# STAT 234 Lecture 6A
# Binomial Distributions, Part 2
# Section 3.5

Yibi Huang
Department of Statistics
University of Chicago

# Binomial Distribution (Review)

Suppose $n$ **independent** Bernoulli trials are to be performed, each of which results in

- a *success* with probability $p$ and
- a *failure* with probability $1 - p$.

If we define

$$X = \text{the number of successes that occur in the } n \text{ trials},$$

then $X$ is said to have a *binomial distribution* with parameters $(n, p)$, denoted as

$$X \sim Bin(n, p).$$

with the *probability mass function (pmf)*

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

# Binomial Conditions (Review)

Conditions required to apply the binomial formula:

1. each trial outcome must be classified as a *success* or a *failure*
2. the probability of success, $p$, must be the same for each trial
3. the number of trials, $n$, must be fixed
4. the trials must be independent
   - Draws made *without replacement* from a population are dependent
   - However, if the sample size (number of trials/draws) is *at most 5%* of the population size, the trials (outcomes of draws) are *approx. independent*.

# Example: Driver's License

Suppose 80% of UChicago undergrads have their driver's license. Among a random sample of 10 UChicago undergrads, what is the probability that exactly 6 have license? Exactly 8?

# Example: Driver's License

Suppose 80% of UChicago undergrads have their driver's license. Among a random sample of 10 UChicago undergrads, what is the probability that exactly 6 have license? Exactly 8?

Let $X$ = the number of undergrads w/ license in a sample of size 10. $X \sim Bin(n = 10, p = 0.8)$

$$\mathrm{P}(X = 6) = \binom{10}{6} \times 0.8^6 \times 0.2^4 = 210 \times 0.8^6 \times 0.2^4 \approx 0.088$$

$$\mathrm{P}(X = 8) = \binom{10}{8} \times 0.8^8 \times 0.2^2 = 45 \times 0.8^8 \times 0.2^2 \approx 0.302.$$

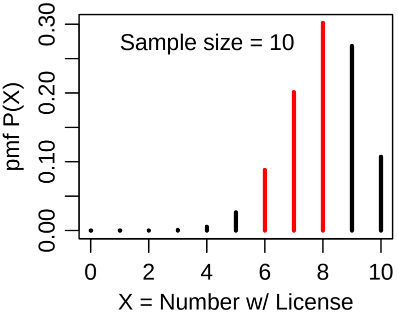
R command for calculating $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ is `dbinom(x, size = n, p)`

```
dbinom(6, size = 10, p = 0.8)
[1] 0.08808038
dbinom(8, size = 10, p = 0.8)
[1] 0.3019899
```

# Example: Driver's License (Cont'd)

In a sample of size 10, what is the probability that 6 to 8 of them have license?

$$\begin{aligned} P(6 \le X \le 8) &= P(X = 6) + P(X = 7) + P(X = 8) \\ &= \binom{10}{6} 0.8^6 0.2^4 + \binom{10}{7} 0.8^7 0.2^3 + \binom{10}{8} 0.8^8 0.2^2 \\ &\approx 0.088 + 0.201 + 0.302 = 0.591 \end{aligned}$$



# Expected Value, Variance, and SD of Binomial Distributions

# Expected Value, Variance and SD of Bin$(n = 1, p)$

A Binomial random variable $X \sim$ Bin$(n, p)$ with $n = 1$ can only take value 0 or 1 with the distribution below

| value of $X$ | 0 | 1 |
|---|---|---|
| probability | $1 - p$ | $p$ |

The expected value, variance, and SD of Bin$(n = 1, p)$ can be calculated as follows.

$$\mathrm{E}(X) = \sum\nolimits_{x=0,1} xp(x) = 0 \cdot (1 - p) + 1 \cdot p = \boxed{p},$$

$$\mathrm{E}(X^2) = \sum\nolimits_{x=0,1} x^2 p(x) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p,$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = p - p^2 = \boxed{p(1 - p)} \quad \text{by the shortcut formula}$$

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \boxed{\sqrt{p(1 - p)}}$$

## Expected Value, Variance and SD of Bin$(n, p)$

For a Binomial random variable $X \sim Bin(n, p)$, the mean, variance, and the SD are respectively

$$\mu = \mathrm{E}(X) = np$$
$$\sigma^2 = \mathrm{Var}(X) = np(1 - p)$$
$$\sigma = \mathrm{SD}(X) = \sqrt{np(1 - p)}$$

Note the SD increases proportionally to $\sqrt{n}$, not $n$.

The proof will be given next week

Suppose 80% of UChicago undergrads have their driver's license. Among a random sample of 63 UChicago undergrads, how many do you expect to have driver's license? With what SD?

$$X \sim \text{Bin}(n = 63, p = 0.8)$$

$$\text{E}(X) = np = 63 \times 0.8 = 50.4,$$

$$\text{SD}(X) = \sqrt{np(1-p)} = \sqrt{63(0.8)(1-0.8)} \approx 3.175$$

Mean and SD of Binomial distributions *might NOT be whole numbers*, and that is alright, these values represent what we would expect to see on average. **Do NOT round the mean and SD to integers**.

# 99.7% of the Probability Are Within 3 SDs from Mean

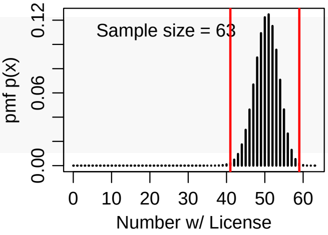
Almost all (99.7%) the probability of a random variables are *within 3 SDs away from the mean (expected value)*.

From the $\mathrm{E}(X)$ and SD just computed, the possible number of students w/ license in a sample of size 63 is likely between

$$50.4 \pm (3 \times 3.175) \approx (40.875, 59.925)$$

For $X \sim$ Bin$(n = 63, p = 0.8)$, $P(41 \leq X \leq 59) \approx 0.9976$ and $P(40 \leq X \leq 60) \approx 0.9992$.

```
sum(dbinom(41:59, size=63, p = 0.8))
[1] 0.9976559
sum(dbinom(40:60, size=63, p = 0.8))
[1] 0.9991908
```



# Proof of the Expected Value of Binomial (Optional)

$$\mathrm{E}(X) = \sum_{x=0}^{n} xP(X = x) = \sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x\binom{n}{x}p^x(1-p)^{n-x}$$

One key step: observe that

$$x\binom{n}{x} = x\frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!(n-x)!} = \frac{n \times (n-1)!}{(x-1)!(n-x)!} = n\binom{n-1}{x-1}.$$

Replacing $x\binom{n}{x}$ in $\mathrm{E}(X)$ with $n\binom{n-1}{x-1}$, we get

$$\begin{aligned}
\mathrm{E}(X) &= \sum\nolimits_{x=1}^{n} x\binom{n}{x}p^x(1-p)^{n-x} \\
&= \sum\nolimits_{x=1}^{n} n\binom{n-1}{x-1}p^x(1-p)^{n-x} \\
&= np\sum\nolimits_{x=1}^{n} \binom{n-1}{x-1}p^{x-1}(1-p)^{n-x} \\
&= np\sum\nolimits_{k=0}^{n-1} \binom{n-1}{k}p^k(1-p)^{n-1-k} \qquad \text{let } k = x-1 \\
&= np(p+1-p)^{n-1} = np
\end{aligned}$$

where $\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k} = (p+1-p)^{n-1}$ comes from the Binomial expansion

$$(a+b)^N = \sum_{k=0}^{N}\binom{N}{k}a^k b^{N-k}$$

with $a = p$, $b = 1-p$, and $N = n-1$.