Statistics

March 17, 2021

Recall hypothesis testing:

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

$\theta \in \Theta_0 \subset \mathbb{R}^k$, $\theta_0 = \begin{pmatrix} \theta_{01} \\ \vdots \\ \theta_{0k} \end{pmatrix}$

$\theta \notin \Theta_0$

Procedure:

1. Collect data $X_1, \dots, X_n = X$
2. Compute an estimate $\hat{\theta}$ (or some some other statistic T).
3. Determine:

R = rejection region

$\beta(\theta)$ = power
$= \mathbb{P}_\theta(X \in R)$ ($R \subset \mathbb{R}^k$ in general)

size $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

Wald Test

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

Collect data:

$X_1, \dots, X_n$ such that $\mathbb{E}(X_j) = \theta_0$ under H_0.

our estimate of θ

Assume that $\hat{\theta} \rightsquigarrow N(\theta_*, , \hat{se}^2)$

true value of θ

Size α Wald test:

Reject H_0 when

Recall: $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$

$$|W| = \left|\frac{\hat{\theta} - \theta_0}{\hat{se}}\right| > z_{\alpha/2}$$

Theorem Asymptotically the Wald test has size α:

$$\mathbb{P}_{\theta_0}(|W| > z_{\alpha/2}) \to \alpha \quad \text{as} \quad n \to \infty.$$

(as our sample size increases.)

Pf: Under H_0, $\frac{\hat{\theta} - \theta_0}{\hat{se}} \rightsquigarrow N(0,1)$, then

$$\mathbb{P}_{\theta_0}(|W| > z_{\alpha/2}) = \mathbb{P}_{\theta_0}\left(\left|\frac{\hat{\theta} - \theta_0}{\hat{se}}\right| > z_{\alpha/2}\right)$$

$$\rightsquigarrow \mathbb{P}(|Z| > z_{\alpha/2}) = \alpha.$$

Note For practical considerations, $\hat{se}$ is used. But se_0 is also ok, if you're able to compute it $\left(se_0^2 = Var_{\theta_0}(\hat{\theta})\right)$.

What is the power of the Wald test?

$$\beta(\theta) = \mathbb{P}_\theta(X \in R)$$

$$= \mathbb{P}_\theta\left(\left|\frac{\hat{\theta} - \theta_0}{\hat{se}}\right| > z_{\alpha/2}\right)$$

$$= 1 - \mathbb{P}_\theta\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta_0}{\hat{se}} < z_{\alpha/2}\right)$$

$$= 1 - \mathbb{P}_\theta\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\hat{se}} + \frac{\theta - \theta_0}{\hat{se}} < z_{\alpha/2}\right)$$

$$= 1 - \mathbb{P}_\theta\left(-z_{\alpha/2} + \frac{\theta_0 - \theta}{\hat{se}} < \frac{\hat{\theta} - \theta}{\hat{se}} < z_{\alpha/2} + \frac{\theta_0 - \theta}{\hat{se}}\right)$$

$$= \boxed{1 - \Phi\left(z_{\alpha/2} + \frac{\theta_0 - \theta}{\hat{se}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\theta_0 - \theta}{\hat{se}}\right)}$$

Example Comparing two means

Data: $X_1, \dots, X_m \leftarrow$ IID Samples from two

$Y_1, \dots, Y_n \leftarrow$ IID different populations.

$$\mathbb{E}(X_j) = \mu_1$$

$$\mathbb{E}(Y_j) = \mu_2 .$$

Want to test

$$H_0: \mu_1 = \mu_2$$

vs. $$H_1: \mu_1 \neq \mu_2$$

Let $\delta = \mu_1 - \mu_2$

$H_0 : \delta = 0$

$H_1 : \delta \neq 0$.

Use the wald test:

Let $\hat{\delta} = \bar{X} - \bar{Y}$

Under suitable conditions on $Var(X_i)$, $Var(Y_j)$.

The sample variances are given by:

$$s_1^2 = \frac{1}{m-1} \sum_1^m (X_i - \bar{X})^2$$

$$s_2^2 = \frac{1}{n-1} \sum_1^n (Y_j - \bar{Y})^2$$

$$\Rightarrow \quad se(\hat{\delta}) = \sqrt{var(\bar{X} - \bar{Y})}$$

$$\approx \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} = \hat{se}$$

$$\Rightarrow \quad \frac{\hat{\delta} - (\mu_1 - \mu_2)}{\hat{se}} \rightsquigarrow N(0,1)$$

Therefore the Wald statistic is given by

$$W = \frac{\hat{\delta} - 0}{\hat{se}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

Pick size α, reject H_0 if $\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} > z_{\alpha/2}$.

<u>For small sample sizes of normal variats:</u>

Recall the Wald assumption:

$$\frac{\hat{\theta} - \theta}{\hat{se}} \rightsquigarrow N(0,1).$$

What if the sample size is small, but normally distributed? ($X_i \sim N(\mu, \sigma^2)$, μ, σ^2 are unknown).

Then let $\hat{\mu} = \bar{X}$ and let

$$s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 , \quad \hat{se}(\hat{\mu}) = \frac{1}{\sqrt{n}} \sqrt{s^2}$$

Consider the quantity:

$$\left.\frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2}}\right\}$$ has a "t-distribution" with n-1 degrees of freedom.

$$\sim t_{n-1}$$

$$\left. f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right) \left(1 + t^2/n\right)^{(n+1)/2}} \right\}$$ approximately normal for large n.

p-value

Instead of reporting "reject H_0" or "retain H_0", you could instead ask if H_0 is rejected for a particular size α.

$$\text{p-value} = \text{smallest } \alpha \text{ at which the test rejects } H_0.$$

$$= \inf\{\alpha : \underbrace{T(X_1,\dots,X_n)}_{\text{test statistic}} \in R_\alpha\}$$

"the p-value is a measure of the evidence against H_0".

Large p-value : Not evidence in favor of H_0.

1. H_0 could be true
2. H_0 is false, but the test has low power.

Thm If we have a test of the form:

Reject H_0 if and only if $T(X_1 \dots X_n) > c_\alpha$,

then
$$\text{p-value} = \sup_{\theta \in \Theta_0} \mathbb{P}\big(T(X_1,\dots,X_n) \geq \underbrace{T(x_1 \dots x_n)}_{\text{observed test statistic}}\big)$$

= probability of observing a more extreme event purely due to random fluctuations when H_0 is true.

For the Wald test: Let w = observed Wald statistic

$$= \frac{\hat{\theta} - \theta_0}{\hat{se}}$$

$$\text{p-value} = \mathbb{P}_{\theta_0}(|W| > |w|) \approx \mathbb{P}(|Z| > |w|) = 2\Phi(-|w|).$$

The p-value is a random variable.

If H_0 is true, then we may observe "reasonable" test statistics or some extreme ones:

$$\mathbb{P}(\text{p-value} < p) = \mathbb{P}_X\left(\mathbb{P}_T(T > T(X_1 \ldots X_n)) < p\right)$$

$$= \mathbb{P}_X\left(1 - F(T(X_1 \ldots X_n)) < p\right)$$

Let T have CDF F

$$= \mathbb{P}(F(T(\ldots)) > 1-p)$$

$$= \mathbb{P}(T > F^{-1}(1-p))$$

$$= 1 - \mathbb{P}(T < F^{-1}(1-p))$$

$$= 1 - F(F^{-1}(1-p)) = 1 - (1-p)$$

$$= p$$

$\Rightarrow$ p-value $\sim$ Unifor$(0,1)$ random variable.

χ^2_ν distribution

If $z_1, \dots, z_k \sim N(0,1)$ and independent, then

$$V = \sum_1^k z_i^2 \sim \chi^2_k$$

$$\mathbb{E}(V) = k \qquad f(v) = \frac{v^{k/2-1} e^{-v/2}}{2^{k/2}\,\Gamma(k/2)}, \quad v > 0.$$

$$Var(V) = 2k$$

Pearson χ^2 test

Setup: Multinomial model / data

Multinomial random variable $X = (X_1, \dots, X_k)$, $\sum_{i=1}^k X_i = n$

X_1: # samples in bin 1

$$\sum_{i=1}^k p_i = 1 \qquad \vec{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}.$$

(extension of binomial random variable)

The MLEs for the p_j are:

$$\hat{p}_j = \frac{X_j}{n}$$

Test: $H_0 : \vec{p} = \vec{p}_0 = \begin{pmatrix} p_{01} \\ p_{02} \\ \vdots \\ p_{0k} \end{pmatrix}$

Pearson χ^2 statistic:

$$T = \sum_{j=1}^k \frac{\left(x_j - \mathbb{E}_{\theta_0}(X_j)\right)^2}{\mathbb{E}_{\theta_0}(X_j)} = \boxed{\sum_{j=1}^k \frac{(x_j - np_{0j})^2}{np_{0j}}}$$

① **Thm** Asymptotically, under H_0, $T \rightsquigarrow \chi^2_{k-1}$

$k-1$ degrees of freedom because $p_1 + p_2 + \dots + p_k = 1$

$\Rightarrow$ $k-1$ values determine the k^{th} value.

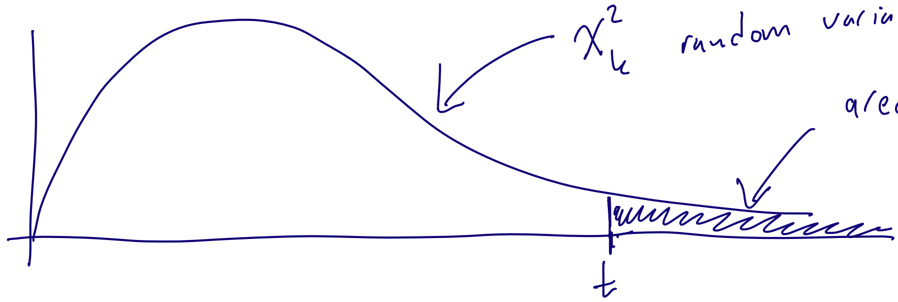
② The test:

Reject H_0 if $T > \chi^2_{k-1,\alpha}$ $\qquad \mathbb{P}(\chi^2_{k-1} > \chi^2_{k-1,\alpha}) = \alpha$

$\Rightarrow$ Asymptotically this test has level α and

p-value $= \mathbb{P}(\chi^2_{k-1} > t)$, when t = observed statistic

This is a one sided test:



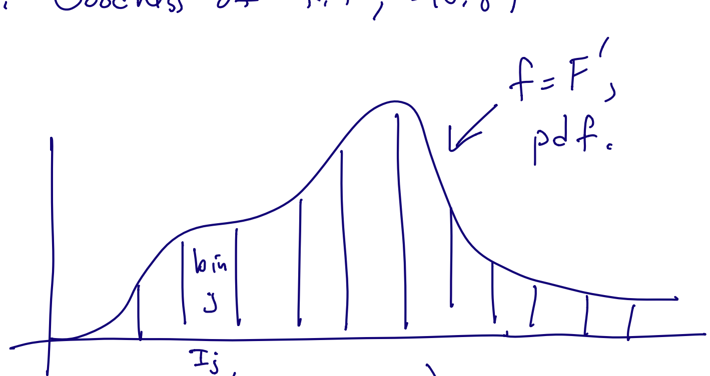
This test can also be used for a continuous distribution

(i.e. Goodness of fit, §10.8)

Collect data $X_1 \dots X_n$ IID

H_0: $X_j \sim F$

H_1: $X_j \not\sim F$.



Let N_j = # of X_j's in bin I_j.

$$p_j = \mathbb{P}(X \in \text{bin } j)$$
$$= \int_{I_j} f(x)\, dx$$

$\Rightarrow \quad N = (N_1, \ldots, N_k) \sim \text{Multinomial}(n, p_1, p_2, \ldots, p_k)$

$$p_j = \int_{I_j} f$$

Define test statistic similarly:

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$$

Under H_0, $Q \sim \chi^2_{k-1}$, apply Pearson test.

The Likelihood Ratio Test

(think how this compares with the Wald test)

Applicable to vector valued parameters:

Consider testing $\quad \theta = (\theta_1, \theta_2, \ldots \theta_k) \in \mathbb{R}^k$

$H_0 : \theta \in \Theta_0 \qquad \Theta_0 \subset \mathbb{R}^k$

$H_1 : \theta \notin \Theta_0$

The Likelihood Ratio Statistic (LRS) (LRT):

$$\lambda = 2 \cdot \log \left(\frac{\sup_{\theta} \mathcal{L}(\theta)}{\sup_{\Theta_0} \mathcal{L}(\theta)} \right) = 2 \log \frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\hat{\theta}_0)}$$

Global MLE

MLE restricted to Θ_0.

where $\mathcal{L}(\theta) = f(x_1, \ldots, x_n; \theta)$

= likelihood of the data with joint pdf f.

Theorem Let $\theta = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_r) \in \mathbb{R}^r$

$$\Theta_0 = \{\theta : \theta_{q+1}, \ldots, \theta_r = \theta_{0q+1}, \ldots, \theta_{0r}\}$$

Under $H_0 : \theta \in \Theta_0$, $\lambda \sim \chi^2_{r-q}$

$$\text{p-value} = \mathbb{P}(\chi^2_{r-q} > \underbrace{\lambda}_{\text{observed value.}})$$

Note: the LRT is applicable to any distribution, relies on the asymptotic normality of the MLE.

Neyman - Pearson Lemma

Consider the special case:

$$\left.\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{array}\right\} \quad \text{What is the most powerful test?}$$

Thm Let

$$T = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{\prod_1^n f(x_i; \theta_1)}{\prod_1^n f(x_i; \theta_0)}$$

Choose k such that $\mathbb{P}_{\theta_0}(T > k) = \alpha$.

Then, the test which rejects H_0 if $T > k$ is the most powerful size α test. (i.e. for a fixed α, this test maximizes $\beta(\theta_1)$.)

Rephrase: Among all tests with the same probability of a Type I error, the Likelihood Ratio Test minimizes the probability of a Type II error.