

9.3: Tests in the Bernoulli Model

Basic Tests

Preliminaries

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample from the Bernoulli distribution with unknown parameter $p \in (0, 1)$. Thus, these are independent random variables taking the values 1 and 0 with probabilities p and $1 - p$ respectively. In the usual language of reliability, 1 denotes *success* and 0 denotes *failure*, but of course these are generic terms. Often this model arises in one of the following contexts:

1. There is an *event* of interest in a basic experiment, with unknown probability p . We replicate the experiment n times and define $X_i = 1$ if and only if the event occurred on run i .
2. We have a population of objects of several different types; p is the unknown proportion of objects of a particular type of interest. We select n objects at random from the population and let $X_i = 1$ if and only if object i is of the type of interest. When the sampling is *with* replacement, these variables really do form a random sample from the Bernoulli distribution. When the sampling is *without* replacement, the variables are dependent, but the Bernoulli model may still be approximately valid if the population size is very large compared to the sample size n . For more on these points, see the discussion of sampling with and without replacement in the chapter on Finite Sampling Models.

In this section, we will construct hypothesis tests for the parameter p . The parameter space for p is the interval $(0, 1)$, and all hypotheses define subsets of this space. This section parallels the section on Estimation in the Bernoulli Model in the Chapter on Interval Estimation.

The Binomial Test

Recall that the number of successes $Y = \sum_{i=1}^n X_i$ has the binomial distribution with parameters n and p , and has probability density function given by

$$\mathbb{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \dots, n\} \quad (9.3.1)$$

Recall also that the mean is $\mathbb{E}(Y) = np$ and variance is $\text{var}(Y) = np(1-p)$. Moreover Y is sufficient for p and hence is a natural candidate to be a test statistic for hypothesis tests about p . For $\alpha \in (0, 1)$, let $b_{n,p}(\alpha)$ denote the quantile of order α for the binomial distribution with parameters n and p . Since the binomial distribution is discrete, only certain (exact) quantiles are possible. For the remainder of this discussion, $p_0 \in (0, 1)$ is a conjectured value of p .

For every $\alpha \in (0, 1)$, the following tests have approximate significance level α :

1. Reject $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ if and only if $Y \leq b_{n,p_0}(\alpha/2)$ or $Y \geq b_{n,p_0}(1 - \alpha/2)$.
2. Reject $H_0 : p \geq p_0$ versus $H_1 : p < p_0$ if and only if $Y \leq b_{n,p_0}(\alpha)$.
3. Reject $H_0 : p \leq p_0$ versus $H_1 : p > p_0$ if and only if $Y \geq b_{n,p_0}(1 - \alpha)$.

Proof

In part (a), H_0 is a simple hypothesis, and under H_0 the test statistic Y has the binomial distribution with parameter n and p_0 . Thus, if H_0 is true, then α is (approximately) the probability of falsely rejecting H_0 by definition of the quantiles. In parts (b) and (c), H_0 specifies a range of values of p . But if H_0 is true, the maximum type 1 probability is (approximately) α and occurs when $p = p_0$.

The test in (a) is the standard, symmetric, two-sided test, corresponding to probability $\alpha/2$ (approximately) in both tails of the binomial distribution under H_0 . The test in (b) is the left-tailed test and the test in (c) is the right-tailed test. As usual, we can generalize the two-sided test by partitioning α between the left and right tails of the binomial distribution in an arbitrary manner.

For any $\alpha, r \in (0, 1)$, the following test has (approximate) significance level α : Reject $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ if and only if $Y \leq b_{n,p_0}(\alpha - r\alpha)$ or $Y \geq b_{n,p_0}(1 - r\alpha)$.

1. $r = \frac{1}{2}$ gives the standard symmetric two-sided test.
2. $r \downarrow 0$ gives the left-tailed test.

3. $r \uparrow 1$ gives the right-tailed test.

Proof

Once again, H_0 is a simple hypothesis and under H_0 , the test statistic Y has the binomial distribution with parameters n and p_0 . Thus if H_0 is true then the probability of falsely rejecting H_0 is α by definition of the quantiles. Parts (a)–(c) follow from properties of the quantile function.

An Approximate Normal Test

When n is large, the distribution of Y is approximately normal, by the central limit theorem, so we can construct an approximate normal test.

Suppose that the sample size n is large. For a conjectured $p_0 \in (0, 1)$, define the test statistic

$$Z = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} \quad (9.3.2)$$

1. If $p = p_0$, then Z has approximately a standard normal distribution.
2. If $p \neq p_0$, then Z has approximately a normal distribution with mean $\sqrt{n} \frac{p-p_0}{\sqrt{p_0(1-p_0)}}$ and variance $\frac{p(1-p)}{p_0(1-p_0)}$

Proof

1. This follows from the DeMoivre-Laplace theorem, the special case of the central limit theorem applied to the binomial distribution. Note that Z is simply the standard score associated with Y .
2. With some fairly simple algebra, we can write

$$Z = \sqrt{n} \frac{p - p_0}{\sqrt{p_0(1-p_0)}} + \sqrt{\frac{p(1-p)}{p_0(1-p_0)}} \frac{Y - np}{\sqrt{np(1-p)}} \quad (9.3.3)$$

The second factor in the second term is again simply the standard score associated with Y and hence this factor has approximately a standard normal distribution. So the result follows from the basic linearity property of the normal distribution.

As usual, for $\alpha \in (0, 1)$, let $z(\alpha)$ denote the quantile of order α for the standard normal distribution. For selected values of α , $z(\alpha)$ can be obtained from the special distribution calculator, or from most statistical software packages. Recall also by symmetry that $z(1-\alpha) = -z(\alpha)$.

For every $\alpha \in (0, 1)$, the following tests have approximate significance level α :

1. Reject $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ if and only if $Z < -z(1-\alpha/2)$ or $Z > z(1-\alpha/2)$.
2. Reject $H_0 : p \geq p_0$ versus $H_1 : p < p_0$ if and only if $Z < -z(1-\alpha)$.
3. Reject $H_0 : p \leq p_0$ versus $H_1 : p \geq p_0$ if and only if $Z > z(1-\alpha)$.

Proof

In part (a), H_0 is a simple hypothesis and under H_0 the test statistic Z has approximately a standard normal distribution. Hence if H_0 is true then the probability of falsely rejecting H_0 is approximately α by definition of the quantiles. In parts (b) and (c), H_0 specifies a range of values of p , and under H_0 the test statistic Z has a nonstandard normal distribution, as described [above](#). The maximum type one error probability is α and occurs when $p = p_0$.

The test in (a) is the symmetric, two-sided test that corresponds to $\alpha/2$ in both tails of the distribution of Z , under H_0 . The test in (b) is the left-tailed test and the test in (c) is the right-tailed test. As usual, we can construct a more general two-sided test by partitioning α between the left and right tails of the standard normal distribution in an arbitrary manner.

For every $\alpha, r \in (0, 1)$, the following test has approximate significance level α : Reject $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ if and only if $Z < z(\alpha - r\alpha)$ or $Z > z(1 - r\alpha)$.

1. $r = \frac{1}{2}$ gives the standard, symmetric two-sided test.
2. $r \downarrow 0$ gives the left-tailed test.

3. $r \uparrow 1$ gives the right-tailed test.

Proof

In part (a), H_0 is again a simple hypothesis, and under H_0 the test statistic Z has approximately a standard normal distribution. So if H_0 is true, the probability of falsely rejecting H_0 is α by definition of the quantiles.

Simulation Exercises

In the proportion test experiment, set $H_0 : p = p_0$, and select sample size 10, significance level 0.1, and $p_0 = 0.5$. For each $p \in \{0.1, 0.2, \dots, 0.9\}$ run the experiment 1000 times and then note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

In the proportion test experiment, repeat the previous exercise with sample size 20.

In the proportion test experiment, set $H_0 : p \leq p_0$, and select sample size 15, significance level 0.05, and $p_0 = 0.3$. For each $p \in \{0.1, 0.2, \dots, 0.9\}$ run the experiment 1000 times and note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

In the proportion test experiment, repeat the previous exercise with sample size 30.

In the proportion test experiment, set $H_0 : p \geq p_0$, and select sample size 20, significance level 0.01, and $p_0 = 0.6$. For each $p \in \{0.1, 0.2, \dots, 0.9\}$ run the experiment 1000 times and then note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

In the proportion test experiment, repeat the previous exercise with sample size 50.

Computational Exercises

In a pole of 1000 registered voters in a certain district, 427 prefer candidate X. At the 0.1 level, is the evidence sufficient to conclude that more than 40% of the registered voters prefer X?

Answer

Test statistic 1.743, critical value 1.282. Reject H_0 .

A coin is tossed 500 times and results in 302 heads. At the 0.05 level, test to see if the coin is unfair.

Answer

Test statistic 4.651, critical values ± 1.961 . Reject H_0 ; the coin is almost certainly unfair.

A sample of 400 memory chips from a production line are tested, and 32 are defective. At the 0.05 level, test to see if the proportion of defective chips is less than 0.1.

Answer

Test statistic -1.333 , critical value -1.645 . Fail to reject H_0 .

A new drug is administered to 50 patients and the drug is effective in 42 cases. At the 0.1 level, test to see if the success rate for the new drug is greater than 0.8.

Answer

Test statistic 0.707, critical value 1.282. Fail to reject H_0 .

Using the M&M data, test the following alternative hypotheses at the 0.1 significance level:

1. The proportion of red M&Ms differs from $\frac{1}{6}$.
2. The proportion of green M&Ms is less than $\frac{1}{6}$.

3. The proportion of yellow M&M is greater than $\frac{1}{6}$.

Answer

1. Test statistic 0.162, critical values ± 1.645 Fail to reject H_0 .
2. Test statistic -4.117 , critical value -1.282 Reject H_0 .
3. Test statistic 8.266, critical value 1.282. Reject H_0 .

The Sign Test

Derivation

Suppose now that we have a basic random experiment with a real-valued random variable U of interest. We assume that U has a continuous distribution with support on an interval of $S \subseteq \mathbb{R}$. Let m denote the quantile of a specified order $p_0 \in (0, 1)$ for the distribution of U . Thus, by definition,

$$p_0 = \mathbb{P}(U \leq m) \quad (9.3.4)$$

In general of course, m is unknown, even though p_0 is specified, because we don't know the distribution of U . Suppose that we want to construct hypothesis tests for m . For a given test value m_0 , let

$$p = \mathbb{P}(U \leq m_0) \quad (9.3.5)$$

Note that p is unknown even though m_0 is specified, because again, we don't know the distribution of U .

Relations

1. $m = m_0$ if and only if $p = p_0$.
2. $m < m_0$ if and only if $p > p_0$.
3. $m > m_0$ if and only if $p < p_0$.

Proof

These results follow since we are assuming that the distribution of U is continuous and is supported on the interval S .

As usual, we repeat the basic experiment n times to generate a random sample $\mathbf{U} = (U_1, U_2, \dots, U_n)$ of size n from the distribution of U . Let $X_i = \mathbf{1}(U_i \leq m_0)$ be the indicator variable of the event $\{U_i \leq m_0\}$ for $i \in \{1, 2, \dots, n\}$.

Note that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a statistic (an observable function of the data vector \mathbf{U}) and is a random sample of size n from the Bernoulli distribution with parameter p .

From the last two results it follows that tests of the unknown quantile m can be converted to tests of the Bernoulli parameter p , and thus the tests developed above apply. This procedure is known as the *sign test*, because essentially, only the sign of $U_i - m_0$ is recorded for each i . This procedure is also an example of a *nonparametric test*, because no assumptions about the distribution of U are made (except for continuity). In particular, we do not need to assume that the distribution of U belongs to a particular parametric family.

The most important special case of the sign test is the case where $p_0 = \frac{1}{2}$; this is the sign test of the median. If the distribution of U is known to be symmetric, the median and the mean agree. In this case, sign tests of the median are also tests of the mean.

Simulation Exercises

In the sign test experiment, set the sampling distribution to normal with mean 0 and standard deviation 2. Set the sample size to 10 and the significance level to 0.1. For each of the 9 values of m_0 , run the simulation 1000 times.

1. When $m = m_0$, give the empirical estimate of the significance level of the test and compare with 0.1.
2. In the other cases, give the empirical estimate of the power of the test.

In the sign test experiment, set the sampling distribution to uniform on the interval $[0, 5]$. Set the sample size to 20 and the significance level to 0.05. For each of the 9 values of m_0 , run the simulation 1000 times.

1. When $m = m_0$, give the empirical estimate of the significance level of the test and compare with 0.05.

2. In the other cases, give the empirical estimate of the power of the test.

In the sign test experiment, set the sampling distribution to gamma with shape parameter 2 and scale parameter 1. Set the sample size to 30 and the significance level to 0.025. For each of the 9 values of m_0 , run the simulation 1000 times.

1. When $m = m_0$, give the empirical estimate of the significance level of the test and compare with 0.025.
2. In the other cases, give the empirical estimate of the power of the test.

Computational Exercises

Using the M&M data, test to see if the median weight exceeds 47.9 grams, at the 0.1 level.

Answer

Test statistic 3.286, critical value 1.282. Reject H_0 .

Using Fisher's iris data, perform the following tests, at the 0.1 level:

1. The median petal length of Setosa irises differs from 15 mm.
2. The median petal length of Verginica irises is less than 52 mm.
3. The median petal length of Versicolor irises is less than 42 mm.

Answer

1. Test statistic 3.394, critical values ± 1.645 Reject H_0 .
2. Test statistic -1.980 , critical value -1.282 Reject H_0 .
3. Test statistic -0.566 , critical value -1.282 Fail to reject H_0 .

This page titled [9.3: Tests in the Bernoulli Model](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.