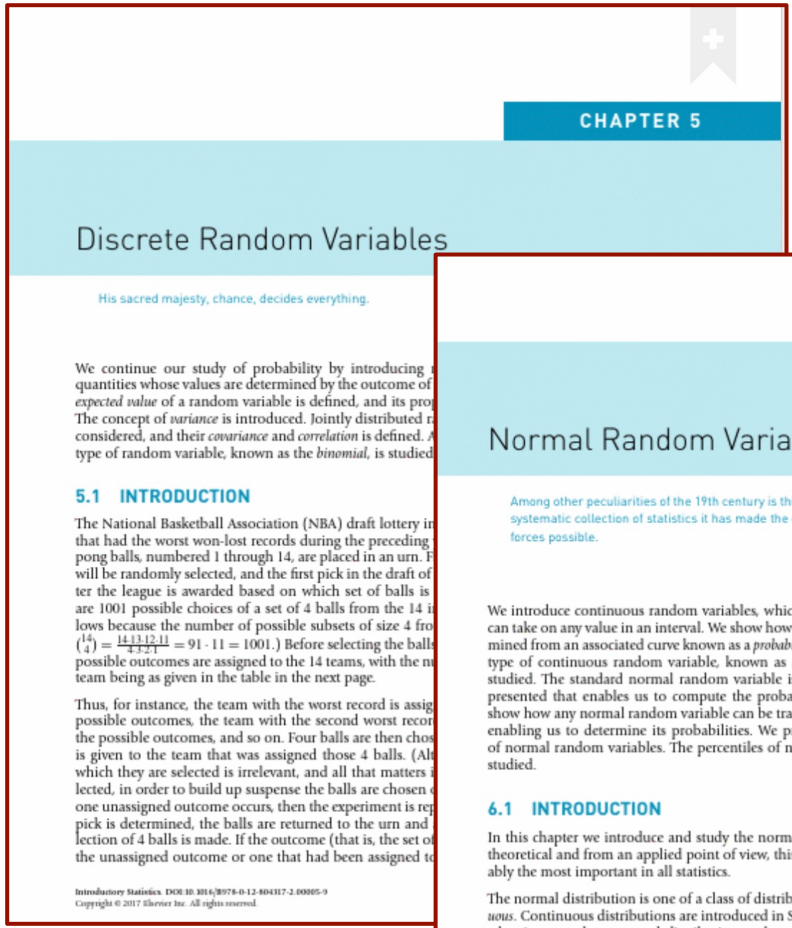
# **Write down** three things you learned from the reading

# (textbook Chapters 5-6)


CHAPTER 5

Discrete Random Variables

His sacred majesty, chance, decides everything.

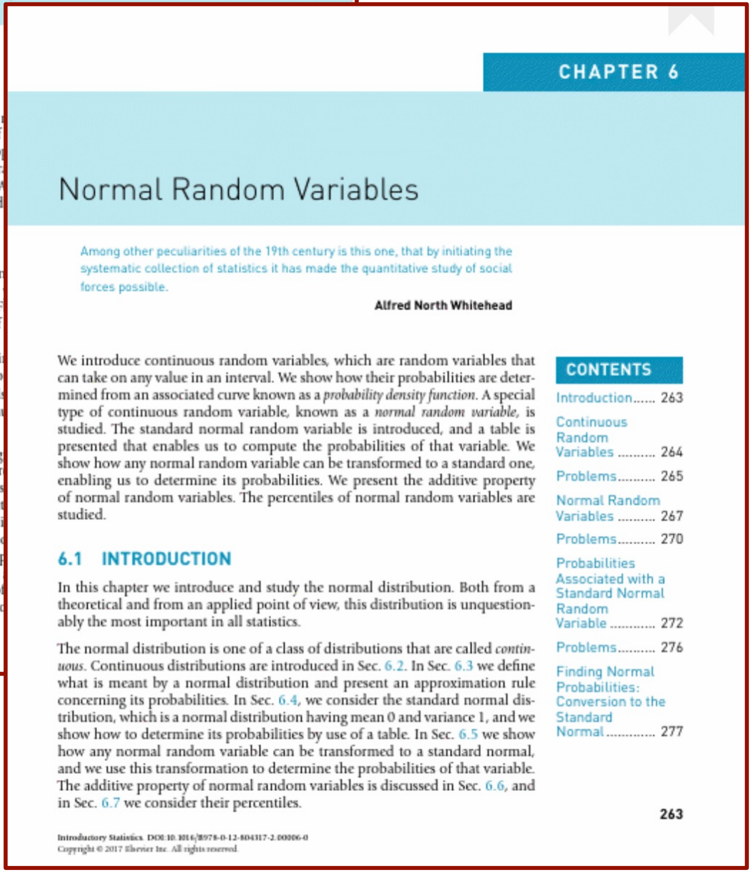
We continue our study of probability by introducing
quantities whose values are determined by the outcome of
*expected value* of a random variable is defined, and its pro
The concept of *variance* is introduced. Jointly distributed r
considered, and their *covariance* and *correlation* is defined.
type of random variable, known as the *binomial*, is studied

5.1 INTRODUCTION

The National Basketball Association (NBA) draft lottery in
that had the worst won-lost records during the preceding
pong balls, numbered 1 through 14, are placed in an urn. F
will be randomly selected, and the first pick in the draft of
ter the league is awarded based on which set of balls is
are 1001 possible choices of a set of 4 balls from the 14 i
lows because the number of possible subsets of size 4 fro
$\binom{14}{4} = \frac{14\cdot13\cdot12\cdot11}{4\cdot3\cdot2\cdot1} = 91 \cdot 11 = 1001$.) Before selecting the balls
possible outcomes are assigned to the 14 teams, with the n
team being as given in the table in the next page.

Thus, for instance, the team with the worst record is assig
possible outcomes, the team with the second worst recor
the possible outcomes, and so on. Four balls are then chos
is given to the team that was assigned those 4 balls. (Alt
which they are selected is irrelevant, and all that matters i
lected, in order to build up suspense the balls are chosen
one unassigned outcome occurs, then the experiment is rep
pick is determined, the balls are returned to the urn and
lection of 4 balls is made. If the outcome (that is, the set of
the unassigned outcome or one that had been assigned t

Introductory Statistics. DOI:10.1016/B978-0-12-804317-2.00005-9



CHAPTER 6

Normal Random Variables

Among other peculiarities of the 19th century is this one, that by initiating the systematic collection of statistics it has made the quantitative study of social forces possible.

**Alfred North Whitehead**

We introduce continuous random variables, which are random variables that can take on any value in an interval. We show how their probabilities are determined from an associated curve known as a *probability density function*. A special type of continuous random variable, known as a *normal random variable*, is studied. The standard normal random variable is introduced, and a table is presented that enables us to compute the probabilities of that variable. We show how any normal random variable can be transformed to a standard one, enabling us to determine its probabilities. We present the additive property of normal random variables. The percentiles of normal random variables are studied.

CONTENTS



6.1 INTRODUCTION

In this chapter we introduce and study the normal distribution. Both from a theoretical and from an applied point of view, this distribution is unquestionably the most important in all statistics.

The normal distribution is one of a class of distributions that are called *continuous*. Continuous distributions are introduced in Sec. 6.2. In Sec. 6.3 we define what is meant by a normal distribution and present an approximation rule concerning its probabilities. In Sec. 6.4, we consider the standard normal distribution, which is a normal distribution having mean 0 and variance 1, and we show how to determine its probabilities by use of a table. In Sec. 6.5 we show how any normal random variable can be transformed to a standard normal, and we use this transformation to determine the probabilities of that variable. The additive property of normal random variables is discussed in Sec. 6.6, and in Sec. 6.7 we consider their percentiles.



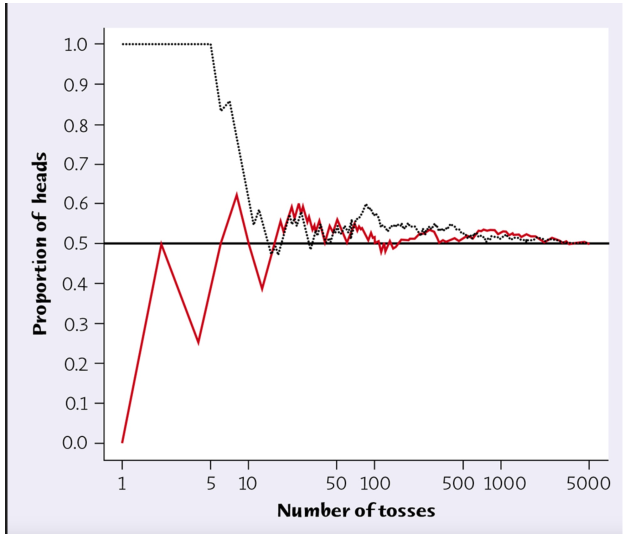
Introductory Statistics. DOI:10.1016/B978-0-12-804317-2.00006-0


*If you couldn't do the reading this week:*

Write three things that you remember from the last lecture

# Previously on 4SSPP109…

- Random processes and their possible outcomes.
- Sample space & events.
- Probability: definition and basic properties
- Conditional probability
- Independence



# 2. Random variables & their distribution


# Random variable (RV)

*A numerical summary of the outcome of a random process.*

**Discrete RV**


- Whether you pass or fail an exam (1 or 0)
- Number of Tube strike closures in this semester

**Continuous RV**

- A person's income
- Minutes played by a player in the 2020-21 Premier League.

# Probability distribution of a discrete RV

- List of all possible outcomes & the probability that each will occur.
- $\Pr(X = x_i)$ → probability that RV $X$ takes on the value $x_i$
- $\sum_{i=1}^{n} \Pr(X = x_i) =$

$$= \Pr(X = x_1) + \Pr(X = x_2) + \Pr(X = x_3) + \cdots = 1$$

  - The probabilities must sum to 1.

# Example: number of Tube strike closures in the Spring semester

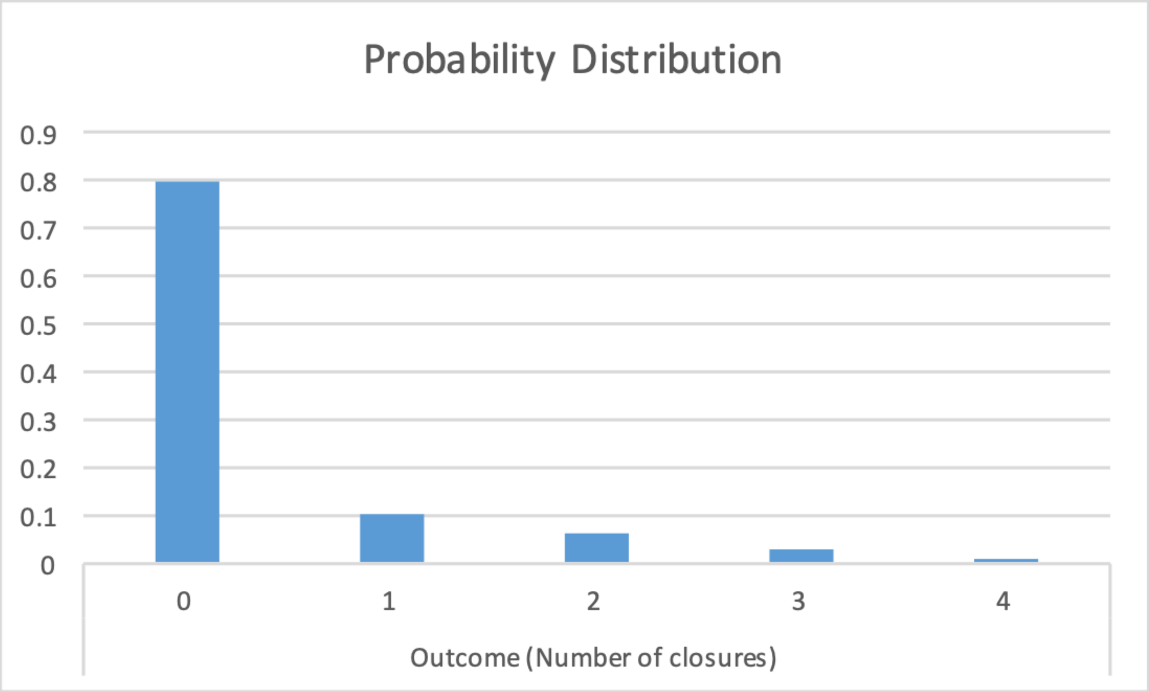
$M$ = number of strikes.

- $\Pr(M = 0)$ → Pr of no strikes.
- $\Pr(M = 1)$ → Pr of 1 strike
- $\Pr(M = 2)$ → Pr of 2 strikes
- ….

# A (hypotetical) probability distribution for Tube strikes

| Outcome (Number of closures) | | | | |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |
| 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

- *How to interpret probabilities here?*
- *What does it mean for the probability of one strike closure to be 10%?*
- *Thought experiment: repeat the same semester many times under the same conditions.*



# A (hypotetical) probability distribution for Tube strikes

| Outcome (Number of closures) | | | | |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |
| 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

**Some possible *events:***

Event A: "not more than one Tube strike closure"

- $\Pr(M = 0 \text{ or } M = 1) = \Pr(M = 0) + \Pr(M = 1) = 0.8 + 0.1 = 0.9$

Event B: "either zero or two closures"

- $\Pr(M = 0 \text{ or } M = 2) = \Pr(M = 0) + \Pr(M = 2) = 0.8 + 0.06 = 0.86$

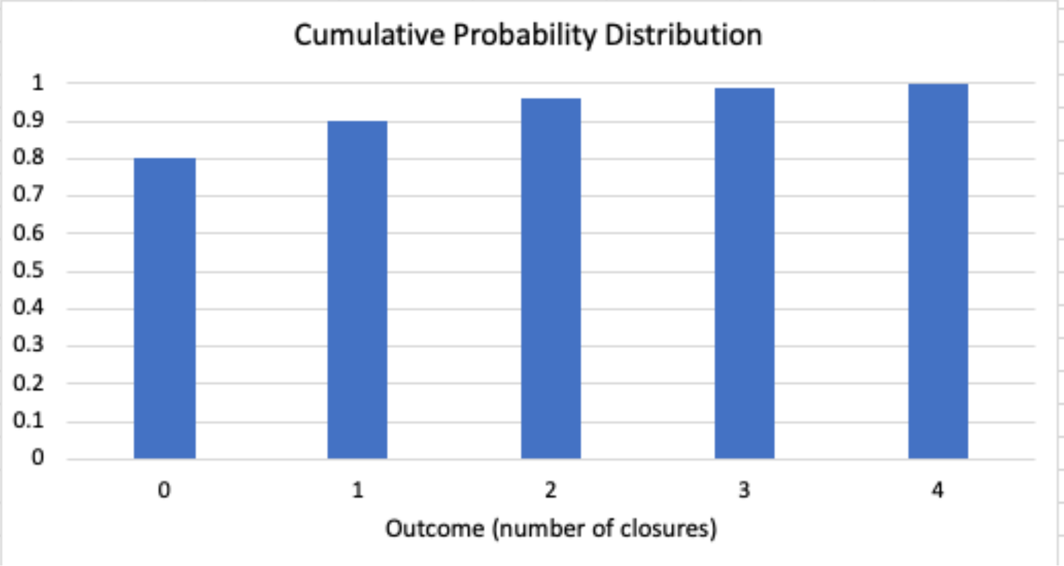
Event C: "not more than two closures"

- $\Pr(M = 0 \text{ or } M=1 \text{ or } M=2) = \Pr(M = 0) + \Pr(M = 1) + \Pr(M = 2)=0.8+0.1+0.06=0.96$

# Cumulative Probability Distribution

- Probability that the RV is *less than or equal to* each possible value.
- In our “Tube strike closures” example:

| | Outcome (Number of closures) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| **Probability Distribution** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |
| **Cumulative probability distribution** | 0.8 | 0.9 | 0.96 | 0.99 | 1 |



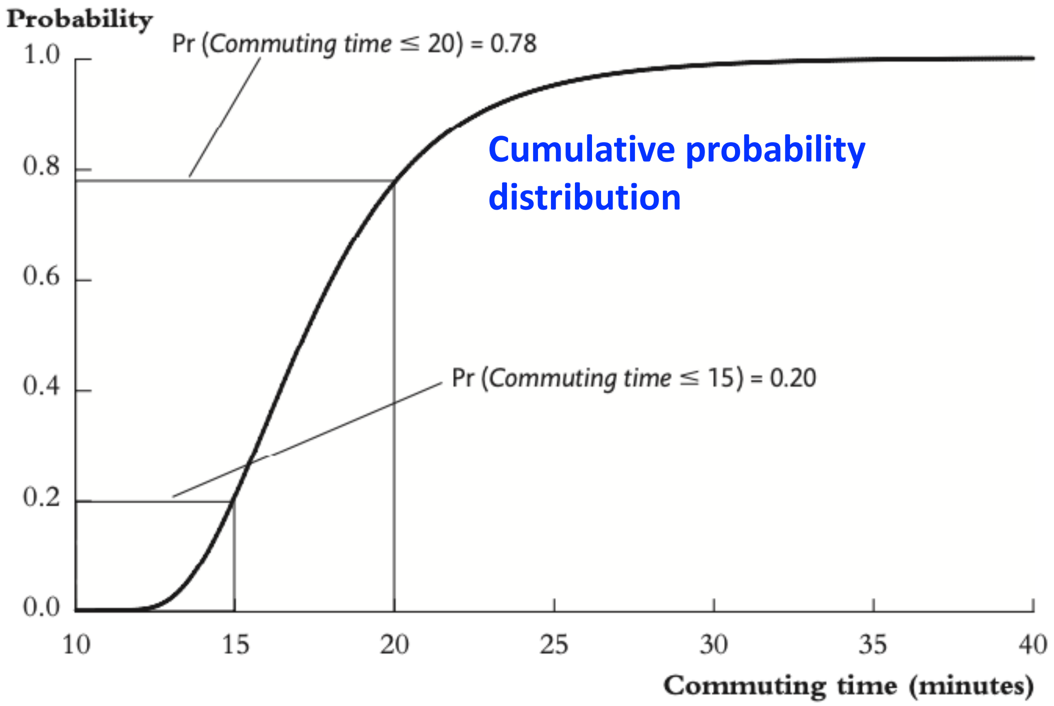
# Binary (or binomial) random variables

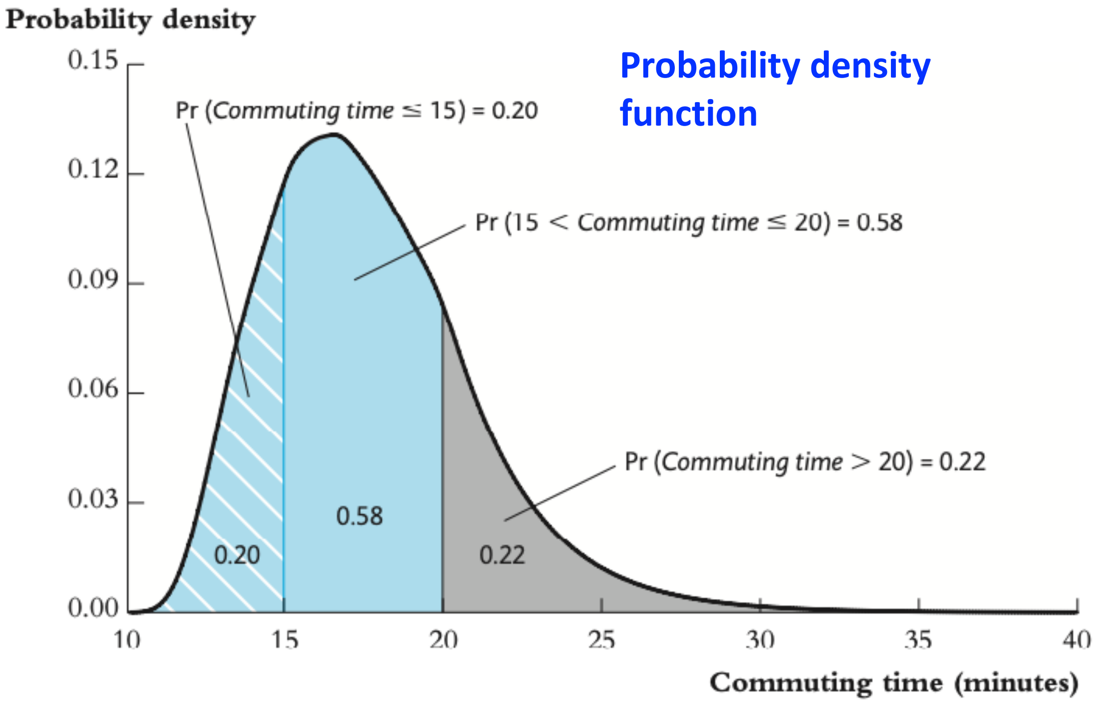
- Possible outcomes: 0 or 1.
  - Result of a coin toss (1=Heads; 0=Tails).
  - Pass/Fail a Quant Methods Exam (1=Pass; 0=Fail)
  - UK citizenship status (1= citizen; 0=non-citizen)


- Its distribution is called 'Bernoulli distribution'.
  - $\Pr(G = 1) \;=\; p$
  - $\Pr(G = 0) \;=\; 1 - p$

# Probability distribution of a continuous RV

- Cumulative probability distribution: $Pr(X \leq x_i)$
- Probability density function (p.d.f.): $Pr(a \leq X \leq b)$
- Example: time it gets someone to commute to work



# Expected value (or *expectation* or *mean*)

- Denoted as $\boldsymbol{E(Y)}$ or $\boldsymbol{\mu_Y}$.
- The long-run average value of Y over many repeated occurrences.
- Weighted average of all possible outcomes, with weights given by probabilities: $E(Y) = y_1p_1 + y_2p_2 + \cdots + \ y_kp_k = \sum_{i=1}^{k} y_ip_i$

Your turn: Compute the expected value of the n. of strike closures

| | Outcome (Number of closures) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| **Probability Distribution** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

# Expected value

- In our Tube strike closures example:

| | Outcome (Number of closures) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| **Probability Distribution** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

$$\boldsymbol{E(Y)} = (0\ x\ 0.80) + (1\ x\ 0.1) + (2\ x\ 0.06) + (3\ x\ 0.03) + (4\ x\ 0.01) = \mathbf{0.35}$$

# Expected value

- If Y is a discrete RV with *k* possible outcomes, and $p_i = P\{Y = y_i\}$:

$$E(Y) = y_1 p_1 + y_2 p_2 + \cdots + y_k p_k = \sum_{i=1}^{k} y_i p_i$$

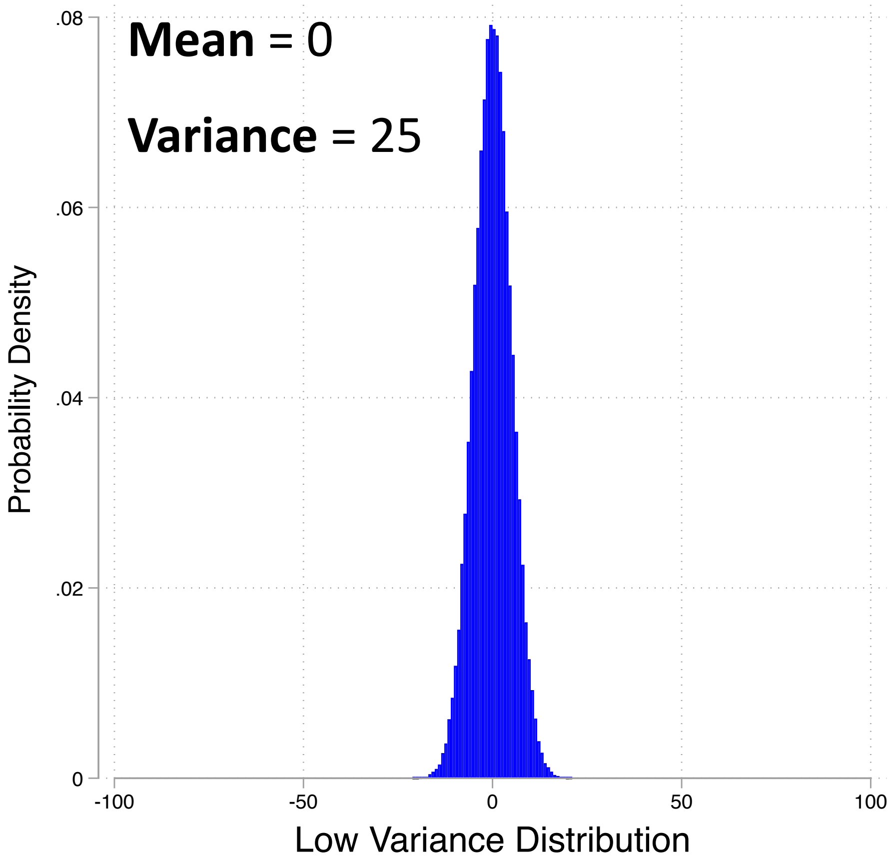
- If G is a binary (or binomial) RV:

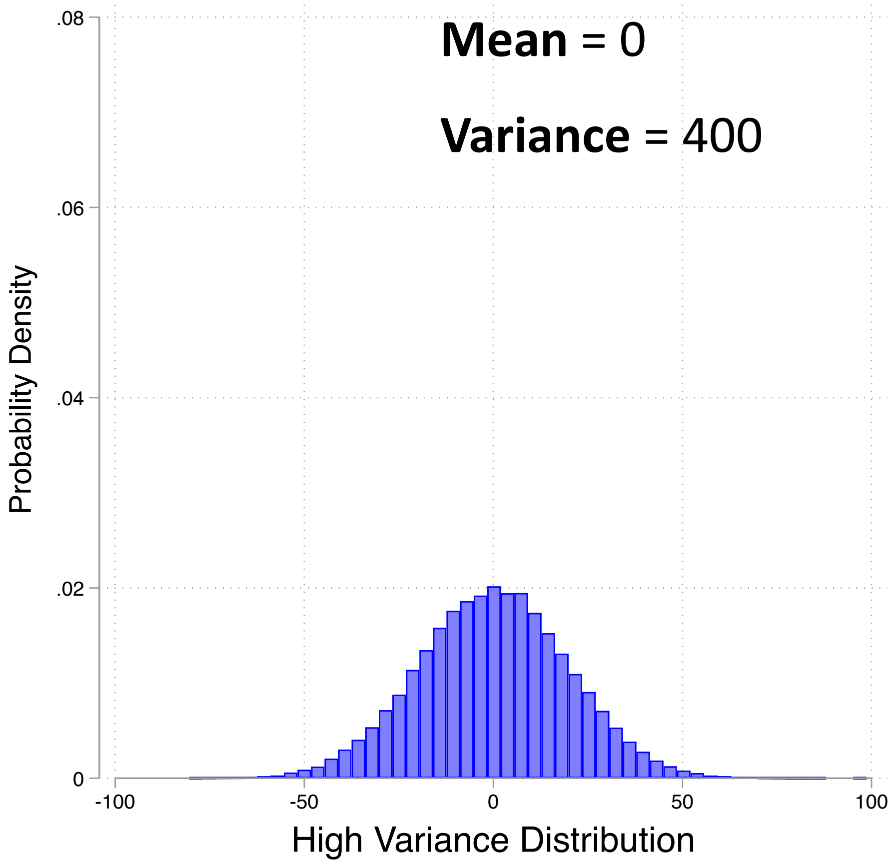
$$E(G) = [0\ x(1-p)] + [1\ x\ p] = p$$

- If X is a continuous RV:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

# Standard deviation and variance



# STANDARD DEVIATION AND VARIANCE

- Measure the dispersion ("spread") of a probability distribution.
- Variance = Mean *squared deviation* of Y from E(Y)
- *Var(Y)* (or $\sigma_Y^2$ ) $= E(Y-\mu_Y)^2 \; \underbrace{= \sum_{i=1}^{k} (y_i - \mu_Y)^2 p_i}$
- Also, *Var(Y)* = $E[Y^2] - \mu_Y^2$
- Standard deviation = square root of the variance
- $\sigma_Y \; [or \; SD(Y)] = \sqrt{Var(Y)}$

(Assuming Y is a discrete RV with *k* possible outcomes)

# STANDARD DEVIATION AND VARIANCE

- Your turn: calculate the variance and standard deviation of the number of campus closures from our example distribution

| | Outcome (Number of closures) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| **Probability Distribution** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

- Remember:
  - Var(y) = $\sum_{i=1}^{k}(y_i - \mu_Y)^2 p_i$
  - $\sigma_Y = \sqrt{Var(Y)}$

- From our previous calculation: $\mu_Y$=0.35

# STANDARD DEVIATION AND VARIANCE

| | Outcome (Number of closures) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| **Probability Distribution** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 |

$$[(0-0.35)^2 * 0.8] + [(1-0.35)^2 * 0.1] + [(2-0.35)^2 * 0.06] + [(3-0.35)^2 * 0.03] + [(4-0.35)^2 * 0.01] = 0.6475$$

- $\mathrm{Var(y)} = \sum_{i=1}^{k} (\boldsymbol{y_i} - \boldsymbol{\mu_Y})^2 \boldsymbol{p_i} = 0.6475$
- $\boldsymbol{\sigma_Y} = \sqrt{\boldsymbol{Var(Y)}} = 0.80$

# VARIANCE OF A BINARY (OR BINOMIAL) RV

- Applying the variance formula to a binary RV:

$$Var(G) = \sigma_G^2 = p(1-p)$$

$$\sigma_G = \sqrt{p(1-p)}$$

# MEAN AND VARIANCE OF LINEAR FUNCTIONS

- A linear function of X: $Y = 2000 + 0.8\,X$
- In general: $Y = a + b\,X$
- Mean: $E(Y) = E(a + b\,X) = a + b\,E(X)$
- Variance: $Var(Y) = Var(a + b\,X) = b^2 Var(X)$