# 3. Tests in the Bernoulli Model

## Preliminaries

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ 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. Usually, 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. 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.

## Tests of $p$

### 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 mean $\mathbb{E}(Y) = n\,p$ and variance $\operatorname{var}(Y) = n\,p\,(1 - p)$. Moreover, recall that $Y$ is sufficient for $p$. For $\alpha \in (0, 1)$, let $b_{n,p}(\alpha)$ denote the quantile of order $\alpha$ for the binomial distribution with parameters $n$ and $p$. Since the binomial distribution is discrete, only certain (exact) quantiles are possible.

1. Show that for any $\alpha \in (0, 1)$ and $r \in (0, 1)$, the following tests have significance level $\alpha$:

a. Reject $H_0$: $p = p_0$ versus $H_1$: $p \ne p_0$ if and only if $Y \le b_{n,p_0}(\alpha - r\,\alpha)$ or $Y \ge b_{n,p_0}(1 - r\,\alpha)$
b. Reject $H_0$: $p \ge p_0$ versus $H_1$: $\mu < p_0$ if and only if $Y \le b_{n,p_0}(\alpha)$

c. Reject $H_0$: $p \le p_0$ versus $H_1$: $p > p_0$ if and only if $Y \ge b_{n,p_0}(1-\alpha)$

As usual, of the two-sided tests in part (a), the unbiased test with $r = \frac{1}{2}$ is most commonly used:

$$\text{Reject } H_0\text{: } p = p_0 \text{ versus } H_1\text{: } p \ne p_0 \text{ if and only if } Y \le b_{n,p_0}\left(\frac{\alpha}{2}\right) \text{ or } Y \ge b_{n,p_0}\left(1-\frac{\alpha}{2}\right).$$

### An Approximate Normal Test

When $n$ is large, the distribution of $Y$ is approximately normal, by the central limit theorem. Thus, an approximate normal test can be constructed using the test statistic

$$Z(p_0) = \frac{Y - n\,p_0}{\sqrt{n\,p_0\,(1-p_0)}}$$

Note that $Z(p_0)$ is the standard score of $Y$ if $p = p_0$. As usual, for $\alpha \in (0, 1)$, let $z(\alpha)$ denote the quantile of order $\alpha$ for the standard normal distribution. For selected values of $\alpha$, $z(\alpha)$ can be obtained from the last row of the table of the $t$ distribution, from the table of the standard normal distribution, from the quantile applet, or from most statistical software packages. Recall also by symmetry that $z(1-\alpha) = -z(\alpha)$

2. Show that for any $\alpha \in (0, 1)$ and $r \in (0, 1)$, the following tests have significance level $\alpha$:

a. Reject $H_0$: $p = p_0$ versus $H_1$: $p \ne p_0$ if and only if $Z(p_0) \le z(\alpha - r\,\alpha)$ or $Z(p_0) \ge z(1 - r\,\alpha)$
b. Reject $H_0$: $p \ge p_0$ versus $H_1$: $p < p_0$ if and only if $Z(p_0) \le z(\alpha)$
c. Reject $H_0$: $p \le p_0$ versus $H_1$: $p > p_0$ if and only if $Z(p_0) \ge z(1-\alpha)$

As usual, of the two-sided tests in part (a), the unbiased test with $r = \frac{1}{2}$ is most commonly used:

$$\text{Reject } H_0\text{: } p = p_0 \text{ versus } H_1\text{: } p \ne p_0 \text{ if and only if } Z(p_0) \le z\left(\frac{\alpha}{2}\right) \text{ or } Z(p_0) \ge z\left(1-\frac{\alpha}{2}\right)$$

### Simulation Exercises

3. 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 $p \in \{0.1, 0.2, ..., 0.9\}$, run the experiment 1000 times, updating every 10 runs, and then note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

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

5. In the proportion test experiment, set $H_0$: $p \le p_0$, and select sample size 15, significance level 0.05, and $p_0 = 0.3$. For each $p \in \{0.1, 0.2, ..., 0.9\}$, run the experiment 1000 times, updating every 10 runs, and then note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

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

7. In the proportion test experiment, set $H_0$: $p \ge p_0$, and select sample size 20, significance level 0.01, and $p_0 = 0.6$. For each $p \in \{0.1, 0.2, ..., 0.9\}$, run the experiment 1000 times, updating every 10 runs, and then note the relative frequency of rejecting the null hypothesis. Graph the empirical power function.

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

**Computational Exercises**

9. 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 that 40% of the registered voters prefer X?

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

11. 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.

12. 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 that 0.8.

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

a. The proportion of red M&Ms differs from $\frac{1}{6}$.
b. The proportion of green M&Ms is less than $\frac{1}{6}$.
c. The proportion of yellow M&M is greater than $\frac{1}{6}$

## The Sign Test

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

$$p = P(U \le m)$$

Suppose that $m$ is unknown and that we want to construct hypothesis tests for $m$. For a given test value $m_0$, let

$$p_0 = P(U \le m_0)$$

14. Show that

a. $m = m_0$ if and only if $p = p_0$.
b. $m < m_0$ if and only if $p < p_0$.
c. $m > m_0$ if and only if $p > p_0$.

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

15. Show that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size $n$ from the Bernoulli distribution with parameter $p$.

From Exercises 14 and 15, tests of the unknown quantile $m$ can be converted to tests of the Bernoulli parameter $p$, and thus the tests developed in the previous subsections 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

16. 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, updating every 10 runs.

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

17. 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, updating every 10 runs.

a. When $m = m_0$, give the empirical estimate of the significance level of the test and compare with 0.05.
b. In the other cases, give the empirical estimate of the power of the test.

18. 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 o $m_0$,

run the simulation 1000 times, updating every 10 runs.

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

### Computational Exercises

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

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

a. The median petal length of Setosa irises differs from 15 mm.
b. The median petal length of Verginica irises is less than 52 mm.
c. The median petal length of Versicolor irises is less than 42 mm.

Virtual Laboratories > 9. Hypothesis Testing > 1 2 3 4 5 6 7
Contents | Applets | Data Sets | Biographies | External Resources | Keywords | Feedback | ©