Lecture 6 - Properties of Random Variables

Sta102 / BME102

Colin Rundel

January 28, 2015

Mean and variance of a discrete RVs

Last time we were introduced to some definitions for calculating the expected value (mean) and variance of a discrete random variable.

- Expected Value

$$E(X) = \sum_x x \cdot P(X = x)$$

- Variance (and Standard Deviation)

$$\begin{aligned} Var(X) &= E\left((X - E(X))^2\right) \\ &= \sum_x (x - E(X))^2 \cdot P(X = x) \\ SD(X) &= \sqrt{Var(X)} \end{aligned}$$

Bernoulli Random Variables

A Bernoulli random variable describes a trial with only two possible outcomes, one of which we will label a success and the other a failure and where the probability of a success is given by the parameter p. (Since it needs to be numeric) the random variable takes the value 1 to indicate a success and 0 to indicate a failure.

Let $X \sim \text{Bern}(p)$ then

$$P(X = x) = \begin{cases} 1 - p & \text{if } x = 0, \\ p & \text{if } x = 1 \end{cases}$$

$$\begin{aligned} E(X) &= p \\ Var(X) &= p(1 - p) \end{aligned}$$

Geometric Random Variables

A Geometric random variable describes the number of (identical) Bernoulli trials that occur before the first success is observed. The distribution has a single parameter, the probability of a success p. There is another slightly different characterization that counts the number of failures before the first success. We will focus on the former for now.

Let $X \sim \text{Geo}(p)$ then

$$P(X = x) = p(1 - p)^{x-1}$$

$$\begin{aligned} E(X) &= 1/p \\ Var(X) &= \frac{1 - p}{p^2} \end{aligned}$$

St. Petersburg Lottery

We start with \$1 on the table and a coin.

At each step: Toss the coin; if it shows Heads, take the money. If it shows Tails, I double the money on the table.

How much would you pay me to play this game? i.e. what is the expected value?

Binomial Distribution

We define a random variable X that reflects the *number of successes* in a *fixed number* of *independent trials*, each with the *same probability of success* as having a binomial distribution.

By definition there are n trials each with probability p of success.

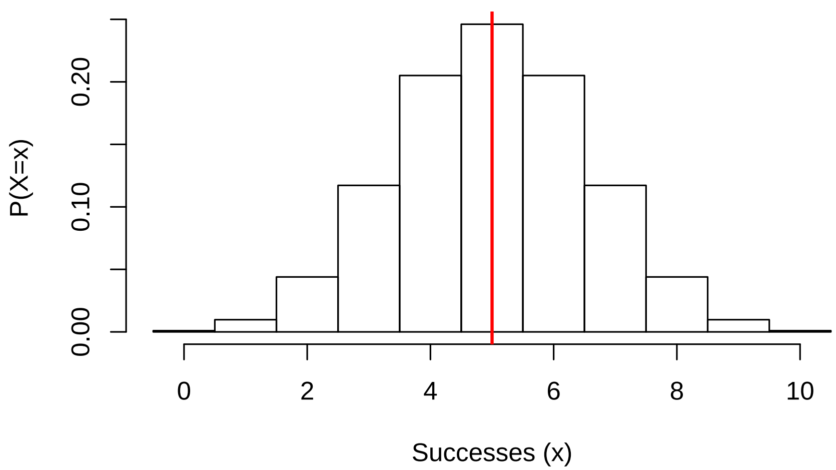
Let $X \sim \text{Binom}(n, p)$ then

$$P(X = x|n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$
$$Var(X) = np(1-p)$$

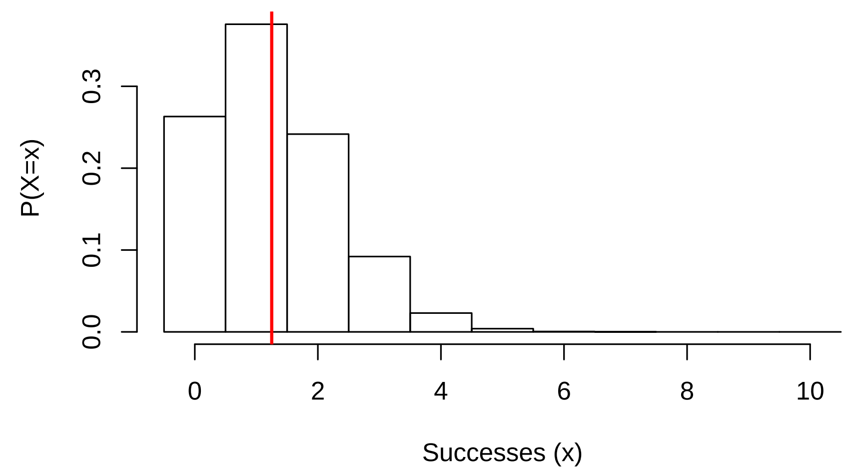
Binomial RVs - Example 1

Let $X \sim \text{Binom}(n = 10, p = 1/2)$,



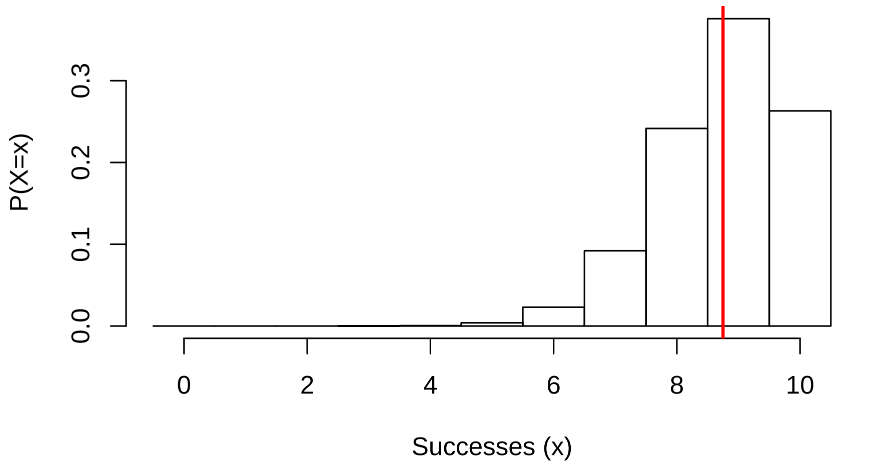
Binomial RVs - Example 2

Let $X \sim \text{Binom}(n = 10, p = 1/8)$,



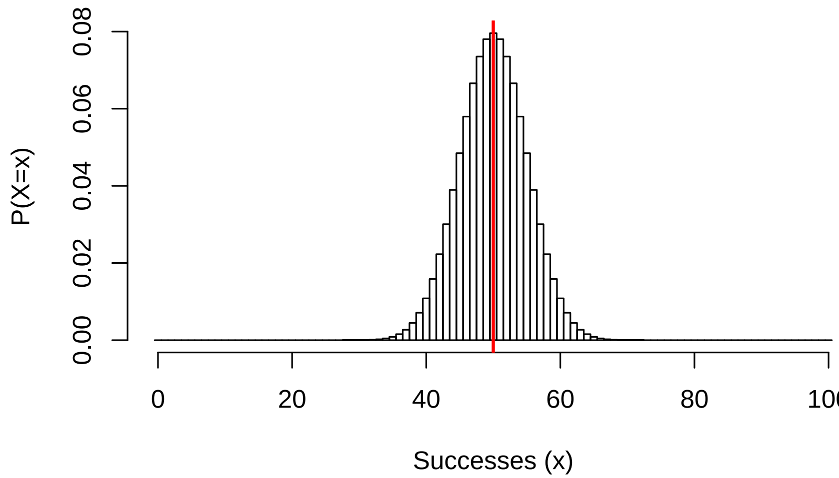
Binomial RVs - Example 3

Let $X \sim \text{Binom}(n = 10, p = 7/8)$,



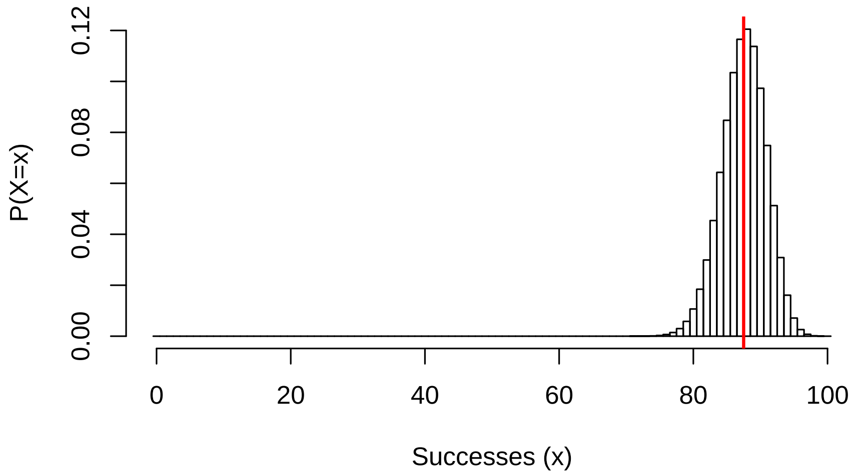
Binomial RVs - Example 4

Let $X \sim \text{Binom}(n = 100, p = 4/8)$,



Binomial RVs - Example 5

Let $X \sim \text{Binom}(n = 100, p = 7/8)$,



Properties of Expected Value

- **Constant** - $E(c) = c$ if c is constant
- **Constant Multiplication** - $E(cX) = cE(X)$
- **Constant Addition** - $E(X + c) = E(X) + c$
- **Addition** - $E(X + Y) = E(X) + E(Y)$
- **Subtraction** - $E(X - Y) = E(X) - E(Y)$
- **Multiplication** - $E(XY) = E(X)E(Y)$ if X and Y are independent.

Constant & Constant Multiplication

Constant

Imagine there is a random variable C that has the value c 100% of the time (e.g. $P(C = c) = 1$)

$$E(C) = \sum_c x\, P(C = x) = c\, P(C = c) = c$$

Constant Multiplication

$$E(cX) = \sum_x cx\, P(X = x) = c \sum_x x\, P(X = x) = cE(X)$$

Constant Addition

Assume X is a discrete random variable and c is some constant value then,

$$\begin{aligned}
E(X + c) &= \sum_x (x + c)\, P(X = x) \\
&= \sum_x \Big(x\, P(X = x) + cP(X = x)\Big) \\
&= \Big(\sum_x x\, P(X = x)\Big) + \Big(\sum_x c\, P(X = x)\Big) \\
&= \Big(\sum_x x\, P(X = x)\Big) + c\Big(\sum_x P(X = x)\Big) \\
&= E(X) + c
\end{aligned}$$

Addition*

Assume X and Y are independent discrete random variables then,

$$\begin{aligned}
E(X + Y) &= \sum_x \sum_y (x + y)\, P(X = x \cap Y = y) \\
&= \sum_x \sum_y (x + y)\, P(X = x)P(Y = y) \\
&= \sum_x \sum_y \Big(x\, P(X = x)P(Y = y) + y\, P(X = x)P(Y = y)\Big) \\
&= \Big(\sum_x \sum_y x\, P(X = x)P(Y = y)\Big) + \Big(\sum_x \sum_y y\, P(X = x)P(Y = y)\Big) \\
&= \Big(\sum_x x\, P(X = x) \sum_y P(Y = y)\Big) + \Big(\sum_y y\, P(Y = y) \sum_x P(X = x)\Big) \\
&= \Big(\sum_x x\, P(X = x)\Big) + \Big(\sum_y y\, P(Y = y)\Big) \\
&= E(X) + E(Y)
\end{aligned}$$

Properties of Variance

- **Constant** - $Var(c) = 0$ if c is constant
- **Constant Multiplication** - $Var(cX) = c^2\ Var(x)$
- **Constant Addition** - $Var(X + c) = Var(X)$
- **Addition** - $Var(X + Y) = Var(X) + Var(Y)$ if X and Y are independent.
- **Subtraction** - $Var(X - Y) = Var(X) + Var(Y)$ if X and Y are independent.

Constant & Constant Multiplication

Constant

Imagine there is a random variable C that has the value c 100% of the time (e.g. $P(C = c) = 1$)

$$
\begin{aligned}
Var(C) &= \sum_{x=c} (x - E(C))^2 \; P(C = x) \\
&= (c - E(C))^2 \; P(C = c) \\
&= (c - c) = 0
\end{aligned}
$$

Constant Multiplication

$$
\begin{aligned}
Var(cX) &= \sum_{x} (cx - E(cX))^2 \; P(X = x) \\
&= \sum_{x} (cx - cE(X))^2 \; P(X = x) \\
&= c^2 \sum_{x} (x - E(X))^2 \; P(X = x) = c^2 Var(x)
\end{aligned}
$$

Constant Addition

Assume X is a discrete random variable and c is some constant value then,

$$
\begin{aligned}
Var(X + c) &= \sum_{x} (x + c - E(X + c))^2 \; P(X = x) \\
&= \sum_{x} (x + c - E(X) - c)^2 \; P(X = x) \\
&= \sum_{x} (x - E(X))^2 \; P(X = x) \\
&= Var(X)
\end{aligned}
$$

Example - Coffee

The average price of a small cup of coffee to go is \$1.40, with a standard deviation of 30¢. An 8.5% tax is added if you take your coffee to stay. Assume that each time you get a coffee to stay you also tip 50¢. What is the mean, variance, and standard deviation of the amount you spend on coffee when to take it to stay?

Let X represent the amount you spend on coffee to go (in ¢), and Y represent the amount you spend on coffee to stay (in ¢). Then,

$$
\begin{aligned}
Y &= X + 0.085X + 50 \\
&= 1.085X + 50
\end{aligned}
$$

Example - Coffee, cont.

We now know that $E(X) = 140$, $SD(X) = 30$, and $Y = 1.085X + 50$,

$$
\begin{aligned}
E(X) &= 140 \\
E(Y) &= E(1.085X + 50) \\
&= 1.085 \; E(X) + 50 \\
&= 1.085 \times 140 + 50 \\
&= 201.9
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= SD(X)^2 = 900 \\
Var(Y) &= Var(1.085X + 50) \\
&= 1.085^2 \; Var(X) \\
&= 1.085^2 \times 900 \\
&= 1059.503
\end{aligned}
$$

$$SD(Y) = \sqrt{1059.503} = 32.55$$

On average you spend \$2.02 per day on coffee, with a standard deviation of \$0.33

Example - Coffee and a Muffin

The average price of a cup of coffee is \$1.40, with a standard deviation of 30¢. The average price of a muffin is \$2.50, with a standard deviation of 15¢. If you get a cup of coffee and a muffin every day for breakfast, what is the mean, variance, and standard deviation of the amount you spend on breakfast daily? Assume that the price of coffee and muffins are independent.

Let X represent the amount you spend on coffee (in ¢), and Y represent the amount you spend on muffins (in ¢).

$$\begin{aligned} E(X) &= 140 \\ E(Y) &= 250 \\ E(X+Y) &= E(X)+E(Y) \\ &= 140+250 \\ &= 390 \end{aligned}$$

$$\begin{aligned} Var(X) &= 30^2 = 900 \\ Var(Y) &= 15^2 = 225 \\ Var(X+Y) &= Var(X)+V(Y) \\ &= 900+225 \\ &= 1125 \end{aligned}$$

Simplifying RVs

Random variables do not work like normal algebraic variables:

$$X_1 + X_2 \neq 2X$$

If we know that X_1 and X_2 have the same distribution then,

$$\begin{aligned} E(X_1+X_2) &= E(X_1)+E(X_2) \\ &= 2E(X_1) \end{aligned}$$

$$\begin{aligned} Var(X_1+X_2) &= Var(X_1)+Var(X_2) \\ &= 2\ Var(X_1) \end{aligned}$$

$$\begin{aligned} E(X_1+X_1) &= E(2X_1) \\ &= 2E(X_1) \end{aligned}$$

$$\begin{aligned} Var(X_1+X_1) &= Var(2X_1) \qquad = 2^2\ Var \\ &= 4\ Var(X_1) \end{aligned}$$

Town Cars

A company has 5 Lincoln Town Cars in its fleet. Historical data show that annual fuel cost for each car is on average \$2,154 with a standard deviation of \$132. What is the mean and the standard deviation of the total annual fuel cost for this fleet?

$$\begin{aligned} E(X_1+X_2+X_3+X_4+X_5) &= E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5) \\ &= 5 \times E(X) = 5 \times 2,154 = \$10,770 \end{aligned}$$

$$\begin{aligned} Var(X_1+X_2+X_3+X_4+X_5) &= Var(X_1)+Var(X_2)+Var(X_3)+Var(X_4)+Var(X_5) \\ &= 5 \times V(X) = 5 \times 132^2 = \$^2 87,120 \end{aligned}$$

$$SD(X_1+X_2+X_3+X_4+X_5) = \sqrt{87,120} = \$295.16$$

Iced Tea

A pitcher is filled with exactly 64 onces of iced tea, you proceed to fill 5 glasses with 8 onces each. If the expected amount you added to each glass is 8 onces with a standard deviation of 0.25 onces, how much iced tea is left in the pitcher (and what is your uncertainty)?

Let X_i be the amount of iced tea added to the ith glass and P be the amount left in the pitcher.

$$P = 64 - X_1 - X_2 - X_3 - X_4 - X_5$$

$$\begin{aligned} E(P) &= E(64 - X_1 - X_2 - X_3 - X_4 - X_5) \\ &= E(64) - E(X_1) - E(X_2) - E(X_3) - E(X_4) - E(X_5) \\ &= 64 - 8 - 8 - 8 - 8 - 8 = 24 \end{aligned}$$

$$\begin{aligned} Var(P) &= Var(64 - X_1 - X_2 - X_3 - X_4 - X_5) \\ &= Var(64) + Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4) + Var(X_5) \\ &= 0 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 2.5 \end{aligned}$$

Properties of Binomial RVs (again)

We can also think of a Binomial random variable as the sum of independent Bernoulli random variables.

Let $X \sim \text{Binom}(n, p)$ then $X = \sum_{i=1}^n Y_i$ where $Y_1, \cdots, Y_n \sim \text{Bern}(p)$.

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n E(Y_i) \\ &= \sum_{i=1}^n p = np \end{aligned}$$

$$\begin{aligned} Var(X) &= Var\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n Var(Y_i) \\ &= \sum_{i=1}^n p(1-p) = np(1-p) \end{aligned}$$