Lecture 7. Simple Hypotheses

Introduction

Let $X_1, \dots, X_n$ be iid, each taking values in $\mathcal{X}$, each with unknown pdf/pmf f, and suppose that we have two hypotheses, H_0 and H_1, about f.

On the basis of data $\mathbf{X} = \mathbf{x}$, we make a choice between the two hypotheses.

Examples

(a) A coin has $\mathbb{P}(\text{Heads}) = \theta$, and is thrown independently n times. We could have $H_0 : \theta = 1/2$ versus $H_1 : \theta = 3/4$.

(b) As in (a), with $H_0 : \theta = 1/2$ as before, but with $H_1 : \theta \neq 1/2$.

(c) Suppose $X_1, \dots, X_n$ are iid discrete rv's. We could have H_0:the distribution is Poisson with unknown mean, and H_1:the distribution is not Poisson. This is a goodness-of-fit test.

(d) General parametric case: $X_1, \dots, X_n$ are iid with density $f(x \mid \theta)$, with $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ where $\Theta_0 \cap \Theta_1 = \emptyset$ (we may or may not have $\Theta_0 \cup \Theta_1 = \Theta$).

(e) We could have $H_0 : f = f_0$ and $H_1 : f = f_1$ where f_0 and f_1 are densities that are completely specified but do not come from the same parametric family.

A **simple hypothesis** H specifies f completely (eg $H_0 : \theta = 1/2$ in (a)).

Otherwise H is a **composite hypothesis** (eg $H_1 : \theta \neq 1/2$ in (b)).

For testing H_0 against an alternative hypothesis H_1, a test procedure has to partition $\mathcal{X}^n$ into two disjoint and exhaustive regions C and $\bar{C}$, such that if $\mathbf{x} \in C$ then H_0 is rejected and if $\mathbf{x} \in \bar{C}$ then H_0 is not rejected.

The **critical region** (or **rejection region**) C defines the test.

When performing a test we may (i) arrive at a correct conclusion, or (ii) make one of two types of error:

(a) we may reject H_0 when H_0 is true (a **Type I error**),

(b) we may not reject H_0 when H_0 is false (a **Type II error**).

NB: When Neyman and Pearson developed the theory in the 1930s, they spoke of 'accepting' H_0. Now we generally refer to '*not rejecting* H_0'.

Testing a simple hypothesis against a simple alternative

When H_0 and H_1 are both simple, let

$$\alpha = \mathbb{P}(\text{Type I error}) = \mathbb{P}(\mathbf{X} \in C \mid H_0 \text{ is true})$$
$$\beta = \mathbb{P}(\text{Type II error}) = \mathbb{P}(\mathbf{X} \notin C \mid H_1 \text{ is true}).$$

We define the **size** of the test to be α.

$1 - \beta$ is also known as the **power** of the test to detect H_1.

Ideally we would like $\alpha = \beta = 0$, but typically it is not possible to find a test that makes both α and β arbitrarily small.

Definition 7.1

- The **likelihood** of a simple hypothesis $H : \theta = \theta^*$ given data $\mathbf{x}$ is $L_{\mathbf{x}}(H) = f_{\mathbf{X}}(\mathbf{x} \mid \theta = \theta^*)$.
- The **likelihood ratio** of two simple hypotheses H_0, H_1, given data $\mathbf{x}$, is $\Lambda_{\mathbf{x}}(H_0; H_1) = L_{\mathbf{x}}(H_1)/L_{\mathbf{x}}(H_0)$.
- A **likelihood ratio test** (LR test) is one where the critical region C is of the form $C = \{\mathbf{x} : \Lambda_{\mathbf{x}}(H_0; H_1) > k\}$ for some k. □

Theorem 7.2

(The Neyman–Pearson Lemma) Suppose $H_0 : f = f_0$, $H_1 : f = f_1$, *where* f_0 *and* f_1 *are continuous densities that are nonzero on the same regions. Then among all tests of size less than or equal to* α, *the test with smallest probability of a Type II error is given by* $C = \{\mathbf{x} : f_1(\mathbf{x})/f_0(\mathbf{x}) > k\}$ *where* k *is chosen such that* $\alpha = \mathbb{P}(\textit{reject } H_0 \,|\, H_0) = \mathbb{P}(\mathbf{X} \in C \,|\, H_0) = \int_C f_0(\mathbf{x})d\mathbf{x}$.

Proof

The given C specifies a likelihood ratio test with size α.

Let $\beta = \mathbb{P}(\mathbf{X} \notin C \,|\, f_1) = \int_{\bar{C}} f_1(\mathbf{x})d\mathbf{x}$.

Let C^* be the critical region of any other test with size less than or equal to α.

Let $\alpha^* = \mathbb{P}(\mathbf{X} \in C^* \,|\, f_0)$, $\beta^* = \mathbb{P}(\mathbf{X} \notin C^* \,|\, f_1)$.

We want to show $\beta \le \beta^*$.

We know $\alpha^* \le \alpha$, ie $\int_{C^*} f_0(\mathbf{x})d\mathbf{x} \le \int_C f_0(\mathbf{x})d\mathbf{x}$.

Also, on C we have $f_1(\mathbf{x}) > kf_0(\mathbf{x})$, while on $\bar{C}$ we have $f_1(\mathbf{x}) \le kf_0(\mathbf{x})$.

Thus

$$\int_{\bar{C}^* \cap C} f_1(\mathbf{x})d\mathbf{x} \ge k \int_{\bar{C}^* \cap C} f_0(\mathbf{x})d\mathbf{x}, \qquad \int_{\bar{C} \cap C^*} f_1(\mathbf{x})d\mathbf{x} \le k \int_{\bar{C} \cap C^*} f_0(\mathbf{x})d\mathbf{x}. \quad (1)$$

Hence

$$\begin{aligned}
\beta - \beta^* &= \int_{\bar{C}} f_1(\mathbf{x})d\mathbf{x} - \int_{\bar{C}^*} f_1(\mathbf{x})d\mathbf{x} \\
&= \int_{\bar{C} \cap C^*} f_1(\mathbf{x})d\mathbf{x} + \int_{\bar{C} \cap \bar{C}^*} f_1(\mathbf{x})d\mathbf{x} - \int_{\bar{C}^* \cap C} f_1(\mathbf{x})d\mathbf{x} - \int_{\bar{C} \cap \bar{C}^*} f_1(\mathbf{x})d\mathbf{x} \\
&\le k \int_{\bar{C} \cap C^*} f_0(\mathbf{x})d\mathbf{x} - k \int_{\bar{C}^* \cap C} f_0(\mathbf{x})d\mathbf{x} \qquad \text{by (??)} \\
&= k \left\{ \int_{\bar{C} \cap C^*} f_0(\mathbf{x})d\mathbf{x} + \int_{C \cap C^*} f_0(\mathbf{x})d\mathbf{x} \right\} \\
&\quad -k \left\{ \int_{\bar{C}^* \cap C} f_0(\mathbf{x})d\mathbf{x} + \int_{C \cap C^*} f_0(\mathbf{x})d\mathbf{x} \right\} \\
&= k(\alpha^* - \alpha) \\
&\le 0.
\end{aligned}$$

□

- We assume continuous densities to ensure that a LR test of exactly size α exists.
- The Neyman–Pearson Lemma shows that α and β cannot both be arbitrarily small.
- It says that the most powerful test (ie the one with the smallest Type II error probability), among tests with size smaller than or equal to α, is the size α likelihood ratio test.
- Thus we should fix $\mathbb{P}$(Type I error) at some level α and then use the Neyman–Pearson Lemma to find the best test.
- Here the hypotheses are not treated symmetrically; H_0 has precedence over H_1 and a Type I error is treated as more serious than a Type II error.
- H_0 is called the **null hypothesis** and H_1 is called the **alternative hypothesis**.
- The null hypothesis is a conservative hypothesis, ie one of "no change," "no bias," "no association," and is only rejected if we have clear evidence against it.
- H_1 represents the kind of departure from H_0 that is of interest to us.

Example 7.3

Suppose that $X_1, \dots, X_n$ are iid $N(\mu, \sigma_0^2)$, where σ_0^2 is known. We want to find the best size α test of $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$, where μ_0 and μ_1 are known fixed values with $\mu_1 > \mu_0$.

$$\begin{aligned}
\Lambda_{\mathbf{x}}(H_0; H_1) &= \frac{(2\pi\sigma_0^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum(x_i - \mu_1)^2\right)}{(2\pi\sigma_0^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum(x_i - \mu_0)^2\right)} \\
&= \exp\left(\frac{(\mu_1 - \mu_0)}{\sigma_0^2} n\bar{x} + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma_0^2}\right) \qquad \text{(check).}
\end{aligned}$$

- This is an increasing function of $\bar{x}$, so for any k,

$$\Lambda_{\mathbf{x}} > k \Leftrightarrow \bar{x} > c \text{ for some } c.$$

- Hence we reject H_0 if $\bar{x} > c$ where c is chosen such that $\mathbb{P}(\bar{X} > c \,|\, H_0) = \alpha$.
- Under H_0, $\bar{X} \sim N(\mu_0, \sigma_0^2/n)$, so $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma_0 \sim N(0,1)$.
- Since $\bar{x} > c \Leftrightarrow z > c'$ for some c', the size α test rejects H_0 if $z = \sqrt{n}(\bar{x} - \mu_0)/\sigma_0 > z_\alpha$.

- Suppose $\mu_0 = 5$, $\mu_1 = 6$, $\sigma_0 = 1$, $\alpha = 0.05$, $n = 4$ and $\mathbf{x} = (5.1, 5.5, 4.9, 5.3)$, so that $\bar{x} = 5.2$.
- From tables, $z_{0.05} = 1.645$.
- We have $z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0} = 0.4$ and this is less than 1.645, so $\mathbf{x}$ is not in the rejection region.
- We do not reject H_0 at the 5%- level; the data are consistent with H_0.
- This does not mean that H_0 is 'true', just that it cannot be ruled out.
- This is called a z-**test**. □

P-values

- In this example, LR tests reject H_0 if $z > k$ for some constant k.
- The size of such a test is $\alpha = \mathbb{P}(Z > k \,|\, H_0) = 1 - \Phi(k)$, and is decreasing as k increases.
- Our observed value z will be in the rejection region $\Leftrightarrow z > k \Leftrightarrow \alpha > p^* = \mathbb{P}(Z > z \,|\, H_0)$.
- The quantity p^* is called the p-**value** of our observed data $\mathbf{x}$.
- For Example 7.3, $z = 0.4$ and so $p^* = 1 - \Phi(0.4) = 0.3446$.
- In general, the p-value is sometimes called the 'observed significance level' of $\mathbf{x}$ and is the probability under H_0 of seeing data that are 'more extreme' than our observed data $\mathbf{x}$.
- Extreme observations are viewed as providing evidence againt H_0.

* The p-value has a Uniform(0,1) pdf under the null hypothesis. To see this for a z-test, note that

$$\mathbb{P}(p* < p \mid H_0) = \mathbb{P}\left([1 - \Phi(Z)] < p \mid H_0\right) = \mathbb{P}(Z > \Phi^{-1}(1-p) \mid H_0)$$
$$= 1 - \Phi\left(\Phi^{-1}(1-p)\right) = 1 - (1-p) = p.$$