The Binomial Distribution

MATH 2441, BCIT

Technical Mathematics for Food Technology

January 31, 2018

Probability Distributions: Concepts

Here are some definitions.

random variable A random variable is a variable (typically represented by X) that has a single numerical value, determined by chance, for each outcome of a procedure.

probability distribution A probability distribution is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

Probability Distributions: Discrete and Continuous

discrete random variable A discrete random variable has a collection of values that is finite or countable.

continuous random variable A continuous random variable has infinitely many values, and the collection of values is not countable.

countability This is best explained by example: the integers are countable, but the real numbers are not.

Discrete Probability Distributions

If there are a finite number of outcomes $X = a_k$ for $k = 1, \ldots, n$, we can list the values of $P(X = a_k)$ in a table.

Example 1: Coin Toss. let $X = 1$ for heads and $X = 0$ for tails. Then

Event	Probability
$X = 1$ or H	0.50
$X = 0$ or T	0.50

When all the probabilities are equal, we call the probability distribution a uniform distribution.

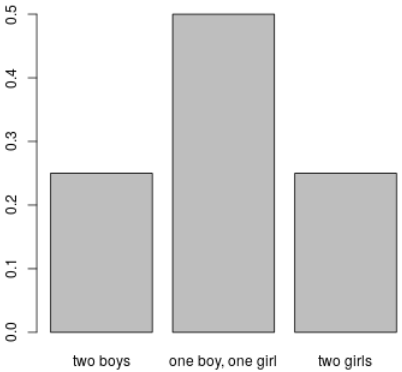
Non-Uniform Discrete Probability Distributions

Some distribution probability distributions are not uniform.

Example 2: Number of Male Children. Consider the two-child family. If X is the random variable corresponding to the number of boys in the family, then the probability distribution table looks as follows (assuming that the probability distribution for one child is uniform).

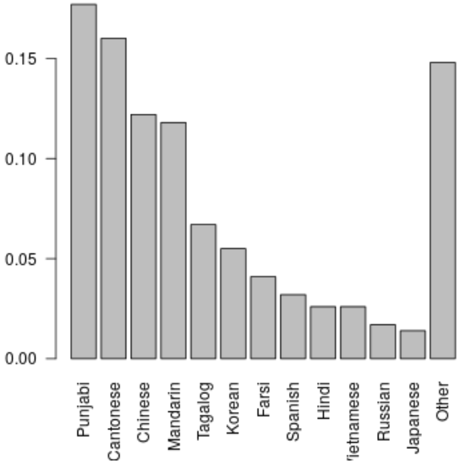
Discrete Probability Distribution Graphs I

Event	Probability
$X = 2$ or two boys	0.25
$X = 1$ or one boy, one girl	0.50
$X = 0$ or two girls	0.25



Discrete Probability Distribution Graphs II

Example 3: Immigrant Languages. Here is the probability distribution for a randomly selected "Vancouverite" (Greater Vancouver) to speak a certain immigrant language at home.



Mean and Variance Formulas

There is a sense in which a probability distribution together with its associated random variable correspond to a population and the property which the random variable picks out. In this spirit, let us define a mean and a variance for a probability distribution.

$$\mu = \sum X \cdot P(X) \tag{1}$$

$$\sigma^2 = \sum (X - \mu)^2 \cdot P(X) \tag{2}$$

$$\sigma^2 = \sum (X^2 \cdot P(X)) - \mu^2 \tag{3}$$

$$\sigma = \sqrt{\sum (X^2 \cdot P(X)) - \mu^2} \tag{4}$$

Die Roll Example

Example 4: Fair Die Roll. Think of rolling a fair die many times. The probability distribution is uniform. The mean is

$$\mu = \sum X \cdot P(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

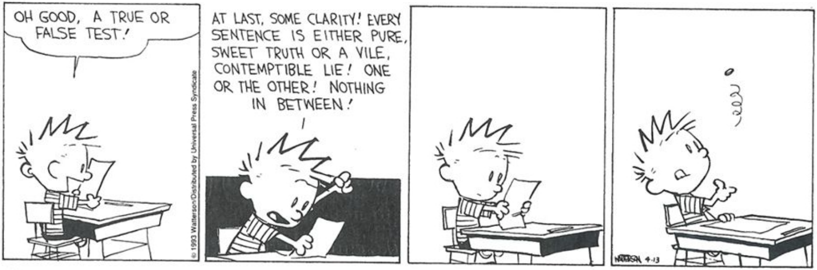
We also call this number the expectation EX of the random variable X. Although you would never expect a die roll to result in "3.5," you would expect the mean of many die rolls to be close to this number. The expected number of boys for one birth is $EX = 0.5$.

The Binomial Probability Distribution

A binomial probability distribution results from a procedure that meets the following requirements.

1. The procedure has a fixed number of trials.
2. The trials must be independent.
3. The outcomes of a trial are binary, i.e. there are only two possible outcomes.
4. The probability of the two outcomes remains constant.

The number of trials is usually labeled n, the two outcomes are called success and failure, and their probabilities on one trial are p and $1 - p$. The random variable keeps track of the number of successes. If, for example, there are 10 trials, then $P(X = 4)$ is the probability of 4 successes out of 10. The number of successes is often labeled x, and we are usually interested in $P(X = x)$.


OH GOOD, A TRUE OR FALSE TEST!
AT LAST, SOME CLARITY! EVERY SENTENCE IS EITHER PURE, SWEET TRUTH OR A VILE, CONTEMPTIBLE LIE! ONE OR THE OTHER! NOTHING IN BETWEEN!
© 1993 Watterson/Distributed by Universal Press Syndicate
WATTERSON 4-13

The Binomial Probability Formula

If n, p, x are as described on the previous slide, then

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} \tag{5}$$

Here are some R commands that help with the binomial distribution.

(1) Find the probability for 7 successes on 12 trials when the probability of success is 40% ($x = 7, n = 12, p = 0.4$):

```
dbinom(7,12,0.4)
0.1009024
```

The Binomial Distribution and R

(2) Do the same for $x = 0, 1, 2, 3, 4, 5$ and $n = 5, p = 0.40$:

```
dbinom(0:5,5,0.4)
0.07776 0.25920 0.34560 0.23040 0.07680 0.01024
```

(3) You can plot this distribution using (notice how skewed the distribution is because of $p = 0.40$):

```
x<-dbinom(0:5,5,0.4)
barplot(x)
```

The Binomial Distribution and R

(4) Often, you want to add the probabilities for a range of x. For example, what is the probability of having strictly fewer than 3 successes on 5 trials with $p = 0.7$?

```
pbinom(2,5,0.4)
0.68256
```

(5) You can list these probabilities as follows:

```
pbinom(0:5,5,0.4)
0.07776 0.33696 0.68256 0.91296 0.98976 1.00000
```

(6) You can simulate a binomial distribution using `rbinom`. Conduct 100 experiments where you perform 12 trials with a success probability $p = 0.4$. Here are the results (number of successes x) with a barplot:

```
x<-rbinom(100,12,0.4)
5 5 5 4 8 6 4 7 7 3 4 2 8 4 6 6 6 6 4 4 3 3 3 4 4 5 8 5
4 7 7 5 6 4 6 3 5 6 4 4 7 5 5 4 7 6 5 3 5 8 5 5 5 6 6 6
5 6 5 3 5 4 4 6 5 4 5 6 6 6 3 5 2 6 7 6 6 2 3 5 3 4 5 3
5 5 5 4 6 1 6 3 3 2 6 4 4 3 4 4
barplot(table(x))
```

Exercises for the Binomial Distribution I

Exercise 1: If you randomly guess on a multiple choice test with four possible answers, what is your probability of getting strictly more than 50% of questions right when there are six questions?

Exercises for the Binomial Distribution II

Exercise 2: The incidence of blue eyes in the population is 12%. In a room with 20 randomly selected people, what is the probability of having three or more people with blue eyes? What is the probability of having strictly fewer than five people with blue eyes?

> *Strictly speaking, the binomial probabilities are only approximate because the selection happens without replacement. If the population is large from which the sample is drawn, then you are allowed to ignore this.*

Exercise 3: Here is the distribution of blood types in Canada.

	O	A	B	AB
Positive	0.390	0.360	0.076	0.025
Negative	0.070	0.060	0.014	0.005

(a) What is the probability of being rhesus factor positive for someone of blood type "A"?

(b) If you meet four randomly selected Canadians, what is the probability that two of them are "O" positive?

(c) In a room with twelve randomly selected Canadians, what is the probability that there are strictly fewer than three people with blood type "B"?

Exercises for the Binomial Distribution IV

Exercise 4: 80.5% of US flights arrive on time. For twelve randomly selected flights, what is the probability that exactly ten of them are on time? What is the probability that between two and four of them are not on time?

Mean and Variance for the Binomial Distribution

There are formulas for the mean and variance of the binomial distribution. Especially the formula for the mean makes immediate sense:

Formulas

mean	μ	$=$	np
variance	σ^2	$=$	npq
standard deviation	σ	$=$	$\sqrt{npq}$

It is a useful rule of thumb to remember that it is unlikely ($< 5\%$) that x is outside of the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$.

Exercise 5: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercise 6: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of “O” positive blood types in a crowd of 100 Canadians.

Exercise 7: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

Exercise 8: What is the rule-of-thumb 95% interval for the following binomial procedures:

1. flipping a fair coin 15 times
2. answering 60 multiple choice questions with four possible answers for each question, where the probability of getting the right answer is 80%
3. randomly answering 60 multiple choice questions with four possible answers for each question
4. The number of "O" positive blood types in a crowd of 100 Canadians.

End of Lesson

Next Lesson: Poisson and Normal Distribution