

Chapter 11: Binomial Distributions

Bernoulli Processes

The Binomial Distribution

Means and Standard Deviations

Chap. 11 - page 1

It is **AS IF** we have a box containing equal-sized slips of paper. A proportion π of slips say success and a proportion $1-\pi$ say failure.

We mix the slips thoroughly and draw one out at random. That is the first trial.

The slip is returned to the box and the slips are thoroughly mixed again. Then another slip is drawn at random, the result recorded and the slip is returned to the box. That is the second trial.

The process is repeated n times.

Chap. 11 - page 3

The Bernoulli Process

(page 308)

categorical yes-no process outcomes

(success-failure, 1 = success, 0 = failure)

Each observation is called a **trial**.

Assumptions for a Bernoulli Process:

- The process is stable and generates a constant long-run proportion of successes.
- One trial does not influence and is not influenced by another, We say the trials are **statistically independent** of one another.

Chap. 11 - page 2

BINS

Binary outcomes

Independent trials

Number n of trials is fixed

Same probability of success on each trial.

Chap. 11 - page 4

Examples:

Service is satisfactory or not for the first person, for the second person, ...

The size of hole in the first washer is within specifications or not. The second washer has a hole within specs or not, ...

Jon answers question 1 correctly or not, then answers question 2 correctly or not, ...

Chap. 11 - page 5

y is a **discrete** variable (a count).

If we were to observe n more trials we would obtain a new value for y (and for p).

We can consider the (sampling) distribution of y (and of p) that would be obtained under repeated sampling.

There is a (theoretical) model for the distribution of y called the **binomial distribution**.

It is convenient to use the notation and language of probability.

We write $Pr(y=1)$ for the probability that $y=1$. In words, this is the probability of one success.

Similarly, $Pr(y=k)$ is the probability that $y=k$ is the probability of k successes.

Chap. 11 - page 7

The Binomial Distribution

(page 311)

Consider a fixed number of trials n and let y be the total number of successes in the n trials.

Let π denote the long-run proportion of successes in the process. This is a parameter of the process. We say that π is the probability of success on one trial.

Let p denote the proportion of successes in our n trials.

If we were to observe n trials we would obtain a value for y (and for p).

y will be one of the values: 0, 1, 2, ..., n and p will be one of the values $0/n=0, 1/n, \dots, n/n=1$.

Chap. 11 - page 6

Theory shows that we have the formula

$$Pr(k \text{ successes}) = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

for any k from 0 to n .

This is called the **Binomial Distribution**.

Chap. 11 - page 8

Examples:

$n = 4$ and $\pi = .3$

$$\begin{aligned}Pr(1 \text{ success}) &= \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k} \\&= \frac{4!}{1!4-1!} 0.3^1 (1-0.3)^{4-1} \\&= 4(0.3)(0.7)^3 \\&= 4(0.3)0.343 \\&= 0.4116\end{aligned}$$

Chap. 11 - page 9

Example:

A quality assurance inspector randomly selects 10 parts from a stable process that produces 20% bad parts.

What is the chance that the inspector gets 10 good parts?

We treat this like a Bernoulli process with $n = 10$ trials, $\pi = 0.2$, and "success" meaning a bad part!

We want the chance of zero successes.

This is

$$Pr(0 \text{ successes}) = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

with $k = 0$, $n = 10$, and $\pi = 0.2$. So

$$Pr(0 \text{ successes}) = \frac{10!}{0!10!} 0.2^0 (1-0.2)^{10-0}$$

Chap. 11 - page 10

$$Pr(0 \text{ successes}) = (1)(1)(1-0.2)^{10}$$

or

$$Pr(0 \text{ successes}) = (0.8)^{10} = 0.10737418$$

or about 11% chance of this happening.

What is the chance that the inspector chooses at most one bad part?

We need to add the chance of no successes to the chance of one success.

The chance of one success is

$$Pr(1 \text{ success}) = \frac{10!}{1!(10-1)!} 0.2^1 (1-0.2)^{10-1}$$

or

$$Pr(1 \text{ success}) = (10)0.2(0.8)^9$$

Chap. 11 - page 11

$$Pr(1 \text{ success}) = (10)0.2(0.8)^9$$

$$Pr(1 \text{ success}) = 0.26843546$$

Adding the chance for no successes and the chance for one success gives

$$0.375809638$$

or about a 38% chance of getting at most one bad part?

Chap. 11 - page 12

Means and Standard Deviations

(page 318)

Consider a general Bernoulli process with long-run success rate π .

Let's evaluate the theoretical mean μ of the process values.

Process: 0, 0, 1, 0, 1, 1, ..., 0 for N trials. The mean (so far) is

$$\begin{aligned}\text{Mean} &= \frac{0 + 0 + 1 + 0 + 1 + 1 + \dots + 0}{N} \\ &= \frac{(\# \text{ of } 0\text{s})0 + (\# \text{ of } 1\text{s})1}{N} \\ &= \frac{(\# \text{ of } 1\text{s})}{N} \\ &\approx \pi \quad \text{for large } N\end{aligned}$$

So we have "shown" that for the Bernoulli process $\mu = \pi$.

Chap. 11 - page 13

The Binomial variable y is just a total of n Bernoulli process values.

So using the results on totals from page 262, we have

$$\mu_y = n\mu = n\pi$$

and

$$\sigma_y = \sqrt{n}\sigma = \sqrt{n}\sqrt{\pi(1-\pi)} = \sqrt{n\pi(1-\pi)}$$

In summary: **The mean and standard deviation of a Binomial Distribution are**

$$\mu_y = n\pi$$

$$\sigma_y = \sqrt{n\pi(1-\pi)}$$

Chap. 11 - page 15

A similar, but slightly longer argument, shows that the process standard deviation is

$$\sigma = \sqrt{\pi(1-\pi)}$$

In summary:

For a Bernoulli process

$$\mu = \pi$$

$$\sigma = \sqrt{\pi(1-\pi)}$$

Chap. 11 - page 14

Proportions (page 321)

The proportion $p = y/n$ is just a mean of n Bernoulli process values.

The results from page 259 become: **For a Bernoulli process, the mean and standard deviation of a proportion p are**

$$\mu_p = \pi$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

As π varies from 0 to 1 the value of

$$\sqrt{\pi(1-\pi)}$$

increases from a low of zero at $\pi = 0$ to a high of 0.5 at $\pi = 0.5$. Then it decreases back to zero at $\pi = 1$.

Thus, the standard deviation of y (or of p) is largest when $\pi = 0.5$.

Chap. 11 - page 16

The Normal Approximation (page 322)

Applying the **Central Limit Effect** to a Bernoulli process mean or proportion we get:

For large n , y is approximately normally distributed with

$$\begin{aligned}\mu_y &= n\pi \\ \sigma_y &= \sqrt{n\pi(1-\pi)}\end{aligned}$$

and (equivalently)

For large n , p is approximately normally distributed with

$$\begin{aligned}\mu_p &= \pi \\ \sigma_p &= \sqrt{\frac{\pi(1-\pi)}{n}}\end{aligned}$$

Chap. 11 - page 17

exact value obtained from the Binomial Distribution formula.

$p \leq 0.2$ is the same as $y/25 \leq 0.2$ which means $y \leq 5$ or $y = 0, 1, 2, 3, 4, \text{ or } 5$.

We need to evaluate the binomial for these six values and add them up.

Actually, let's use Minitab:

Cumulative Distribution Function
Binomial with $n = 25$ and $\pi = 0.4$

k	$P(y \leq k)$
0	0.0000
1	0.0001
2	0.0004
3	0.0024
4	0.0095
5	0.0294

so that the exact answer is 0.0294.

Chap. 11 - page 19

Examples:

Let $n = 25$ and $\pi = 0.4$. Let's approximate the chance that $p \leq 0.2$.

First we get the mean and standard deviation for the distribution of p . They are

$$\begin{aligned}\mu_p &= \pi = 0.4 \\ \sigma_p &= \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.4(1-0.4)}{25}} = 0.098\end{aligned}$$

So

$$z = \frac{0.2 - 0.4}{0.098} = -2.04$$

and we need the area below this z -value. From the normal table we have area 0.0207 or a little more than 2%.

We can compare this approximation to the

Chap. 11 - page 18

The Iowa Poll

The Iowa Poll is based on random samples of about 800 people (out of the hundreds of thousands of adults in Iowa).

The January 22, 1995 poll used $n = 804$.

One question asked was:

"From your understanding of the way things work in Iowa, do you think that most people convicted of first-degree murder eventually gain their freedom or do you think most of them serve the rest of their lives in prison?"

Suppose that in the whole population 50% have such a belief. What is the chance of getting 64% or more with that belief **in our sample?**

Chap. 11 - page 20

Here $n = 804$ and $\pi = 0.5$
 First we get the mean and standard deviation
 for the distribution of p . They are

$$\mu_p = \pi = 0.5$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.5(1-0.5)}{804}} = 0.0176$$

So

$$z = \frac{0.64 - 0.5}{0.0176} = 7.95$$

This is an enormously large z-value! The area
 above this value is zero to many places and
 the outcome ($p \geq 0.64$) is extremely
 improbable.

In the poll 64% responded that they think
 most gain their freedom. (This is, in fact, NOT
 SO!)

We know that the proportion has
 (approximately) a normal distribution with
 mean $\pi = 0.503$ and standard deviation

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

$$= \sqrt{\frac{0.503(1-0.503)}{592}} = 0.0205$$

In standardized terms, our observed p is

$$z = \frac{0.466 - 0.503}{0.0205} = -1.80$$

which not an unusually small standard normal
 value.

**Omit Section 11.11, page 327, for now.
 We'll come back to it before Chapter 17.**

Missouri Lotto

Pick 6 (out of 1, 2, ..., 48)

A Kansas City resident thought too many
 consecutive numbers were coming up in the
 lottery picks.

After lengthy calculation it can be shown that
 the chance of at least two consecutive
 numbers in one run of the Pick 6 (out of 48)
 lottery is 0.503.

During the period July 1989 through March
 11, 1995, the Pick 6 was run 592 times. Of
 these, 276 produced at least two consecutive
 numbers. Is this too many, too few, about
 right?

These data give a proportion of
 $p = 276/592 = 0.466$
 which is lower than expected.

Testing Hypotheses about π (page 327)

A **statistical hypothesis** is a statement about
 a process or population parameter.

In a 1978 issue of *Consumers Reports*
 they reported on a taste test of *Miller High
 Life* versus *Lowenbrau*.

They are brewed by the same company
 but *Lowenbrau* costs a lot more. Can
 people tell the difference?

A panel of 24 staffers were assembled.
 Panelists were given three glasses of beer
 to taste of which two were from the same
 bottle and asked to identify which glass
 contained the beer that was different.

We treat the 24 tasters as 24 Bernoulli trials.

“Success” corresponds to correct identification of the different beer.

The symbol π denotes the proportion of “average beer drinkers” who would correctly choose the different beer under the same conditions as the taste test.

If the panelists were only guessing which of the three glasses of beer was the correct one, they would be expected to get the right answer about one-third of the time, or 8 of the 24.

We interpret this as the claim that $\pi = 1/3$. This is a statistical hypothesis. Others might claim that $\pi > 1/3$. This another, competing hypothesis. Our goal in testing statistical hypotheses is to see how much support the data provide for various hypotheses.

If $H_0: \pi = 1/3$ is true, the mean number of correct beer identifications is $n\pi = 24(1/3) = 8$.

According to *Consumer's Reports* “only 11 of the total of 24 judgements were correct.”

Is this enough to declare that average beer drinkers are better than would be with random guessing?

Decision Rule: Reject the null hypothesis of random guessing if (and only if) the number of correct judgements y is “large enough.”

How large is large enough?

Null Hypothesis: The hypothesis of no difference, no change, or no effect—written H_0 and read as H-nought or H-zero. (page 329)

In our example $H_0: \pi = 1/3$.

Alternative Hypothesis: This hypothesis competes with the null hypothesis. It is denoted by H_1 .

In our example $H_1: \pi > 1/3$.

A statistical test is designed to assess the strength of evidence against the null hypothesis.

Consider the following table:

Decision	Truth	
	H_0 True	H_0 False
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Type I error: Rejecting a true null hypothesis. (page 329)

Type II error: Failing to reject a false null hypothesis. (page 329)

The probability of a Type I error is called the **significance level** of the statistical test.

The significance level is usually denoted by the Greek letter α (in fractional terms).

We would like this to be small. We specify "How large is large" by making the significance level "small enough."

Typical values chosen for α are 0.05 or 0.01.

Specifically, we reject H_0 if (and only if) $y \geq y_0$ where y_0 is chosen to make the significance level small enough.

y_0 is called the **critical value** of the test.

The binomial distribution permits us to find y_0 .

Type II Errors (page 331)

Remember: A **type II error** occurs when we fail to reject a false null hypothesis.

In our example $H_1: \pi > 1/3$ so there are many false null hypotheses. (infinitely many!)

Any value of π greater than $1/3$ corresponds to a particular alternative hypothesis and a false null hypothesis.

Our decision rule says: Reject H_0 if, and only if, we observe 13 or more correct judgements.

We will fail to reject H_0 if, and only if, we observe fewer than 13 correct judgements.

In our example, using the binomial distribution with $n = 24$ and $\pi = 1/3$ we find that

$$Pr(y \geq 12) = 0.06766$$

$$\text{and } Pr(y \geq 13) = 0.02844.$$

So we can set the significance level at $\alpha = 0.02844$ by taking $y_0 = 13$.

Our decision rule is then: Reject H_0 if, and only if, we observe 13 or more correct judgements.

In fact, only 11 correct judgements were made. *Consumer Reports* said "...we don't consider that to be statistically significant evidence that a beer drinker can tell domestic Lowenbrau from Miller High Life."

Since $11 < 13$ the evidence is considered insufficient to refute the null hypothesis.

What is the chance that we will observe fewer than 13 correct judgements **when the alternative is true?**

This is the probability of a Type II error.

Since there are many particular alternatives, there are many values for Type II errors and, in fact there is a whole curve of Type II error probabilities.

Example:

Find the probability of a Type II error when π is 0.70.

This is just the chance that we observe fewer than 13 correct judgements when $\pi = 0.70$.

Use the Binomial distribution with $n = 24$, $\pi = 0.70$, and get the sum of the probabilities from 0 up to 12.

Using Minitab we get a probability of a Type II error of 0.0314 when the particular alternative $\pi = 0.70$ is true.

Similarly we could obtain the probability of a Type II error for any value of π greater than 1/3 and make a table or graph of these probabilities.

p -values (page 331)

Rather than setting a significance level and critical value for y , some statisticians prefer an alternative approach based on p -values.

The **p -value** of a test procedure is the probability of obtaining an outcome at least as extreme as the outcome actually observed.

The alternative hypothesis determines the direction of being even more extreme.

The p -value is calculated assuming the null hypothesis is true.

Small p -value provide evidence against the null hypothesis.

If the p -value is small enough, the evidence is said to be **statistically significant**.

A common rule is to require the p -value to be 0.05 or less for statistical significance.

In our example we obtained 11 correct judgements. If the null hypothesis is true, what is the chance of seeing 11 or more correct judgements? This is the p -value of the test.

Using the Binomial with $\pi = 1/3$ we find this probability to be

$$p\text{-value} = Pr(y \geq 11) = 0.1401$$

which is not too small.

Consumer's Reports said "Although 11 correct choices were an improvement over pure guesswork, we don't consider that to be statistically significant evidence that a beer drinker can tell domestic Lowenbrau from Miller High Life."