## Example

Let $X_1$, $X_2$, $X_3$, ... represent the outcomes of successive tosses of my 40¢ coin.

Q: What is the probability of five heads in a row?

$$P(\ X_1=1 \text{ and } X_2=1 \text{ and } X_3=1 \text{ and } X_4=1 \text{ and } X_5=1\ ) = ?$$

If we <u>assume</u> that the trials are independent, then $X_1$, $X_2$, ..., $X_5$ are independent random variables, and we have five independent events, $X_1=1$, $X_2=1$, etc. The probability that all five events occur is the product of the individual probabilities,

$$P(X_1=1)P(X_2=1)...P(X_5=1) = \theta\theta\theta\theta\theta = \theta^5$$

where $\theta$ (the Greek letter "theta") represents the probability that the coin will fall heads on any one toss. For a fair coin $\theta=1/2$.

Q: What is probability of observing the pattern HHHTH?

P( $X_1$=1 and $X_2$=1 and $X_3$=1 and $X_4$=0 and $X_5$=1 )

$= P(X_1=1)P(X_2=1)P(X_3=1)P(X_4=0)P(X_5=1)$

$= \theta\theta\theta(1-\theta)\theta = \theta^4(1-\theta)$

Q: If we observe five heads in a row, what is the probability of Tails on the next toss?

P($X_6$=0 | $X_1$=1 and $X_2$=1 and $X_3$=1 and $X_4$=1 and $X_5$=1 )

$= P(X_1=1 \cap X_2=1 \cap X_3=1 \cap X_4=1 \cap X_5=1 \cap X_6=0)$
$/P(X_1=1 \cap X_2=1 \cap X_3=1 \cap X_4=1 \cap X_5=1)$

$= \theta^5(1-\theta)/\theta^5 = 1-\theta$

Since the trials are independent, what happens on the first five trials does not affect the probability on the sixth trial. So,

P($X_6$=0 | $X_1$=1 and $X_2$=1 and $X_3$=1 and $X_4$=0 and $X_5$=1 )

$= P(X_6=0) = 1-\theta$

## Another Example

Let $X_1$ represent the outcome of a toss of my 40¢ coin, and let the next flip be determined by $X_1$.

If $X_1 = 1$ (first toss is heads), then I toss the 40¢ piece again, but if $X_1 = 0$ (tails), then I toss a quarter.

Let $X_2$ represent the outcome of the second toss.

$$P(X_1 = 1) = \theta$$
$$P(X_2 = 1) = ?$$

When $X_1 = 1$ the second toss is also made with the 40¢ piece, so

$$P(X_2 = 1 \mid X_1 = 1) = \theta$$

But when $X_1 = 0$ the second toss is made with a quarter, so

$$P(X_2 = 1 \mid X_1 = 0) = 1/2$$

Therefore,

$$P(X_2 = 1) = P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) + P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0)$$

$$= \theta^2 + (1/2)(1-\theta)$$

(by the Law of Total Probability)

Because

$$P(X_2 = 1) \neq P(X_2 = 1 \mid X_1 = 1) \qquad \text{(in general)}$$

$X_1$ and $X_2$ are <u>not</u> independent random variables.
(Except in the trivial case when $\theta = 1/2$)

## Bernoulli Trials

Many random trials have the same form as the experiment of tossing a 40¢ coin. The simple probability model describing trials of this form is used so often that it has a name:

Bernoulli Trial Model: The random variable *X* has only two possible values, 0 and 1, so the sample space is simply S = {0,1}. The probability distribution is completely determined by one number, $P(X=1) = \theta$. (The probability of the other value, 0, is just $1-\theta$.)

Bernoulli trial models are used to represent random processes ("random trials") that have only two possible results:

girl or boy
heads or tails
responds or does not respond
age < 60 or age > 60
exposed or not exposed
diseased or not
dead or alive

One of the results is coded "0", and the other "1."

This is a "**Bernoulli trial probability model**":

(1) $X$ is the random variable
(2) $S = \{0,1\}$
(3) $P(X = 1) = \theta$, $P(X = 0) = 1-\theta$

We say:

- $X$ is a "Bernoulli random variable."
- $X$ has a Bernoulli probability distribution

**Note:**
This simple model has *two* variables:

(1) The random variable $X$, taking values 0 or 1.
(2) Another variable, $\theta$, which represents the probability that $X = 1$.

We use different types of letters, Latin and Greek, to represent these two variables in order to emphasize that they are different kinds of variables: $X$ is a random variable, and $\theta$ is an ordinary (non-random) variable.

- When $X$ is the result of tossing a quarter, with "1" representing "heads", $\theta = 1/2$.
- When $X$ is the sex of a baby, with "1" representing "boy", $\theta = 0.512$.
- When $X$ is the result of tossing a 40¢ piece, the value of $\theta$ is unknown.

## Binomial Probability Distribution

Single Bernoulli trials are not very interesting—too simple. Here are some more interesting random processes, along with some questions that we might want to answer. What are the appropriate probability models for these problems?

(1) A couple will have three babies. (What is the probability that they will have exactly one boy?)

(2) A couple will have five babies. (What is the probability that they will have three boys?)

(3) Fifty babies are born in a hospital. ( What is the probability that half are boys? What is the probability that half or more of them are boys?)

The questions all refer to a random variable, say Y, representing the number of boys born:

| | |
|---|---|
| (1) Sample space S = {0,1,2,3} | $P(Y = 1) = ?$ |
| (2) Sample space S = {0,1,2,3,4,5} | $P(Y = 3) = ?$ |
| (3) Sample space S = {0,1,2, ... ,50} | $P(Y = 25) = ?$ |
| | $P(Y \geq 25) = ?$ |

Consider the first question: P($Y$ = 1) = ?

We can describe this process in terms of three simpler ones: The couple has their first child,
... then their second,
... then their third.

These are three <u>independent</u> <u>Bernoulli</u> <u>trials</u>, with outcomes 0 (girl) or 1 (boy) and "success probability" θ = 0.512 (boy):

$X_1$ = sex of the first-born
$X_2$ = sex of the second-born
$X_3$ = sex of the third-born

⇨ **$Y = X_1 + X_2 + X_3$ ; $Y$ is the number of successes in the three independent Bernoulli trials.**

| $Y$ | P($Y$) | ($X_1$,$X_2$,$X_3$) | P($X_1$,$X_2$,$X_3$) |
|---|---|---|---|
| 0 | $(1-\theta)^3$ | (0,0,0) | $(1-\theta)^3$ |
| | | (1,0,0) | $\theta(1-\theta)^2$ |
| 1 | $3\theta(1-\theta)^2$ | (0,1,0) | $\theta(1-\theta)^2$ |
| | | (0,0,1) | $\theta(1-\theta)^2$ |
| | | (1,1,0) | $\theta^2(1-\theta)$ |
| 2 | $3\theta^2(1-\theta)$ | (1,0,1) | $\theta^2(1-\theta)$ |
| | | (0,1,1) | $\theta^2(1-\theta)$ |
| 3 | $\theta^3$ | (1,1,1) | $\theta^3$ |
| | 1 | | 1 |

These are the correct probabilities for Y if

(a) The three trials, $X_1$, $X_2$, and $X_3$, are independent, and
(b) $P(X_1 = 1) = P(X_2 = 1) = P(X_3 = 1) = \theta$. (All have the same probability of success)

In using this model for *Y*, we're assuming that (a) and (b) are true (or at least approximately true).

*This means, for example, that we're assuming that none of the babies are identical twins. Why?*

*Because if the first two are identical twins, then $X_1$ and $X_2$ will not be independent:* $P(X_1 = 1 \text{ and } X_2 = 0) = 0$, not $\theta(1-\theta)$.

*Or to put it another way, the conditional probability,* $P(X_1 = 1 \mid X_2 = 1) = 1$*, does not equal the unconditional probability,* $P(X_1 = 1) = \theta$.

How about the second process (Couple has five babies)?

Once Again, the number of boys, $Y$, is a sum of <u>independent</u> <u>Bernoulli</u> random variables, $X_1$, $X_2$,...,$X_5$.

| $Y$ | P($Y$) | ($X_1$,$X_2$, ...$X_5$) | P($X_1$,$X_2$,...$X_5$) |
|---|---|---|---|
| 0 | $(1-\theta)^5$ | (0,0,0,0,0) | $(1-\theta)^5$ |
| | | (1,0,0,0,0) | $\theta(1-\theta)^4$ |
| | | (0,1,0,0,0) | $\theta(1-\theta)^4$ |
| 1 | $5\theta(1-\theta)^4$ | (0,0,1,0,0) | $\theta(1-\theta)^4$ |
| | | (0,0,0,1,0) | $\theta(1-\theta)^4$ |
| | | (0,0,0,0,1) | $\theta(1-\theta)^4$ |
| | | (1,1,0,0,0) | $\theta^2(1-\theta)^3$ |
| | | (1,0,1,0,0) | $\theta^2(1-\theta)^3$ |
| | | (1,0,0,1,0) | $\theta^2(1-\theta)^3$ |
| | | (1,0,0,0,1) | $\theta^2(1-\theta)^3$ |
| 2 | $10\theta^2(1-\theta)^3$ | (0,1,1,0,0) | $\theta^2(1-\theta)^3$ |
| | | (0,1,0,1,0) | $\theta^2(1-\theta)^3$ |
| | | (0,1,0,0,1) | $\theta^2(1-\theta)^3$ |
| | | (0,0,1,1,0) | $\theta^2(1-\theta)^3$ |
| | | (0,0,1,0,1) | $\theta^2(1-\theta)^3$ |
| | | (0,0,0,1,1) | $\theta^2(1-\theta)^3$ |
| 3 | etc. | ... | ... |

There are five ways to have one boy, all with the same probability, $\theta(1-\theta)^4$, so

$$P(Y = 1) = 5\,\theta(1-\theta)^4.$$

There are ten ways to have two boys, all with the same probability, $\theta^2(1-\theta)^3$, so

$$P(Y = 2) = 10\,\theta^2(1-\theta)^3.$$

Etc....

Since we know that $\theta = 0.512$ we have:

$$\begin{aligned}
P(Y = 0) &= (1-\theta)^5 &&= 0.028\\
P(Y = 1) &= 5\theta(1-\theta)^4 &&= 0.145\\
P(Y = 2) &= 10\theta^2(1-\theta)^3 &&= 0.305\\
P(Y = 3) &= 10\theta^3(1-\theta)^2 &&= 0.320\\
P(Y = 4) &= 5\theta^4(1-\theta) &&= 0.168\\
P(Y = 5) &= \theta^5 &&= \underline{0.035}\\
& && \quad\ 1.000
\end{aligned}$$

Again, these are correct probabilities for $Y$ if the five Bernoulli trials are

(a) independent, and
(b) have the same success probability, $\theta = 0.512$.

This probability model can also be used to answer the third question concerning the number of boys born in a hospital having 50 births per day.

*The number of boys is a random variable, Y, which is the sum of fifty independent Bernoulli random variables.*

For any probability model that has this form, where *Y* is the number of successes in some fixed number, n, of independent Bernoulli trials, with probability of success θ on each trial, the random variable *Y* has possible values S = {0,1,2,...,n}, and

$$P(Y = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k} \qquad k = 0,1,2, ...,n$$

where $\binom{n}{k}$ is the number of different possible sequences of k ones and (n-k) zeroes (we'll see how to calculate this later).

⇨ ***In this case we say Y* is a "binomial random variable" or *Y* has a binomial probability distribution.**

A **Binomial probability model** consists of:

(1) $Y$ is number of "successes" out of n trials
(2) S = {0, 1, 2, ..., n}

(3) $P(Y = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ for k = 0,1,2, ...,n

Mathematically we say that

*If $X_1,\ldots,X_n$ are independent Bernoulli random variables with common success probability $\theta$, then*

$Y = \sum_{i=1}^{n} X_i$ *has a binomial probability distribution (with parameters n and $\theta$).*

For the hospital with n = 50 births, the number of males, *Y*, has a binomial probability distribution with parameters n = 50 and θ = 0.512.

The probability that exactly half of the babies will be boys is

$$P(Y = 25) = \binom{50}{25}(0.512)^{25}(1- 0.512)^{25} = 0.111$$

In this case, every sequence of 25 boys and 25 girls, such as 1101100... (first two babies are boys, third is a girl, then two more boys, etc.) has a <u>very</u> low probability,

$$(0.512)^{25}(1- 0.512)^{25} = 8.7547666 \times 10^{-16}.$$

But there are <u>very</u> many of these sequences,

$$\binom{50}{25} = 1.26410606 \times 10^{14},$$

and the probability of observing one of them is

$$(8.7547666 \times 10^{-16})(1.26410606 \times 10^{14})=0.111$$

What is the probability of <u>this</u> sequence: 1111...111000...0000 ? (First 25 babies born are all boys, then next 25 are all girls)

What is the probability of this one: 1010101010...101010 ?

What is the probability of this sequence: 1101100010...001 ? (This one has no pattern that is obvious.)

The probability that half or more of the 50 babies are boys is

$$P(Y \geq 25) = P(Y = 25) + P(Y = 26) + \cdots + P(Y = 50) = \sum_{k=25}^{50} \mathrm{P}(Y = k)$$

(why?)

$$\sum_{k=25}^{50} \mathrm{P}(Y = k) = 0.111 + 0.112 + 0.104 + 0.090 + ... + 0 + 0 = 0.622$$

**Application?**

If the number of babies is 500, instead of 50, the probability that half or more are boys increases to 0.72, and if the number of babies is 5000, the probability of 2500 or more boys is 0.96.

Calculations like these were used in what is sometimes cited as the first published description of the reasoning behind statistical tests. That was "An Argument for Divine Providence Taken From the Constant Regularity of the Births of Both Sexes". (John Arbuthnot, 1710)

Arbuthnot looked at the register of births for the city of London, and found that for 82 consecutive years there had been more male than female births. He reasoned that this was either the result of chance or of "art"; i.e., the process might be a random one, but if it is not, then the result must be determined by "Divine Providence".

He figured that if the process was determined by chance, then

$$P(82 \text{ consecutive "boy" years}) = (1/2)^{82} = 2.068 \times 10^{-25}$$

Because this probability is so small, Arbuthnot rejected the hypothesis that the process is a random one, and concluded that "it follows that it is art, not Chance, that governs" in the distribution of sexes.

Arbuthnot assumed that if it were a random, or "chance", process, then the probability that the number of boys would exceed the number of girls in a year must be 1/2.

*Let's see …* assuming just for illustration that there are 10,000 births in a year. The number of male births, say *Y*, has a binomial distribution, and the probability of a "male year" (more than half of those born are boys), $P(Y > 5000)$, is

$$\sum_{k=5001}^{10,000} P(Y=k) = \sum_{k=5001}^{10,000} \binom{10,000}{k} \theta^{k} (1-\theta)^{10,000-k}$$

If on each birth the probability of a boy is $\theta = 0.5$, this probability turns out to be 0.496, so the probability of 82 consecutive "male years" is $(0.496)^{82} = \boldsymbol{1.072 \times 10^{-25}}$.

*But*, if on each birth the probability of a boy is not 0.5, but 0.512, then the probability of a "male year" turns out to be 0.99159, so the probability of 82 consecutive "male years" is $(0.99159)^{82} = \boldsymbol{0.500}$. In that case the fact that there were 82 consecutive "male years" is no more improbable or surprising than getting "heads" on a coin toss -- it is just the sort of result that you would expect to see if the process were truly random.

**The Number** $\binom{n}{k}$

$\binom{n}{k}$ is the number of ways of choosing the k trials (from the n) where the "successes" will occur.

$\binom{n}{k}$ is read "n choose k."

Possible Sequences

$$\left.\begin{array}{l} (\underbrace{1,1,\ldots,1}_{k\ 1\text{'s}},\underbrace{0,0,\ldots,0}_{n-k\ 0\text{'s}}) \\ \\ (1,0,0,1,\ldots..0,0,1) \\ \\ (\underbrace{0,0,\ldots,0}_{n-k},\underbrace{1,1,\ldots,1}_{k}) \end{array}\right\} \binom{n}{k} \text{ sequences}$$

If we want to choose a committee of k members from a population of n people, there are

$\binom{n}{k}$ different committees that we can choose.

*Note: Same person cannot appear twice—a committee of three cannot consist of you, me, and me. Once a person has been selected to be on the committee, they can't be selected again.*

Here is an example with n = 5. Population: A, B, C, D, E

| | | |
|---|---|---|
| k = 1 | A B C D E | $\binom{5}{1}=5$ |
| k =2 | AB AC AD AE<br>BC BD BE DE<br>CD CE | $\binom{5}{2}=10$ |
| k = 3 | ABC ABD ABE<br>ACD ACE ADE<br>BCD BCE BDE<br>CDE | $\binom{5}{3}=10$ |
| k = 4 | ABCD ABCE ABDE ACDE<br>BCDE | $\binom{5}{4}=5$ |
| k = 5 | ABCDE | $\binom{5}{5}=1$ |

**The number of different samples of size k that can be chosen from a population of size n (Read "n choose k")**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

**where n! = n(n-1)(n-2)...(2)(1).**

(with the special definition: 0! = 1 )

$$\binom{5}{2} = \frac{5!}{2!\ 3!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(3 \times 2 \times 1)} = 10$$

$$\binom{10}{3} = \frac{10!}{3!\ 7!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} = 120.$$

$$\binom{n}{k} = {}_{n}C_{k}$$

(Alternative notation used on some calculators)

**Example** (Tversky and Kahneman)

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50%, sometimes lower.

For a period of 1 year, each hospital recorded the days on which more than 60% of the babies born were boys. Which hospital do you think recorded more such days?

- ☞ The larger hospital
- ☞ The smaller hospital
- ☞ About the same

**Modeling the question:**
The number of boys born in the larger hospital, say *X*, has a binomial(45, 1/2) probability distribution, and the number born in the smaller hospital, *Y*, has a binomial(15, 1/2) distribution.

$X \sim Bin(n=45, \theta=1/2)$
$Y \sim Bin(15, 1/2)$

Now 60% of 45 is 27, while 60% of 15 is 9.

Thus the probability that boys are more than 60% of the babies born in one day at the larger hospital is:

$$P(X > 27) = \sum_{k=28}^{45} P(X=k) = \sum_{k=28}^{45} \binom{45}{k} (0.5)^k (0.5)^{45-k} = 0.068$$

The probability that boys are more than 60% of the babies born in one day at the smaller hospital is:

$$P(Y > 9) = \sum_{k=10}^{15} P(Y=k) = \sum_{k=10}^{15} \binom{15}{k} (0.5)^k (0.5)^{15-k} = 0.151$$

On any given day, the probability that more than 60% of the babies born in the small hospital will be boys is 0.151, more than twice the probability in the large hospital, which is only 0.068.

In a year we can expect to see this happen (more than 60% of the babies born in one day are boys) in the small hospital on about

$$(0.151)(365) = 55.1 \text{ days.}$$

In the large hospital it will happen on only about

$$(0.068)(365) = 24.8 \text{ days.}$$

⇨ **In general, we say the expected number of success from a binomial random variable, say $Y \sim Bin(n, \theta)$, is simply $n\theta$. Mathematically we write $E(Y) = n\theta$.**

(We'll learn more about expected values later.)

## Probabilities and Observed Proportions

> <u>*Very important distinction*</u>*: When we observe a binomial random variable, $Y$, and see the value $Y = y$, the observed proportion of successes, $y/n$, is not the probability of success, $\theta$.*

I tossed my 40¢ piece 10 times and observed 5 heads. This does not establish that $\theta = 5/10 = 0.5$. When I tossed it 5 more times, I observed 2 heads, for a total of 7 heads in 15 trials. This does not prove that $\theta = 7/15$, either.

Every toss represents evidence about the probability, $\theta$. If I make a large number of tosses, I will have very strong evidence, and the proportion of successes will probably be very close to $\theta$.

(This is what statistics, as opposed to probability, is all about— how to interpret and use evidence of this sort.)

The proportion of heads in a large number of trials is a random variable. The probability that this random variable will be close to θ, say within 10%, gets larger and larger as the sample size increases. **But it never reaches 1 and the probability that the proportion of heads will <u>not</u> be close to θ never reaches zero.** *The proportion is always only an estimate of θ, not θ itself.*

No value of n is "large enough to satisfy the frequency definition of probability." No matter how large n is, one more trial will <u>change</u> the <u>proportion</u> of successes (from y/n to either y/(n+1) or (y+1)/(n+1)).

The proportion of successes represents an <u>estimate</u> of the probability of success, and when the number of trials is large, this estimate will usually be very close to the probability (we shall soon learn why).

But because there is always a chance that the proportion of trials will <u>not</u> be close to the true probability, there is an important distinction between the estimate (the observed proportion) and the thing being estimated (the probability θ).