Chapter 5.8 Binomial Probabilities and the Normal Curve

Jim Albert and Monika Hu

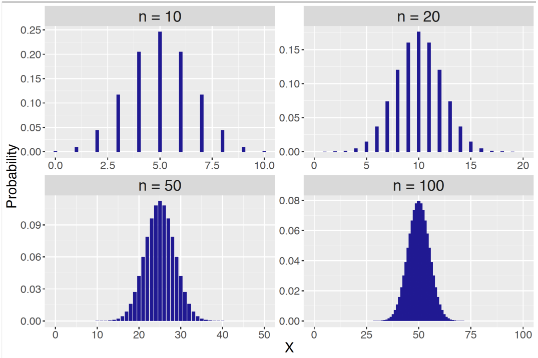
Chapter 5 Continuous Random Variables

A Binomial Story

- Suppose that half of one's student body is female and one takes a sample survey of n students to learn if they are interested in using a new proposed recreational sports complex.
- Let X denote the number of females in the sample.
- We know that X will be distributed Binomial with parameters n and $p = 1/2$.
- Let's explore the shape of these binomial distributions for different sample sizes.

Binomial Shapes with $p = 0.5$

- Here are Binomial probabilities for probability of success $p = 0.1$ and sample sizes $n = 10$, 20, 50, and 100.

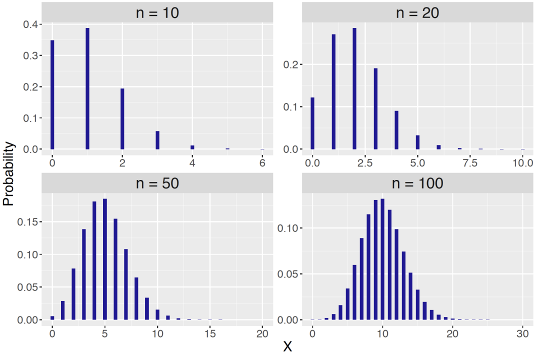


Comments

- Note that each distribution is symmetric about the mean $\mu = np$.
- Also the shape of the distribution seems to resemble a Normal curve as the number of trials n increases.

Binomial Shapes with $p = 0.1$

- Same binomial story but suppose the probability of success is $p = 0.1$ instead of $p = 0.5$.
- Below figure shows the probability distributions again for the sample sizes $n = 10$, 20, 50, and 100.



Comments

- Again note that as n increases, the probabilities become more Normal-shaped and the Normal curve seems to be a good match for $n = 100$.

Normal approximation to binomial

- Have a Binomial random variable X with n trials and probability of success p
- As the number of trials n approaches infinity, the distribution of the standardized score

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{1}$$

approaches a Normal distribution with mean 0 and standard deviation 1.

- It means, that for a large number of trials, one can approximate a Binomial random variable X by a Normal random variable with

$$\mu = np, \;\; \sigma = \sqrt{np(1-p)}. \tag{2}$$

Example

- Suppose that 10% of the student body would use the new recreational sports complex.
- One takes a random sample of 100 students — what's the probability that 5 or fewer students in the sample would use the new facility?
- The random variable X in this problem is the number of students in the sample that would use the facility.
- X has a Binomial distribution with $n = 100$ and $p = 0.1$.

Approximation

- The exact binomial distribution is shown below.
- We can approximate by a Normal curve with $\mu = 100(0.1) = 10$ and $\sigma = \sqrt{100(0.1)(0.9)} = 3$
- Note that it is a pretty good fit to the histogram.

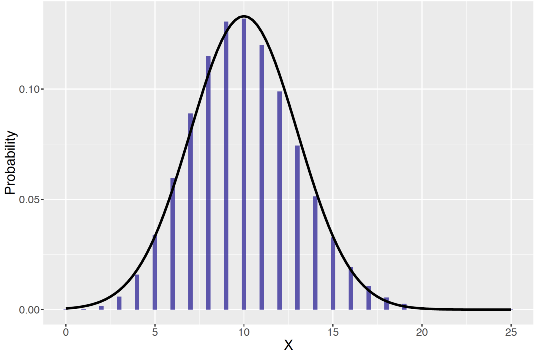


Figure 1: Histogram of Binomial probabilities, with the

Binomial Computations Using a Normal Curve

- One is interested in the probability that at most 5 students use the facility – $P(X \leq 5)$.
- This probability is approximated by the area under a Normal(10, 3) curve between $X = 0$ and $X = 5$.
- Using the R `pnorm()` function, we compute this Normal curve area to be

```r
pnorm(5, 10, 3) - pnorm(0, 10, 3)
```

```
## [1] 0.04736129
```

How accurate is this normal approximation?

- Compare with an exact binomial calculation. Using the `pbinom()`, we find the probability that X is at most 5 is

```r
pbinom(5, size = 100, prob = 0.10)
```

```
## [1] 0.05757689
```

- Here one sees that the Normal approximation gives a similar answer to the exact Binomial computation.