Exercise 6: Statistical inference (III)

Jianhui Gao

July 16, 2025

Part 1: Wald, Score, and likelihood ratio test statistics

Write out the likelihood function, and derive the test statistics of the Wald, Score, and likelihood ratio test.

1. $X_i \overset{\text{i.i.d.}}{\sim} f(x \mid \theta)$

$$f(x \mid \theta) = \theta \exp(-x\theta)\mathbb{I}\{x > 0\}$$

Solution

The likelihood function, $L(\theta|\mathbf{x})$, is the joint density of the observed data $\mathbf{x} = (x_1, ..., x_n)$. Due to the i.i.d. assumption, it is the product of the individual PDFs:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta e^{-x_i\theta} = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$$

The log-likelihood function, $\ell(\theta|\mathbf{x})$ is:

$$\ell(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x}) = n\log(\theta) - \theta \sum_{i=1}^{n} x_i$$

To derive the test statistics, we first need the Maximum Likelihood Estimator (MLE), the Score function, and the Fisher Information.

The first derivative of the log-likelihood.

$$U(\theta) = \frac{\partial \ell}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

We find the MLE by setting the score function to zero and solving for θ.

$$\frac{n}{\hat{\theta}} - \sum_{i=1}^{n} x_i = 0 \implies \hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

The negative expected value of the second derivative of the log-likelihood.

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{n}{\theta^2}$$

Since the second derivative is not a function of the data, the expectation is straightforward:

$$I(\theta) = -E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = -E\left[-\frac{n}{\theta^2}\right] = \frac{n}{\theta^2}$$

All three statistics asymptotically follow a chi-squared distribution with 1 degree of freedom, χ_1^2, under the null hypothesis H_0. The Wald test measures the distance between the MLE ($\hat{\theta}$) and the hypothesized value (θ_0), scaled by the curvature of the likelihood function at the MLE. The formula is $W = (\hat{\theta} - \theta_0)^2 I(\hat{\theta})$.

$$
\begin{aligned}
I(\hat{\theta}) &= \frac{n}{\hat{\theta}^2} = \frac{n}{(1/\bar{x})^2} = n\bar{x}^2 \\
W &= \left(\frac{1}{\bar{x}} - \theta_0\right)^2 (n\bar{x}^2) \\
&= \left(\frac{1 - \theta_0\bar{x}}{\bar{x}}\right)^2 (n\bar{x}^2) \\
&= n(1 - \theta_0\bar{x})^2
\end{aligned}
$$

The Score test evaluates how close the slope of the log-likelihood function is to zero at the hypothesized value θ_0. The formula is $S = [U(\theta_0)]^2[I(\theta_0)]^{-1}$.

$$
\begin{aligned}
U(\theta_0) &= \frac{n}{\theta_0} - \sum x_i = \frac{n}{\theta_0} - n\bar{x} = n\left(\frac{1 - \theta_0\bar{x}}{\theta_0}\right) \\
I(\theta_0) &= \frac{n}{\theta_0^2} \\
S &= \left[n\left(\frac{1 - \theta_0\bar{x}}{\theta_0}\right)\right]^2 \left(\frac{n}{\theta_0^2}\right)^{-1} \\
&= \frac{n^2(1 - \theta_0\bar{x})^2}{\theta_0^2} \cdot \frac{\theta_0^2}{n} \\
&= n(1 - \theta_0\bar{x})^2
\end{aligned}
$$

The LR test compares the maximized value of the likelihood function under the null hypothesis, $L(\theta_0)$, with its globally maximized value, $L(\hat{\theta})$. The formula is $LR = 2[\ell(\hat{\theta}) - \ell(\theta_0)]$.

$$
\begin{aligned}
\ell(\hat{\theta}) &= n\log(\hat{\theta}) - \hat{\theta}\sum x_i = n\log\left(\frac{1}{\bar{x}}\right) - \left(\frac{1}{\bar{x}}\right)(n\bar{x}) = -n\log(\bar{x}) - n \\
\ell(\theta_0) &= n\log(\theta_0) - \theta_0\sum x_i = n\log(\theta_0) - n\theta_0\bar{x} \\
LR &= 2\left[(-n\log(\bar{x}) - n) - (n\log(\theta_0) - n\theta_0\bar{x})\right] \\
&= 2n\left[-\log(\bar{x}) - 1 - \log(\theta_0) + \theta_0\bar{x}\right] \\
&= 2n\left[\theta_0\bar{x} - 1 - (\log(\bar{x}) + \log(\theta_0))\right] \\
&= 2n[\theta_0\bar{x} - 1 - \log(\theta_0\bar{x})]
\end{aligned}
$$

Bonus: For the exponential distribution model we analyzed, the Wald and Score test statistics simplify to the identical algebraic formula, while the Likelihood Ratio (LR) statistic has a distinct form.

2. $X_i \overset{\text{i.i.d.}}{\sim} f(x \mid \theta)$

$$
f(x \mid \theta) = \theta c^\theta x^{-(\theta+1)} \mathbb{I}\{x > c\} \quad \text{(Pareto distribution)}
$$

where c is a known constant and θ is unknown.

Solution

The likelihood function, $L(\theta|\mathbf{x})$, is the product of the individual PDFs:

$$
L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \theta c^\theta x_i^{-(\theta+1)} = \theta^n c^{n\theta} \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}
$$

The log-likelihood, $\ell(\theta|\mathbf{x})$, is:

$$\begin{aligned}\ell(\theta|\mathbf{x}) &= \log\left(\theta^n c^{n\theta}\left(\prod_{i=1}^n x_i\right)^{-(\theta+1)}\right)\\ &= n\log(\theta) + n\theta\log(c) - (\theta+1)\sum_{i=1}^n \log(x_i)\\ &= n\log(\theta) + \theta\left(n\log(c) - \sum_{i=1}^n \log(x_i)\right) - \sum_{i=1}^n \log(x_i)\end{aligned}$$

The first derivative of the log-likelihood with respect to θ.

$$U(\theta) = \frac{\partial \ell}{\partial \theta} = \frac{n}{\theta} + n\log(c) - \sum_{i=1}^n \log(x_i)$$

Set the score function to zero to find the MLE.

$$\frac{n}{\hat{\theta}} + n\log(c) - \sum_{i=1}^n \log(x_i) = 0$$

$$\frac{n}{\hat{\theta}} = \sum_{i=1}^n \log(x_i) - n\log(c) = \sum_{i=1}^n (\log(x_i) - \log(c)) = \sum_{i=1}^n \log\left(\frac{x_i}{c}\right)$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \log(x_i/c)}$$

The negative expected value of the second derivative of the log-likelihood.

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{n}{\theta^2}$$

Since the second derivative is a constant with respect to the data, the Fisher Information is:

$$I(\theta) = -E\left[-\frac{n}{\theta^2}\right] = \frac{n}{\theta^2}$$

All three statistics asymptotically follow a χ_1^2 distribution under H_0. The Wald test is based on the distance between $\hat{\theta}$ and θ_0.

$$\begin{aligned}W &= (\hat{\theta} - \theta_0)^2 I(\hat{\theta})\\ &= (\hat{\theta} - \theta_0)^2\left(\frac{n}{\hat{\theta}^2}\right)\\ &= n\left(\frac{\hat{\theta} - \theta_0}{\hat{\theta}}\right)^2 = n\left(1 - \frac{\theta_0}{\hat{\theta}}\right)^2\end{aligned}$$

The Score test is based on the slope of the log-likelihood at θ_0.

$$\begin{aligned}U(\theta_0) &= \frac{n}{\theta_0} + n\log(c) - \sum \log(x_i) = \frac{n}{\theta_0} - \sum \log(x_i/c) = \frac{n}{\theta_0} - \frac{n}{\hat{\theta}}\\ S &= [U(\theta_0)]^2 [I(\theta_0)]^{-1}\\ &= \left[n\left(\frac{1}{\theta_0} - \frac{1}{\hat{\theta}}\right)\right]^2 \left(\frac{n}{\theta_0^2}\right)^{-1}\\ &= n^2 \frac{(\hat{\theta} - \theta_0)^2}{\theta_0^2 \hat{\theta}^2} \frac{\theta_0^2}{n} = n\frac{(\hat{\theta} - \theta_0)^2}{\hat{\theta}^2} = n\left(1 - \frac{\theta_0}{\hat{\theta}}\right)^2\end{aligned}$$

The LR test compares the log-likelihood values at $\hat{\theta}$ and θ_0.

$$\begin{aligned}
\ell(\theta) &= n\log\theta - \theta\frac{n}{\hat{\theta}} - \sum\log(x_i) \\
\ell(\hat{\theta}) &= n\log(\hat{\theta}) - \hat{\theta}\frac{n}{\hat{\theta}} - \sum\log(x_i) = n\log(\hat{\theta}) - n - \sum\log(x_i) \\
\ell(\theta_0) &= n\log(\theta_0) - \theta_0\frac{n}{\hat{\theta}} - \sum\log(x_i) \\
\ell(\hat{\theta}) - \ell(\theta_0) &= (n\log\hat{\theta} - n) - (n\log\theta_0 - n\frac{\theta_0}{\hat{\theta}}) \\
&= n\left[\log\left(\frac{\hat{\theta}}{\theta_0}\right) - 1 + \frac{\theta_0}{\hat{\theta}}\right] \\
LR = 2[\ell(\hat{\theta}) - \ell(\theta_0)] &= 2n\left[\log\left(\frac{\hat{\theta}}{\theta_0}\right) + \frac{\theta_0}{\hat{\theta}} - 1\right]
\end{aligned}$$

Part 2: Test equivalence

Let θ be a scalar parameter and suppose we test

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Let W be the Wald test statistic and let λ be the likelihood ratio test statistic. Show that these tests are equivalent in the sense that

$$\frac{W}{\lambda} \xrightarrow{\text{P}} 1$$

as $n \to \infty$. Hint: Use a Taylor expansion of the log-likelihood $\ell(\theta)$ to show that

$$\lambda \approx \left(\sqrt{n}\left(\widehat{\theta} - \theta_0\right)\right)^2 \left(-\frac{1}{n}\ell''(\widehat{\theta})\right)$$

Solution

The **Likelihood Ratio (LR) statistic**, λ, is:

$$\lambda = 2(\ell(\hat{\theta}) - \ell(\theta_0))$$

The **Wald statistic**, W, using the observed Fisher information, is:

$$W = (\hat{\theta} - \theta_0)^2 \left(-\ell''(\hat{\theta})\right)$$

where $\ell(\theta)$ is the log-likelihood, $\ell''(\theta)$ is its second derivative, and $\hat{\theta}$ is the MLE.

We perform a second-order Taylor expansion of $\ell(\theta_0)$ around the MLE $\hat{\theta}$, including the remainder term in **little-o notation**:

$$\ell(\theta_0) = \ell(\hat{\theta}) + \ell'(\hat{\theta})(\theta_0 - \hat{\theta}) + \frac{1}{2}\ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2 + o_p\left((\hat{\theta} - \theta_0)^2\right)$$

The term $o_p\left((\hat{\theta} - \theta_0)^2\right)$ is a random quantity that converges to zero in probability faster than $(\hat{\theta} - \theta_0)^2$. By definition of the MLE, $\ell'(\hat{\theta}) = 0$, which simplifies the expansion to:

$$\ell(\theta_0) = \ell(\hat{\theta}) + \frac{1}{2}\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2 + o_p\left((\hat{\theta} - \theta_0)^2\right)$$

Rearranging the equation above allows us to express λ explicitly.

$$\ell(\hat{\theta}) - \ell(\theta_0) = -\frac{1}{2}\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2 - o_p\left((\hat{\theta} - \theta_0)^2\right)$$

Multiplying by 2 gives the expression for λ:

$$\lambda = 2(\ell(\hat{\theta}) - \ell(\theta_0)) = (\hat{\theta} - \theta_0)^2\left(-\ell''(\hat{\theta})\right) - 2 \cdot o_p\left((\hat{\theta} - \theta_0)^2\right)$$
$$\lambda = W + o_p\left((\hat{\theta} - \theta_0)^2\right)$$

Finally, we analyze the ratio $\frac{W}{\lambda}$ as $n \to \infty$.

$$\frac{W}{\lambda} = \frac{W}{W + o_p\left((\hat{\theta} - \theta_0)^2\right)}$$

Divide the numerator and denominator by W:

$$\frac{W}{\lambda} = \frac{1}{1 + \frac{o_p\left((\hat{\theta} - \theta_0)^2\right)}{W}}$$

We examine the remainder term in the denominator:

$$\frac{o_p\left((\hat{\theta} - \theta_0)^2\right)}{W} = \frac{o_p\left((\hat{\theta} - \theta_0)^2\right)}{(\hat{\theta} - \theta_0)^2\left(-\ell''(\hat{\theta})\right)}$$

By standard MLE theory, $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} I(\theta_0)$, where $I(\theta_0)$ is a positive constant. This implies $-\ell''(\hat{\theta}) = O_p(n)$. Therefore, the term simplifies to:

$$\frac{o_p(1)}{-\ell''(\hat{\theta})} = \frac{o_p(1)}{O_p(n)} \xrightarrow{P} 0$$

The remainder term vanishes as $n \to \infty$. The denominator of our ratio converges to 1:

$$1 + \frac{o_p\left((\hat{\theta} - \theta_0)^2\right)}{W} \xrightarrow{P} 1 + 0 = 1$$

This explicitly shows that:

$$\frac{W}{\lambda} \xrightarrow{\text{P}} 1$$

Justification: Convergence to Fisher Information

The key result that $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} I(\theta_0)$ is established in two main steps:

1. **Law of Large Numbers:** The term $-\frac{1}{n}\ell''(\theta_0)$, evaluated at the true parameter θ_0, is an average of i.i.d. random variables. By the Law of Large Numbers, this average converges in probability to its expectation, which is the definition of the Fisher Information, $I(\theta_0)$.

$$-\frac{1}{n}\ell''(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta_0) \xrightarrow{P} -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta_0)\right] = I(\theta_0)$$

2. **Consistency of the MLE:** The MLE is a consistent estimator, meaning $\hat{\theta} \xrightarrow{P} \theta_0$. Because the function is continuous, the **Continuous Mapping Theorem** allows us to substitute the consistent estimator $\hat{\theta}$ for the true parameter θ_0 in the limit, and the convergence result still holds.

This confirms that $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} I(\theta_0)$, which is a positive constant.