Hypothesis Testing II

Maxim Panov

Skoltech

November, 2021

Skoltech

Outline

Outline

Wald's Criterion

Pearson's chi-squared (χ^2) test

Likelihood-ratio Test

Permutation Test

Multiple Testing

Wald's Criterion

Let

- θ — scalar parameter;
- $\widehat{\theta}_n$ — its estimate;
- $\widehat{se}$ — standard error estimate $\widehat{\theta}_n$.

Hypothesis:

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

Definition (Wald's criterion of size α)

If $\widehat{\theta}_n$ is an asymptotically normal parameter estimate θ, i.e.

$$W_n = \frac{\widehat{\theta}_n - \theta_0}{\widehat{se}} \rightsquigarrow \mathcal{N}(0, 1), \quad n \to \infty,$$

then the hypothesis H_0 is rejected if $|W_n| > z_{\alpha/2}$.

Wald's Criterion

Theorem

Asymptotically, the size of the Wald's test is α, i.e.

$$W(\theta_0) = \mathbb{P}(|W_n| > z_{\alpha/2} \mid \theta_0) \to \alpha, \quad n \to \infty.$$

Proof.

Provided that $\theta = \theta_0$, due to the asymptotic normality of the estimate: $\frac{\widehat{\theta}_n - \theta_0}{\widehat{se}} \rightsquigarrow \mathcal{N}(0,1)$. Therefore, the probability of rejecting the main hypothesis when it is actually correct is:

$$\mathbb{P}(|W_n| > z_{\alpha/2} \mid \theta_0) = \mathbb{P}\left(\frac{|\widehat{\theta}_n - \theta_0|}{\widehat{se}} > z_{\alpha/2} \,\middle|\, \theta_0\right) \to$$
$$\to \mathbb{P}(|Z| > z_{\alpha/2}) = \alpha.$$

$\square$

Example: Comparison of Mean Values

- Consider $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ – two independent samples from general populations.
- Their means are equal to μ_1 and μ_2 correspondingly.
- Let also $\widehat{s}_1^2$ and $\widehat{s}_2^2$ be sample variances.
- Let $\delta = \mu_1 - \mu_2$.

$$H_0 \colon \delta = 0 \ \text{ vs. } \ H_1 \colon \delta \neq 0; \qquad \widehat{\delta} = \bar{X} - \bar{Y}; \qquad \widehat{se} = \sqrt{\frac{\widehat{s}_1^2}{m} + \frac{\widehat{s}_2^2}{n}}.$$

Hypothesis H_0 is rejected if $|W| > z_{\alpha/2}$, where

$$W = \frac{\widehat{\delta} - 0}{\widehat{se}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\widehat{s}_1^2}{m} + \frac{\widehat{s}_2^2}{n}}}.$$

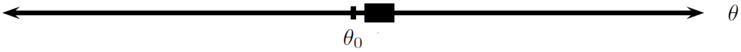
Wald's Criterion

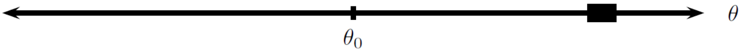
Theorem

Wald's criterion of size α rejects the hypothesis $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta \neq \theta_0$, if and only if $\theta_0 \notin C_n$, where

$$C_n = (\widehat{\theta}_n - \widehat{se}\ z_{\alpha/2}, \widehat{\theta}_n + \widehat{se}\ z_{\alpha/2}).$$

Thus, testing a hypothesis is equivalent to checking whether the value of θ_0 falls within the confidence interval.





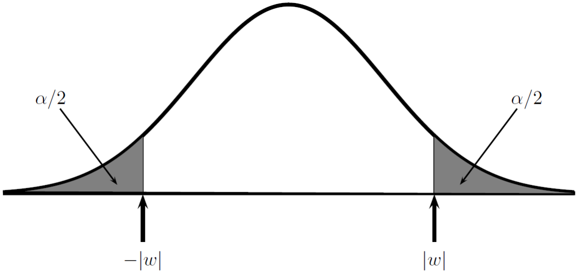
P-values for Wald's criterion

Theorem

Let $w = \frac{\widehat{\theta}_n - \theta_0}{\widehat{se}}$ *be an observed value of Wald's statistic* W*.*
Then

$$\text{p-value} = \mathbb{P}(|W_n| > |w| \mid \theta_0) \simeq \mathbb{P}(|Z| > |w|) = 2\Phi(-|w|),$$

where $Z \sim \mathcal{N}(0, 1)$*.*



Wald's Criterion

Pearson's chi-squared (χ^2) test

Likelihood-ratio Test

Permutation Test

Multiple Testing

Pearson's chi-squared (χ^2) test

- Consider the data $X_1, \dots, X_n$ having a multinomial distribution (n, p).
- Likelihood function the becomes

$$\mathcal{L}_n(p) = \prod_{j=1}^{k} p_j^{N_j}.$$

Definition (Pearson's (χ^2) statistic)

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

Chi-square distribution (χ^2)

Definition (Chi-square distribution (χ^2))

Let $Z_1, \dots, Z_k$ be independent standard normally distributed random variables. Consider

$$V = \sum_{i=1}^{k} Z_i^2,$$

then $V \sim \chi_k^2$ is chi-square distributed with k degrees of freedom.

The corresponding density:

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

Let us define the upper quantile $\chi^2_{k,\alpha} = F^{-1}(1-\alpha)$ with F being distribution function, i.e.

$$\mathbb{P}(\chi_k^2 > \chi^2_{k,\alpha}) = \alpha.$$

Pearson's chi-squared (χ^2) test

Pearson's (χ^2) statistic:

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

Theorem (Convergence to χ^2)

Under H_0 it holds

$$Q_n \rightsquigarrow \chi^2_{k-1}, \quad n \to \infty.$$

Then,

$$\text{p-value} = \mathbb{P}(\chi^2_{k-1} > q),$$

where q is an observed value of the statistic Q_n.

Example: Mendel's Peas

- Mendel bred peas with round yellow seeds and wrinkled green seeds.
- There are four types of descendents: round yellow, wrinkled yellow, round green, and wrinkled green.
- The number of each type is multinomial with probability $p = (p_1, p_2, p_3, p_4)$.

Example: Mendel's Peas

- Mendel bred peas with round yellow seeds and wrinkled green seeds.
- There are four types of descendents: round yellow, wrinkled yellow, round green, and wrinkled green.
- The number of each type is multinomial with probability $p = (p_1, p_2, p_3, p_4)$.

Mendel's theory of inheritance predicts that p is equal to

Example: Mendel's Peas

- Mendel bred peas with round yellow seeds and wrinkled green seeds.
- There are four types of descendents: round yellow, wrinkled yellow, round green, and wrinkled green.
- The number of each type is multinomial with probability $p = (p_1, p_2, p_3, p_4)$.

Mendel's theory of inheritance predicts that p is equal to

$$p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right).$$

Example: Mendel's Peas

- In $n = 556$ trials Mendel observed $X = (315, 101, 108, 32)$.
- We will test
$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0.$$
- Since, $np_{01} = 312.75, np_{02} = np_{03} = 104.25$, and $np_{04} = 34.75$, the test statistic is
$$Q_n = \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.47.$$
- For $\alpha = 0.05$ it hold that $\chi^2_{3,\alpha} = 7.815$.
- Since $0.47 < 7.815$ we conclude that the data do not contradict the theory.

Wald's Criterion

Pearson's chi-squared (χ^2) test

Likelihood-ratio Test

Permutation Test

Multiple Testing

Likelihood-ratio Test

Definition

Consider two competing hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1.$$

Let $\widehat{\theta}$ be an MLE estimate and $\widehat{\theta}_0$ be an MLE estimate when $\theta \in \Theta_0$
Likelihood-ratio statistics are:

$$\lambda_n = 2 \log \frac{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)} = 2 \log \frac{\mathcal{L}_n(\widehat{\theta})}{\mathcal{L}_n(\widehat{\theta}_0)},$$

where $\mathcal{L}_n(\theta)$ is the likelihood of the data corresponding to the model with parameter θ.

Theorem

Let us assume that $\theta = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_r)$ *and let*

$$\Theta_0 = \{\theta\colon (\theta_{q+1}, \ldots, \theta_r) = (\theta_{0,q+1}, \ldots, \theta_{0,r})\}.$$

Let λ *be the likelihood-ratio test.*
Under the hypothesis H_0*, i.e.* $\theta \in \Theta_0$ *we obtain*

$$\lambda_n \rightsquigarrow \chi^2_{r-q,\alpha},$$

where $(r - q)$ *is the dimension of* Θ *minus the dimension of* Θ_0*.*
p-value *for the criterion is* $\mathbb{P}(\chi^2_{r-q} > \lambda_n)$*.*

Example

Consider $\theta = (\theta_1, \ldots, \theta_5)$, it is necessary to check that $\theta_4 = \theta_5 = 0$. Then the limiting distribution has 2 degrees of freedom.

Example: Mendel's Peas Revisited

Example (Mendel's Peas $1/2$)

Likelihood-ratio statistics for

$$H_0\colon p = p_0 \quad \text{vs.} \quad H_1\colon p \neq p_0$$

takes the form:

$$\lambda_n = 2\log\frac{\mathcal{L}_n(\widehat{p})}{\mathcal{L}_n(\widehat{p}_0)} = 2\sum_{j=1}^{4} X_j \log\frac{\widehat{p}_j}{\widehat{p}_{0j}} = 2\cdot\left[315\log\left(\frac{315/556}{9/16}\right) + \right.$$

$$\left. +101\log\left(\frac{101/556}{3/16}\right) + 108\log\left(\frac{108/556}{3/16}\right) + 32\log\left(\frac{32/556}{1/16}\right)\right] = 0.48.$$

Example: Mendel's Peas Revisited

Example (Mendel's Peas $2/2$)

- Under the hypothesis H_1 there are 4 parameters.
- Since the sum of the parameters must be equal to 1, the dimension of the parameter space is 3.
- Under the hypothesis H_0 there are no free parameters, so the number of degrees of freedom is 3 and χ^2_3 is the limiting distribution.

$$\text{p-value} = \mathbb{P}(\chi^2_3 > 0.48) = 0.92.$$

Remark

Usually both the χ^2 test and the likelihood-ratio test give approximately the same results, provided that the sample size is large enough.

Neyman-Pearson Test for the Case of Two Simple Hypotheses

Lemma (Neyman-Pearson)

$$H_0 : \theta = \theta_0 \quad \textit{vs.} \quad H_1 : \theta = \theta_1.$$

Neyman-Pearson statistics:

$$T_n = \frac{\mathcal{L}_n(\theta_1)}{\mathcal{L}_n(\theta_0)} = \frac{\prod_{i=1}^{n} f(x_i; \theta_1)}{\prod_{i=1}^{n} f(x_i; \theta_0)}. \tag{1}$$

- *Suppose that H_0 is rejected for $T_n > t$.*
- *Choose $t = t_\alpha$ so that $\mathbb{P}_{\theta_0}(T > t_\alpha) = \alpha$.*

Then, the Neyman-Pearson test (based on the (1) statistics) will be the most powerful one $W(\theta_1)$ among all α size tests.

Permutation Test

Definition (Problem formulation)

Permutation criterion is used to check if two distributions are different.
Consider $X_1, \ldots, X_m \sim F_X$ and $Y_1, \ldots, Y_n \sim F_Y$ — two independent samples. We want to solve:

$$H_0 \colon F_X = F_Y \quad \text{vs.} \quad H_1 \colon F_X \neq F_Y\,.$$

Permutation criterion is "exact" in the sense that it does not use the assumption of asymptotic convergence to a normal distribution.

Permutation Test

1. Let us denote by $T(x_1, \ldots, x_m, y_1, \ldots, y_n)$ some test statistics like

$$T(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = |\bar{X}_m - \bar{Y}_n|.$$

2. Set $N = m + n$ and consider all $N!$ Permutations of the combined sample $X_1, \ldots, X_m, Y_1, \ldots, Y_n$.
3. For each of the permutations, calculate the value of the statistics T.
4. Let us denote these values $T_1, \ldots, T_{N!}$.

Theorem (Permutation criterion)

If H_0 is true, then for fixed ordered values $\{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$ statistic values T are uniformly distributed on the set $T_1, \ldots, T_{N!}$.

Permutation Test

Theorem

Let us denote as the permutation distribution of the statistic T such that:

$$P_0(T = T_i) = \frac{1}{N!}, \quad i = 1, \dots, N!$$

Consider t_{obs}, the value of the statistic that was obtained in the experiment. Then:

$$\text{p-value} = \mathbb{P}(T > t_{obs} \mid f) = \frac{1}{N!} \sum_{j=1}^{N!} \mathbb{I}(T_j > t_{obs}), \quad f \in F_0.$$

Example: Permutation Test

Example

Suppose that $(X_1, X_2, Y_1) = (1, 9, 3)$. Let $T(X_1, X_2, Y_1) = |\bar{X} - \bar{Y}| = 2$, then

Permutation	Value T	Probability
(1,9,3)	2	1/6
(9,1,3)	2	1/6
(1,3,9)	7	1/6
(3,1,9)	7	1/6
(3,9,1)	5	1/6
(9,3,1)	5	1/6

$$\text{p-value} = \mathbb{P}(T > 2) = 4/6.$$

Multiple Testing

Consider m different cases of hypothesis testing:

$$H_{0i} \quad \text{vs.} \quad H_{1i}, \quad i = 1, \ldots, m.$$

Let $P_1, \ldots, P_m$ denote the $\mathrm{p\text{-}values}$ of these checks.

Definition (Holm–Bonferroni method)

For given $\mathrm{p\text{-}values}$ $P_1, \ldots, P_m$ the main hypothesis H_{0i} is rejected if

$$P_i < \frac{\alpha}{m}.$$

Multiple Testing

Theorem

When applying the Holm–Bonferroni method, the probability of incorrectly rejecting any of the main hypotheses is less than or equal to α.

Proof.

Consider the notation

- R is the event that at least one of the underlying hypotheses has been falsely rejected;
- R_i is the event that the i-th main hypothesis was falsely rejected.

Then

$$\mathbb{P}(R) = \mathbb{P}\left(\bigcup_{i=1}^{m} R_i\right) \leq \sum_{i=1}^{m} \mathbb{P}(R_i) = \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha.$$

$\square$

False Discovery Rate, FDR

	H_0 not rejected	H_0 rejected	$\sum$
H_0 is true	U	V	m_0
H_0 is false	T	S	m_1
$\sum$	$m - R$	R	m

Definition

Based on the table, the proportion (FDP) and frequency (FDR) of the occurrence of false deviations (among deviations in general):

$$FDP = \begin{cases} \frac{V}{R}, & R > 0, \\ 0, & R = 0, \end{cases}$$

$$FDR = \mathbb{E}(FDP).$$

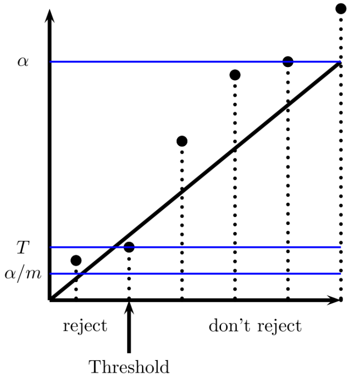
$Benjamini - Hochberg$ method

1. Let $P_{(1)} < \cdots < P_{(m)}$ be the observations $\mathrm{p\text{-}value}$, sorted in ascending order.
2. Let $C_m = 1$ if $P_{(1)}, \ldots, P_{(m)}$ are independent, otherwise $C_m = \sum\limits_{i=1}^{m} \frac{1}{i}$.
3. Let's define: $I_i = \frac{i\alpha}{C_m m}$, $R = \max\{i \colon P_{(i)} < I_i\}$.
4. $T = P_{(R)}$ is method threshold.
5. If $P_i \leq T$ then H_{0i} is rejected.

Theorem

When applying this method, regardless of how many main hypotheses are actually true and regardless of the distribution $P_1, \ldots, P_m$, in the case of an incorrect main hypothesis, the following holds:

$$FDR = \mathbb{E}(FDP) \leq \frac{m_0}{m}\alpha \leq \alpha.$$



Thank you for your attention!