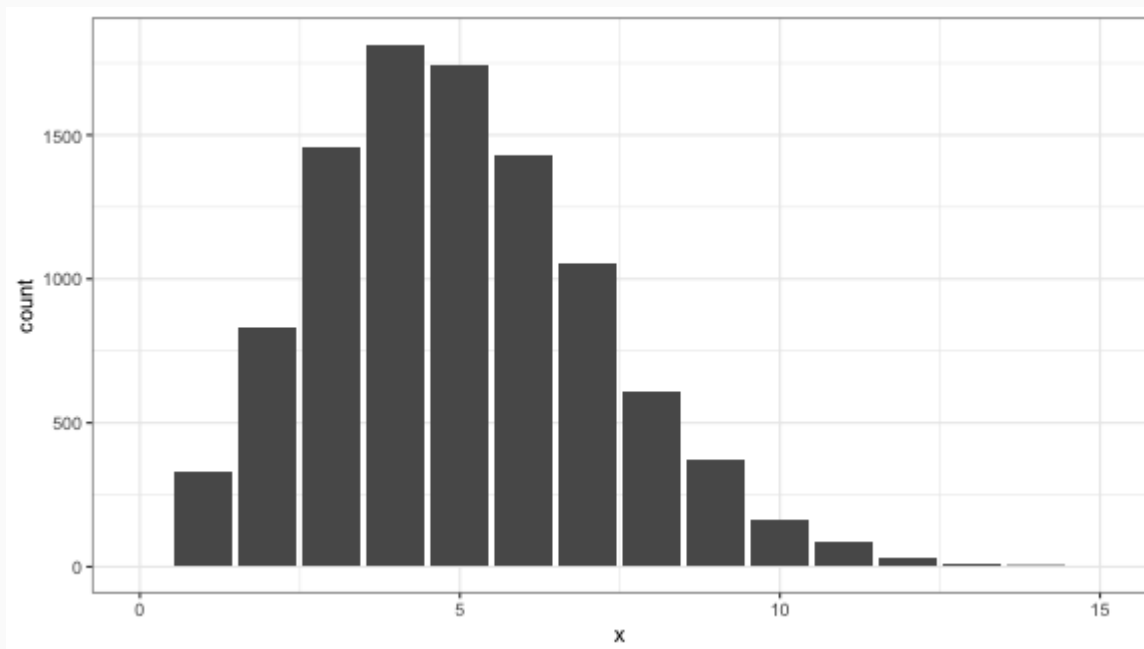
What is its *expected value*?

$$\begin{aligned} E(U) &= E\left(\frac{\partial}{\partial\theta}\log(f(x|\theta))\right) \\ &= \int \frac{\partial}{\partial\theta}\log(f(x|\theta))f(x|\theta_0)\mathrm{d}x \\ &= \int \frac{\partial}{\partial\theta}f(x|\theta)\frac{1}{f(x|\theta)}f(x|\theta_0)\mathrm{d}x \\ &= \int \frac{\partial}{\partial\theta}f(x|\theta)\mathrm{d}x \\ &= \frac{\partial}{\partial\theta}\int f(x|\theta)\mathrm{d}x \\ &= \frac{\partial}{\partial\theta}1 \\ &= 0 \end{aligned}$$

Poisson RV

Let $X \sim \text{Poi}(\theta_0)$.

```
n <- 10000  
theta <- 5  
x <- rpois(n, theta)
```



Score RV

Define $U = \frac{\partial}{\partial \theta} \log f(X|\theta)$.

$$\begin{aligned} U &= \frac{\partial}{\partial \theta} \log \frac{1}{x!} \theta^x e^{-\theta} \\ &= \frac{\partial}{\partial \theta} (-\log(x!) + x \log(\theta) - \theta) \\ &= 0 + x \frac{1}{\theta} - 1 \end{aligned}$$

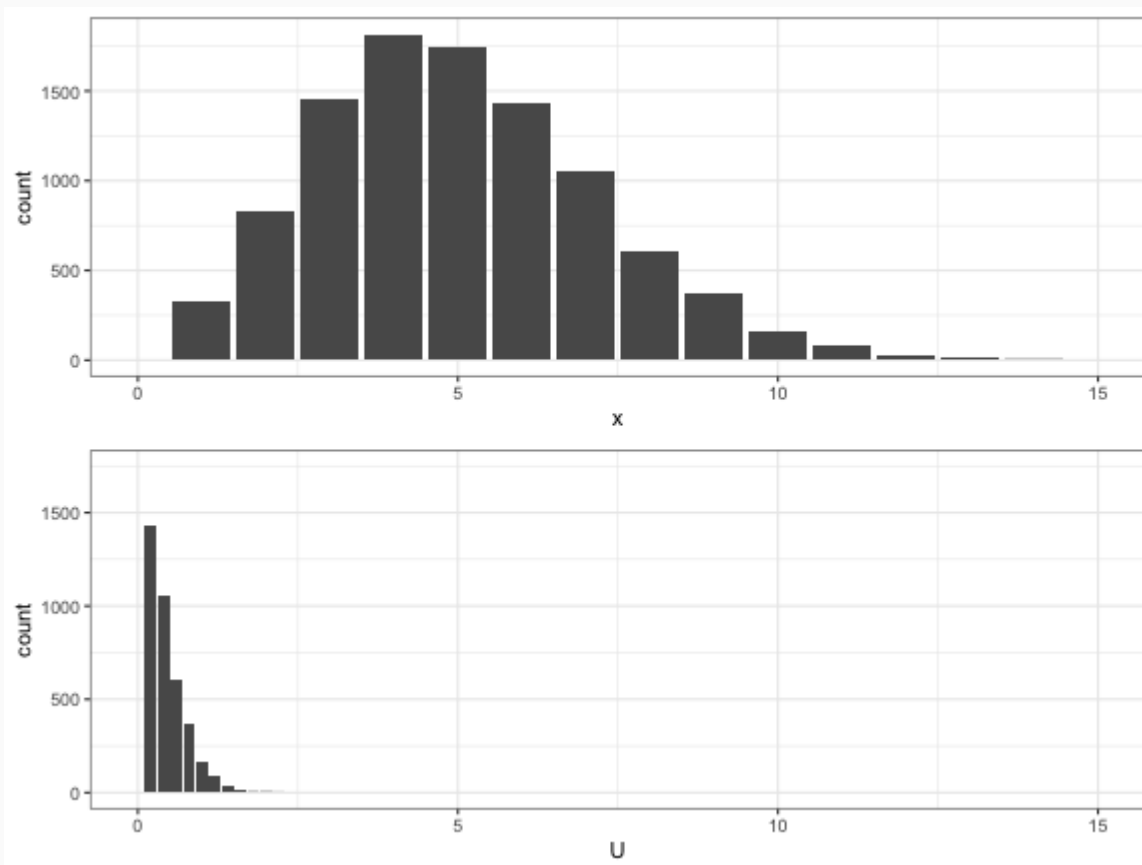
When is $E(U) = E(X \frac{1}{\theta} - 1) = 0$?

$$\begin{aligned} E(X \frac{1}{\theta} - 1) &= E(X) \frac{1}{\theta} - 1 \\ &= \theta_0 \frac{1}{\theta} - 1 \end{aligned}$$

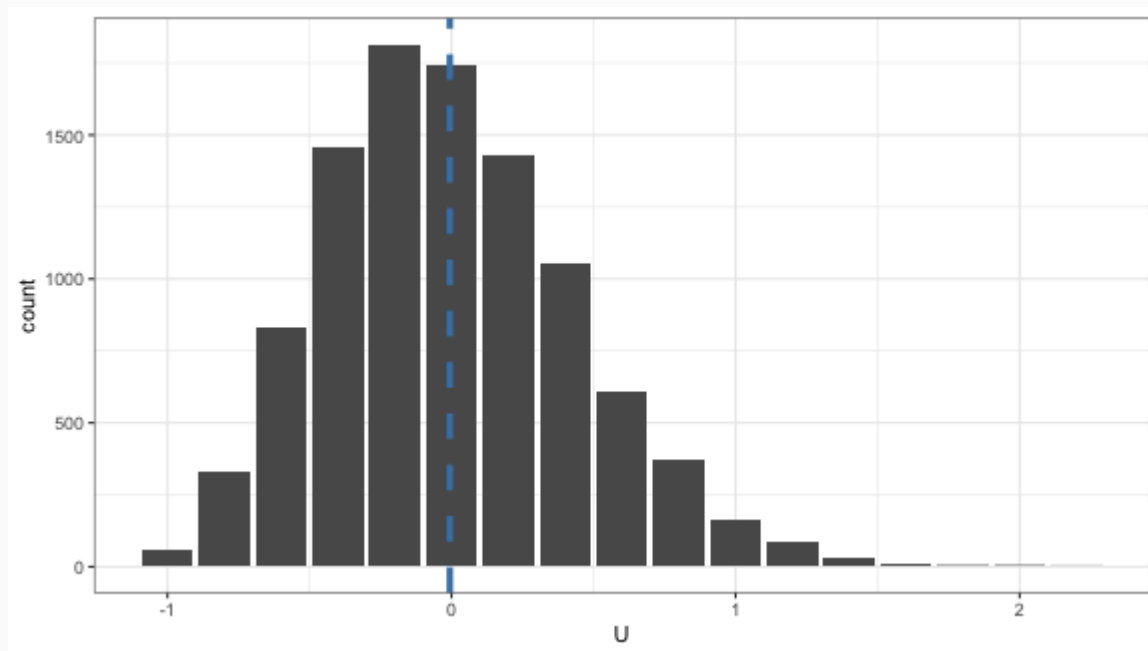
When we use θ_0 in our density.

Distribution of X vs U

```
U <- x / theta - 1
```



Distribution of U



Finding the variance of U

Recall $Var(U) = E(U^2) - E(U)^2$, so we seek $E(U^2)$. Begin by writing down our previous result, that the expected value is zero, and take derivatives of both sides.

$$\begin{aligned}\frac{\partial}{\partial \theta} 0 &= \frac{\partial}{\partial \theta} \int \frac{\partial}{\partial \theta} \log(f(x|\theta)) f(x|\theta_0) dx \\ 0 &= \int \frac{\partial^2}{\partial \theta^2} \log(f(x|\theta)) f(x|\theta_0) dx + \int \frac{\partial}{\partial \theta} \log(f(x|\theta)) \frac{\partial}{\partial \theta} f(x|\theta_0) dx \\ 0 &= E \left(\frac{\partial^2}{\partial \theta^2} \log(f(x|\theta)) \right) + \int \frac{\partial}{\partial \theta} \log(f(x|\theta)) \frac{\partial}{\partial \theta} f(x|\theta_0) \frac{1}{f(x|\theta)} f(x|\theta) dx \\ 0 &= E \left(\frac{\partial^2}{\partial \theta^2} \log(f(x|\theta)) \right) + \int \frac{\partial}{\partial \theta} \log(f(x|\theta_0)) \frac{\partial}{\partial \theta} \log(f(x|\theta_0)) f(x|\theta_0) dx \\ &\rightarrow Var(U) = E(U^2) = -E \left(\frac{\partial^2}{\partial \theta^2} \log(f(x|\theta)) \right) = I(\theta)\end{aligned}$$

Score variance

For $X \sim \text{Pois}(\theta)$,

$$\begin{aligned} \text{Var}(U) &= -E \left(\frac{\partial^2}{\partial \theta^2} \log(f(X|\theta)) \right) \\ &= -E \left(\frac{\partial}{\partial \theta} \left(x \frac{1}{\theta} - 1 \right) \right) \\ &= -E \left(-x \frac{1}{\theta^2} \right) = \frac{1}{\theta}. \end{aligned}$$

```
1/theta
```

```
## [1] 0.2
```

```
var(U)
```

```
## [1] 0.1961535
```

A CLT for \bar{U}

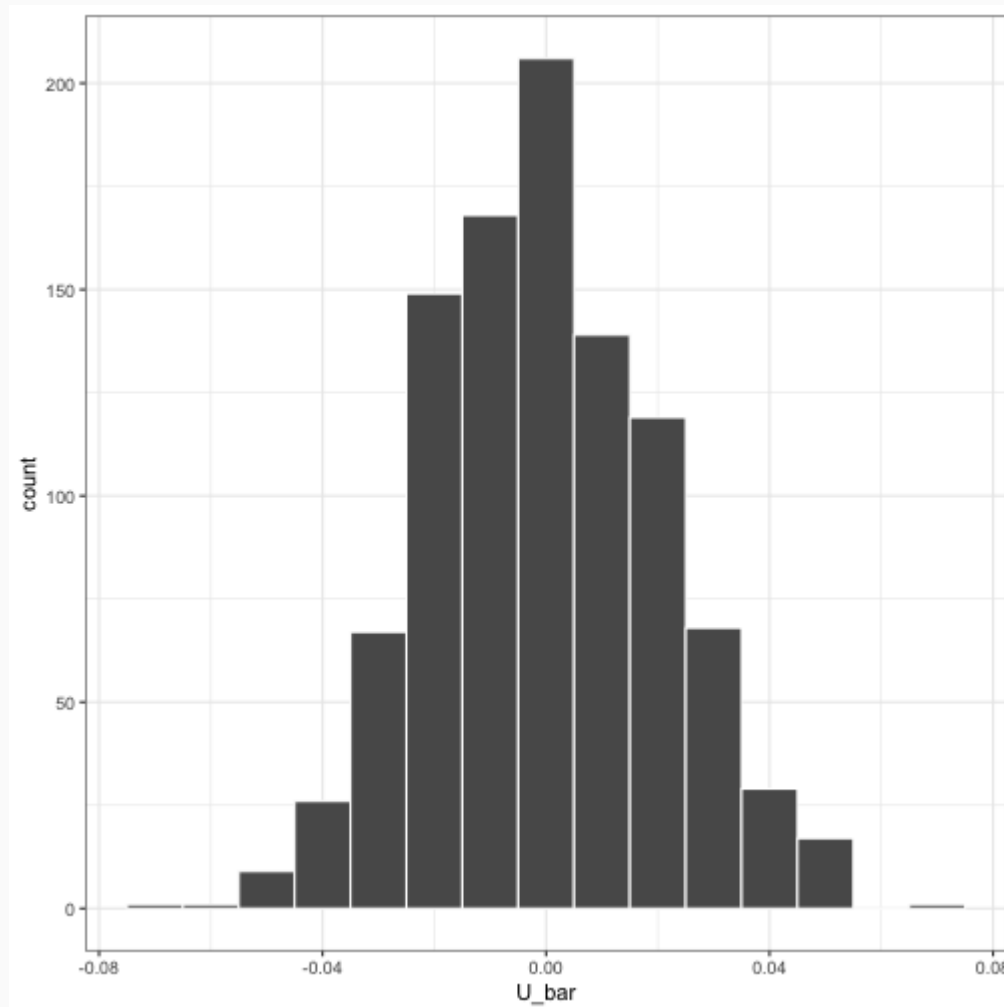
$U(X)$ is a function of a single RV. If we have an iid sample of size n - X_1, X_2, \dots, X_n - we have a corresponding iid sample $U(X_1), U(X_2), \dots, U(X_n)$, each with mean 0 and variance $I(\theta)$, by the CLT,

$$\frac{\frac{1}{n} \sum_{i=1}^n U_i - 0}{\sqrt{I(\theta)/n}} \xrightarrow{D} N(0, 1)$$

CLT for \bar{U}

```
n <- 500
theta <- 5
it <- 1000
U_bar <- rep(NA, it)
for (i in 1:it) {
  x <- rpois(n, theta)
  U <- x / theta - 1
  U_bar[i] <- mean(U)
}
```

CLT for \bar{U}



CLT for \bar{U}

```
var(U_bar)
```

```
## [1] 0.0004181934
```

```
(1 / theta) / n
```

```
## [1] 4e-04
```


The Asymptotic Normality of the MLE

Theorem: Let X_1, X_2, \dots, X_n be an iid sample from a regular family with parameter θ . Let $\hat{\theta}^{MLE}$ be the solution to the equation

$$\frac{\partial}{\partial \theta} \log(f(x_1, x_2, \dots, x_n | \theta)) = 0$$

then

$$\sqrt{nI(\theta)}(\hat{\theta}^{MLE} - \theta) \xrightarrow{D} N(0, 1)$$

Proof: Consider $U_i = \frac{\partial}{\partial \theta} \log(f(X_i|\theta))$ as a function of both X_i and θ . If we sum the U_i and expand around the true value:

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(x_i|\hat{\theta})) \approx \sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(x_i|\theta)) + \left[\sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log(f(x_i|\theta)) \right] (\hat{\theta} - \theta)$$

Since $\hat{\theta}$ is the MLE, the term on the left is 0. Now we can write this as a function of U_i and rearrange:

$$\begin{aligned} \sum_{i=1}^n U_i &= - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log(f(x_i|\theta)) (\hat{\theta} - \theta) \\ \frac{\frac{1}{n} \sum_{i=1}^n U_i}{\sqrt{I(\theta)/n}} &= \frac{-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log(f(x_i|\theta))}{\sqrt{I(\theta)/n}} (\hat{\theta} - \theta) \end{aligned}$$

We recognize the LHS as an RV that converges to the standard normal, so what is the RHS?

By the LLN,

$$-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log(f(x_i|\theta)) \rightarrow I(\theta),$$

so we can rewrite the RHS as

$$\frac{I(\theta)}{(I(\theta)/n)^{1/2}}(\hat{\theta} - \theta) = \sqrt{nI(\theta)}(\hat{\theta} - \theta) \xrightarrow{D} N(0, 1)$$

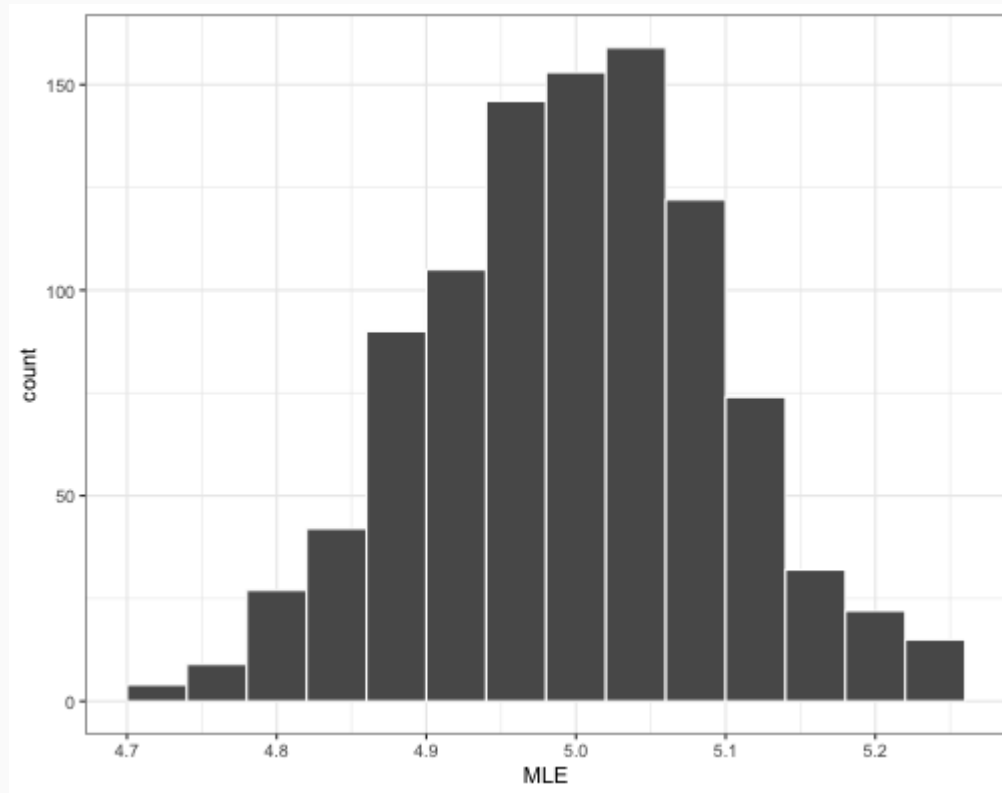
or

$$\hat{\theta} \xrightarrow{D} N\left(\theta, \frac{1}{\sqrt{nI(\theta)}}\right)$$

CLT for $\hat{\theta}^{MLE}$

```
it <- 1000
MLE <- rep(NA, it)
for (i in 1:it) {
  n <- 500
  theta <- 5
  x <- rpois(n, theta)
  MLE[i] <- mean(x)
}
```

CLT for $\hat{\theta}^{MLE}$



CLT for $\hat{\theta}^{MLE}$

```
z <- sqrt(n * (1 / theta)) * (MLE - theta)
```

