Lecture 25

25.1 Goodness-of-fit for composite hypotheses.

(Textbook, Section 9.2)

Suppose that we have a sample of random variables $X_1, \ldots, X_n$ that can take a finite number of values $B_1, \ldots, B_r$ with unknown probabilities

$$p_1 = \mathbb{P}(X = B_1), \ldots, p_r = \mathbb{P}(X = B_r)$$

and suppose that we want to test the hypothesis that this distribution comes from a parameteric family $\{\mathbb{P}_\theta : \theta \in \Theta\}$. In other words, if we denote $p_j(\theta) = \mathbb{P}_\theta(X = B_j)$, we want to test:

$$\begin{cases} H_1 : & p_j = p_j(\theta) \text{ for all } j \leq r \text{ for some } \theta \in \Theta \\ H_2 : & \text{otherwise.} \end{cases}$$

If we wanted to test H_1 for one particular fixed θ we could use the statistic

$$T = \sum_{j=1}^{r} \frac{(\nu_j - np_j(\theta))^2}{np_j(\theta)},$$

and use a simple χ^2 test from last lecture. The situation now is more complicated because we want to test if $p_j = p_j(\theta), j \leq r$ at least for some $\theta \in \Theta$ which means that we have many candidates for θ. One way to approach this problem is as follows.

(Step 1) Assuming that hypothesis H_1 holds, i.e. $\mathbb{P} = \mathbb{P}_\theta$ for some $\theta \in \Theta$, we can find an estimate θ^* of this unknown θ and then

(Step 2) try to test whether indeed the distribution $\mathbb{P}$ is equal to $\mathbb{P}_{\theta^*}$ by using the statistics

$$T = \sum_{j=1}^{r} \frac{(\nu_j - np_j(\theta^*))^2}{np_j(\theta^*)}$$

in χ^2 test.

This approach looks natural, the only question is what estimate θ^* to use and how the fact that θ^* also depends on the data will affect the convergence of T. It turns out that if we let θ^* be the maximum likelihood estimate, i.e. θ that maximizes the likelihood function

$$\varphi(\theta) = p_1(\theta)^{\nu_1} \ldots p_r(\theta)^{\nu_r}$$

then the statistic

$$T = \sum_{j=1}^{r} \frac{(\nu_j - np_j(\theta^*))^2}{np_j(\theta^*)} \to \chi^2_{r-s-1}$$

converges to χ^2_{r-s-1} distribution with $r - s - 1$ degrees of freedom, where s is the dimension of the parameter set Θ. Of course, here we assume that $s \leq r - 2$ so that we have at least one degree of freedom. Very informally, by dimension we understand the number of free parameters that describe the set Θ, which we illustrate by the following examples.

1. The family of Bernoulli distributions $B(p)$ has only one free parameter $p \in [0, 1]$ so that the set $\Theta = [0, 1]$ has dimension $s = 1$.

2. The family of normal distributions $N(\mu, \sigma^2)$ has two free parameters $\mu \in \mathbb{R}$ and $\sigma^2 \geq 0$ and the set $\Theta = \mathbb{R} \times [0, \infty)$ has dimension $s = 2$.

3. Let us consider a family of all distributions on the set $\{0, 1, 2\}$. The distribution

$$\mathbb{P}(X = 0) = p_1, \mathbb{P}(X = 1) = p_2, \mathbb{P}(X = 2) = p_3$$

is described by parameters p_1, p_2 and p_3. But since they are supposed to add up to 1, $p_1 + p_2 + p_3 = 1$, one of these parameters is not free, for example, $p_3 = 1 - p_1 - p_2$. The remaining two parameters belong to a set

$$p_1 \in [0, 1], \;\; p_2 \in [0, 1 - p_1]$$

shown in figure 25.1, since their sum should not exceed 1 and the dimension of this set is $s = 2$.

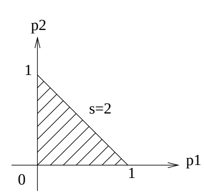

Figure 25.1: Free parameters of a three point distribution.

Example. (textbook, p.545) Suppose that a gene has two possible alleles A_1 and A_2 and the combinations of theses alleles define there possible genotypes A_1A_1, A_1A_2 and A_2A_2. We want to test a theory that

$$\left.\begin{array}{rl} \text{Probability to pass } A_1 \text{ to a child} & = \theta : \\ \text{Probability to pass } A_2 \text{ to a child} & = 1-\theta : \end{array}\right\}$$

and the probabilities of genotypes are given by

$$\begin{aligned} p_1(\theta) &= \mathbb{P}(A_1A_1) = \theta^2 \\ p_2(\theta) &= \mathbb{P}(A_1A_2) = 2\theta(1-\theta) \\ p_3(\theta) &= \mathbb{P}(A_2A_2) = (1-\theta)^2 \end{aligned} \tag{25.1}$$

Suppose that given the sample $X_1, \ldots, X_n$ of the population the counts of each genotype are ν_1, ν_2 and ν_3. To test the theory we want to test the hypotheses

$$\begin{cases} H_1 : & p_1 = p_1(\theta),\ p_2 = p_2(\theta),\ p_3 = p_3(\theta) \text{ for some } \theta \in [0,1] \\ H_2 : & \text{otherwise.} \end{cases}$$

First of all, the dimension of the parameter set is $s = 1$ since the family of distributions in (25.1) are described by one parameter θ. To find the MLE θ^* we have to maximize the likelihood function

$$p_1(\theta)^{\nu_1} p_2(\theta)^{\nu_2} p_3(\theta)^{\nu_3}$$

or, equivalently, maximize the log-likelihood

$$\begin{aligned} \log p_1(\theta)^{\nu_1} p_2(\theta)^{\nu_2} p_3(\theta)^{\nu_3} &= \nu_1 \log p_1(\theta) + \nu_2 \log p_2(\theta) + \nu_3 \log p_3(\theta) \\ &= \nu_1 \log \theta^2 + \nu_2 \log 2\theta(1-\theta) + \nu_3 \log(1-\theta)^2. \end{aligned}$$

To find the critical point we take the derivative, set it equal to 0 and solve for θ which gives (we omit these simple steps):

$$\theta^* = \frac{2\nu_1 + \nu_2}{2n}.$$

Therefore, under the null hypothesis H_1 the statistic

$$\begin{aligned} T &= \frac{(\nu_1 - np_1(\theta^*))^2}{np_1(\theta^*)} + \frac{(\nu_2 - np_2(\theta^*))^2}{np_2(\theta^*)} + \frac{(\nu_3 - np_3(\theta^*))^2}{np_3(\theta^*)} \\ &\to \chi^2_{r-s-1} = \chi^2_{3-1-1} = \chi^2_1 \end{aligned}$$

converges to χ^2_1 distribution with one degree of freedom. If we take the level of significance $\alpha = 0.05$ and find the threshold c so that

$$0.05 = \alpha = \chi^2_1(T > c) \Rightarrow c = 3.841$$

then we can use the following decision rule:

$$\begin{cases} H_1 : & T \leq c = 3.841 \\ H_2 : & T > c = 3.841 \end{cases}$$

□

General families.

We could use a similar test when the distributions $\mathbb{P}_\theta, \theta \in \Theta$ are not necessarily supported by a finite number of points $B_1, \ldots, B_r$ (for example, continuous distributions). In this case if we want to test the hypotheses

$$\begin{cases} H_1 : & \mathbb{P} = \mathbb{P}_\theta \text{ for some } \theta \in \Theta \\ H_2 : & \text{otherwise} \end{cases}$$

we can discretize them as we did in the last lecture (see figure 25.2), i.e. consider a family of distributions

$$p_j(\theta) = \mathbb{P}_\theta(X \in I_j) \text{ for } j \leq r,$$

and instead consider derivative hypotheses

$$\begin{cases} H_1 : & p_j = p_j(\theta) \text{ for some } \theta, j = 1, \cdots, r \\ H_2 : & \text{otherwise.} \end{cases}$$

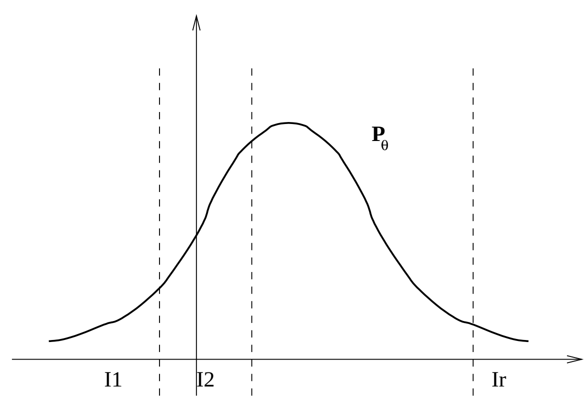

Figure 25.2: Goodness-of-fit for Composite Hypotheses.