Lecture 19
Likelihood Ratio Test

Chao Song

College of Ecology

Lanzhou University

December 2, 2024

Power of a hypothesis test

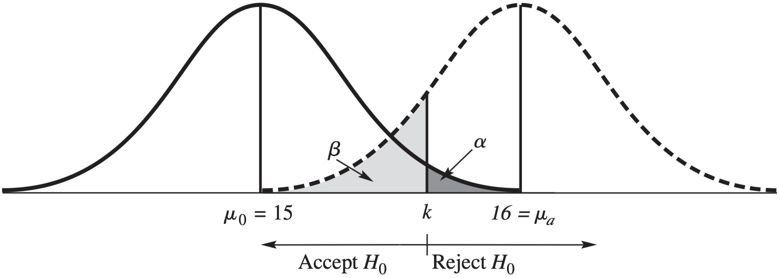
Recall the two types of errors associated with a hypothesis test.

	Not reject H_0	Reject H_0
H_0 is true	Correct	Type I error (α)
H_0 is false	Type II error (β)	Correct

While we control rate of the type I error by setting the significance level α, we typically do not know the rate of type II error. Ideally, we want to minimize type II error β. Or equivalently, we want to maximize the probability of rejecting H_0 when it is not true, $1-\beta$, which we call the **power of a test**.

Power of a hypothesis test

The power of a statistical test typically depends on the significance level, the true value of the parameter, and the sample size.



Graphic illustration of type I, type II errors and statistical power.

Power of a hypothesis test

Example: Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from $N(\mu, \sigma^2)$. When σ^2 is unknown, we use a t-test to test the hypothesis $\mathrm{H_0}$: $\mu = 0$ against $\mathrm{H_a}$: $\mu \neq 0$. What is the power of this test?

Using a t-test, we reject $\mathrm{H_0}$ at the α level if

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \geqslant t_{\frac{\alpha}{2}}(n-1) \quad or \quad T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \leqslant -t_{\frac{\alpha}{2}}(n-1)$$

That is, we reject $\mathrm{H_0}$ if

$$\overline{X} \geqslant \mu_0 + t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \quad or \quad \overline{X} \leqslant \mu_0 - t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}$$

Power of a hypothesis test

The power of the test is

$$P\left(\overline{X} \geqslant \mu_0 + t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \text{ or } \overline{X} \leqslant \mu_0 - t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \,\middle|\, \mu\right)$$

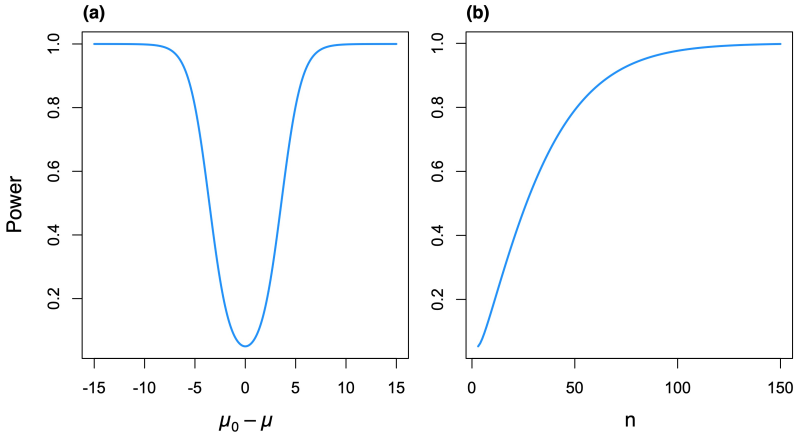
Note that when $\mu \neq \mu_0$, $T = \frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$, the power of the test is

$$P\left(T \geqslant \frac{\mu_0 - \mu}{s/\sqrt{n}} + t_{\frac{\alpha}{2}}(n-1) \text{ or } T \leqslant \frac{\mu_0 - \mu}{s/\sqrt{n}} - t_{\frac{\alpha}{2}}(n-1)\right),$$

the value of which depends on μ, n, and α.

Power of a hypothesis test

Power of a one sample t-test when (a) sample size $n = 10$ and (b) $\mu_0 - \mu = 2$.
For both panels, $\alpha = 0.05$ and the sample standard deviation $s = 2$.



Best critical region

For a particular hypothesis test H_0: $\theta = \theta_0$, we define a critical region C of size α as $P(C|\theta_0) = \alpha$. What is the best way to define such a critical region?

Recall the two type of errors associated with hypothesis testing. The significance level α determines the rate of type I error. Thus, for a critical region with pre-specified α, we want to minimize type II error.

A critical region of size α for H_0: $\theta = \theta_0$ is the **best critical region** if, for every other critical region D of size α, we have

$$P(C|\theta = \theta_1) \geqslant P(D|\theta = \theta_1)$$

where $\theta_1 \neq \theta$. That is, when H_0 is not true, the probability of rejecting H_0 with the use of critical region C is at least as great as the corresponding probability with the use of any other critical region D of the same size α.

Best critical region

Neyman–Pearson Lemma: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with PDF or PMF $f(x|\theta)$, where θ_0 and θ_1 are two possible values of θ. Let $L(\theta)$ be the likelihood function, ie.,

$$L(\theta) = f(X_1|\theta)f(X_2|\theta)\cdots f(X_n|\theta).$$

If there exist a positive constant k and a region C such that

- $P\big[(X_1, X_2, \ldots, X_n) \in C|\theta_0\big] = \alpha$;
- $\dfrac{L(\theta_0)}{L(\theta_1)} \leqslant k \textit{ for } (X_1, X_2, \ldots, X_n) \in C$;
- $\dfrac{L(\theta_0)}{L(\theta_1)} \geqslant k \textit{ for } (X_1, X_2, \ldots, X_n) \notin C$

then C is the best critical region of size α for testing $\mathrm{H_0}$: $\theta = \theta_0$ against $\mathrm{H_a}$: $\theta = \theta_1$.

Most powerful test

A test defined by a best critical region is the **most powerful test** because it has the greatest value of power compared with other tests with the same significance level α. A test is called a **uniformly most powerful test** if it is the most powerful test against each possible hypothesis in H_a.

- Neyman-Pearson Lemma suggests that we can find a most powerful test for a single point null and alternative hypotheses based on the ratio of likelihood. However, for composite hypotheses, the uniformly most powerful test may not exist.
- Neyman-Pearson Lemma requires that the likelihood function does not contain unknown parameters.
- Nonetheless, the lemma suggests that likelihood ratio may be a general way for constructing hypothesis testing even thought it is not always the most powerful.

Likelihood ratio test

Let Ω be the set of all possible values of parameter θ given by either $\mathrm{H_0}$ or $\mathrm{H_a}$. Let ω be a subset of Ω and ω' be its complement. The null and alternative hypotheses can be stated as

$$H_0 : \ \theta \in \omega, \quad H_a : \ \theta \in \omega'$$

Let $L(\hat{\omega})$ be the maximum of the likelihood function with respect to θ when $\theta \in \omega$ and $L(\hat{\Omega})$ be the maximum of the likelihood function with respect to θ when $\theta \in \Omega$. To test $\mathrm{H_0}$ against $\mathrm{H_a}$, the critical region is the set of points in the sample space for which

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} \leqslant k,$$

where $0 < k < 1$ and k is selected so that the test has a desired significance level α.

Likelihood ratio test

Intuitively, $L(\hat{\Omega})$ represents the best explanation for the observed data when either H_0 or H_1 is true, i.e., $\theta \in \Omega = \omega \cup \omega'$. Similarly. $L(\hat{\omega})$ represents the best explanation for the observed data when H_0 is true. When $L(\hat{\omega}) = L(\hat{\Omega})$, the best explanation for the observed data can be found inside ω and we should not reject H_0. However, if $L(\hat{\omega}) < L(\hat{\Omega})$, the best explanation of data can be found in ω' and we should reject H_0 and favor H_a.

In fact, many of the hypothesis tests we discussed in previous lectures are likelihood ratio tests, although we did not explicitly derive it from the principle of likelihood ratio.

Likelihood ratio test

Example: Suppose a random sample $X_1, X_2, \ldots, X_n$ arises from a normal population $N(\mu, \sigma^2)$, where both μ and σ^2 are unknonwn. Construct the likelihood ratio test of H_0: $\mu = \mu_0$ against H_a: $\mu \neq \mu_0$.

For this test, the parameter spaces are

$$\omega = \{(\mu, \sigma^2) : \mu = \mu_0,\ 0 < \sigma^2 < \infty\}$$
$$\Omega = \{(\mu, \sigma^2) : -\infty < \mu < \infty,\ 0 < \sigma^2 < \infty\}$$

If $(\mu, \sigma^2) \in \Omega$, the maximum likelihood estimates are $\hat{\mu} = \overline{X}$ and $\hat{\sigma^2} = (1/n)\sum_{i=1}^n (X_i - \overline{X})^2$. Thus

$$L(\hat{\Omega}) = \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi(\frac{1}{n})\sum_{i=1}^n (X_i - \overline{X})^2}} \exp\left[-\frac{(X_i - \overline{X})^2}{(\frac{2}{n})\sum_{i=1}^n (X_i - \overline{X})^2}\right]\right]$$
$$= \left[\frac{1}{2\pi(\frac{1}{n})\sum_{i=1}^n (X_i - \overline{X})^2}\right]^{n/2} \exp\left[-\frac{\sum_{i=1}^n (X_i - \overline{X})^2}{(\frac{2}{n})\sum_{i=1}^n (X_i - \overline{X})^2}\right]$$

Likelihood ratio test

$$L(\hat{\Omega}) = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^{n}(X_i - \overline{X})^2}\right]^{n/2}$$

If $(\mu, \sigma^2) \in \omega$, $\mu = \mu_0$ and the maximum likelihood estimate is $\hat{\sigma^2} = (1/n)\sum_{i=1}^{n}(X_i - \mu_0)^2$. Thus,

$$L(\hat{\omega}) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi(\frac{1}{n})\sum_{i=1}^{n}(X_i - \mu_0)^2}} \exp\left[-\frac{(X_i - \mu_0)^2}{(\frac{2}{n})\sum_{i=1}^{n}(X_i - \mu_0)^2}\right]\right]$$

$$= \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^{n}(X_i - \mu_0)^2}\right]^{n/2}$$

The likelihood ratio is

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \mu_0)^2}\right]^{n/2}$$

Likelihood ratio test

Note that

$$\sum_{i=1}^{n}(X_i-\mu_0)^2=\sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu_0)^2=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu_0)^2$$

Make the substitution in the denominator of λ, we have

$$\lambda=\left[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu_0)^2}\right]^{n/2}=\left[1+\frac{n(\overline{X}-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right]^{-n/2}$$

The likelihood ratio test given by $\lambda\leqslant k$ is

$$\left[1+\frac{n(\overline{X}-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right]^{-n/2}\leqslant k$$

Likelihood ratio test

Solving the inequality, we have

$$\frac{(\overline{X}-\mu_0)^2}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]/n} \geqslant (n-1)(k^{-2/n}-1)$$

Or, equivalently,

$$\left(\frac{\overline{X}-\mu_0}{s/\sqrt{n}}\right)^2 \geqslant (n-1)(k^{-2/n}-1)$$

where s^2 is the sample variance.

This is clearly equivalent to the two tailed t-test for testing the mean of a normal population where we reject H_0 if

$$T \geqslant t_{\alpha/2}(n-1) \quad or \quad T \leqslant -t_{\alpha/2}(n-1).$$

where the test statistic T is calculated as $\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$

Likelihood ratio test

The likelihood ratio method does not always produce a test statistic with a known probability distribution. How do we use likelihood ratio test then?

Wilk's theorem: Let r_0 and r be the number of free parameters under ω and Ω, respectively. Under regularity conditions, $-2\ln(\lambda)$ asymptotically approaches $\chi^2(r - r_0)$ as sample size approaches ∞.

- The theorem gives us a general way of hypothesis testing. When sample size is large, we compare $-2\ln(\lambda)$ to a chi-square distribution with appropriate degrees of freedom. We reject the null hypothesis if the test statistic $-2\ln(\lambda)$ exceeds the critical value.
- The regularity conditions mainly involve the existence of derivatives of the likelihood function with respect to the parameters and the condition that the region over which the likelihood function is positive does not depend on unknown parameters. These conditions are satisfied for almost all distributions we discussed in this class.

Likelihood ratio test

Example: Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with unknown parameter λ. Test H_0: $\lambda = \lambda_0$ against H_a: $\lambda \neq \lambda_0$.

Recall that the maximum likelihood estimate of λ for a Poisson distribution is $\hat{\lambda} = \overline{X}$. Under H_0, no parameter needs to be estimated, thus

$$L(\hat{\omega}) = \prod_{i=1}^{n} \frac{\lambda_0^{X_i}}{X_i!} e^{-\lambda_0} = \frac{\lambda_0^{\sum_{i=1}^{n} X_i} e^{-n\lambda_0}}{\prod_{i=1}^{n} X_i!}$$

$$L(\hat{\Omega}) = \prod_{i=1}^{n} \frac{\overline{X}^{X_i}}{X_i!} e^{-\overline{X}} = \frac{\overline{X}^{\sum_{i=1}^{n} X_i} e^{-n\overline{X}}}{\prod_{i=1}^{n} X_i!}$$

Likelihood ratio test

Let Λ be the likelihood ratio test statistic.

$$\Lambda = -2\ln\left(\frac{L(\hat{\omega})}{L(\hat{\Omega})}\right) = -2\ln\left(\frac{\lambda_0^{\sum_{i=1}^{n} X_i} e^{-n\lambda_0}}{\overline{X}^{\sum_{i=1}^{n} X_i} e^{-n\overline{X}}}\right)$$

$$= 2n\Big(\overline{X}\ln\left(\frac{\overline{X}}{\lambda_0}\right) - \overline{X} + \lambda_0\Big)$$

Here, $\Lambda \sim \chi^2(1)$. We reject H_0: $\lambda = \lambda_0$ if $\Lambda \geqslant \chi^2_\alpha(1)$.