

9.6: Chi-Square Tests

In this section, we will study a number of important hypothesis tests that fall under the general term *chi-square tests*. These are named, as you might guess, because in each case the test statistics has (in the limit) a chi-square distribution. Although there are several different tests in this general category, they all share some common themes:

- In each test, there are one or more underlying multinomial samples. Of course, the multinomial model includes the Bernoulli model as a special case.
- Each test works by comparing the *observed frequencies* of the various outcomes with *expected frequencies* under the null hypothesis.
- If the model is *incompletely specified*, some of the expected frequencies must be estimated; this reduces the degrees of freedom in the limiting chi-square distribution.

We will start with the simplest case, where the derivation is the most straightforward; in fact this test is equivalent to a test we have already studied. We then move to successively more complicated models.

The One-Sample Bernoulli Model

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$. Thus, these are independent random variables taking the values 1 and 0 with probabilities p and $1 - p$ respectively. We want to test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, where $p_0 \in (0, 1)$ is specified. Of course, we have already studied such tests in the Bernoulli model. But keep in mind that our methods in this section will generalize to a variety of new models that we have not yet studied.

Let $O_1 = \sum_{j=1}^n X_j$ and $O_0 = n - O_1 = \sum_{j=1}^n (1 - X_j)$. These statistics give the number of times (frequency) that outcomes 1 and 0 occur, respectively. Moreover, we know that each has a binomial distribution; O_1 has parameters n and p , while O_0 has parameters n and $1 - p$. In particular, $\mathbb{E}(O_1) = np$, $\mathbb{E}(O_0) = n(1 - p)$, and $\text{var}(O_1) = \text{var}(O_0) = np(1 - p)$. Moreover, recall that O_1 is sufficient for p . Thus, any good test statistic should be a function of O_1 . Next, recall that when n is large, the distribution of O_1 is approximately normal, by the central limit theorem. Let

$$Z = \frac{O_1 - np_0}{\sqrt{np_0(1 - p_0)}} \quad (9.6.1)$$

Note that Z is the standard score of O_1 under H_0 . Hence if n is large, Z has approximately the standard normal distribution under H_0 , and therefore $V = Z^2$ has approximately the chi-square distribution with 1 degree of freedom under H_0 . As usual, let χ_k^2 denote the quantile function of the chi-square distribution with k degrees of freedom.

An approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_1^2(1 - \alpha)$.

The test above is equivalent to the unbiased test with test statistic Z (the approximate normal test) derived in the section on Tests in the Bernoulli model.

For purposes of generalization, the critical result in the next exercise is a special representation of V . Let $e_0 = n(1 - p_0)$ and $e_1 = np_0$. Note that these are the expected frequencies of the outcomes 0 and 1, respectively, under H_0 .

V can be written in terms of the observed and expected frequencies as follows:

$$V = \frac{(O_0 - e_0)^2}{e_0} + \frac{(O_1 - e_1)^2}{e_1} \quad (9.6.2)$$

This representation shows that our test statistic V measures the discrepancy between the expected frequencies, under H_0 , and the observed frequencies. Of course, large values of V are evidence in favor of H_1 . Finally, note that although there are two terms in the expansion of V in Exercise 3, there is only one degree of freedom since $O_0 + O_1 = n$. The observed and expected frequencies could be stored in a 1×2 table.

The Multi-Sample Bernoulli Model

Suppose now that we have samples from several (possibly) different, independent Bernoulli trials processes. Specifically, suppose that $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,n_i})$ is a random sample of size n_i from the Bernoulli distribution with unknown success parameter $p_i \in (0, 1)$ for each $i \in \{1, 2, \dots, m\}$. Moreover, the samples $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$ are independent. We want to test hypotheses about the unknown parameter vector $\mathbf{p} = (p_1, p_2, \dots, p_m)$. There are two common cases that we consider below, but first let's set up the essential notation

that we will need for both cases. For $i \in \{1, 2, \dots, m\}$ and $j \in \{0, 1\}$, let $O_{i,j}$ denote the number of times that outcome j occurs in sample \mathbf{X}_i . The observed frequency $O_{i,j}$ has a binomial distribution; $O_{i,1}$ has parameters n_i and p_i while $O_{i,0}$ has parameters n_i and $1 - p_i$.

The Completely Specified Case

Consider a specified parameter vector $\mathbf{p}_0 = (p_{0,1}, p_{0,2}, \dots, p_{0,m}) \in (0, 1)^m$. We want to test the null hypothesis $H_0 : \mathbf{p} = \mathbf{p}_0$, versus $H_1 : \mathbf{p} \neq \mathbf{p}_0$. Since the null hypothesis specifies the value of p_i for each i , this is called the *completely specified case*. Now let $e_{i,0} = n_i(1 - p_{i,0})$ and let $e_{i,1} = n_i p_{i,0}$. These are the expected frequencies of the outcomes 0 and 1, respectively, from sample \mathbf{X}_i under H_0 .

If n_i is large for each i , then under H_0 the following test statistic has approximately the chi-square distribution with m degrees of freedom:

$$V = \sum_{i=1}^m \sum_{j=0}^1 \frac{(O_{i,j} - e_{i,j})^2}{e_{i,j}} \quad (9.6.3)$$

Proof

This follows from the result [above](#) and independence.

As a rule of thumb, “large” means that we need $e_{i,j} \geq 5$ for each $i \in \{1, 2, \dots, m\}$ and $j \in \{0, 1\}$. But of course, the larger these expected frequencies the better.

Under the large sample assumption, an approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_m^2(1 - \alpha)$.

Once again, note that the test statistic V measures the discrepancy between the expected and observed frequencies, over all outcomes and all samples. There are $2m$ terms in the expansion of V in Exercise 4, but only m degrees of freedom, since $O_{i,0} + O_{i,1} = n_i$ for each $i \in \{1, 2, \dots, m\}$. The observed and expected frequencies could be stored in an $m \times 2$ table.

The Equal Probability Case

Suppose now that we want to test the null hypothesis $H_0 : p_1 = p_2 = \dots = p_m$ that all of the success probabilities are the same, versus the complementary alternative hypothesis H_1 that the probabilities are not all the same. Note, in contrast to the previous model, that the null hypothesis does not specify the value of the common success probability p . But note also that under the null hypothesis, the m samples can be combined to form one large sample of Bernoulli trials with success probability p . Thus, a natural approach is to estimate p and then define the test statistic that measures the discrepancy between the expected and observed frequencies, just as before. The challenge will be to find the distribution of the test statistic.

Let $n = \sum_{i=1}^m n_i$ denote the total sample size when the samples are combined. Then the overall sample mean, which in this context is the overall sample proportion of successes, is

$$P = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} X_{i,j} = \frac{1}{n} \sum_{i=1}^m O_{i,1} \quad (9.6.4)$$

The sample proportion P is the best estimate of p , in just about any sense of the word. Next, let $E_{i,0} = n_i(1 - P)$ and $E_{i,1} = n_i P$. These are the *estimated* expected frequencies of 0 and 1, respectively, from sample \mathbf{X}_i under H_0 . Of course these estimated frequencies are now *statistics* (and hence random) rather than parameters. Just as before, we define our test statistic

$$V = \sum_{i=1}^m \sum_{j=0}^1 \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (9.6.5)$$

It turns out that under H_0 , the distribution of V converges to the chi-square distribution with $m - 1$ degrees of freedom as $n \rightarrow \infty$.

An approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_{m-1}^2(1 - \alpha)$.

Intuitively, we lost a degree of freedom over the [completely specified case](#) because we had to estimate the unknown common success probability p . Again, the observed and expected frequencies could be stored in an $m \times 2$ table.

The One-Sample Multinomial Model

Our next model generalizes the [one-sample Bernoulli model](#) in a different direction. Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sequence of multinomial trials. Thus, these are independent, identically distributed random variables, each taking values in a set S with k elements. If we want, we can assume that $S = \{0, 1, \dots, k-1\}$; the one-sample Bernoulli model then corresponds to $k = 2$. Let f denote the common probability density function of the sample variables on S , so that $f(j) = \mathbb{P}(X_i = j)$ for $i \in \{1, 2, \dots, n\}$ and $j \in S$. The values of f are assumed unknown, but of course we must have $\sum_{j \in S} f(j) = 1$, so there are really only $k-1$ unknown parameters. For a given probability density function f_0 on S we want to test $H_0 : f = f_0$ versus $H_1 : f \neq f_0$.

By this time, our general approach should be clear. We let O_j denote the number of times that outcome $j \in S$ occurs in sample \mathbf{X} :

$$O_j = \sum_{i=1}^n \mathbf{1}(X_i = j) \quad (9.6.6)$$

Note that O_j has the binomial distribution with parameters n and $f(j)$. Thus, $e_j = n f_0(j)$ is the expected number of times that outcome j occurs, under H_0 . Our test statistic, of course, is

$$V = \sum_{j \in S} \frac{(O_j - e_j)^2}{e_j} \quad (9.6.7)$$

It turns out that under H_0 , the distribution of V converges to the chi-square distribution with $k-1$ degrees of freedom as $n \rightarrow \infty$. Note that there are k terms in the expansion of V , but only $k-1$ degrees of freedom since $\sum_{j \in S} O_j = n$.

An approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_{k-1}^2(1-\alpha)$.

Again, as a rule of thumb, we need $e_j \geq 5$ for each $j \in S$, but the larger the expected frequencies the better.

The Multi-Sample Multinomial Model

As you might guess, our final generalization is to the multi-sample multinomial model. Specifically, suppose that $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,n_i})$ is a random sample of size n_i from a distribution on a set S with k elements, for each $i \in \{1, 2, \dots, m\}$. Moreover, we assume that the samples $(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$ are independent. Again there is no loss in generality if we take $S = \{0, 1, \dots, k-1\}$. Then $k=2$ reduces to the [multi-sample Bernoulli model](#), and $m=1$ corresponds to the [one-sample multinomial model](#).

Let f_i denote the common probability density function of the variables in sample \mathbf{X}_i , so that $f_i(j) = \mathbb{P}(X_{i,l} = j)$ for $i \in \{1, 2, \dots, m\}$, $l \in \{1, 2, \dots, n_i\}$, and $j \in S$. These are generally unknown, so that our vector of parameters is the vector of probability density functions: $\mathbf{f} = (f_1, f_2, \dots, f_m)$. Of course, $\sum_{j \in S} f_i(j) = 1$ for $i \in \{1, 2, \dots, m\}$ so there are actually $m(k-1)$ unknown parameters. We are interested in testing hypotheses about \mathbf{f} . As in the multi-sample Bernoulli model, there are two common cases that we consider below, but first let's set up the essential notation that we will need for both cases. For $i \in \{1, 2, \dots, m\}$ and $j \in S$, let $O_{i,j}$ denote the number of times that outcome j occurs in sample \mathbf{X}_i . The observed frequency $O_{i,j}$ has a binomial distribution with parameters n_i and $f_i(j)$.

The Completely Specified Case

Consider a given vector of probability density functions on S , denoted $\mathbf{f}_0 = (f_{0,1}, f_{0,2}, \dots, f_{0,m})$. We want to test the null hypothesis $H_0 : \mathbf{f} = \mathbf{f}_0$, versus $H_1 : \mathbf{f} \neq \mathbf{f}_0$. Since the null hypothesis specifies the value of $f_i(j)$ for each i and j , this is called the *completely specified case*. Let $e_{i,j} = n_i f_{0,i}(j)$. This is the expected frequency of outcome j in sample \mathbf{X}_i under H_0 .

If n_i is large for each i , then under H_0 , the test statistic V below has approximately the chi-square distribution with $m(k-1)$ degrees of freedom:

$$V = \sum_{i=1}^m \sum_{j \in S} \frac{(O_{i,j} - e_{i,j})^2}{e_{i,j}} \quad (9.6.8)$$

Proof

This follows from the [one-sample multinomial case](#) and independence.

As usual, our rule of thumb is that we need $e_{i,j} \geq 5$ for each $i \in \{1, 2, \dots, m\}$ and $j \in S$. But of course, the larger these expected frequencies the better.

Under the large sample assumption, an approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_{m(k-1)}^2(1-\alpha)$.

As always, the test statistic V measures the discrepancy between the expected and observed frequencies, over all outcomes and all samples. There are mk terms in the expansion of V in Exercise 8, but we lose m degrees of freedom, since $\sum_{j \in S} O_{i,j} = n_i$ for each $i \in \{1, 2, \dots, m\}$.

The Equal PDF Case

Suppose now that we want to test the null hypothesis $H_0 : f_1 = f_2 = \dots = f_m$ that all of the probability density functions are the same, versus the complementary alternative hypothesis H_1 that the probability density functions are not all the same. Note, in contrast to the previous model, that the null hypothesis does not specify the value of the common success probability density function f . But note also that under the null hypothesis, the m samples can be combined to form one large sample of multinomial trials with probability density function f . Thus, a natural approach is to estimate the values of f and then define the test statistic that measures the discrepancy between the expected and observed frequencies, just as before.

Let $n = \sum_{i=1}^m n_i$ denote the total sample size when the samples are combined. Under H_0 , our best estimate of $f(j)$ is

$$P_j = \frac{1}{n} \sum_{i=1}^m O_{i,j} \quad (9.6.9)$$

Hence our estimate of the expected frequency of outcome j in sample \mathbf{X}_i under H_0 is $E_{i,j} = n_i P_j$. Again, this estimated frequency is now a *statistic* (and hence random) rather than a parameter. Just as before, we define our test statistic

$$V = \sum_{i=1}^m \sum_{j \in S} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (9.6.10)$$

As you no doubt expect by now, it turns out that under H_0 , the distribution of V converges to a chi-square distribution as $n \rightarrow \infty$. But let's see if we can determine the degrees of freedom heuristically.

The limiting distribution of V has $(k-1)(m-1)$ degrees of freedom.

Proof

There are km terms in the expansion of V . We lose m degrees of freedom since $\sum_{j \in S} O_{i,j} = n_i$ for each $i \in \{1, 2, \dots, m\}$. We must estimate all but one of the probabilities $f(j)$ for $j \in S$, thus losing $k-1$ degrees of freedom.

An approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi_{(k-1)(m-1)}^2(1-\alpha)$.

A Goodness of Fit Test

A *goodness of fit test* is an hypothesis test that an unknown sampling distribution is a particular, specified distribution or belongs to a parametric family of distributions. Such tests are clearly fundamental and important. The [one-sample multinomial model](#) leads to a quite general goodness of fit test.

To set the stage, suppose that we have an observable random variable X for an experiment, taking values in a general set S . Random variable X might have a continuous or discrete distribution, and might be single-variable or multi-variable. We want to test the null hypothesis that X has a given, completely specified distribution, or that the distribution of X belongs to a particular parametric family.

Our first step, in either case, is to sample from the distribution of X to obtain a sequence of independent, identically distributed variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Next, we select $k \in \mathbb{N}_+$ and partition S into k (disjoint) subsets. We will denote the partition by $\{A_j : j \in J\}$ where $\#(J) = k$. Next, we define the sequence of random variables $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ by $Y_i = j$ if and only if $X_i \in A_j$ for $i \in \{1, 2, \dots, n\}$ and $j \in J$.

\mathbf{Y} is a multinomial trials sequence with parameters n and f , where $f(j) = \mathbb{P}(X \in A_j)$ for $j \in J$.

The Completely Specified Case

Let H denote the statement that X has a given, completely specified distribution. Let f_0 denote the probability density function on J defined by $f_0(j) = \mathbb{P}(X \in A_j | H)$ for $j \in J$. To test hypothesis H , we can formally test $H_0 : f = f_0$ versus $H_1 : f \neq f_0$, which of course, is precisely the problem we solved in the [one-sample multinomial model](#).

Generally, we would partition the space S into as many subsets as possible, subject to the restriction that the expected frequencies all be at least 5.

The Partially Specified Case

Often we don't really want to test whether X has a completely specified distribution (such as the normal distribution with mean 5 and variance 9), but rather whether the distribution of X belongs to a specified parametric family (such as the normal). A natural course of action in this case would be to estimate the unknown parameters and then proceed just as above. As we have seen before, the expected frequencies would be statistics E_j because they would be based on the estimated parameters. As a rule of thumb, we lose a degree of freedom in the chi-square statistic V for each parameter that we estimate, although the precise mathematics can be complicated.

A Test of Independence

Suppose that we have observable random variables X and Y for an experiment, where X takes values in a set S with k elements, and Y takes values in a set T with m elements. Let f denote the joint probability density function of (X, Y) , so that $f(i, j) = \mathbb{P}(X = i, Y = j)$ for $i \in S$ and $j \in T$. Recall that the marginal probability density functions of X and Y are the functions g and h respectively, where

$$g(i) = \sum_{j \in T} f(i, j), \quad i \in S \quad (9.6.11)$$

$$h(j) = \sum_{i \in S} f(i, j), \quad j \in T \quad (9.6.12)$$

Usually, of course, f , g , and h are unknown. In this section, we are interested in testing whether X and Y are independent, a basic and important test. Formally then we want to test the null hypothesis

$$H_0 : f(i, j) = g(i) h(j), \quad (i, j) \in S \times T \quad (9.6.13)$$

versus the complementary alternative H_1 .

Our first step, of course, is to draw a random sample $(\mathbf{X}, \mathbf{Y}) = ((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$ from the distribution of (X, Y) . Since the state spaces are finite, this sample forms a sequence of multinomial trials. Thus, with our usual notation, let $O_{i,j}$ denote the number of times that (i, j) occurs in the sample, for each $(i, j) \in S \times T$. This statistic has the binomial distribution with trial parameter n and success parameter $f(i, j)$. Under H_0 , the success parameter is $g(i) h(j)$. However, since we don't know the success parameters, we must estimate them in order to compute the expected frequencies. Our best estimate of $f(i, j)$ is the sample proportion $\frac{1}{n} O_{i,j}$. Thus, our best estimates of $g(i)$ and $h(j)$ are $\frac{1}{n} N_i$ and $\frac{1}{n} M_j$, respectively, where N_i is the number of times that i occurs in sample \mathbf{X} and M_j is the number of times that j occurs in sample \mathbf{Y} :

$$N_i = \sum_{j \in T} O_{i,j} \quad (9.6.14)$$

$$M_j = \sum_{i \in S} O_{i,j} \quad (9.6.15)$$

Thus, our estimate of the expected frequency of (i, j) under H_0 is

$$E_{i,j} = n \frac{1}{n} N_i \frac{1}{n} M_j = \frac{1}{n} N_i M_j \quad (9.6.16)$$

Of course, we define our test statistic by

$$V = \sum_{i \in S} \sum_{j \in T} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \quad (9.6.17)$$

As you now expect, the distribution of V converges to a chi-square distribution as $n \rightarrow \infty$. But let's see if we can determine the appropriate degrees of freedom on heuristic grounds.

The limiting distribution of V has $(k-1)(m-1)$ degrees of freedom.

Proof

There are km terms in the expansion of V . We lose one degree of freedom since $\sum_{i \in S} \sum_{j \in T} O_{i,j} = n$. We must estimate all but one of the probabilities $g(i)$ for $i \in S$, thus losing $k-1$ degrees of freedom. We must estimate all but one of the probabilities $h(j)$ for $j \in T$, thus losing $m-1$ degrees of freedom.

An approximate test of H_0 versus H_1 at the α level of significance is to reject H_0 if and only if $V > \chi^2_{(k-1)(m-1)}(1-\alpha)$.

The observed frequencies are often recorded in a $k \times m$ table, known as a *contingency table*, so that $O_{i,j}$ is the number in row i and column j . In this setting, note that N_i is the sum of the frequencies in the i th row and M_j is the sum of the frequencies in the j th column. Also, for historical reasons, the random variables X and Y are sometimes called *factors* and the possible values of the variables *categories*.

Computational and Simulation Exercises

Computational Exercises

In each of the following exercises, specify the number of degrees of freedom of the chi-square statistic, give the value of the statistic and compute the P -value of the test.

A coin is tossed 100 times, resulting in 55 heads. Test the null hypothesis that the coin is fair.

Answer

1 degree of freedom, $V = 1$, $P = 0.3173$

Suppose that we have 3 coins. The coins are tossed, yielding the data in the following table:

	Heads	Tails
Coin 1	29	21
Coin 2	23	17
Coin 3	42	18

1. Test the null hypothesis that all 3 coin are fair.
2. Test the null hypothesis that coin 1 has probability of heads $\frac{3}{5}$; coin 2 is fair; and coin 3 has probability of heads $\frac{2}{3}$.
3. Test the null hypothesis that the 3 coins have the same probability of heads.

Answer

1. 3 degree of freedom, $V = 11.78$, $P = 0.008$.
2. 3 degree of freedom, $V = 1.283$, $P = 0.733$.
3. 2 degree of freedom, $V = 2.301$, $P = 0.316$.

A die is thrown 240 times, yielding the data in the following table:

Score	1	2	3	4	5	6
Frequency	57	39	28	28	36	52

1. Test the null hypothesis that the die is fair.
2. Test the null hypothesis that the die is an ace-six flat die (faces 1 and 6 have probability $\frac{1}{4}$ each while faces 2, 3, 4, and 5 have probability $\frac{1}{8}$ each).

Answer

1. 5 degree of freedom, $V = 18.45$, $P = 0.0024$
2. 5 degree of freedom, $V = 5.383$, $P = 0.3709$

Two dice are thrown, yielding the data in the following table:

Score	1	2	3	4	5	6
Die 1	22	17	22	13	22	24
Die 2	44	24	19	19	18	36

1. Test the null hypothesis that die 1 is fair and die 2 is an ace-six flat.
2. Test the null hypothesis that all the dice have have the same probability distribuiton.

Answer

1. 10 degree of freedom, $V = 6.2$, $P = 0.798$.
2. 5 degree of freedom, $V = 7.103$, $P = 0.213$.

A university classifies faculty by rank as *instructors*, *assistant professors*, *associate professors*, and *full professors*. The data, by faculty rank and gender, are given in the following contingency table. Test to see if faculty rank and gender are independent.

Faculty	Instructor	Assistant Professor	Associate Professor	Full Professor
Male	62	238	185	115
Female	118	122	123	37

Answer

3 degrees of freedom, $V = 70.111$, $P \approx 0$.

Data Analysis Exercises

The Buffon trial data set gives the results of 104 repetitions of Buffon's needle experiment. The number of crack crossings is 56. In theory, this data set should correspond to 104 Bernoulli trials with success probability $p = \frac{2}{\pi}$. Test to see if this is reasonable.

Answer

1 degree of freedom, $V = 4.332$, $P = 0.037$.

Test to see if the alpha emissions data come from a Poisson distribution.

Answer

We partition of \mathbb{N} into 17 subsets: $\{0, 1\}$, $\{x\}$ for $x \in \{2, 3, \dots, 16\}$ and $\{17, 18, \dots\}$ There are 15 degrees of freedom. The estimated Poisson parameter is 8.367, $V = 9.644$, $P = 0.842$.

Test to see if Michelson's velocity of light data come from a normal distribution.

Answer

Using the following partition of \mathbb{R} : $\{(-\infty, 750), [750, 775), [775, 800), [800, 825), [825, 850), [850, 875), [875, 900), [900, 925), [925, 950), [950, 975), [975, 1000), [1000, \infty)\}$ have 8 degrees of freedom, $V = 11.443$, $P = 0.178$.

Simulation Exercises

In the simulation exercises below, you will be able to explore the goodness of fit test empirically.

In the dice goodness of fit experiment, set the sampling distribution to fair, the sample size to 50, and the significance level to 0.1. Set the test distribution as indicated below and in each case, run the simulation 1000 times. In case (a), give the empirical estimate of the significance level of the test and compare with 0.1. In the other cases, give the empirical estimate of the power of the test. Rank the distributions in (b)-(d) in increasing order of apparent power. Do your results seem reasonable?

1. fair
2. ace-six flats
3. the symmetric, unimodal distribution
4. the distribution skewed right

In the dice goodness of fit experiment, set the sampling distribution to ace-six flats, the sample size to 50, and the significance level to 0.1. Set the test distribution as indicated below and in each case, run the simulation 1000 times. In case (a), give the empirical estimate of the significance level of the test and compare with 0.1. In the other cases, give the empirical estimate of the power of the test. Rank the distributions in (b)-(d) in increasing order of apparent power. Do your results seem reasonable?

1. fair

2. ace-six flats
3. the symmetric, unimodal distribution
4. the distribution skewed right

In the dice goodness of fit experiment, set the sampling distribution to the symmetric, unimodal distribution, the sample size to 50, and the significance level to 0.1. Set the test distribution as indicated below and in each case, run the simulation 1000 times. In case (a), give the empirical estimate of the significance level of the test and compare with 0.1. In the other cases, give the empirical estimate of the power of the test. Rank the distributions in (b)-(d) in increasing order of apparent power. Do your results seem reasonable?

1. the symmetric, unimodal distribution
2. fair
3. ace-six flats
4. the distribution skewed right

In the dice goodness of fit experiment, set the sampling distribution to the distribution skewed right, the sample size to 50, and the significance level to 0.1. Set the test distribution as indicated below and in each case, run the simulation 1000 times. In case (a), give the empirical estimate of the significance level of the test and compare with 0.1. In the other cases, give the empirical estimate of the power of the test. Rank the distributions in (b)-(d) in increasing order of apparent power. Do your results seem reasonable?

1. the distribution skewed right
2. fair
3. ace-six flats
4. the symmetric, unimodal distribution

Suppose that D_1 and D_2 are different distributions. Is the power of the test with sampling distribution D_1 and test distribution D_2 the same as the power of the test with sampling distribution D_2 and test distribution D_1 ? Make a conjecture based on your results in the previous three exercises.

In the dice goodness of fit experiment, set the sampling and test distributions to fair and the significance level to 0.05. Run the experiment 1000 times for each of the following sample sizes. In each case, give the empirical estimate of the significance level and compare with 0.05.

1. $n = 10$
2. $n = 20$
3. $n = 40$
4. $n = 100$

In the dice goodness of fit experiment, set the sampling distribution to fair, the test distributions to ace-six flats, and the significance level to 0.05. Run the experiment 1000 times for each of the following sample sizes. In each case, give the empirical estimate of the power of the test. Do the powers seem to be converging?

1. $n = 10$
2. $n = 20$
3. $n = 40$
4. $n = 100$