

10.18 Example (Mendel's peas). Mendel bred peas with round yellow seeds and wrinkled green seeds. There are four types of progeny: round yellow, wrinkled yellow, round green, and wrinkled green. The number of each type is multinomial with probability $p = (p_1, p_2, p_3, p_4)$. His theory of inheritance predicts that p is equal to

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

In $n = 556$ trials he observed $X = (315, 101, 108, 32)$. We will test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$. Since, $np_{01} = 312.75$, $np_{02} = np_{03} = 104.25$, and $np_{04} = 34.75$, the test statistic is

$$\begin{aligned} \chi^2 &= \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} \\ &\quad + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.47. \end{aligned}$$

The $\alpha = .05$ value for a χ^2_3 is 7.815. Since 0.47 is not larger than 7.815 we do not reject the null. The p-value is

$$\text{p-value} = \mathbb{P}(\chi^2_3 > .47) = .93$$

which is not evidence against H_0 . Hence, the data do not contradict Mendel's theory.³■

In the previous example, one could argue that hypothesis testing is not the right tool. Hypothesis testing is useful to see if there is evidence to reject H_0 . This is appropriate when H_0 corresponds to the status quo. It is not useful for proving that H_0 is true. Failure to reject H_0 might occur because H_0 is true, but it might occur just because the test has low power. Perhaps a confidence set for the distance between p and p_0 might be more useful in this example.

10.5 The Permutation Test

The permutation test is a nonparametric method for testing whether two distributions are the same. This test is “exact,” meaning that it is not based on large sample theory approximations. Suppose that $X_1, \dots, X_m \sim F_X$ and $Y_1, \dots, Y_n \sim F_Y$ are two independent samples and H_0 is the hypothesis that

³There is some controversy about whether Mendel's results are “too good.”

the two samples are identically distributed. This is the type of hypothesis we would consider when testing whether a treatment differs from a placebo. More precisely we are testing

$$H_0 : F_X = F_Y \quad \text{versus} \quad H_1 : F_X \neq F_Y.$$

Let $T(x_1, \dots, x_m, y_1, \dots, y_n)$ be some test statistic, for example,

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

Let $N = m + n$ and consider forming all $N!$ permutations of the data $X_1, \dots, X_m, Y_1, \dots, Y_n$. For each permutation, compute the test statistic T . Denote these values by $T_1, \dots, T_{N!}$. Under the null hypothesis, each of these values is equally likely.⁴ The distribution \mathbb{P}_0 that puts mass $1/N!$ on each T_j is called the **permutation distribution** of T . Let t_{obs} be the observed value of the test statistic. Assuming we reject when T is large, the p-value is

$$\text{p-value} = \mathbb{P}_0(T > t_{\text{obs}}) = \frac{1}{N!} \sum_{j=1}^{N!} I(T_j > t_{\text{obs}}).$$

10.19 Example. Here is a toy example to make the idea clear. Suppose the data are: $(X_1, X_2, Y_1) = (1, 9, 3)$. Let $T(X_1, X_2, Y_1) = |\bar{X} - \bar{Y}| = 2$. The permutations are:

permutation	value of 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

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

Usually, it is not practical to evaluate all $N!$ permutations. We can approximate the p-value by sampling randomly from the set of permutations. The fraction of times $T_j > t_{\text{obs}}$ among these samples approximates the p-value.

⁴More precisely, under the null hypothesis, given the ordered data values, $X_1, \dots, X_m, Y_1, \dots, Y_n$ is uniformly distributed over the $N!$ permutations of the data.

Algorithm for Permutation Test

1. Compute the observed value of the test statistic

$$t_{\text{obs}} = T(X_1, \dots, X_m, Y_1, \dots, Y_n).$$
2. Randomly permute the data. Compute the statistic again using the permuted data.
3. Repeat the previous step B times and let T_1, \dots, T_B denote the resulting values.
4. The approximate p-value is

$$\frac{1}{B} \sum_{j=1}^B I(T_j > t_{\text{obs}}).$$

10.20 Example. DNA microarrays allow researchers to measure the expression levels of thousands of genes. The data are the levels of messenger RNA (mRNA) of each gene, which is thought to provide a measure of how much protein that gene produces. Roughly, the larger the number, the more active the gene. The table below, reproduced from Efron et al. (2001) shows the expression levels for genes from ten patients with two types of liver cancer cells. There are 2,638 genes in this experiment but here we show just the first two. The data are log-ratios of the intensity levels of two different color dyes used on the arrays.

	Type I					Type II				
Patient	1	2	3	4	5	6	7	8	9	10
Gene 1	230	-1,350	-1,580	-400	-760	970	110	-50	-190	-200
Gene 2	470	-850	-.8	-280	120	390	-1730	-1360	-1	-330
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Let's test whether the median level of gene 1 is different between the two groups. Let ν_1 denote the median level of gene 1 of Type I and let ν_2 denote the median level of gene 1 of Type II. The absolute difference of sample medians is $T = |\hat{\nu}_1 - \hat{\nu}_2| = 710$. Now we estimate the permutation distribution by simulation and we find that the estimated p-value is .045. Thus, if we use a $\alpha = .05$ level of significance, we would say that there is evidence to reject the null hypothesis of no difference. ■

In large samples, the permutation test usually gives similar results to a test that is based on large sample theory. The permutation test is thus most useful for small samples.

10.6 The Likelihood Ratio Test

The Wald test is useful for testing a scalar parameter. The likelihood ratio test is more general and can be used for testing a vector-valued parameter.

10.21 Definition. *Consider testing*

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \notin \Theta_0.$$

The likelihood ratio statistic is

$$\lambda = 2 \log \left(\frac{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)} \right) = 2 \log \left(\frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\hat{\theta}_0)} \right)$$

where $\hat{\theta}$ is the MLE and $\hat{\theta}_0$ is the MLE when θ is restricted to lie in Θ_0 .

You might have expected to see the maximum of the likelihood over Θ_0^c instead of Θ in the numerator. In practice, replacing Θ_0^c with Θ has little effect on the test statistic. Moreover, the theoretical properties of λ are much simpler if the test statistic is defined this way.

The likelihood ratio test is most useful when Θ_0 consists of all parameter values θ such that some coordinates of θ are fixed at particular values.

10.22 Theorem. *Suppose that $\theta = (\theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_r)$. Let*

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

Let λ be the likelihood ratio test statistic. Under $H_0 : \theta \in \Theta_0$,

$$\lambda(x^n) \rightsquigarrow \chi_{r-q, \alpha}^2$$

where $r - q$ is the dimension of Θ minus the dimension of Θ_0 . The p-value for the test is $\mathbb{P}(\chi_{r-q}^2 > \lambda)$.

For example, if $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ and we want to test the null hypothesis that $\theta_4 = \theta_5 = 0$ then the limiting distribution has $5 - 3 = 2$ degrees of freedom.