Multiple Linear Regression

Inference for the MLE

Math 392

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:

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

The log-likelihood function l:

$$l(\theta) = \log(f(x_1|\theta)) + \log(f(x_1|\theta)) + \ldots + \log(f(x_1|\theta))$$

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

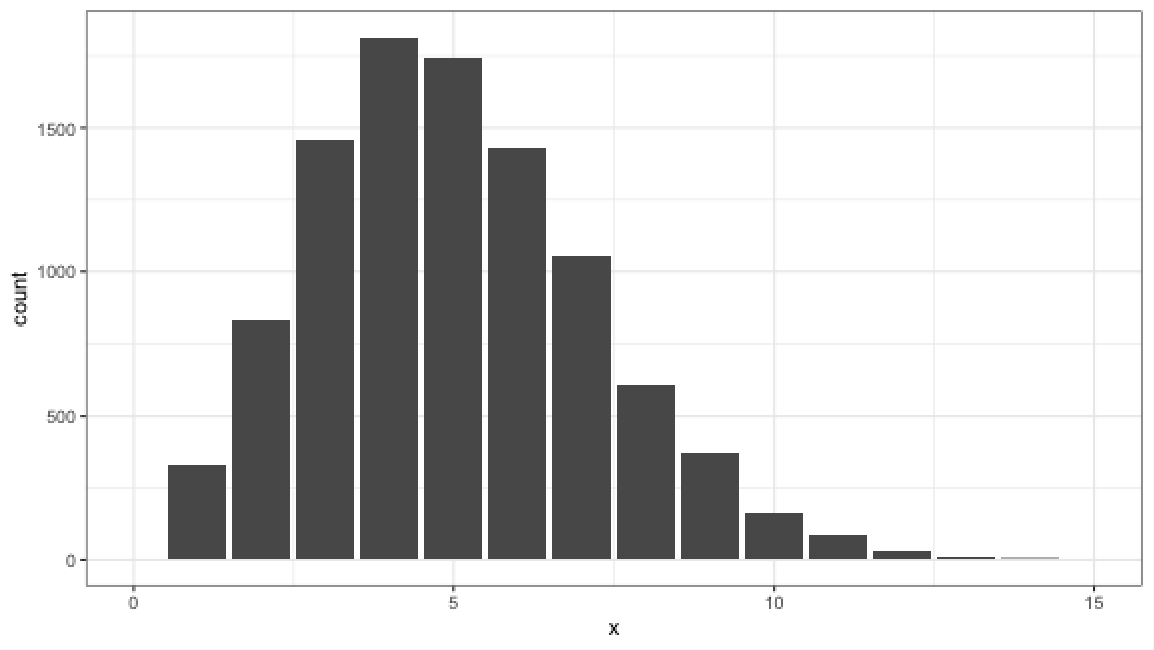
What is its *expected value*?

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

Poisson RV

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

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



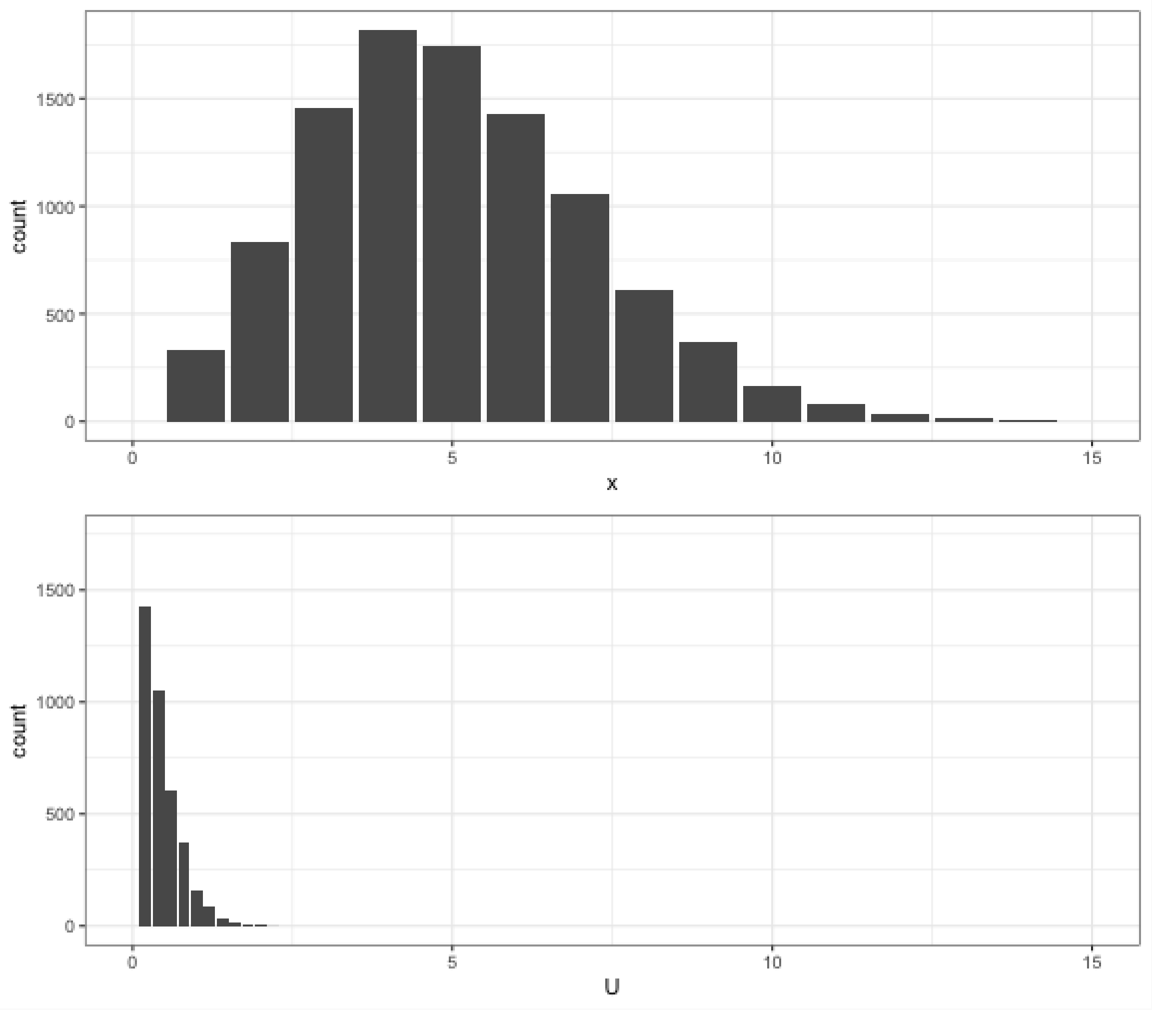
Score RV

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

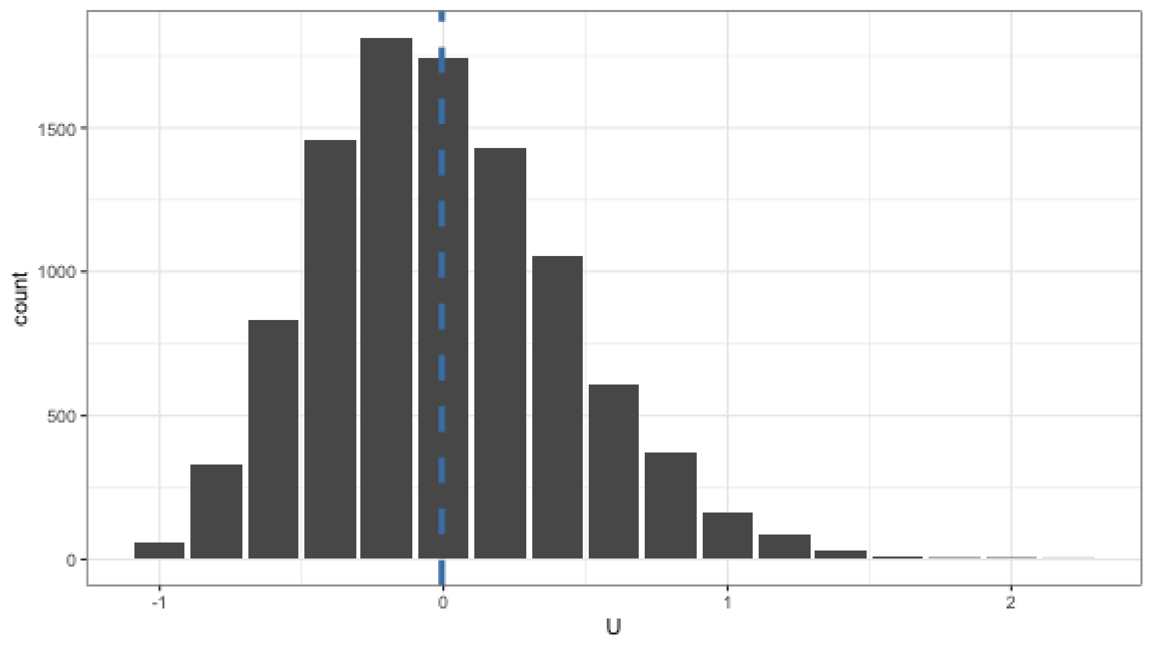
$$
\begin{aligned}
U &= \frac{\partial}{\partial\theta} \log \frac{1}{x!} \theta^x e^{-\theta} \\
&= \frac{\partial}{\partial\theta} (-\log(x!) + x \log(\theta) - \theta) \\
&= 0 + x \frac{1}{\theta} - 1
\end{aligned}
$$

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

Distribution of X vs U



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.

$$\begin{aligned}
\frac{\partial}{\partial\theta}0 &= \frac{\partial}{\partial\theta}\int \frac{\partial}{\partial\theta}\log(f(x|\theta))f(x|\theta)\mathrm{d}x \\
0 &= \int \frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))f(x|\theta)\mathrm{d}x + \int \frac{\partial}{\partial\theta}\log(f(x|\theta))\frac{\partial}{\partial\theta}f(x|\theta)\mathrm{d}x \\
0 &= E\left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right) + \int \frac{\partial}{\partial\theta}\log(f(x|\theta))\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}f(x|\theta)\mathrm{d}x \\
0 &= E\left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right) + \int \frac{\partial}{\partial\theta}\log(f(x|\theta))\frac{\partial}{\partial\theta}\log(f(x|\theta))f(x|\theta)\mathrm{d}x
\end{aligned}$$

$$\rightarrow Var(U) = E(U^2) = -E\left(\frac{\partial^2}{\partial\theta^2}\log(f(x|\theta))\right) = I(\theta)$$

Score variance

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

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

```
1/theta
```

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

```
var(U)
```

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

A CLT for $\bar{U}$

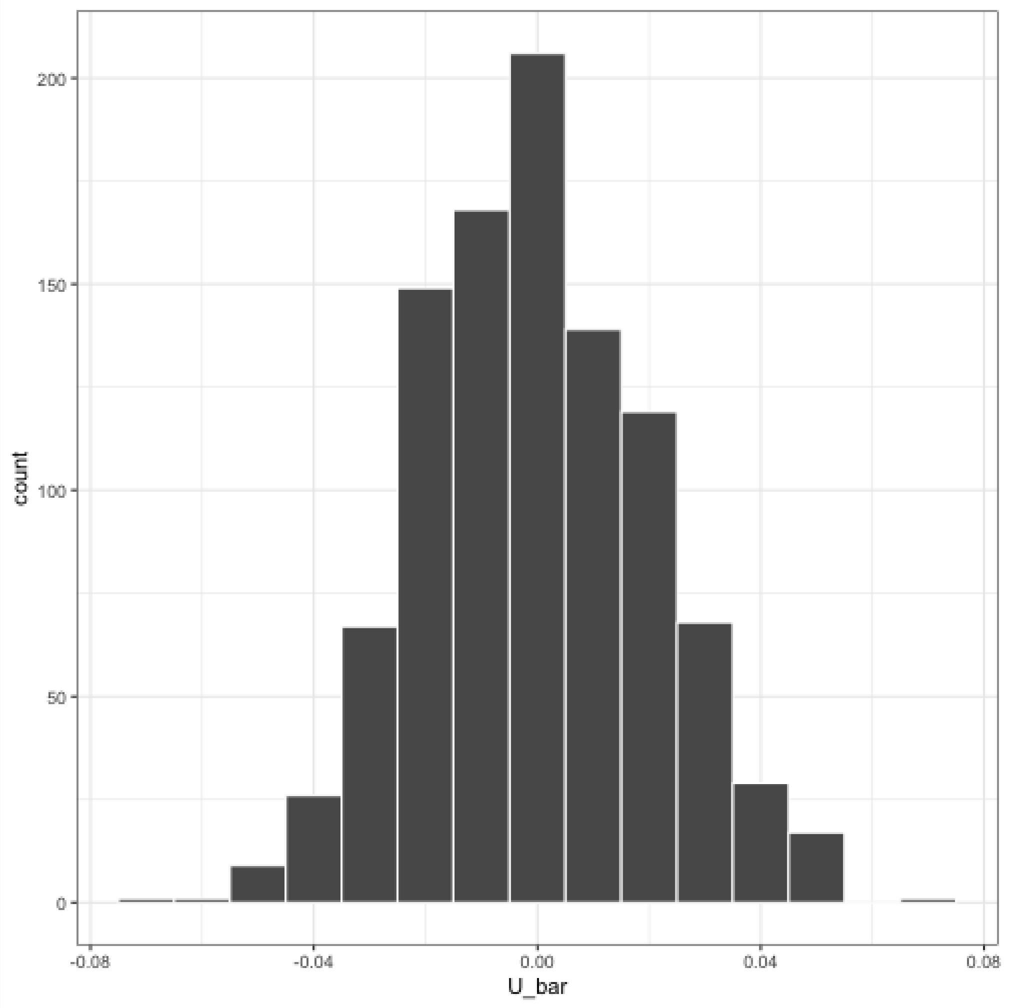
$U(X)$ is a function of a single RV. If we have an iid sample of size n - $X_1, X_2, \ldots, X_n$ - we have a corresponding iid sample $U(X_1), U(X_2), \ldots, U(X_n)$, each with mean 0 and variance $I(\theta)$, by the CLT,

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

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 $\bar{U}$

```r
var(U_bar)
```

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

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

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

The Asymptotic Normality of the MLE

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

$$\frac{\partial}{\partial\theta}\log(f(x1, x2, \ldots, x_n|\theta)) = 0$$

then

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

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:

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

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

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

By the LLN,

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

so we can rewrite the RHS as

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

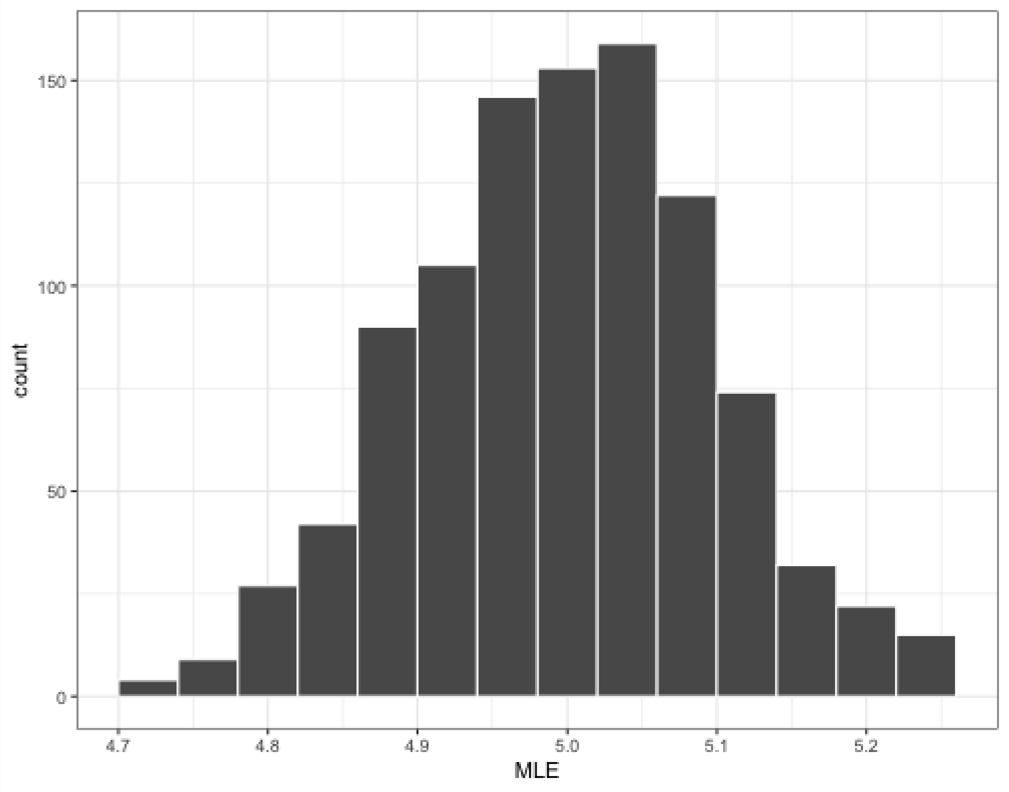
or

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

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

```r
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)
```