Inference about a Population Proportion (π)

AAO unit 28; Baldi & Moore, Ch 19

Sahir Bhatnagar and James Hanley

EPIB 607
Department of Epidemiology, Biostatistics, and Occupational Health
McGill University

sahir.bhatnagar@mcgill.ca
https://sahirbhatnagar.com/EPIB607/

October 16, 2018

Binomial Model for Sampling Variability of Proportion/Count in a Sample

The Binomial Distribution: what it is

- It is the $n+1$ probabilities $p_0, p_1, ..., p_y, ..., p_n$ of observing $0, 1, 2, \ldots, n$ "positives" in n independent realizations of a Bernoulli random variable Y:

$$Y = \begin{cases} 1 & P(Y=1) = \pi \\ 0 & P(Y=0) = 1-\pi \end{cases}$$

  The number is the sum of n i.i.d. Bernoulli random variables. (*such as in SRS of n individuals*)

The Binomial Distribution: what it is

- It is the $n+1$ probabilities $p_0, p_1, ..., p_y, ..., p_n$ of observing $0, 1, 2, \ldots, n$ "positives" in n independent realizations of a Bernoulli random variable Y:

$$Y = \begin{cases} 1 & P(Y=1) = \pi \\ 0 & P(Y=0) = 1-\pi \end{cases}$$

  The number is the sum of n i.i.d. Bernoulli random variables. (*such as in SRS of n individuals*)

- Each of the n observed elements is binary (0 or 1)

The Binomial Distribution: what it is

- It is the $n+1$ probabilities $p_0, p_1, ..., p_y, ..., p_n$ of observing $0, 1, 2, \ldots, n$ "positives" in n independent realizations of a Bernoulli random variable Y:

$$Y = \begin{cases} 1 & P(Y=1) = \pi \\ 0 & P(Y=0) = 1 - \pi \end{cases}$$

  The number is the sum of n i.i.d. Bernoulli random variables. (*such as in SRS of n individuals*)

- Each of the n observed elements is binary (0 or 1)

- There are 2^n possible *sequences* … but only $n+1$ possible *values*, i.e. $0/n,\ 1/n,\ \ldots,\ n/n$ *(can think of y as sum of n Bernoulli random variables)*

- Note: it is better to work in same scale as the parameter, i.e., in [0,1]. Not the [0,n] count scale.

The Binomial Distribution: what it is

- Apart from (n), the probabilities p_0 to p_n depend on only 1 parameter:
  - the probability that a selected individual will be "positive" i.e.,
  - the proportion of "positive" individuals in sampled population
- Usually denote this (un-knowable) proportion by π

Author	**Parameter**	**Statistic**
Clayton & Hills	π	$p = D/N$
Hanley et al.	π	$p = y/n$
M&M, Baldi & Moore	p	$\hat{p} = y/n$
Miettinen	P	$p = y/n$

- Shorthand: $Y \sim \text{Binomial}(n, \pi)$.

Example

- Suppose a woman plans to have 3 children.
- Suppose at each birth,

$$P(\text{female child}) = 1/2$$

  and the sex of the child at each birth is independent of the sex at any previous birth.
- What is the probability of having all daughters?

The binomial distribution

F (1/2) M (1/2)

FF (1/4) $\underbrace{\text{FM} \quad \text{MF}}_{(1/2)}$ MM (1/4)

FFF (1/8) $\underbrace{\text{FFM FMF MFF}}_{(3/8)}$ $\underbrace{\text{FMM MFM MMF}}_{(3/8)}$ MMM (1/8)

The binomial distribution

Let Y be the number of daughters a woman will have, n the number of children she will have, and p the probability of a daughter at any birth. Then:

$$P(Y = k \text{ out of } n) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)}$$

where $n! = 1 \times 2 \times 3 \times ... \times (n-1) \times n$, and $0! = 1$.

The binomial distribution

Let Y be the number of daughters a woman will have, n the number of children she will have, and p the probability of a daughter at any birth. Then:

$$P(Y = k \text{ out of } n) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)}$$

where $n! = 1 \times 2 \times 3 \times ... \times (n-1) \times n$, and $0! = 1$.

The binomial distribution

The binomial distribution has the following four assumptions:

1. There are n trials.
2. Each trial has two possible outcomes, "success" or "failure." (Usually, when coding the outcomes, we take 1 to be the success outcome.)
3. The probability of success, p, is the same for each trial.
4. Each trial is independent.