

# PH 712 Probability and Statistical Inference

## Part IX: Hypothesis Testing

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/11/18 13:05:52

---

Is it a squirrel?



Figure 1: Flying Squirrel (Photograph by Joel Sartore)

- Make a decision between two hypotheses  $H_0$  : YES and  $H_1$ : NO.
  - Checking necessary conditions under  $H_0$
- It is a binary classification problem.

### Problem formalization

- Assumptions
  - $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x | \theta)$ 
    - \*  $\theta$  is fixed and unknown BUT is believed to be inside  $\Theta$
  - To make a decision on  $\theta$  between two hypotheses  $H_0 : \theta \in \Theta_0$  and  $H_1 : \theta \in \Theta_1$ 
    - \*  $\Theta_0 \cup \Theta_1 = \Theta$
    - \*  $\Theta_0 \cap \Theta_1 = \emptyset$
- Four possible outcomes
  - True positive (TP):  $H_0$  is wrong (i.e.,  $H_1$  is true) and we reject  $H_0$  (i.e., accept  $H_1$ );
  - False positive (FP, type I error):  $H_0$  is true (i.e.,  $H_1$  is wrong) but we reject  $H_0$  (i.e., accept  $H_1$ );
  - True negative (TN):  $H_0$  is true (i.e.,  $H_1$  is wrong) and we accept  $H_0$  (i.e., reject  $H_1$ );
  - False negative (FN, type II error):  $H_0$  is wrong (i.e.,  $H_1$  is true) but we accept  $H_0$  (i.e., reject  $H_1$ ).
  - E.g., in the context of identifying the animal,
    - \* TP: it is NOT a squirrel and is NOT identified as a squirrel
    - \* FP: it is a squirrel but is NOT identified as a squirrel
    - \* TN: it is a squirrel and is identified as a squirrel
    - \* FN: it is NOT a squirrel but is identified as a squirrel

	Accept $H_0$	Reject $H_0$
$H_0$ is true	True negative (TN)	False positive (FP, type I error)
$H_0$ is false	False negative (FN, type II error)	True positive (TP)

- Different objectives leading to different strategies:
  - Minimizing the misclassification rate:  $\Pr(\text{FP}) + \Pr(\text{FN})$ 
    - \* Commonly adopted by classification techniques
  - Controlling the false discovery rate (FDR):  $\Pr(\text{FP})/\{\Pr(\text{FP}) + \Pr(\text{TP})\}$ 
    - \* For sequential or simultaneous testing
  - Minimizing  $\Pr(\text{FN})$  with  $\Pr(\text{FP})$  capped; specifically, minimizing  $\Pr(\text{type II error})$  with  $\Pr(\text{type I error}) \leq \alpha$ 
    - \* Leading to the optimal hypothesis test

## Formalizing the hypothesis test

- A test, say  $\phi$ , is an indicator function

$$\phi(x_1, \dots, x_n) = \mathbf{1}_R(x_1, \dots, x_n) = \begin{cases} 0, & (x_1, \dots, x_n) \notin R \\ 1, & (x_1, \dots, x_n) \in R \end{cases}$$

- Input: the sample or its realization
- Output: the action after observing the input, i.e., 0 (accepting  $H_0$ ) or 1 (rejecting  $H_0$ )
- *Rejection region*:  $R$ , the set corresponding to the rejection of  $H_0$ 
  - \*  $R$  is typically specified in terms of the realization of a *test statistic*; e.g., if  $R = \{(x_1, \dots, x_n) : \bar{x} \geq 3\}$ , then  $\bar{X}$  is a test statistic.
- Each test corresponds to a unique rejection region
  - Two tests are equivalent  $\Leftrightarrow$  their rejection regions are identical

## Uniformly most powerful (UMP) level $\alpha$ test (CB Sec 8.3.2)

- *Power function*: given a test  $\phi$  and its rejection region  $R$ , the power function  $\beta_\phi(\theta)$  is the probability of rejecting  $H_0$ : for all  $\theta \in \Theta$ ,

$$\beta_\phi(\theta) = \Pr\{(X_1, \dots, X_n) \in R\} = \Pr\{\phi(X_1, \dots, X_n) = 1\}$$

- Preferring large  $\beta_\phi(\theta)$  for all  $\theta \in \Theta_1$  and small  $\beta_\phi(\theta)$  for all  $\theta \in \Theta_0$ , because
  - \*  $\Pr(\text{type I error}) = \beta_\phi(\theta_0)$  if  $H_0$  is correct ( $\theta_0 \in \Theta_0$ )
  - \*  $\Pr(\text{type II error}) = 1 - \beta_\phi(\theta_0)$  if  $H_1$  is correct ( $\theta_0 \in \Theta_1$ )
  - \* But  $\theta_0$  is unknown
- A test  $\phi$  is of *size*  $\alpha \Leftrightarrow \sup_{\theta \in \Theta_0} \beta_\phi(\theta) = \alpha$ 
  - $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$ : the supremum of  $\beta_\phi(\theta)$  in  $\Theta_0 \Leftrightarrow$  the maximum of  $\beta_\phi(\theta)$  in the closure of  $\Theta_0$
  - $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) = \alpha \Rightarrow \Pr(\text{type I error}) \leq \alpha$
- A test  $\phi$  is of *level*  $\alpha \Leftrightarrow \sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \alpha$ 
  - $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$ : the supremum of  $\beta_\phi(\theta)$  in  $\Theta_0 \Leftrightarrow$  the maximum of  $\beta_\phi(\theta)$  in the closure of  $\Theta_0$
  - $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \alpha \Rightarrow \Pr(\text{type I error}) \leq \alpha$
- Let  $\phi$  be a level  $\alpha$  test for  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$ . If  $\beta_\phi(\theta) \geq \beta_{\phi'}(\theta)$  for all  $\theta \in \Theta_1$  and any other test  $\phi'$  of level  $\alpha$ , then  $\phi$  is a UMP level  $\alpha$  test.

## Example Lec9.1

- $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma^2)$  with unknown  $\theta$  and known  $\sigma$ . Consider a test for  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta \neq \theta_0$  with rejection region  $\{(x_1, \dots, x_n) : \sqrt{n}|\bar{x} - \theta_0|/\sigma > c\}$ .
  1. Elaborate the power function.
  2. Find sample size  $n$  and threshold  $c$  if one desires that the type I error rate is 5% and the type II error rate at  $\theta_0 + \sigma$  is 25%.

## Likelihood ratio test (LRT, CB Sec. 8.2.1 & 10.3.1)

- Hypotheses:  $H_0 : \theta \in \Theta_0$  vs.  $H_1 : \theta \in \Theta_1$ 
  - $\Theta = \Theta_0 \cup \Theta_1$
  - $\Theta_0 \cap \Theta_1 = \emptyset$
- Test statistic

$$\lambda(X_1, \dots, X_n) = \frac{L(\hat{\theta}_{ML,0})}{L(\hat{\theta}_{ML})}$$

- $\hat{\theta}_{ML,0}$ : MLE of  $\theta$  under  $H_0$
- $\hat{\theta}_{ML}$ : MLE of  $\theta \in \Theta$
- Rejection region

$$R = \{(x_1, \dots, x_n) : \lambda(x_1, \dots, x_n) \leq c_\alpha\},$$

where  $c_\alpha$  is chosen to make sure the size is  $\alpha$ , i.e.,

$$\sup_{\theta \in \Theta_0} \Pr\{\lambda(X_1, \dots, X_n) \leq c_\alpha\} = \alpha.$$

- Essential but challenging to know the distribution of  $\lambda(X_1, \dots, X_n)$  under  $H_0$
- Implementation
  1. Confirm the value of  $\alpha$ ;
  2. Figure out  $\hat{\theta}_{ML,0}$  and  $\hat{\theta}_{ML}$ .
  3. Solve equation
 
$$\sup_{\theta \in \Theta_0} \Pr\{\lambda(X_1, \dots, X_n) \leq c_\alpha\} = \alpha$$
 for  $c_\alpha$ ;
  4. Construct the rejection region  $\{(x_1, \dots, x_n) : \lambda(x_1, \dots, x_n) \leq c_\alpha\}$ .
- Why is LRT promoted?
  - Neyman-Pearson Lemma (CB Thm 8.3.12): LRT is the UMP level  $\alpha$  test for simple hypotheses ( $H_0 : \theta = \theta_0$  vs  $H_1 : \theta = \theta_1$ )
  - Karlin-Rubin theorem (CB Thm 8.3.17): under certain conditions, LRT is the UMP level  $\alpha$  test for one-sided hypotheses ( $H_0 : \theta \leq \theta_0$  (or  $\theta = \theta_0$ ) vs  $H_1 : \theta > \theta_0$  OR  $H_0 : \theta \geq \theta_0$  (or  $\theta = \theta_0$ ) vs  $H_1 : \theta < \theta_0$ )
- Special cases
  - Equivalent to the  $Z$ -test if 1) the sample is iid normal with known variance and 2) the mean is to be tested
  - Equivalent to the  $t$ -test if 1) the sample is iid normal with unknown variance and 2) the mean is to be tested
  - Equivalent to the  $F$ -test if 1) the sample is iid normal with the mean and variance both unknown and 2) the variance is to be tested

## LRT (con'd)

- Asymptotic rejection region (CB Thm 10.3.3)

$$R \approx \{(x_1, \dots, x_n) : -2 \ln \lambda(x_1, \dots, x_n) \geq \chi_{\nu, 1-\alpha}^2\} = \{(x_1, \dots, x_n) : \lambda(x_1, \dots, x_n) \leq \exp(-\chi_{\nu, 1-\alpha}^2/2)\},$$

where  $\chi_{\nu, 1-\alpha}^2$  is the  $(1 - \alpha)$  quantile of  $\chi^2(\nu)$ , i.e.,  $F_{\chi^2(\nu)}(\chi_{\nu, 1-\alpha}^2) = 1 - \alpha$ .

- (CB Thm 10.3.1) Because, as  $n \rightarrow \infty$ , under  $H_0$ ,

$$-2 \ln \lambda(X_1, \dots, X_n) \approx \chi^2(\nu),$$

where  $\nu$  = the difference of numbers of free parameters between  $\Theta_0$  and  $\Theta$ .

- Implementation (asymptotic)
  1. Confirm the value of  $\alpha$ ;
  2. Figure out  $\hat{\theta}_{0,ML}$  and  $\hat{\theta}_{ML}$ ;
  3. Check  $\nu$ , the difference of numbers of free parameters between  $\Theta_0$  and  $\Theta$ ;
  4. Construct the asymptotic rejection region  $\{x_1, \dots, x_n : -2 \ln \lambda(x_1, \dots, x_n) \geq \chi_{\nu, 1-\alpha}^2\}$ .

## CB Ex 8.2

- For a given city in a given year, assume that the number of automobile accidents follows a Poisson distribution. In past years the average number of accidents per year was 15, and this year it was 10. Is it justified to claim that the accident rate has dropped?
- Demo report: Testing hypotheses  $H_0 : \underline{\quad}$  vs.  $H_1 : \underline{\quad}$ , we carried on the  $\underline{\quad}$  test and obtained  $\underline{\quad}$  as the value of test statistic. The corresponding rejection region is  $\underline{\quad}$ . So, at the  $\underline{\quad}$  level, there was/wasn't a strong statistical evidence against  $H_0$ , i.e., we believed that  $\underline{\quad}$ .

## $p$ -value (CB Sec 8.3.4)

- Motivation
  - Recall that a rejection region  $R$  consists of a test statistic (e.g.,  $\lambda(X_1, \dots, X_n)$  for LRT) and critical point (e.g.,  $c_\alpha$  for LRT)
    - \* The test statistic NOT uniquely defined
    - \* The critical point varying with the definition of test statistic
  - Would like to fix the critical point to be  $\alpha$  by defining a test statistic  $p(X_1, \dots, X_n)$  (i.e.,  $p$ -value) such that the following set is equivalent to  $R$

$$\{(x_1, \dots, x_n) : p(x_1, \dots, x_n) \leq \alpha\}$$

- \* More convenient in communication because the critical point is  $\alpha$  by default
- NOT always well-defined
- (CB Thm 8.3.27) If  $H_0$  is rejected when a test statistic  $T(x_1, \dots, x_n)$  is too large, then

$$p(x_1, \dots, x_n) = \sup_{\theta \in \Theta_0} \Pr\{T(X_1, \dots, X_n) \geq T(x_1, \dots, x_n)\}.$$

## CB Ex 8.2

- For a given city in a given year, assume that the number of automobile accidents follows a Poisson distribution. In past years the average number of accidents per year was 15, and this year it was 10. Is it justified to claim that the accident rate has dropped?
- Demo report: Testing hypotheses  $H_0 : \underline{\quad}$  vs.  $H_1 : \underline{\quad}$ , we carried on the  $\underline{\quad}$  test and obtained  $\underline{\quad}$  as the  $p$ -value. So, at the  $\underline{\quad}$  level, there was/wasn't a strong statistical evidence against  $H_0$ , i.e., we believed that  $\underline{\quad}$ .

## Example Lec9.2

- $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ . Consider  $H_0 : \mu = \mu_0$  vs  $H_1 : \mu \neq \mu_0$ .
  1. Verify that the size  $\alpha$  LRT rejects  $H_0$  when  $|\bar{x} - \mu_0| > t_{n-1, 1-\alpha/2}(s/\sqrt{n})$ , where  $s = \sqrt{(n-1)^{-1} \sum_i (x_i - \bar{x})^2}$ .
  2. Find the the expression of  $p$ -value for LRT.

## Wald test (CB pp. 493)

- Testing  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta \neq \theta_0$
- With an estimator  $\hat{\theta}$  such that  $(\hat{\theta} - \theta_0)/\sqrt{\text{var}(\hat{\theta})} \approx \mathcal{N}(0, 1)$  under  $H_0$  as  $n \rightarrow \infty$
- Test statistic:  $(\hat{\theta} - \theta_0)/\sqrt{\text{var}(\hat{\theta})}$ 
  - Asymptotically equivalent to LRT for this two sided test if  $\hat{\theta} = \hat{\theta}_{\text{ML}}$
  - Substitute  $\widehat{\text{var}}(\hat{\theta})$  for  $\text{var}(\hat{\theta})$  if  $\text{var}(\hat{\theta})$  is well approximated by  $\widehat{\text{var}}(\hat{\theta})$  (obtained by the delta methods/bootstrap)
- Level  $\alpha$  Wald rejection region:  $\{(x_1, \dots, x_n) : |\hat{\theta} - \theta_0|/\sqrt{\text{var}(\hat{\theta})} \geq \Phi_{1-\alpha/2}^{-1}\}$

- $\Phi_{1-\alpha/2}^{-1}$ : the  $(1 - \alpha/2)$  quantile of  $\mathcal{N}(0, 1)$
- $p\text{-value} = 2\Phi\left(-|\hat{\theta} - \theta_0|/\sqrt{\text{var}(\hat{\theta})}\right)$ 
  - $\Phi(\cdot)$ : cdf of  $\mathcal{N}(0, 1)$

### Score test (CB pp. 494)

- Testing  $H_0 : \theta = \theta_0$  vs.  $H_1 : \theta \neq \theta_0$
- Test statistic:  $\ell'(\theta_0)/\sqrt{I_n(\theta_0)}$  ( $\approx \mathcal{N}(0, 1)$  under  $H_0$  as  $n \rightarrow \infty$ )
- Level  $\alpha$  score rejection region:  $\{(x_1, \dots, x_n) : |\ell'(\theta_0)|/\sqrt{I_n(\theta_0)} \geq \Phi_{1-\alpha/2}^{-1}\}$ .
- $p\text{-value} = 2\Phi\left(-|\ell'(\theta_0)|/\sqrt{I_n(\theta_0)}\right)$

### CB Examples 10.3.5 & 10.3.6

- $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p)$ ,  $p \in (0, 1)$ . Derive LRT, Wald and score tests for  $H_0 : p = p_0$  versus  $H_1 : p \neq p_0$ .