

QCB 408 / 508 – Notes on Week 5

Yushi Tang

2020-03-11

Contents

1	Summary	1
2	Details	1
2.1	Hypothesis Testing I: General Framework	1
2.2	Hypothesis Testing II: Generalized Likelihood Ratio Test	5
2.3	Hypothesis Testing III: Pivotal Statistic Hypothesis Testing	9
2.4	Bayesian Inference: General Framework	11
3	References	11

1 Summary

For this week, we mainly introduced the fundamental elements of hypothesis testing and the basic idea of Bayesian inference. We discussed two specific scenarios for conducting hypothesis testing in class, including the generalized likelihood ratio test and pivotal statistic hypothesis test, as the practical examples for this topic. In this note, I also added several extra examples and settings for test design. Since the p-value is an important concept for hypothesis testing which could also be easily misused, I made more supplementary information about p-value, especially in the setting of multiple hypothesis tests. Last but not least, we touched gently the general framework of Bayesian Inference in class. I annotated the class material for Bayesian Inference with some of my own thoughts. Please check the content table above for the complete list of topics. All material could be categorized into two parts:

- 1). The concepts and examples from **Class**:
 - Definition 1.1, 1.7, 1.8, 1.9, 1.10, 1.12, 2.1, 2.2, 2.4, 3.1, 3.2, 3.4, 4.1
 - Example 1.2, 2.6, 3.3, 3.5, 3.6
- 2). The concepts and examples from **Yushi**:
 - Definition 1.4, 1.5,
 - Example 1.3, 1.6, 1.11, 1.13, 2.3, 2.5, 2.7, 4.2
 - Remark 1.12, 2.1, 2.3 (proof), 4.3

2 Details

2.1 Hypothesis Testing I: General Framework

Motivation. Generally speaking, we want to test whether our parameters follow some specific laws or not. As for simple cases, we aim at making probabilistic statements about the value of the parameters of interest with respect to categories (or sets) of interests.

Remark. When we say simple, we refer to hypothesis that could be defined in term of a single value. On the opposite side, we have complex hypothesis testing, which refer to testing hypothesis that are defined in term of a collection of values.

Definition 1.1 (Simple Hypothesis). Let $\mathbf{X} = (X_1, X_2, \dots, X_n) \sim \mathcal{F}_\theta$. A simply hypothesis test could be defined as H_0 in testing

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

where θ_0 is the null value and θ_1 is the alternative value.

Example 1.2 (Binomial Test). Suppose $\mathbf{X} = (X_1, X_2, \dots, X_n) \stackrel{i.i.d.}{\sim} \text{Binomial}(p)$. A simple hypothesis test could be testing

$$H_0 : p = 1 \text{ vs. } H_1 : p = 2$$

where $p_0 = 1$ is the null value and $p_1 = 2$ is the alternative value.

Example 1.3 (Normal Test). Suppose $\mathbf{X} = (X_1, \dots, X_n) \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$. A simple hypothesis test could be testing

$$H_0 : \theta = 1/2 \text{ vs. } H_1 : \theta = 2$$

Remark. Simple hypothesis testing is basically a classification problem. We observed the data. Then our goal is to collect the information showing that the alternative hypothesis H_1 (could be considered as category 1) is favored than the null hypothesis H_0 (could be considered as category 0) or the other way. The key is to use the likelihood to evaluate which category is preferred. It is reasonable to conduct all inference through the likelihood function.

Definition 1.4 (Hypothesis Test) Under the same setting in the definition for symple hypothesis, let $\mathbf{X} = (X_1, \dots, X_n) \sim \mathcal{F}_\theta$, $\theta \in \Theta$, where the p.d.f./p.m.f. could be written as $f(\mathbf{x}|\theta)$, we can construct

$$\begin{aligned} &\text{Null Hypothesis } H_0 : \theta \in \Theta_0 \text{ vs.} \\ &\text{Alternative Hypothesis } H_1 : \theta \in \Theta_1, \end{aligned}$$

where $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$, and $\Theta_0 \cap \Theta_1 = \emptyset$. Often we have $\Theta_1 = \Theta_0^c = \Theta - \Theta_0$. Under this setting, we define the hypothesis test as a rule that directs us to not reject H_0 or reject H_0 defined by specifying a rejection region / significance region, Γ_c , of sample outcomes. Now we can make a decision based on observed data as

$$\begin{aligned} \mathbf{x} \in \Gamma_c &\Rightarrow \text{reject } H_0 \\ \mathbf{x} \notin \Gamma_c &\Rightarrow \text{not reject } H_0 \end{aligned}$$

Remark. Now we can consider simple hypotheses as H_0 where Θ_0 contains a single element.

Definition 1.5 (Composite Hypothesis). Under the setting of Definition 1.4, If Θ_0 contains more than one possible value in Θ , the H_0 is called a **composite hypothesis**.

Example 1.6 (Binomial Test Revision). Consider the same setting in Example 1.2, a composite hypothesis test could be testing

$$H_0 : p \leq 1 \text{ vs. } H_1 : p > 2$$

Since we touched the concept of significance regions in Definition 1.4, we can now define it formally.

Definition 1.7 (Significance Regions). Define

$$\Gamma_c = \{\mathbf{x} : \mathcal{L}(\theta; \mathbf{x}) / \mathcal{L}(\theta_0; \mathbf{x}) \geq c\}, \quad c \geq 0 \quad (1.1)$$

where

$$\Gamma_{c'} \supseteq \Gamma_c \text{ if } c' \leq c.$$

Then Γ_c as a function of $c \geq 0$ forms a nested set of significance region. In other word, if we write

$$S(\underline{\mathbf{x}}) = \frac{\mathcal{L}(\theta; \underline{\mathbf{x}})}{\mathcal{L}(\theta_0; \underline{\mathbf{x}})}, \quad (1.2)$$

then we acquire $S(x) \geq c$ which would capture Γ_c as the critical region.

Remark. $S(\underline{\mathbf{x}}) = \mathcal{L}(\theta; \underline{\mathbf{x}})/\mathcal{L}(\theta_0; \underline{\mathbf{x}})$ is an example of a test statistic.

Table 1: Four scenarios of hypothesis testing

	Test is significant	Test is not significant
H_0 is true	false positive	true negative
H_1 is true	true positive	false negative

Definition 1.8 (False Positive Rate / Type I Error Rate). First, we define the false positive as type I error. Then, define

$$\text{FP rate} := \text{Type I error rate} = \int_{\tau_c} f(\underline{\mathbf{x}}; \theta_0) d\mathbf{x} = \mathbb{P}(\underline{\mathbf{x}} \in \Gamma_c; \theta = \theta_0) \quad (1.3)$$

Ideally, the type I error should be small.

Definition 1.9 (False Negative Rate / Type II Error Rate). Similar as before, we can define the false negative as type II error. Then, define

$$\text{FN rate} := \text{Type II error rate} = 1 - \int_{\tau_c} f(x; \theta_1) dx = 1 - \mathbb{P}(\underline{\mathbf{x}} \in \Gamma_c; \theta = \theta_1) \quad (1.4)$$

Ideally, we want both Type I and Type II errors to have low probability.

Definition 1.10 (Power).

$$\text{Power} := 1 - \text{FN rate} = 1 - \text{Type II error rate} = \mathbb{P}(\underline{\mathbf{x}} \in \Gamma_c; \theta = \theta_1) \quad (1.5)$$

Example 1.11 (Power Function and Type I Error). A practical setting could be counting the total number of successes in 5 Bernoulli trials as X . Suppose $X \sim \text{Binomial}(5, \theta)$ and we want to test

$$H_0 : \theta \leq 1/2 \text{ vs. } H_1 : \theta > 1/2$$

where $\Theta_0 = [0, 1/2]$, $\Theta_1 = (1/2, 1]$. Suppose we define significance region as $\Gamma_c = \{3, 4, 5\}$, then we can derive the power function as

$$\begin{aligned} \beta(\theta) &= \mathbb{P}(\underline{\mathbf{x}} \in \{3, 4, 5\}; \theta) = \sum_{x=3}^5 \binom{5}{x} \theta^x (1 - \theta)^{5-x} \\ \text{Type I Error} = \alpha &= \sup_{\theta \in \Theta_0} \beta(\theta) = \sup_{\theta \in \Theta_0} \sum_{x=3}^5 \binom{5}{x} \theta^x (1 - \theta)^{5-x} = 0.5 \end{aligned}$$

Theorem 1.12 (Neyman-Pearson Lemma) Let $\Gamma_c = \{\underline{\mathbf{x}} : \mathcal{L}(\theta_1; \underline{\mathbf{x}})/\mathcal{L}(\theta_0; \underline{\mathbf{x}}) \geq c\}$ as the “most powerful” significant regions. Given a false positive rate, the highest power possible is Γ_c

Remark 1.12. The Neyman-Pearson Lemma establishes existence and uniqueness of an optimal testing procedure in the context of simple hypothesis testing. It provides the theoretical basis to construct optimal test statistics for the hypothesis testing problems.

Example 1.13 (Normal Testing with Neyman-Pearson Lemma). Let $\mathbf{X} = (X_1, \dots, X_n) \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is known and $\theta = \mu \in \Theta = \mathbb{R}$. With Neyman-Pearson Lemma, we can develop the test statistic as

$$\begin{aligned} \frac{\mathcal{L}(\theta_1; \mathbf{x})}{\mathcal{L}(\theta_0; \mathbf{x})} &= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_1)^2]}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta_0)^2]} \\ &= \frac{\exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \theta_1)^2]}{\exp[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} (\bar{x} - \theta_0)^2]} \\ &= \exp\left\{-\frac{n}{2\sigma^2} [(\bar{x} - \theta_1)^2 - (\bar{x} - \theta_0)^2]\right\} \\ &= \exp\left\{\frac{n}{2\sigma^2} [(\theta_1 - \theta_0)(\bar{x} - \frac{\theta_1 + \theta_0}{2})]\right\}. \end{aligned}$$

Then the “most powerful” significant region could be written as

$$\Gamma_c = \left\{ \exp\left\{\frac{n}{2\sigma^2} [(\theta_1 - \theta_0)(\bar{x} - \frac{\theta_1 + \theta_0}{2})]\right\} \geq c \right\} = \left\{ \bar{x} \geq \frac{2\sigma^2 \log c}{n(\theta_1 - \theta_0)} + \frac{1}{2}(\theta_1 + \theta_0) \right\}$$

where c is chosen so that the test has size / type I error as α .

2.2 Hypothesis Testing II: Generalized Likelihood Ratio Test

Motivation. Let Θ be the set of all possible θ , Θ_0 be the null values, and Θ_1 be the alternative values. Directly we have $\Theta_0 \cap \Theta_1 = \emptyset$, and $\Theta_0 \cup \Theta_1 = \Theta$. We want to find an efficient way to test

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

Definition 2.1 (Null Distribution). The null distribution is the probability distribution of the test-statistic when H_0 is true. We can derive an approximate null distribution by taking $n \rightarrow \infty$ (i.e. $\theta \in \Theta_0$).

Definition 2.2 (Generalized likelihood ratio test). The generalized likelihood ratio test has the test statistic as

$$\lambda(\underline{x}) = \frac{\max_{\theta \in \Theta} \mathcal{L}(\theta; \underline{x})}{\max_{\theta \in \Theta_0} \mathcal{L}(\theta; \underline{x})} \quad (2.1)$$

which is equivalent to the test statistic as

$$\lambda(\underline{x}) = \frac{\mathcal{L}(\hat{\theta}_{MLE}; \underline{x})}{\mathcal{L}(\hat{\theta}_{0,MLE}; \underline{x})} \quad (2.2)$$

When H_0 is true, the test statistic will converge in distribution to a χ^2 -distribution. i.e.

$$2 \log(\lambda(\underline{x})) \xrightarrow{d} \chi_\nu^2 \quad (2.3)$$

where $\nu = \dim(\Theta) - \dim(\Theta_0)$.

Remark 2.1. Here ([paper link](#)) might be the 7 ways to show the pearson independent test and χ^2 approximation. To answer Prof. Storey's question in the class, my favorite one is using the characteristic function to prove. Please correct me if this is not what we talked about during the class.

Remark 2.2. There is usually (i.e. almost every time we can see in real life) a one-to-one correspondence between the test-statistic and the significant regions s.t. the larger the test-statistic is, the smaller the significant region would be.

Remark 2.3. Under the setting of generalized likelihood ratio test, we have

$$\text{Type I error rate} = \int_{\Gamma_c} f(\underline{x}; \theta_0) d\mathbf{x} \approx \int_{\Gamma_c} g(\underline{x}) d\mathbf{x} \text{ as } n \rightarrow \infty \quad (2.4)$$

where $\Gamma_c = \{\underline{x} : \lambda(\underline{x}) \geq c\}$ and $g(\underline{x})$ is the p.d.f for χ_ν^2 -distribution.

We can actually prove equation (2.3) as below.

Proof. Let $Q_n = 2 \log(\lambda(\underline{x}))$, then we can write Q_n as

$$Q_n = \log(\lambda(\underline{x})) = 2[\ell(\hat{\theta}|\underline{x}) - \ell(\theta_0|\underline{x})]$$

where ℓ denotes the log-likelihood function. We can expand $\ell(\theta_0|\underline{x})$ in a Taylor series about $\theta_0 = \hat{\theta}$. According to Taylor's theorem, there is a value $\tilde{\theta}$ between θ_0 and $\hat{\theta}$ s.t.

$$\ell(\theta_0|\underline{x}) = \ell(\hat{\theta}|\underline{x}) + \frac{\partial}{\partial \theta} \ell(\hat{\theta}|\underline{x})(\theta_0 - \hat{\theta}) + \frac{1}{2} \frac{\partial^2}{\partial \theta^2} \ell(\tilde{\theta}|\underline{x})(\theta_0 - \hat{\theta})^2.$$

Since $\frac{\partial}{\partial \theta} \ell(\hat{\theta}|\underline{x}) = 0$ for MLE, then we have

$$Q = -\frac{\partial^2}{\partial \theta^2} \ell(\tilde{\theta}|\underline{x})(\hat{\theta} - \theta_0)^2 = n(-\frac{1}{n} \frac{\partial^2}{\partial \theta^2} \ell(\tilde{\theta}|\underline{x})) \xrightarrow{\mathbb{P}} \mathcal{I}(\theta_0) \quad (2.5)$$

where $\mathcal{I}(\theta_0) = \lim_{n \rightarrow \infty} n^{-1} I_n(\theta_0)$ is the limiting average information under H_0 . Considering that

$$\mathcal{I}(\theta_0)[\sqrt{n}(\hat{\theta} - \theta_0)]^2 \xrightarrow{d} \chi_1^2,$$

together with the Slutsky's theorem, we can derive that under H_0

$$Q \xrightarrow{d} \chi_1^2,$$

□

Example 2.3 (Likelihood Ratio Test for Exponential Distribution). Suppose $\mathbf{X} = (X_1, \dots, X_n) \stackrel{i.i.d.}{\sim} \text{Exponential}(\lambda)$, where we can write the p.d.f. as $f_X(x; \lambda) = \exp\{-x/\lambda\}/\lambda$. Consider testing

$$H_0 : \lambda = 1 \text{ vs. } H_1 : \lambda \neq 1,$$

let us try to develop a likelihood ratio test. We can then write the log-likelihood function as

$$\ell(\lambda; \mathbf{x}) = -n \log(\lambda) - \sum_{i=1}^n \mathbf{x}_i / \lambda,$$

and the score function as

$$U(\lambda; \mathbf{x}) = \frac{\partial}{\partial \lambda} \ell(\lambda; \mathbf{x}) = -\frac{n}{\lambda} + \frac{\sum_{i=1}^n \mathbf{x}_i}{\lambda^2}$$

Setting the score function to 0, we derive the MLE as

$$\hat{\lambda} = \bar{x},$$

check the second derivative of log-likelihood function as $\frac{\partial^2}{\partial \lambda^2} \ell(\hat{\lambda}; \mathbf{x}) = -\frac{n}{\bar{x}} < 0$, thus we obtained the MLE. Then taking the negative partial derivative of the score, we have the observed information as

$$-\frac{\partial^2}{\partial \lambda^2} \ell(\lambda; \mathbf{x}) = -\frac{n}{\lambda^2} + \frac{2 \sum_{i=1}^n \mathbf{x}_i}{\lambda^3}$$

Under the MLE, we have

$$-\frac{\partial^2}{\partial \lambda^2} \ell(\hat{\lambda}; \mathbf{x}) = \frac{n}{\bar{x}}$$

which satisfies

$$-n^{-1} \frac{\partial^2}{\partial \lambda^2} \ell(\hat{\lambda}; \mathbf{x}) = \bar{x}^{-1} \xrightarrow{\mathbb{P}} \mathcal{I}(\lambda_0),$$

then we have

$$\mathcal{I}(\lambda_0)[\sqrt{n}(\hat{\lambda} - \lambda_0)] \xrightarrow{d} \chi_1^2$$

we can then develop the test statistic for the likelihood ratio test as

$$Q = 2[\ell(\hat{\lambda}; \mathbf{x}) - \ell(\lambda_0; \mathbf{x})] = 2[-n \log(\bar{x}) - n + n\bar{x}] = 2n[\bar{x} - 1 - \log(\bar{x})] \xrightarrow{d} \chi_1^2$$

under H_0 .

Definition 2.4 (P-Value). The p-value is the minimum type I error rate obtainable when calling a test significant based on data \mathbf{x} . More formally, let $X_1^*, X_2^*, \dots, X_n^* \sim \mathcal{F}_{\theta_0}$, if test statistic $S(\mathbf{x})$ satisfies that the larger $S(\mathbf{x})$ would yield smaller significant region. In other word, we can write p-value here as

$$p - \text{value}(\mathbf{x}) = \mathbb{P}(S(\mathbf{x}^*) \geq S(\mathbf{x}); \theta_0) \quad (2.5)$$

where we take $S(\mathbf{X}^*)$ as random and $S(\mathbf{X})$ as observed.

Remark. Inappropriate interpretation of p-value is very common, even in some high-profile scientific papers and reports. Here we provide some intuitive interpretation for p-value. If we generate data from the null distribution, p-value could be considered as the probability that the test statistic based on generated data is greater or equal to the observed test statistics.

Example 2.5 (P-Value for Normal Distribution Test). Let $\mathbf{X} = (X_1, \dots, X_n) \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \mathbb{R}$. Consider testing hypothesis

$$H_0 : \theta \leq 2 \text{ vs. } H_1 : \theta > 2$$

with the significance region as $\Gamma_c = \{\mathbf{x} : \bar{x} \geq c\}$ with size $\alpha = 0.5$. We can summarize the testing results by p-value as

$$\begin{aligned} p\text{-value} &= \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[\bar{X} \geq \bar{x}_{obs}] \\ &= \mathbb{P}[\bar{X} \geq \bar{x}_{obs} | \theta = 2] \\ &= \mathbb{P}(\bar{X} - 2 \geq \bar{x}_{obs} - 2 | \theta = 2) \\ &= \mathbb{P}\left(\frac{\bar{X} - 2}{\sqrt{1/n}} \geq \frac{\bar{x}_{obs} - 2}{\sqrt{1/n}} | \theta = 2\right) \\ &= \mathbb{P}\left(Z \geq \frac{\bar{x}_{obs} - 2}{\sqrt{1/n}}\right) \end{aligned}$$

where \bar{X} is a $\mathcal{N}(\theta, 1/n)$ r.v., \bar{x}_{obs} is the observed value of \bar{X} from the sample, and $Z \sim \mathcal{N}(0, 1)$ is a standard normal distributed r.v.

Then, if $n = 20$ and we observe $\bar{x}_{obs} = 2.95$, we can calculate the p-value as

$$p\text{-value} = \mathbb{P}\left[Z \geq \frac{2.95 - 2}{\sqrt{1/20}}\right] = \mathbb{P}(Z \geq 4.25) = 0.00001.$$

If we observe $\bar{x}_{obs} = 2.38$, then we can calculate the p-value as

$$p\text{-value} = \mathbb{P}\left[Z \geq \frac{2.38 - 2}{\sqrt{1/20}}\right] = \mathbb{P}(Z \geq 1.70) = 0.045.$$

From this example, we can see that p-value actually gives us a quantitative measure of how unusual our observed data are under the assumption that the null hypothesis is true.

Example 2.6 (Test for Hardy Weinberg Equilibrium, HWE) Recall the general setting of *HWE* as below. We observe that $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{F}_\theta$, where $X_i \in \{0, 1, 2\}$, $\theta = \{r_0, r_1\}$ which is a two-dimensional parameter space. We have

$$\mathbb{P}(X = k) = r_k, \quad k = 0, 1, 2$$

where $r_2 = 1 - r_0 - r_1$, $\dim(\Theta) = 2$. Now we might want to construct the null hypothesis. Let $p = \frac{r_1}{2} + r_2$, our goal is to test

$$H_0 : X \sim \text{Binomial}(2, p) \text{ vs. } H_1 : X \sim \text{Multinomial}(1, (r_0, r_1, r_2))$$

i.e. we are trying to test

$$H_0 : \theta \in \{p^2, 2p(1-p)\}, \text{ where } p = \frac{r_1}{2} + r_2 \text{ vs. } H_1 : \text{not } H_0$$

For the observed data, we define that

$$n_0 := \#\{X_i = 0\}, \quad n_1 := \#\{X_i = 1\}, \quad n_2 := \#\{X_i = 2\}, \quad n = n_0 + n_1 + n_2$$

Under the unconstrained way of calculation (i.e. considering inbreeding while conducting calculation), we derive the MLE as

$$\text{MLE} : \hat{r}_k = \frac{n_k}{n}, \quad k = 0, 1, 2.$$

Under the constrained way of calculation, we derive the MLE as

$$\text{MLE} : \hat{r}_0^0 = (1 - \hat{p}^2), \quad \hat{r}_1^0 = 2\hat{p}(1 - \hat{p}), \quad \hat{r}_2^0 = \hat{p}^2, \quad \text{where } \hat{p} = \frac{n_1/2 + n_2}{n}$$

Then we have the likelihood function as

$$\mathcal{L}(\underline{\mathbf{x}}; \underline{\mathbf{x}}) \propto r_0^{n_0} r_1^{n_1} r_2^{n_2}$$

Now we can write the test statistic as

$$\lambda(\underline{\mathbf{x}}) = \frac{\hat{r}_0^{n_0} \hat{r}_1^{n_1} \hat{r}_2^{n_2}}{(\hat{r}_0^0)^{n_0} (\hat{r}_1^0)^{n_1} (\hat{r}_2^0)^{n_2}} = \frac{\hat{r}_0^{n_0} \hat{r}_1^{n_1} \hat{r}_2^{n_2}}{(1 - \hat{p})^{2n_0} (2\hat{p}(1 - \hat{p}))^{n_1} (\hat{p}^2)^{n_2}}$$

Then the null distribution should be approximately a χ_ν^2 -distribution, i.e.

$$2 \log(\lambda(\underline{\mathbf{x}})) \sim \chi_\nu^2$$

where $\nu = \dim(\Theta) - \dim(\Theta_0) = 2 - 1 = 1$. Note that here $\dim(\Theta_0) = 1$ since there is only one allele frequency so that we only have one-dimensional parameter.

Example 2.7 (P-value for Multiple Hypothesis Testing). This is my favorite topic about p-value and hypothesis testing. I also think this is a good summary example about all we learned in this week so I post it here. Consider a multiple hypothesis testing problem, which is very common in statistical genetics analysis, where H_0 is simple and $\{T_k : 1 \leq k \leq K\}$ are K independent continuously distributed test statistics for H_0 . Let $\alpha(T_k)$ denote the p-value for the k -th test statistic. We can show that - (a) $\alpha(T_1) \sim \text{Uniform}(0, 1)$ - (b) $-2 \sum_{k=1}^K \log(\alpha(T_k)) \sim \chi_{2K}^2$

Proof. For (a), recall that for a simple hypothesis, the p-value for $T_1 = t$ could be written as

$$\alpha(t) = \mathbb{P}[T_1(X) > t] = 1 - F_{T_1}(t)$$

where $F_{T_1}(t)$ is the c.d.f. of the statistic T_1 . Obviously, the probability that the p-value is outside $[0, 1]$ should be 0. For $t \in (0, 1)$ we have

$$\begin{aligned} \mathbb{P}[\alpha(T_1) < t] &= \mathbb{P}[1 - F_{T_1}(T_1) \leq t] \\ &= \mathbb{P}[F_{T_1}(T_1) \geq 1 - t] \\ &= \mathbb{P}[T_1 \geq F_{T_1}^{-1}(1 - t)] \\ &= 1 - \mathbb{P}[T_1 < F_{T_1}^{-1}(1 - t)] \\ &= 1 - F_{T_1}(F_{T_1}^{-1}(1 - t)) \\ &= 1 - (1 - t) \\ &= t \end{aligned}$$

which is the p.d.f. of a standard Uniform random variable.

For (b), let $X \sim \text{Uniform}(0, 1)$ and $Y = -2 \log(X)$. Then we have

$$\begin{aligned} \mathbb{P}[Y \leq y] &= \mathbb{P}[-2 \log(X) \leq y] \\ &= \mathbb{P}[X \geq e^{-y/2}] \\ &= (1 - e^{-y/2}) \mathbb{1}_{0 < y < 1} \cdot e^{-y/2} \\ &= (1 - e^{-y/2}) \mathbb{1}_{y > 0} \end{aligned}$$

which is the exponential distribution as $Y \sim \text{Exponential}(2) \stackrel{d}{=} \text{Gamma}(1, 2)$. Since each $\alpha(T_k)$ is a standard Uniform according to (a), the $\sum_{k=1}^K -2 \log(\alpha(T_k))$ is a sum of K Gamma r.v.'s with parameters (1, 2), i.e. $\text{Gamma}(K, 2) \stackrel{d}{=} \chi_{2K}^2$. \square

2.3 Hypothesis Testing III: Pivotal Statistic Hypothesis Testing

Definition 3.1 (Pivotal Statistic). The pivotal statistic is a statistic whose distribution does not depend on some specific parameters.

Remark. A statistic should quantify the evidence that the alternative is favored than the null one.

Definition 3.2 (Pivotal Statistic Hypothesis Testing). Suppose we have a pivotal statistic of the form

$$\frac{\hat{\theta} - \theta}{\widehat{se}(\hat{\theta})} \sim \mathcal{N}(0, 1), \quad (3.1)$$

we can develop a pivotal statistic hypothesis test by testing

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

which is a two-side test where θ_0 is a known value. We can design another useful statistic based on this, say observed statistic, as

$$|Z| = \left| \frac{\hat{\theta} - \theta_0}{\widehat{se}(\hat{\theta})} \right| \quad (3.2)$$

Remark. When $\theta = \theta_0$, we have $\frac{\hat{\theta} - \theta}{\widehat{se}(\hat{\theta})} \sim \mathcal{N}(0, 1)$. We can also say that:

$$\frac{\hat{\theta} - \theta}{\widehat{se}(\hat{\theta})} \sim \mathcal{N}(0, 1) \quad (3.3)$$

Remark. In (3.2), larger $|Z|$ means more evidence against H_0 is favored than H_1 .

Example 3.3 (Poisson Pivotal Statistic Hypothesis Testing). Consider $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} \text{Poisson}(\lambda)$, we want to test

$$H_0 : \lambda = 5 \text{ vs. } H_1 : \lambda \neq 5$$

where we take the parameter under null as $\lambda_0 = 5$. We can develop the pivotal statistic in general form as

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\hat{\lambda}/n}} \sim \mathcal{N}(0, 1), \text{ where } \hat{\lambda} = \bar{x}$$

When taking $\lambda = 5$ here, we have the specific test form as

$$\frac{\hat{\lambda} - 5}{\sqrt{\hat{\lambda}/n}} \sim \mathcal{N}(0, 1), \text{ where } \hat{\lambda} = \bar{x}$$

and

$$\frac{\hat{\lambda} - 5}{\sqrt{5/n}} \sim \mathcal{N}(0, 1), \text{ where } \hat{\lambda} = \bar{x}$$

Definition 3.4 (Type I Error Rate for Pivotal Statistic Hypothesis Testing). Suppose c is our significant cut-off, i.e. we detect our hypothesis as significant when $|Z| \geq c$. Let $Z^* \sim \mathcal{N}(0, 1)$, then our type I error rate now is $\mathbb{P}(|Z^*| \geq c)$

Remark. Correspondingly, we can write the p-value as $\mathbb{P}(|Z^*|)$, which implies the significant regions as $\mathbb{P}_c = \{Z : |Z| \geq c\}$.

Remark. We have some general requirements for the pivotal statistic hypothesis test.

1. A statistic that quantify evidence against H_0 in favor of H_1 . This requirement is assessed in term of power. Generally, we want higher power which is greater or equal to the type I error rate.

2. The (appropriate) distribution of the statistic when H_0 is true.

Another important question to ask here is what would the sampling distribution of the p-value be? If we have the true null distribution identified, then the p-value should follow a Uniform(0,1) distribution when our H_0 is true, which is written as

$$\mathbb{P}(p \leq \alpha) = \alpha \text{ when } H_0 \text{ is true} \quad (3.4)$$

To make the error more conservative, we may want

$$\mathbb{P}(p \leq \alpha) \leq \alpha \quad (3.5)$$

Example 3.5 (One-Side Test). Considering testing

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0,$$

we can design our test statistic as

$$Z = \frac{\hat{\theta} - \theta}{\hat{se}(\theta)}$$

where we have the significant region as

$$\Gamma_c = \{z : z > c\}$$

Example 3.6 (Two-Sample Pivotal Statistic). Suppose $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{F}_\theta$, and $Y_1, Y_2, \dots, Y_m \stackrel{i.i.d.}{\sim} \mathcal{F}_\gamma$, where X and Y are independent. Then we have the MLE for θ and γ as $\hat{\theta}_n$ and $\hat{\gamma}_m$ in accordingly, together with their estimated standard error as $\hat{se}(\hat{\theta}_n)$ and $\hat{se}(\hat{\gamma}_m)$. Our goal is to conduct inference on $\theta - \gamma$ (i.e. we are looking at whether θ and γ are equal or not). Considering testing

$$H_0 : \theta - \gamma = 0 \text{ vs. } H_1 : \theta - \gamma \neq 0,$$

we can develop test statistic as

$$\frac{\hat{\theta}_n - \hat{\gamma}_m - (\theta - \gamma)}{\sqrt{\hat{se}(\hat{\theta}_n)^2 + \hat{se}(\hat{\gamma}_m)^2}} \sim \mathcal{N}(0, 1)$$

2.4 Bayesian Inference: General Framework

Motivation. Frequentist inference is straightforward for computation. However, the hypothesis testing is hard for frequentist inference. If we have a prior probability distribution on the parameters, we can make the hypothesis testing easier based on likelihood function. This leads us to the idea of Bayesian inference.

Remark 4.1 (Bayesian Inference). A prior distribution is assumed for the parameters of interest under the setting of Bayesian inference.

Example 4.1. Let random variable $P \sim \text{Uniform}(0, 1)$. Suppose we have the data generating distribution as $X|P = p \sim \text{Binomial}(n, p)$. Now we can write down the p.d.f. of P given X as

$$f(p|X = x) = \frac{\mathbb{P}(X = x|P = p) \cdot f(p)}{\mathbb{P}(X = x)} \quad (4.1)$$

where $f(p) = 1 \cdot \mathbb{1}_{p \in [0, 1]}$ which is the prior distribution. We name the conditional p.d.f. $f(p|X = x)$ as the posterior distribution, i.e. updated distribution of p given $X = x$. Note that for the denominator we have

$$\mathbb{P}(X = x) = \int \mathbb{P}(X = x|P = p^*)f(p^*)dp^* \quad (4.2)$$

which is actually difficult to calculate in the real world.

More generally, we will assume that $(X_1, X_2, \dots, X_n)|\theta \stackrel{i.i.d.}{\sim} \mathcal{F}_\theta$ and θ is drawn from a prior distribution as $\theta \sim \mathcal{F}_\tau$. Then the essential task is to calculate the posterior distribution of $\theta|\underline{X} = \underline{x}$ as

$$f(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)f(\theta)}{f(\underline{x})} = \frac{f(\underline{x}|\theta)f(\theta)}{\int f(\underline{x}|\theta^*)f(\theta^*)d\theta^*} \quad (4.3)$$

If there is indeed a “true” θ , then if certain regularity conditions are met, then $f(\theta|\underline{x})$ will concentrate around the true θ as $n \rightarrow \infty$.

Example 4.2 (Bernoulli-Uniform). Suppose $\mathbf{X} = (X_1, \dots, X_n) \stackrel{i.i.d.}{\sim} \text{Bernoulli}(\theta)$, where $\theta \in (0, 1)$. The prior here is $\theta \sim \text{Uniform}(0, 1)$, which is a Beta(1,1) distribution. We can show that the posterior is also a Beta distribution. Here, we can write the likelihood of the prior / prior density as

$$\pi(\theta) = \mathbb{1}_{0 \leq \theta \leq 1}, \quad \mathcal{L}(\theta|\underline{x}) = \theta^s(1 - \theta)^{n-s}$$

where $s = \sum_{i=1}^n x_i$. We can write the posterior as

$$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int f(\underline{x}|\theta)\pi(\theta)d\theta} = \frac{\theta^s(1 - \theta)^{n-s} \mathbb{1}_{0 \leq \theta \leq 1}}{\int_0^1 \theta^s(1 - \theta)^{n-s}d\theta} = B(s+1, n-s+1)^{-1} \theta^s(1 - \theta)^{n-s} \mathbb{1}_{0 \leq \theta \leq 1}$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. The posterior is the density of the Beta(s+1, n-s+1) distribution.

Remark 4.3. The basis of Bayesian Inference is the Bayes Theorem. As the consequence of Bayes Theorem, we can conclude that - The posterior distribution would only involve the data through likelihood. Therefore, the sufficiency and likelihood principles are automatically satisfied. - Posterior distribution is distribution of θ given all the observed data. - If we have two sets of data \underline{x} and \underline{y} s.t. $\mathcal{L}(\theta|\underline{x}) \propto \mathcal{L}(\theta|\underline{y})$, then we can claim that \underline{x} and \underline{y} will yield the same posterior distribution given the same priors, i.e. $\pi(\theta|\underline{x}) = \pi(\theta|\underline{y})$.

3 References

- [1] Bickel, P. J. and K. A. Doksum (2015): Mathematical Statistics, Prentice Hall.
- [2] Casella, G. and R. L. Berger (2002): Statistical Inference, Duxbury.
- [3] Lehmann, E. L. (1999): Elements of Large-Sample Theory, Springer.

```

> sessionInfo()
R version 3.6.2 (2019-12-12)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS Catalina 10.15.3

Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods    base

other attached packages:
[1] forcats_0.5.0  stringr_1.4.0  dplyr_0.8.4    purrr_0.3.3
[5] readr_1.3.1    tidyr_1.0.2    tibble_2.1.3   ggplot2_3.2.1
[9] tidyverse_1.3.0 knitr_1.28

loaded via a namespace (and not attached):
[1] tidyselect_1.0.0 xfun_0.12      haven_2.2.0    lattice_0.20-38
[5] colorspace_1.4-1 vctrs_0.2.3    generics_0.0.2 htmltools_0.4.0
[9] yaml_2.2.1        rlang_0.4.4    pillar_1.4.3   withr_2.1.2
[13] glue_1.3.1        DBI_1.1.0      dbplyr_1.4.2   modelr_0.1.6
[17] readxl_1.3.1      lifecycle_0.1.0 munsell_0.5.0  gtable_0.3.0
[21] cellranger_1.1.0 rvest_0.3.5    evaluate_0.14  fansi_0.4.1
[25] broom_0.5.5       Rcpp_1.0.3     scales_1.1.0   backports_1.1.5
[29] jsonlite_1.6.1    fs_1.3.2       hms_0.5.3      digest_0.6.25
[33] stringi_1.4.6     grid_3.6.2     cli_2.0.2      tools_3.6.2
[37] magrittr_1.5      lazyeval_0.2.2 crayon_1.3.4    pkgconfig_2.0.3
[41] xml2_1.2.2        reprex_0.3.0   lubridate_1.7.4 assertthat_0.2.1
[45] rmarkdown_2.1     httr_1.4.1     rstudioapi_0.11 R6_2.4.1
[49] nlme_3.1-142      compiler_3.6.2

```