Lesson 17:
Chapter 9 Section 2

Caleb Moxley

BSC Mathematics

4 November 15

§9.2 Goodness of Fit

In §9.1, we compared the categorical-variable distributions of c populations using the χ^2 test.

§9.2 Goodness of Fit

In §9.1, we compared the categorical-variable distributions of c populations using the χ^2 test. We'll consider a slight variation of this test for this section. In particular, we'll investigate whether or not a sample is drawn from a population with a hypothesized distribution.

§9.2 Goodness of Fit

In §9.1, we compared the categorical-variable distributions of c populations using the χ^2 test. We'll consider a slight variation of this test for this section. In particular, we'll investigate whether or not a sample is drawn from a population with a hypothesized distribution. This test will also be a chi-squared right-tailed test.

§9.2 Goodness of Fit

In §9.1, we compared the categorical-variable distributions of c populations using the χ^2 test. We'll consider a slight variation of this test for this section. In particular, we'll investigate whether or not a sample is drawn from a population with a hypothesized distribution. This test will also be a chi-squared right-tailed test. The test statistic will also be computed in a similar manner.

§9.2 Goodness of Fit

In §9.1, we compared the categorical-variable distributions of c populations using the χ^2 test. We'll consider a slight variation of this test for this section. In particular, we'll investigate whether or not a sample is drawn from a population with a hypothesized distribution. This test will also be a chi-squared right-tailed test. The test statistic will also be computed in a similar manner.

Note: This test will be discussed in the discrete probability distribution case only!

§9.2 Goodness of Fit

Definition (expected counts)

If you want to test that a sample comes from a population with a particular distribution, you'll need the expected counts. If the hypothesized distribution is

A_i	A_1	A_2	$\dots$	A_k
$P(A_i)$	$P(A_1)$	$P(A_2)$	$\dots$	$P(A_k)$

§9.2 Goodness of Fit

Definition (expected counts)

If you want to test that a sample comes from a population with a particular distribution, you'll need the expected counts. If the hypothesized distribution is

A_i	A_1	A_2	$\dots$	A_k
$P(A_i)$	$P(A_1)$	$P(A_2)$	$\dots$	$P(A_k)$

Then then expected count in the A_i event is $n \cdot P(A_i)$.

Example (binomial)

Say we think a the number females in a 3-child family has a $B(3, 0.5)$ distribution. What would be the expected counts?

§9.2 Goodness of Fit

Example (binomial)

Say we think a the number females in a 3-child family has a $B(3, 0.5)$ distribution. What would be the expected counts?

X	0	1	2	3
$P(X)$	0.125	0.375	0.375	0.125

Example (binomial)

Say we think a the number females in a 3-child family has a $B(3, 0.5)$ distribution. What would be the expected counts?

X	0	1	2	3
$P(X)$	0.125	0.375	0.375	0.125

Thus, with a sample of 3-child families of size 1000, we would expect to see

Example (binomial)

Say we think a the number females in a 3-child family has a $B(3, 0.5)$ distribution. What would be the expected counts?

X	0	1	2	3
$P(X)$	0.125	0.375	0.375	0.125

Thus, with a sample of 3-child families of size 1000, we would expect to see 125 families with 0 female children, 375 families with 1 female child, 375 families with 2 female children, and 125 families with 3 female children.

§9.2 Goodness of Fit

Definition (χ^2 goodness of fit test)

To test the null hypothesis that a population has a particular distribution against the alternative hypothesis that it does not, calculate the χ^2 statistic

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}.$$

The P-value will be calculated using the χ^2 distribution with degrees of freedom $k - 1$ where k is the number of categories in the hypothesized distribution. The test is right-tailed.

Definition (χ^2 goodness of fit test)

To test the null hypothesis that a population has a particular distribution against the alternative hypothesis that it does not, calculate the χ^2 statistic

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}.$$

The P-value will be calculated using the χ^2 distribution with degrees of freedom $k-1$ where k is the number of categories in the hypothesized distribution. The test is right-tailed.

To use this test, make sure all expected counts are at least 5 or more.

Example (uniform random variable goodness of fit)

Is the following sample of weights from a population with a continuous uniform distribution where the upper and lower limits are 10 and 60? Use $\alpha = 0.1$.

Example (uniform random variable goodness of fit)

Is the following sample of weights from a population with a continuous uniform distribution where the upper and lower limits are 10 and 60? Use $\alpha = 0.1$.

Data Set 1

Example (uniform random variable goodness of fit)

Is the following sample of weights from a population with a continuous uniform distribution where the upper and lower limits are 10 and 60? Use $\alpha = 0.1$.

Data Set 1

To check this, let's compute the expected and observed counts in each of the categories below.

Example (uniform random variable goodness of fit)

Is the following sample of weights from a population with a continuous uniform distribution where the upper and lower limits are 10 and 60? Use $\alpha = 0.1$.

Data Set 1

To check this, let's compute the expected and observed counts in each of the categories below. (Note: We'll use five equal categories to keep the expected counts 5 or more.)

Example (uniform random variable goodness of fit)

Is the following sample of weights from a population with a continuous uniform distribution where the upper and lower limits are 10 and 60? Use $\alpha = 0.1$.

Data Set 1

To check this, let's compute the expected and observed counts in each of the categories below. (Note: We'll use five equal categories to keep the expected counts 5 or more.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Example (uniform random variable goodness of fit, con't.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Thus, the test statistic is

Example (uniform random variable goodness of fit, con't.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Thus, the test statistic is

$$X^2 = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{4}{5} + \frac{25}{5} = 6.4.$$

Example (uniform random variable goodness of fit, con't.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Thus, the test statistic is

$$X^2 = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{4}{5} + \frac{25}{5} = 6.4.$$

And so the P-value is $P(\chi^2 > 6.4) = 0.1712$.

Example (uniform random variable goodness of fit, con't.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Thus, the test statistic is

$$X^2 = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{4}{5} + \frac{25}{5} = 6.4.$$

And so the P-value is $P(\chi^2 > 6.4) = 0.1712$. And we fail to reject our null hypothesis.

Example (uniform random variable goodness of fit, con't.)

Category	10-20	20-30	30-40	40-50	50-60
Expected Count	5	5	5	5	5
Observed Count	4	4	3	4	10

Thus, the test statistic is

$$X^2 = \frac{1}{5} + \frac{1}{5} + \frac{1}{5} + \frac{4}{5} + \frac{25}{5} = 6.4.$$

And so the P-value is $P(\chi^2 > 6.4) = 0.1712$. And we fail to reject our null hypothesis. Our data is not significant enough to reject the claim that the data came from a population with a continuous uniform distribution with bound 10 and 60.

Example (general goodness of fit)

Do the observed counts below seem to suggest that the sample came from a distribution given below? Use $\alpha = 0.5$.

§9.2 Goodness of Fit

Example (general goodness of fit)

Do the observed counts below seem to suggest that the sample came from a distribution given below? Use $\alpha = 0.5$.

Category	1-5	5-10	10-15	15-20
Observed Count	8	14	10	8

Category	1-5	5-10	10-15	15-20
P(Category)	0.1	0.4	0.4	0.1

Example (general goodness of fit)

Do the observed counts below seem to suggest that the sample came from a distribution given below? Use $\alpha = 0.5$.

Category	1-5	5-10	10-15	15-20
Observed Count	8	14	10	8

Category	1-5	5-10	10-15	15-20
P(Category)	0.1	0.4	0.4	0.1

Well, the test statistic is $X^2 = 10.5$ and the P-value is 0.0148.

Example (general goodness of fit)

Do the observed counts below seem to suggest that the sample came from a distribution given below? Use $\alpha = 0.5$.

Category	1-5	5-10	10-15	15-20
Observed Count	8	14	10	8

Category	1-5	5-10	10-15	15-20
P(Category)	0.1	0.4	0.4	0.1

Well, the test statistic is $X^2 = 10.5$ and the P-value is 0.0148. Thus, we reject H_0. There is enough evidence to warrant the rejection of the claim that the underlying distribution of the population from which this sample was taken is not given by the table above.