Lecture 16: Score and Wald Tests
Lecture Outline

• Score Tests
• Wald Tests
• Comparison of the 3 Likelihood Tests
The Score Function

• Let $L(\theta)$ be the likelihood, and $l(\theta) = \log[L(\theta)]$ be the log-likelihood.

• In order to calculate the MLEs, what do we need to calculate?

• The score function, $U(\theta)$, is the derivative of the log-likelihood w.r.t. $\theta$:

$$U(\theta) = \frac{\partial}{\partial \theta} \log[L(\theta)]$$

• We’ve been using the score function all along, we just never named it as such.
Distribution of the Score Function

• What is the expectation of the score function, $U(\theta)$? Why?

$$E_x[U(\theta)] = \int \frac{\partial}{\partial \theta} \log[L(\theta)] f(x \mid \theta) dx = \int \left\{ \frac{\partial}{\partial \theta} \log[f(x \mid \theta)] \right\} f(x \mid \theta) dx$$

$$= \int \left\{ \frac{\partial f(x \mid \theta)}{\partial \theta} / f(x \mid \theta) \right\} f(x \mid \theta) dx = \int \frac{\partial f(x \mid \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int f(x \mid \theta) dx = 0$$

• What is the true variance of the score function?

$$Var_x[U(\theta)] = E_x[U(\theta)^2] = E_x \left[ \left( \frac{\partial}{\partial \theta} \log[L(\theta)] \right)^2 \right] = I_n(\theta)$$

• What is the variance of the score function if $H_0$ is true? $I_n(\theta_0)$

• Is it reasonable to assume that the score function, as a random variable, converges to a normal distribution? Why?
The Score Test

• To test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$

• The score test, $S(\theta_0)$, is then:

$$S(\theta_0) = \frac{[U(\theta_0)]^2}{I_n(\theta_0)}$$

• When the null hypothesis is true, $S(\theta_0) \rightarrow \chi^2_{df = 1}$.

• What is $I_n(\theta_0)$? It’s the usual expected Fisher information, $I_n(\theta_0)$, with $\theta_0$ plugged in for $\theta$ in the result.

*Often, the score statistic is written as the square root of the formula above, and then it follows a standard normal dist.
Example: Score Test

• Let \( X_1, \ldots, X_n \sim Expo(\lambda) \). We’d like to test:
  \( H_0: \lambda = \lambda_0 \) vs. \( H_A: \lambda \neq \lambda_0 \).
• The likelihood function is then:
  \[
l(\lambda) = n \log(\lambda) - \lambda \sum X_i = n \log(\lambda) - n \lambda \bar{X}\]
• So the score function is:
  \[
  U(\lambda) = \frac{\partial l(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - n\bar{X}
  \]
• And fisher’s information is:
  \[
  I_n(\lambda) = -E\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2}\right) = -E\left(-\frac{n}{\lambda^2}\right) = \frac{n}{\lambda^2}
  \]
Example: Score Test

• \( H_0: \lambda = \lambda_0 \) vs. \( H_A: \lambda \neq \lambda_0 \). Let’s calculate the score stat:

\[
S(\lambda_0) = \frac{[U(\lambda_0)]^2}{I_n(\lambda_0)} = \left[ \frac{n}{\lambda} - n\bar{X} \right]^2 = \left[ \frac{n}{\lambda^2} \right]^{\lambda=\lambda_0} = \left[ n\left( \frac{1}{\lambda_0} - \bar{X} \right) \right]^2
\]

\[
= \frac{\lambda_0^2}{n} n^2 \left[ \frac{1}{\lambda_0} - \bar{X} \right]^2 = n \left[ \lambda_0 \left( \frac{1}{\lambda_0} - \bar{X} \right) \right]^2 = n(1 - \lambda_0 \bar{X})^2
\]

• Why is this a reasonable test statistic? What is the critical region given \( \alpha = 0.05 \)?
Example: Score Test in R

• The survival time for patients in a diet study was measured [it is reasonable to assume $X_i \sim Expo(\lambda)$]. Based on their age, the expected average survival should be 60 months. There were 1537 patients and the sample mean was 60.54 months.

• By hand, perform a score test to determine whether patients in our diet study had a significantly different average survival compared to what is expected based on their age.

• Use R to calculate an approximate 95% confidence interval for $\lambda$. 
Score Based Confidence Intervals

• Just like for likelihood ratio tests, in order to build a confidence interval from a score test, we will have to invert the test for various value of $\theta_0$ in the null hypothesis.

• Why can’t we just invert the formula for the score statistic $S(\theta_0)$?

• Just like the LRT, it’s not estimating $\theta$ directly.

• So we will have to recalculate the score statistic for varying values of $\theta_0$, and see if that value of $\theta_0$ is rejected or not based on the score statistic, $S(\theta_0)$. 
Example: Score Test in R

```r
# Read the data
f=file.choose()
data=read.csv(f,header=T)
attach(data)
xbar=mean(survtime)
n=length(survtime)
lambda0=1/60

score=n*(1-lambda0*xbar)^2
p.value=1-pchisq(score,df=1)
> score
[1] 0.1257059
> p.value
[1] 0.7229265

lambda0=1/(550:650/10)
scores=n*(1-lambda0*xbar)^2
> ci=lambda0[scores<qchisq(0.95,df=1)]
> c(min(ci),max(ci))
[1] 0.01569859 0.01733102
```
Score Test for Multiple Parameters

• To test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ for a $p$-dimensional restricted parameter vector, $\theta_0$.

• The score test, $S(\theta_0)$, is then:

$$S(\theta_0) = U^T(\hat{\theta}_0)\left[I_n^{-1}(\hat{\theta}_0)\right]U(\hat{\theta}_0)$$

• When the null hypothesis is true, $S(\theta_0) \rightarrow \chi^2_{df = p}$.

• What the heck is $\hat{\theta}_0$?

• It’s the maximum likelihood estimate of the other parameters in $\theta$ that are not restricted in the null hypothesis $\theta_0$. 
Lecture Outline

• Score Tests
• Wald Tests
• Comparison of the 3 Likelihood Tests
The Wald Test

- The 3rd likelihood based test is called the Wald test.
- The **Wald test statistic**, $W$, can be used to test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ and is calculated to be:

$$W = \frac{[\hat{\theta} - \theta_0]^2}{1 / I_n(\hat{\theta})} = I_n(\hat{\theta})[\hat{\theta} - \theta_0]^2$$

where $\hat{\theta}$ is the MLE, and $I_n(\hat{\theta})$ is the expected Fisher information evaluated at the MLE (the usual estimate we have been using for the confidence interval).

- When the null hypothesis is true, $W \rightarrow \chi^2_{df = 1}$.
- The Wald test is easy to calculate. It is different in that it does not calculate the standard error assuming the null hypothesis is true.
Wald-like Confidence Intervals

• Finally, we have a closed form solution to calculate a confidence interval…hooray!
• To calculate a Wald-like confidence interval, use the formula:

\[ \hat{\theta} \pm z^* \sqrt{\frac{1}{I(\hat{\theta})}} \]

• Guess what, it’s the formula we’ve been using all along ☺️
Example: Wald Test

• Let \( X_1, \ldots, X_n \sim \text{Expo}(\lambda) \). We’d like to test:
  \( H_0: \lambda = \lambda_0 \) vs. \( H_A: \lambda \neq \lambda_0 \).

• What is the MLE? Set the score function to zero and solve to get:
  \( \hat{\lambda} = \frac{1}{\bar{X}} \)

• Expected Fisher’s Information is
  \[ I_n(\lambda) = -E\left( \frac{\partial^2 l(\lambda)}{\partial \lambda^2} \right) = -E\left( -\frac{n}{\lambda^2} \right) = \frac{n}{\lambda^2} \]

• Evaluate this at the MLE and we get:
  \[ I_n(\hat{\lambda}) = \frac{n}{\hat{\lambda}^2} = \frac{n}{(1/\bar{X})^2} = n\bar{X}^2 \]
Example: Wald Test (cont.)

• \( H_0: \lambda = \lambda_0 \) vs. \( H_A: \lambda \neq \lambda_0 \). Let’s calculate the Wald stat:

\[
W = I(\hat{\lambda})[\hat{\lambda} - \lambda_0]^2 = n\bar{X}^2[1 / \bar{X} - \lambda_0]^2
\]

• Why is this a reasonable test statistic? What is the critical region given \( \alpha = 0.05 \)?

• Let’s calculate the Wald-based 95% confidence interval:

\[
\hat{\lambda} \pm z^* \sqrt{\frac{1}{I(\hat{\lambda})}} = \frac{1}{\bar{X}} \pm 1.96 \sqrt{\frac{1}{n\bar{X}^2}}
\]
Example: a Wald Test in R

```r
lambda0 = 1/60
wald = n * xbar^2 * (1/xbar - lambda0)^2
p.value = 1 - pchisq(wald, df = 1)
wald
p.value
c(1/xbar - 1.96 * sqrt(1/n/xbar^2),
   1/xbar + 1.96 * sqrt(1/n/xbar^2))

> wald
[1] 0.1257059
> p.value
[1] 0.7229265
> c(1/xbar - 1.96 * sqrt(1/n/xbar^2),
   + 1/xbar + 1.96 * sqrt(1/n/xbar^2))
[1] 0.01569152 0.01734306
```

- Why are the Wald and Score tests and intervals so similar?
The Wald Test

• What are the advantages and disadvantages to the Wald test?

**Big Advantage:**
Easy to calculate & Confidence Interval has a closed form

**BIG Disadvantage:**
Even in the test, the standard error is estimated under the MLE (the alternative hypothesis). This can lead to some screwy results (like when the sample proportion is estimated to be, or close to, 0 or 1).
Wald Test for Multiple Parameters

• To test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ for a $p$-dimensional restricted parameter vector, $\theta_0$.

• The *Wald test*, $W$, is then:

$$W = (\hat{\theta} - \theta_0)\left[I_n(\hat{\theta})\right](\hat{\theta} - \theta_0)$$

• When the null hypothesis is true, $W \rightarrow \chi^2_{df = p}$. 
Lecture Outline

• Score Tests
• Wald Tests
• Comparison of the 3 Likelihood Tests
LR vs. Score vs. Wald

• What’s the point of all these testing procedures?
• The 3 testing procedures we have covered are general methods to form test statistics based on the likelihood function in one way or another.
• These are all asymptotic results. Why can we use the normal distribution (or $\chi^2$ distribution) for these procedures?
• The Central Limit Theorem!
• What is the difference between these 3 types of tests? What are the assumptions for each? What are they all doing?
• A picture is worth a thousand words…
LR vs. Score vs. Wald
Take Home Message

- The **score function**, $U(\theta)$, is just the derivative of the log-likelihood function:

$$U(\theta) = \frac{\partial}{\partial \theta} \log[L(\theta)]$$

- The $\chi^2$ **Score Test** is calculated to be:

$$S(\theta_0) = \left[ U(\theta_0) \right]^2 \frac{I_n(\theta_0)}{n}$$

- The $\chi^2$ **Wald Test** can be used to test hypotheses.

$$W = I_n(\hat{\theta}) \left[ \hat{\theta} - \theta_0 \right]^2$$

- Score based CI’s are calculated by inverting the score test, and Wald-based CI’s have a nice closed form 😊

- All 3 types of tests (LR, Score, and Wald) are based on the likelihood function. They are all asymptotically equivalent.