The Binomial distribution

Cécile Ané

Stat 371

Spring 2006

Outline

1. Examples
2. General case
3. Assumptions
4. Mean and Standard deviation

Coin tossing example

Consider tossing a fair coin 3 times independently.
Possible outcomes:

HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

Let $Y =$ # of Heads.

$$\begin{aligned}\mathbb{P}\{Y=3\} &= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}\\ \mathbb{P}\{Y=2\} &= 3\,(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}) = \frac{3}{8}\end{aligned}$$

Other example: experiment to mutate a gene in bacteria. The experiment was repeated 3 times independently with 3 colonies.

- Heads = got a mutation
- Tail = failure, no mutation.

Tossing an unfair coin

Same setting, but $p = \mathbb{P}\{\text{Heads}\} \neq 0.5$.
Mutagenesis experiment: $p = \mathbb{P}\{\text{getting a mutation}\}$.

HHH
HHT
HTH
HTT
THH
THT
TTH
TTT

$$\begin{aligned}\mathbb{P}\{Y=3\} &= p^3\\ \mathbb{P}\{Y=2\} &= p\,p\,(1-p) + p\,(1-p)\,p + (1-p)\,p\,p\\ &= 3p^2(1-p)\end{aligned}$$

How do we get $\mathbb{P}\{Y=j\}$, $j=1,2,3$ without counting all possibilities?

General case

We have n **coin tosses**, with **probability** p **of Heads**. We conduct the trials independently. What is the probability of **getting exactly** j **Heads**?

$$\begin{aligned}\mathbb{P}\{Y=j\} &= \frac{n!}{j!(n-j)!}\,p^j(1-p)^{n-j}\\ &= {}_nC_j\;p^j(1-p)^{n-j}\end{aligned}$$

with

- $p^j = \overbrace{p\,p\ldots p}^{j \text{ times}}$, $p^0 = 1$,
- and factorial notation: $j! = j(j-1)\ldots 2.1$ and $0! = 1$
  Ex: $4! = 4.3.2.1 = 24$.
- cf. Table 2 p.674 for the ${}_nC_j$ numbers.

General case

$$\mathbb{P}\{Y=j\} = \frac{n!}{j!(n-j)!}\,p^j(1-p)^{n-j}$$

Examples:
$p = .5$ (fair coin), $n = 3$ tosses, $j = 2$ Heads. We get

$$\mathbb{P}\{Y=2\} = \frac{3!}{2!\,1!}(.5)^2(.5)^1 = \frac{3.2.1}{2.1\;1}(.5)^3 = 3/8$$

as before. In general, for any p, we get

$$\mathbb{P}\{Y=2\} = \frac{3!}{2!\,1!}p^2(1-p)^1 = 3\,p^2(1-p)$$

as before.

Calculation examples

- 2 successes out of 7 trials, probability of success is $p = .6$: $n = 7, j = 2$

$$\begin{aligned}\mathbb{P}\{Y = 2\} &= \frac{7!}{2!\,5!}(.6)^2(.4)^5 = \frac{7.6\;5!}{1.2\;5!}(.6)^2(.4)^5 \\ &= 21(.6)^2(.4)^5 = .0774\end{aligned}$$

- 5 failures out of 7 trials, probability of failure is $p = .4$: $n = 7, j = 5$

$$\begin{aligned}\mathbb{P}\{Y = 5\} &= \frac{7!}{5!\,2!}(.4)^5(.6)^2 = \frac{7.6\;5!}{5!\;2.1}(.4)^5(.6)^2 \\ &= 21(.4)^5(.6)^2 = .0774\end{aligned}$$

same result...

Calculation examples

A new drug is available. Its success rate is $1/6$: probability that a patient is improved. I try it independently on 6 patients.

Probability that at least one patient improves?
$p = 1/6$, $n = 6$, $j = 1, 2, 3, 4, 5$ or 6.

$$\begin{aligned}\mathbb{P}\{\text{at least one improves}\} &= \mathbb{P}\{Y = 1 \text{ or } Y = 2 \text{ or} \ldots \text{ or } Y = 6\} \\ &= \mathbb{P}\{Y = 1\} + \mathbb{P}\{Y = 2\} + \cdots + \mathbb{P}\{Y = 6\} \\ &= 1 - \mathbb{P}\{Y = 0\} \\ &= 1 - \frac{6!}{0!6!}\,(1/6)^0(5/6)^6 = 1 - (5/6)^6 \\ &= .665\end{aligned}$$

Notation

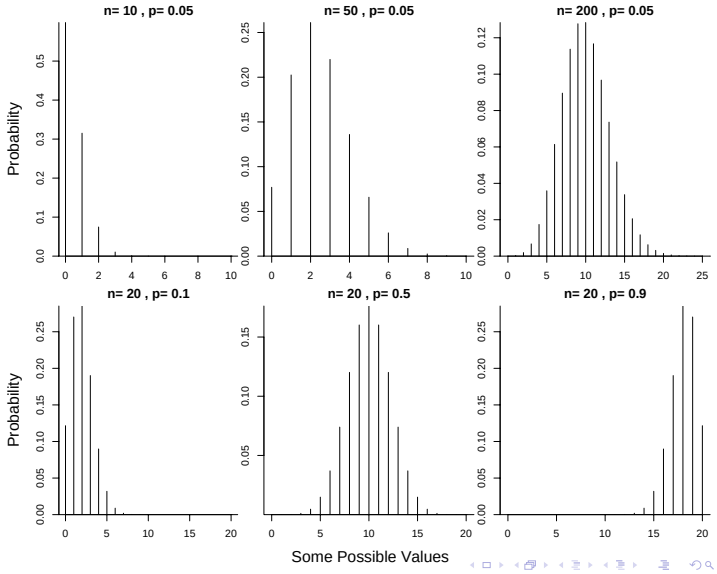
Consider a random variable Y where $Y = \#$ of successes. Suppose we have n trials. We write

$$Y \sim \mathcal{B}(n, p)$$

$\mathcal{B}$ for binomial. In the last example we had $Y \sim \mathcal{B}(6, 1/6)$. This notation is a shorthand for this distribution table:

y	0	1	2	3	4	5	6
$\mathbb{P}\{Y = y\}$	0.335	0.402	0.200	0.054	0.008	0.0006	0.00002

describes Y's probability distribution, i.e. its probability mass function.



Underlying assumptions

- Trials have exactly 2 outcomes
- The probability of success p is the same for all trials.
- Number of trials n is fixed in advance. If new patients are enrolled until one of them at least gets improved, then the binomial is not the correct distribution for Y.
- All trials are independent.

Underlying assumptions

Are assumptions met for

- $Y = \#$ kids with a cold, out of 20 in Mrs. Smith's kindergarten class
- $Y = \#$ days with rain next week
- $Y = \#$ of students who answered "7" to the last question on the survey (pick a random number)
- Consider April 1, May 1, June 1, . . . , October 1. (7 dates) $Y = \#$ days with rain out of these 7 days.
- Drug trial on rats housed 3 in a cage

The binomial is a **model**. It is not reality. It is a way to provide structure on real world phenomena.

Mean and Standard deviation

If $Y \sim \mathcal{B}(n,p)$ then

$$\mu = \mathbb{E}Y = np \qquad \text{and}$$

$$\sigma^2 = np(1-p) \qquad \text{i.e. } \sigma = \sqrt{np(1-p)}$$

Coin tossing: $p = 0.5$.

- $n = 100$ tosses. Mean: 50 Heads, standard deviation $\sqrt{n}/2 = 5$ Heads.

  What outcome would you predict: between 40 and 60 Heads? between 30 and 70 Heads?

- $n = 10,000$ tosses. $\mu = 5,000$ and $s = 100/2 = 50$.

  What would you predict: between 4,500 and 5,500 Heads? between 4900 and 5100 Heads? between 4990 and 5010 Heads?