DISCRETE RANDOM VARIABLES

The binomial distribution

Binomial distribution

- A BINOMIAL EXPERIMENT has the following characteristics:
  - A fixed (predetermined) number of identical "trials"
  - Each trial results in one of two possible outcomes ("success" or "failure")
  - The probability of "success" is the same for all trials ($P(\text{"}success\text{"}) = p$)
  - The trials are independent
- If Y is the total number (out of the n trials) of successes, then we say that Y has a binomial distribution.
- Notation:

$$Y \sim Binom(n, p)$$

Examples

- Example 5: Ten patients are recruited; at random each one is given an experimental treatment and researchers record whether each patient's symptoms improve after 24 hours. Let Y be the number of patients who show improvement. Does Y have a binomial distribution?

  n = 10

  YES

- Example 6: There are 30 patients in a registry. Eight are selected at random, each one is tested for hypertension. Let Y be the number of patients with hypertension. Does Y have a binomial distribution?

  $\frac{7}{30}$ $\frac{7}{29}$ ~~indep~~

  NO

Derivation of the binomial distribution

- For n trials of "S" and "F", what does the sample space S look like?

$n=4$ $\{SSSS,\ SSSF,\ SSFS,\ \ldots\}$

- What are the simple events associated with the event $Y = y$ (i.e., exactly y successes out of n)?

$n=4$
$y=2$

SSFF SFSF SFFS
FSSF FSFS FFSS

$\frac{6}{2}$

Derivation of the binomial distribution

- How many simple events in the event $Y = y$?

How many ways to group y S and n-y F ?

$$C_y^n = \binom{n}{y} = \frac{n!}{y!\,(n-y)!}$$

"counting"

- What is the probability of each simple event in the event $Y = y$?

$$\underbrace{S \quad S \quad S \quad \ldots \quad S}_{y} \quad \underbrace{F \quad F \quad \ldots \quad F}_{n-y}$$

$$p \quad p \quad p \quad \quad p \quad 1-p \quad 1-p \quad \quad 1-p$$

indep. so can multiply

$$p^y (1-p)^{n-y}$$

Derivation of the binomial distribution

- What is $P(Y = y)$?

simple events with $Y = y$ is $\binom{n}{y}$

Prob. of each event is $p^y (1-p)^{n-y}$

Add these probabilities

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

Example 7

- In the US, type O+ is the most common blood type, comprising 37% of the population. If 10 donors are selected from the population at random, what is the probability that exactly two of them will be O+?

10 repeated trials $n = 10$

+ "success" O+ "failure" = anything else

same prob. of success? Y $0.37 = p$

indep? YES

$$P(Y=2) = \binom{10}{2}(0.37)^2(0.63)^8 = \frac{\cancel{10}^{5} \cdot 9 \cdot \cancel{8!}}{\cancel{8!} \cdot \cancel{2} \cdot 1}(0.37)^2 \times (0.63)^8$$

$$45(0.37)^2(0.63)^8 = \ldots$$

Cumulative binomial probabilities

- Single probabilities ($P(Y = y)$) are relatively simple to calculate; cumulative probabilities can take some work.
- $P(Y \leq a) = \sum_{i=0}^{a} P(Y = i) = P(Y = 0) + P(Y = 1) + \cdots + P(Y = a)$
- Textbook: Appendix 3 Table 1 gives cumulative binomial probabilities
- Also R: (previous example) To get the probability that fewer than five are O+ you could do

$P(Y \leq 4)$ y

```
> pbinom(q=4, size=10, prob=0.37)
[1] 0.7060723
```

cumulative

n   p

$P(Y \leq y)$ if $Y \sim \text{binomial}(n, p)$

Mean and variance of a binomial random variable

- Theorem 7: If Y is a random variable based on n trials and success probability p, then

$$E[Y] = \boxed{np}; \qquad Var(Y) = np(1-p).$$

$y=0$ term is 0

Proof $E[Y] \overset{def}{=} \sum_y y\,P(Y=y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=1}^{n} y \frac{n!}{y!\,(n-y)!} p^y (1-p)^{n-y}$

$(y-1)!$

$$= \sum_{y=1}^{n} \frac{n!}{(y-1)!\,(n-y)!}\, p\, p^{y-1} (1-p)^{n-y} = np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!\,(n-y)!}\, p^{y-1} (1-p)^{n-y}$$

change of variables
$x = y-1$
$y=1 \to x=0$
$y=n \to x=n-1$

$$= np \underbrace{\sum_{x=0}^{n-1} \frac{(n-1)!}{x!\,(n-1-x)!}\, p^x (1-p)^{n-1-x}}_{\text{sum of all prob for " Binom}(n-1,p)} = 1$$

$= np$ $\square$

Example 8

- For the blood type example, if Y is the number of O+ donors, find $E[Y]$, $Var(Y)$, and σ.

$p = 0.37 \qquad n = 10$

$$E[Y] = np$$
$$= 10(0.37) = 3.7$$

$$Var(Y) = np(1-p) = 10(0.37)(0.63) =$$

$$\sigma = \sqrt{Var(Y)} = \sqrt{10(0.37)(0.63)} =$$

X_1 X_2 X_3 X_n

S F F . . . S

1 0 0 . . . 1

Bernoulli distribution

- A special case of the binomial distribution is the BERNOULLI DISTRIBUTION in which $n = 1$, i.e., the only outcomes are 1 (with probability p) and 0 (with probability $1 - p$).

  S F

- If $X_1, X_2, \ldots, X_n \overset{ind.}{\sim} Bernoulli(p)$ then $Y = \sum_{i=1}^{n} X_i \sim Binom(n, p)$.
- $E[X_i] = p;\ \ Var(X_i) = p(1 - p)$

$E[Y] = np$

$n = 1 \quad E[X_i] = p \quad Var(X_i) = p(1-p)$