# Discrete Distributions, Section 2.1-2.3

Grethe Hystad

September 16, 2012

# Random Variables

Numerical outcomes such as

- the number of students who received A in a course
- the number of heads observed in flipping a coin 50 times.
- number of accidents at a particular street

whose values can change from experiment to experiment are called random variables.

**Definition**

A random variable is a real valued function whose domain is the sample space:

$$X : S \to \mathbb{R}.$$

We use capital letters near the end of the alphabet, e.g. X,Y,Z for random variables.

# Random Variables

## Example

- Suppose a fair coin is tossed 4 times.

# Random Variables

## Example

- Suppose a fair coin is tossed 4 times.
- There are $2^4 = 16$ possible outcomes.

# Random Variables

## Example

- Suppose a fair coin is tossed 4 times.
- There are $2^4 = 16$ possible outcomes.
- Let X be the number of heads observed. X is a random variable and can take the values 0,1,2,3, and 4.

# Random Variables

## Example

- Suppose a fair coin is tossed 4 times.
- There are $2^4 = 16$ possible outcomes.
- Let X be the number of heads observed. X is a random variable and can take the values 0,1,2,3, and 4.
- Since each outcomes is equally likely, the probability that 0 head is observed is $P(X=0) = \frac{\binom{4}{0}}{16} = \frac{1}{16}$.
  The probability that exactly 1 head is observed is $P(X=1) = \frac{\binom{4}{1}}{16} = \frac{1}{4}$.
  The probability that exactly 2 heads are observed is $P(X=2) = \frac{\binom{4}{2}}{16} = \frac{3}{8}$.

# Random Variables

## Example

The probability that exactly 3 heads are observed is

$P(X = 3) = \frac{\binom{4}{3}}{16} = \frac{1}{4}$.

$P(X = 4) = \frac{\binom{4}{4}}{16} = \frac{1}{16}$.

# Random Variables

**Example**

The probability that exactly 3 heads are observed is
$P(X = 3) = \frac{\binom{4}{3}}{16} = \frac{1}{4}$.
$P(X = 4) = \frac{\binom{4}{4}}{16} = \frac{1}{16}$.

- The values that random variables can assume are denoted by lower case letters, such as x,y,z.

## Random Variables

**Example**

The probability that exactly 3 heads are observed is

$P(X = 3) = \frac{\binom{4}{3}}{16} = \frac{1}{4}$.

$P(X = 4) = \frac{\binom{4}{4}}{16} = \frac{1}{16}$.

- The values that random variables can assume are denoted by lower case letters, such as x,y,z.
- We write $P(X = x) = f(x)$ for the probability that X takes on the value x.

# Random Variables

**Example**

The probability that exactly 3 heads are observed is
$P(X = 3) = \frac{\binom{4}{3}}{16} = \frac{1}{4}$.
$P(X = 4) = \frac{\binom{4}{4}}{16} = \frac{1}{16}$.

- The values that random variables can assume are denoted by lower case letters, such as x,y,z.
- We write $P(X = x) = f(x)$ for the probability that X takes on the value x.
- X has the probability distribution:

| $x$ | 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| $P(X = x)$ | $\frac{1}{16}$ | $\frac{1}{4}$ | $\frac{3}{8}$ | $\frac{1}{4}$ | $\frac{1}{16}$ | 1 |

# Random Variables

- The probability that 1 or 2 heads are observed is $P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$

# Random Variables

- The probability that 1 or 2 heads are observed is $P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$
- The probability that at most 2 heads are observed is $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} = \frac{11}{16}$

# Random Variables

- The probability that 1 or 2 heads are observed is $P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$
- The probability that at most 2 heads are observed is $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} = \frac{11}{16}$
- The probability that more than 2 heads are observed is $P(X > 2) = P(X = 3) + P(X = 4) = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}$.

# Random Variables

- The probability that 1 or 2 heads are observed is $P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$
- The probability that at most 2 heads are observed is $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} = \frac{11}{16}$
- The probability that more than 2 heads are observed is $P(X > 2) = P(X = 3) + P(X = 4) = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}$.
- The probability that more than 2 heads are observed can also be found as $P(X > 2) = 1 - P(X \leq 2) = 1 - \frac{11}{16} = \frac{5}{16}$.

# Random Variables

- The probability that 1 or 2 heads are observed is $P(X = 1 \cup X = 2) = P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$
- The probability that at most 2 heads are observed is $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{16} + \frac{1}{4} + \frac{3}{8} = \frac{11}{16}$
- The probability that more than 2 heads are observed is $P(X > 2) = P(X = 3) + P(X = 4) = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}$.
- The probability that more than 2 heads are observed can also be found as $P(X > 2) = 1 - P(X \leq 2) = 1 - \frac{11}{16} = \frac{5}{16}$.
- The probability that more than 1 head but at most 3 heads are observed is $P(1 < X \leq 3) = P(X = 2) + P(X = 3) = \frac{5}{8}$.

# What does this notation mean?

Let A be some subset of the range of a discrete random variable. For example $A = \{\text{is greater than } 2\}$, $A = \{\text{between } 0 \text{ and } 2\}$. We write

$$\{e \in S | X(e) \in A\} \tag{1}$$

for the set of outcomes e in S such that, $X(e)$, the value of the random variable is in the subset A. We typically write (1) as $(X \in A)$. For example,
$\{X \text{ is greater than } 2\} = (X > 2) = \{e \in S | X(e) > 2\}$
$\{X \text{ is between } 0 \text{ and } 2\} = (0 < X < 2) = \{e \in S | 0 < X(e) < 2\}$.

# Discrete Random Variables

**Definition**

A random variable $X : S \to \mathbb{R}$ is said to be discrete if its range,

$$\{X(e) | e \in S\},$$

is either finite or infinite countable, i.e., X can take on only a finite number- or a countable infinite number- of possible values x.

## Probability Function

### Definition

Let $X(S) \subset \mathbb{R}$ be a random variable. The *Probability function* of X, $f : X(S) \to [0, 1]$ is defined as

$$f(x) = P(\{e \in S | X(e) = x\})$$

for any $x \in X(S)$. Thus, $f(x)$ is the probability that X equals x, and we write

$$f(x) = P(X = x).$$

The probability function, $f(x)$, assigns probability to each value x of X so that the following conditions are satisfied:

- $P(X = x) = f(x) \geq 0$ for all $x \in X(S)$
- $\sum_{x \in X(S)} P(X = x) = 1$, where the sum is over all possible values of x.

## Example

- Lets go back to the previous example: Toss a fair coin four times. Let X be the number of heads observed.
- X has the following probability distribution:

| $x$ | 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| $f(x)$ | $\frac{1}{16}$ | $\frac{1}{4}$ | $\frac{3}{8}$ | $\frac{1}{4}$ | $\frac{1}{16}$ | 1 |

- The mean, $\mu$, of X which is the weighted average of the numbers $0, 1, 2, 3, 4$ in which weights are $f(0)$, $f(1)$, $f(2)$, $f(3)$, $f(4)$ respectively is:

$$\mu = \sum_{x=0}^{4} xf(x) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{16} = 2$$

## Definition

The mean of a random variable X with probability mass function $f(x)$ is given by:

$$\mu = \sum_x x f(x).$$

## Example

- The variance, $\sigma^2$, of X is the weighted average of $(x-\mu)^2$:

$$\begin{aligned}\sigma^2 &= \sum_{x=0}^{4} (x-\mu)^2 f(x) \\ &= (0-2)^2 \cdot \frac{1}{16} + (1-2)^2 \cdot \frac{1}{4} + (2-2)^2 \cdot \frac{3}{8} \\ &+ (3-2)^2 \cdot \frac{1}{4} + (4-4)^2 \cdot \frac{1}{16} = 1\end{aligned}$$

# Variance

- If X is a finite discrete random variable with mean value $\mu$ and X takes on the value x, then $x - \mu$ is called the deviation of x from $\mu$.
- For example a deviation of $+1$ means x is 1 unit to the right of $\mu$ and $-1$ means 1 unit to the left of $\mu$. In order to not worry about the sign difference, we work with the square of the deviation, $(x - \mu)^2$, instead.

## Definition

The variance of a random variable X with mean value $\mu$ and probability mass function $f(x)$ is given by

$$\sigma^2 = \sum_x (x - \mu)^2 f(x).$$

# Standard deviation

The units associated with $\sigma^2$ are the square of the units of measurement for X. The smallest value for $\sigma^2$ is zero, which happens when X takes on a constant value with probability 1. As the points with positive probability spread out, the variance becomes larger.

**Definition**

The standard deviation, $\sigma$, of a random variable $X$ is the square root of the variance.

The standard deviation maintains the original units of measure.

# Variance

The next theorem gives a convenient formula for the variance of X.

## Theorem

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

## Definition

- $\sum_x x^r f(x)$ is called the $r^{th}$ moment about the origin.
- $\mu$ is the first moment about the origin.
- $\sum_x x^2 f(x)$ is the second moment about the origin.

## Sample

**Example**

- Simulate flipping a fair coin 4 times and count the number of heads, $X$. Perform this experiment $n = 100$ times. We obtained the following frequencies:

In R:

$> x = c(0, 1, 2, 3, 4)$

$> f = c(1/16, 1/4, 3/8, 1/4, 1/16)$

$> data = sample(x, 100, replace = TRUE, prob = f)$

# Sample

## Example

- Simulate flipping a fair coin 4 times and count the number of heads, $X$. Perform this experiment $n = 100$ times. We obtained the following frequencies:
-

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $n(x)$: frequency | 7 | 23 | 32 | 30 | 8 |
| $h(x)$: relative frequency | $\frac{7}{100}$ | $\frac{23}{100}$ | $\frac{32}{100}$ | $\frac{30}{100}$ | $\frac{8}{100}$ |

In R:

$> x = c(0, 1, 2, 3, 4)$
$> f = c(1/16, 1/4, 3/8, 1/4, 1/16)$
$> data = sample(x, 100, replace = TRUE, prob = f)$

# Sample

## Example

The mean for this data (sample mean, average) is:

$$\begin{aligned}\bar{x} &= \frac{0 \cdot 7 + 1 \cdot 23 + 2 \cdot 32 + 3 \cdot 30 + 4 \cdot 8}{100} \\ &= 0 \cdot \frac{7}{100} + 1 \cdot \frac{23}{100} + 2 \cdot \frac{32}{100} + 3 \cdot \frac{30}{100} + 4 \cdot \frac{8}{100} \\ &= \sum_{x=0}^{4} x h(x) = 2.09.\end{aligned}$$

# Sample mean

## Definition

The sample mean of the observations, $x_1$, $x_2$,...,$x_n$, is:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Notice that we are placing equal weight $\frac{1}{n}$ of each observed value. We have

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{x \in X(S)} x n(x) = \sum_{x \in X(S)} x\frac{n(x)}{n} = \sum_{x \in X(S)} x h(x),$$

where $h(x) = \frac{n(x)}{n}$ and $n(x)$ is the frequency of $x$.

# Sample variance for the previous example

## Example

The sample variance, $s^2$, averages $(x - \bar{x})^2$:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\sum_{x \in X(S)}(x - \bar{x})^2 h(x)$$

$$= \frac{100}{99}\sum_{x=0}^{4}(x - 2.09)^2 h(x) = \frac{100}{99}\Big[(0 - 2.09)^2 \cdot \frac{7}{100} + (1 - 2.09)^2 \cdot \frac{23}{100}$$

$$+ (2 - 2.09)^2 \cdot \frac{32}{100} + (3 - 2.09)^2 \cdot \frac{30}{100} + (4 - 2.09)^2 \cdot \frac{8}{100}\Big]$$

$$= 1.13.$$

Notice that we divided by $n - 1$ instead of $n$. More about that later.

# Sample variance

## Definition

The sample variance of the observations, $x_1$, $x_2$,...,$x_n$, is:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

# R calculation

```
> x=c(rep(0,7), rep(1,23), rep(2,32),rep(3,30), rep(4,8))
> x
  [1] 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [38] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3
 [75] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4
> length(x)
[1] 100
> mean(x)
[1] 2.09
> var(x)
[1] 1.133232
> sd(x)
[1] 1.064534
```

# Sample versus population

- In statistics, a population is the entire group of subjects that we are interested in information about. A population gives all values for a random variable taken in the long run with mean, $\mu$, and variance, $\sigma^2$.
- A sample is an observed value for a variable and is a part of the population from which we collect information. The sample is used to get information about the entire population.
- The sample mean, $\bar{x}$, is an estimate of the population mean, $\mu$.
- The sample variance, $s^2$, is an estimate of the population variance, $\sigma^2$.

# Expectations, section 2.2

## Example

- Suppose you and your friend are matching balanced coins. Each of you flip a fair coin. If the upper faces match, you win \$1; if they do not match, you lose \$1. The probability of winning a match is $\frac{1}{2}$, so in the long run you should expect to win about half of the time.

# Expectations, section 2.2

## Example

- Suppose you and your friend are matching balanced coins. Each of you flip a fair coin. If the upper faces match, you win \$1; if they do not match, you lose \$1. The probability of winning a match is $\frac{1}{2}$, so in the long run you should expect to win about half of the time.
- On average, you will expect to win the following amount per game in the long run, $(-1)\frac{1}{2} + (1)\frac{1}{2} = 0$ which is the expected value of your winnings. This is a fair game.

# Expected Value

## Definition

The expected value of a discrete random variable with probability mass function $f(x)$, is given by

$$E(X) = \sum_x xf(x).$$

(The sum is over all $x$ for which $f(x) > 0$). We usually denote $E(X) = \mu$.

Note that in order for the expected value of a discrete random variable X to exist, we must have

$$\sum_x |x|f(x) < \infty \quad \text{(converges absolutely).}$$

## Example

- Now suppose you instead win \$5 per game of matching coins. Your expected winnings per game is $(-5)\frac{1}{2} + 5\frac{1}{2} = 0$.
- Now if X represents your winnings per game when you are playing for \$1, then $5X$ represents your winnings per game when you play for \$5.

For a random variable, we can form functions of X, say $u(X)$.

## Theorem

*If $X$ is a discrete random variable with probability distribution $f(x)$ and if $u(x)$ is any real-valued function of $X$, then*

$$E[u(X)] = \sum_x u(x)f(x).$$

## Example

- Now suppose you win \$1 if the match is tails and \$2 if the match is heads. You loose \$1.50 if the coins do not match.
- Then your expected winnings per game is $(-1.5)\frac{1}{2} + (1)\frac{1}{4} + (2)\frac{1}{4} = 0$.

- In the first game, it is more likely for the random variable to take on values closer to the mean than in the second experiment. For example, in the first experiment, the random variable take on one of the values $-1$ or 1 with probability 1. Thus, it deviates by 1 unit from the mean.
- In the second experiment, the positive values average 1.5 units from the expected value of 0 as do the negative values.

# Variance and Standard deviation

## Definition

The variance of a random variable X with expected value $\mu$ is given by

$$V(X) = E[(X - \mu)^2].$$

We usually denote $V(x) = \sigma^2$.

## Definition

The standard deviation of a random variable X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[(x - \mu)^2]}.$$

The standard deviation maintains the original units of measure.

# Variance and Standard Deviation

## Example

For the first coin experiment with $\mu = 0$, we have

- $\sigma^2 = E[(x-\mu)^2] = (-1)^2\frac{1}{2} + (1)^2\frac{1}{2} = 1$
- $\sigma = 1.$

For the second game,

- $\sigma^2 = (-1.5)^2\frac{1}{2} + (1)^2\frac{1}{4} + 2^2\frac{1}{4} = 2.375$
- $\sigma = \sqrt{2.375} = 1.54.$

# Expectations, example

**Example**

Let X have p.m.f $f(x) = \frac{x}{5}$, $x = 2, 3$. Let $Y = X^3$. Find $E(Y)$.

**Solution**

$E(Y) = E(X^3) = \sum_{x=2}^{3} x^3 \frac{x}{5} = \frac{2^4}{5} + \frac{3^4}{5} = \frac{97}{5}$

# Expectation; linear operator

- If k is a constant, then $E(k) = k$.

# Expectation; linear operator

- If k is a constant, then $E(k) = k$.
- If k is a constant and v is a function, then $E[kv(x)] = kE[v(x)]$.

# Expectation; linear operator

- If k is a constant, then $E(k) = k$.
- If k is a constant and v is a function, then $E[kv(x)] = kE[v(x)]$.
- If $k_1, k_2$ are constants, $v_1$ and $v_2$ are functions, then $E[k_1 v_1(X) + k_2 v_2(X)] = k_1 E[v_1(X)] + k_2 E[v_2(X)]$.

# Expectation; linear operator

- If k is a constant, then $E(k) = k$.
- If k is a constant and v is a function, then $E[kv(x)] = kE[v(x)]$.
- If $k_1, k_2$ are constants, $v_1$ and $v_2$ are functions, then $E[k_1 v_1(X) + k_2 v_2(X)] = k_1 E[v_1(X)] + k_2 E[v_2(X)]$.
- In general: For constants $k_1, ..., k_m$, functions $v_1, ...., v_m$, we have $E[\sum_{i=1}^{m} k_i v_i(X)] = \sum_{i=1}^{m} k_i E[v_i(X)]$.

# Expectation; linear operator

- If k is a constant, then $E(k) = k$.
- If k is a constant and v is a function, then $E[kv(x)] = kE[v(x)]$.
- If $k_1, k_2$ are constants, $v_1$ and $v_2$ are functions, then $E[k_1 v_1(X) + k_2 v_2(X)] = k_1 E[v_1(X)] + k_2 E[v_2(X)]$.
- In general: For constants $k_1, ..., k_m$, functions $v_1, ...., v_m$, we have $E[\sum_{i=1}^{m} k_i v_i(X)] = \sum_{i=1}^{m} k_i E[v_i(X)]$.
- Thus $E$ is a linear operator.

## Theorem

*For any random variable $X$ and constants $a$ and $b$,*

*1.* $E(aX + b) = aE(X) + b$

*2.* $V(aX + b) = a^2 V(X)$

**Theorem**

*For any random variable X and constants a and b,*

1. $E(aX + b) = aE(X) + b$
2. $V(aX + b) = a^2 V(X)$

**Theorem**

*If X is a random variable with mean* $\mu$*, then*

$$V(X) = E(X^2) - \mu^2.$$

# Summary

Let X and Y be random variables defined on a sample space and let a be a constant. Then

- $E(X+Y) = E(X) + E(Y)$
- $E(aX) = aE(X)$
- $E(a) = a$
- $E(aX+b) = aE(X) + b$
- $V(X+b) = V(X)$
- $V(aX) = a^2 V(X)$
- $V(aX+b) = a^2 V(X)$
- If X and Y are independent, then $V(X+Y) = V(X) + V(Y)$.

## Example

A company is calculating the cost of equipment maintenance for its budget. In order to decide the budget, the number of machine malfunctions per month is considered. Based on past performance, the distribution of X is given as follows:

| x | $f(x)$ |
|---|---|
| 0 | 0.1 |
| 1 | 0.3 |
| 2 | 0.4 |
| 3 | 0.2 |

Determine the expected value, variance and standard deviation of the number of malfunctions per month.

## Solution

*The expected number of malfunctions per month is*

$$E(x) = \sum_x xf(x) = 0(0.1) + 1(0.3) + 2(0.4) + 3(0.2) = 1.7$$

*The variance is*

$$\begin{aligned} V(x) &= \sum_x (x - 1.7)^2 f(x) \\ &= (0 - 1.7)^2(0.1) + (1 - 1.7)^2(0.3) + \\ &+ (2 - 1.7)^2(0.4) + (3 - 1.7)^2(0.2) \\ &= 0.81 \ \textit{ times squared per month.} \end{aligned}$$

*The standard deviation is*

$$\sigma = \sqrt{0.81} = 0.9 \ \textit{ times per month.}$$

# R-commands

$> x = c(0:3)$
$> f = c(0.1, 0.3, 0.4, 0.2)$
$> E = sum(x * f)$ # Expectations
$> V = sum((x - E)^2 * p)$ # Variance
$> s = sqrt(V)$ # Standard deviation

## Example

The company in the previous example wants to consider the cost of maintenance. It costs \$100 to pay for a representative to check out malfunctions, and the parts cost an average \$80 to correct each malfunction. Find the expected value, variance and mean of the monthly cost of visits by the representative.

## Solution

*The monthly cost is*

$$80X + 100,$$

*where* $X$ *is the number of malfunctions per month. Then the expected cost per month is*

$$E(80X + 100) = 80E(X) + 100 = 80(1.7) + 100 = 236 \ \ dollar.$$

# Solution continue

## Solution

*The variance is*

$$V(80X+100) = 80^2 V(X) = 80^2(0.81) = 5184 \ \ \textit{dollar}^2 \ \ \textit{per month.}$$

*The standard deviation is*

$$\sigma = \sqrt{5184} = 72 \ \ \textit{dollar per month.}$$

### Example

In a lottery, a player selects 3 numbers from 1 to 20. During the draw, 3 numbers are drawn at random and without replacement. To win, all 3 numbers must match those drawn in any order. It cost $5 to by a ticket in the lottery. If the player wins, the pay off is $1000 minus the $5 paid for the ticket. Let X be the pay off to the bettor. Find $E(X)$.

## Solution

- *The number of possible winning numbers is the combination of* 3 *objects selected from* 20*, that is,* $\binom{20}{3} = 1140$*. Thus the probability of winning is* $\frac{1}{1140}$*.*
- *We have* $f(x) = \begin{cases} \frac{1}{1140} & \text{if } x = 995; \\ 1 - \frac{1}{1140} & \text{if } x = -5. \end{cases}$
- $E(X) = \frac{1}{1140} \cdot 995 + (1 - \frac{1}{1140}) \cdot (-5) \approx \$ - 4.1.$

## Example

Let $u(x) = (x - b)^2$. Find the value of $b$ for which $E[(x - b)^2]$ is a minimum.

## Example

Let $u(x) = (x - b)^2$. Find the value of $b$ for which $E[(x - b)^2]$ is a minimum.

## Solution

$g(b) = E[(x - b)^2] = E[X^2 - 2bX + b^2] = E(X^2) - 2bE(X) + b^2$.
*We have* $g'(b) = -2E(X) + 2b = 0$. *Hence* $b = E(X)$.
*Since* $g''(b) = 2 > 0$, $b = E(X)$ *gives the minimum value of* $E[(x - b)^2]$.

## Example

Let $u(x) = (x-b)^2$. Find the value of $b$ for which $E[(x-b)^2]$ is a minimum.

## Solution

$g(b) = E[(x-b)^2] = E[X^2 - 2bX + b^2] = E(X^2) - 2bE(X) + b^2$.
*We have* $g'(b) = -2E(X) + 2b = 0$. *Hence* $b = E(X)$.
*Since* $g''(b) = 2 > 0$, $b = E(X)$ *gives the minimum value of* $E[(x-b)^2]$.

- Homework problem: Find the value of $c$ for which $E[|x-c|]$ is a minimum.

# "Standardized" form of X

Suppose $E(X) = \mu$ and $V(X) = \sigma^2$ for a random variable X.
Define the "standardized" form of X, as

$$Y = \frac{X - \mu}{\sigma}.$$

Then $E(Y) = 0$ and $V(Y) = 1$.
To prove this is left as homework.

# 2.2 and 2.3; Special Discrete Distributions; The Cumulative Probability Distribution (CDF)

## Definition

The *(cumulative) distribution function*, $F(x)$, for a random variable X evaluated at x is defined by

$$F(x) = P(X \leq x).$$

If X is discrete,

$$F(x) = \sum_{k=-\infty}^{x} f(k),$$

where $f(x)$ is the probability mass function.

# Cumulative Probability Distribution (CDF)

- $P(X > x) = 1 - P(X \leq x)$

## Cumulative Probability Distribution (CDF)

- $P(X > x) = 1 - P(X \leq x)$
- If $a < b$, then $(X \leq b) \setminus (X \leq a) = (a < X \leq b)$

# Cumulative Probability Distribution (CDF)

- $P(X > x) = 1 - P(X \leq x)$
- If $a < b$, then $(X \leq b) \setminus (X \leq a) = (a < X \leq b)$
- $$\begin{aligned} P(a < X \leq b) &= P((X \leq b) \setminus (X \leq a)) \\ &= P(X \leq b) - P(X \leq a) = F(b) - F(a). \end{aligned}$$

# Properties of the Distribution Function

A function, F, is a Cumulative Distribution Function iff

- $\lim_{x \to -\infty} F(x) = 0$

# Properties of the Distribution Function

A function, F, is a Cumulative Distribution Function iff

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$

# Properties of the Distribution Function

A function, F, is a Cumulative Distribution Function iff

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- F is nondecreasing; that is if $a < b$, $F(a) \leq F(b)$.

# Properties of the Distribution Function

A function, F, is a Cumulative Distribution Function iff

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- F is nondecreasing; that is if $a < b$, $F(a) \leq F(b)$.
- F is right-hand continuous; that is

$$\lim_{x \to x_0^+} F(x) = F(x_0).$$

## Example

Five lightbulb, 2 of which are defect are being mixed. Two of the lightbulb are randomly selected. Let X denote the number of defective lightbulb picked.

(A) Determine the probability mass function for X.

(B) Determine the Cumulative Distribution function for X.

## Solution

- *The probability of selecting 0 defect lightbulbs is*
  $f(0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$

## Example

Five lightbulb, 2 of which are defect are being mixed. Two of the lightbulb are randomly selected. Let X denote the number of defective lightbulb picked.
(A) Determine the probability mass function for X.
(B) Determine the Cumulative Distribution function for X.

## Solution

- *The probability of selecting 0 defect lightbulbs is*
  $f(0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$
- *The probability of selecting 1 defect lightbulb is*
  $f(1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{3}{5}$

## Example

Five lightbulb, 2 of which are defect are being mixed. Two of the lightbulb are randomly selected. Let X denote the number of defective lightbulb picked.
(A) Determine the probability mass function for X.
(B) Determine the Cumulative Distribution function for X.

## Solution

- *The probability of selecting 0 defect lightbulbs is* $f(0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$
- *The probability of selecting 1 defect lightbulb is* $f(1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{3}{5}$
- *The probability of selecting 2 defect lightbulb is* $f(2) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1}{10}$.

## Example

Five lightbulb, 2 of which are defect are being mixed. Two of the lightbulb are randomly selected. Let X denote the number of defective lightbulb picked.
(A) Determine the probability mass function for X.
(B) Determine the Cumulative Distribution function for X.

## Solution

- *The probability of selecting 0 defect lightbulbs is* $f(0) = \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} = \frac{3}{10}$
- *The probability of selecting 1 defect lightbulb is* $f(1) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{3}{5}$
- *The probability of selecting 2 defect lightbulb is* $f(2) = \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}} = \frac{1}{10}$.

# Solution continue

The probability mass function is given by,

| x | $f(x)$ |
|---|---|
| 0 | $\frac{3}{10}$ |
| 1 | $\frac{3}{5}$ |
| 2 | $\frac{1}{10}$ |

(B) The cumulative distribution function is

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.30, & 0 \leq x < 1 \\ 0.90, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

# Bernoulli Distribution

- Experiments with two possible outcomes.
- A selected lightbulb is either defective or no defective.
- You either win or loose.
- A cat is either pregnant or not pregnant.

Such experiments are called Bernoulli trials.

# The Bernoulli Distribution

- Suppose one outcome of a Bernoulli trial is indentified to be a success and the other a failure. Define the random variable X as,
  - $X = 1$ if the outcome is a success.
  - $X = 0$ if the outcome is a failure.

## The Bernoulli Distribution

- Suppose one outcome of a Bernoulli trial is indentified to be a success and the other a failure. Define the random variable X as,
  - $X = 1$ if the outcome is a success.
  - $X = 0$ if the outcome is a failure.
- Let p denote the probability of observing a success. Then $(1 - p)$ is the probability of observing a failure.

# The Bernoulli Distribution

- Suppose one outcome of a Bernoulli trial is indentified to be a success and the other a failure. Define the random variable X as,
  - $X = 1$ if the outcome is a success.
  - $X = 0$ if the outcome is a failure.
- Let p denote the probability of observing a success. Then $(1 - p)$ is the probability of observing a failure.
- X is said to have a Bernoulli distribution.

# The Bernoulli Distribution

- Suppose one outcome of a Bernoulli trial is indentified to be a success and the other a failure. Define the random variable X as,
  - $X = 1$ if the outcome is a success.
  - $X = 0$ if the outcome is a failure.
- Let p denote the probability of observing a success. Then $(1 - p)$ is the probability of observing a failure.
- X is said to have a Bernoulli distribution.
- The Bernoulli random variable is a bulding block for other probability distributions such as the binomial distribution.

# The Bernoulli Distribution

**Theorem**

*X has a Bernoulli distribution if its p.m.f. is given by*

$$f(x) = p^x(1-p)^{1-x}, \ \ x = 0, 1 \ \ \textit{for} \ \ 0 < p < 1.$$

*We have*

$$E(X) = p \quad \textit{and} \quad Var(X) = p(1-p).$$

$E(X) = \sum_x f(x) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1-p) + 1(p) = p.$

$$\begin{aligned} Var(X) = E(X^2) - [E(X)]^2 &= \sum_x x^2 f(x) - p^2 \\ &= 0^2(1-p) + 1^2(p) - p^2 \\ &= p - p^2 = p(1-p). \end{aligned}$$

# The Binomial Distribution

- Suppose we perform n independent Bernoulli trials.
- Each trials have probability p of success.
- Let the random variable X be the number of successes in the n trials.

The distribution of X is called the Binomial distribution.

# The Binomial Distribution

- Suppose we insepct n items independently and record values for $X_1$, $X_2$,..., $X_n$, where

$$X_i = \begin{cases} 1, & \text{if the } i^{th} \text{ item is defective} \\ 0, & \text{if the } i^{th} \text{ item is not defective} \end{cases}$$

- Then

$$X = \sum_{i=1}^{n} X_i$$

  denotes the number of defective items among the n items.
- We assume that $P(X_i = 1) = p$ for each $i$.

# The Binomial Distribution

Suppose $X_1$, $X_2$, ..., $X_n$ are n independent Bernoulli random variables with

$$X = \sum_{i=1}^{n} X_i.$$

Then

$$E(X) = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np.$$

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

# The Binomial Distribution

**Theorem**

*X has a Binomial Distribution if its p.m.f. is given by*

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \;\; x = 0, 1, 2, ..., n \;\; \text{for} \;\; 0 \leq p \leq 1.$$

*We have*

$$E(X) = np \;\; \text{and} \;\; Var(X) = np(1-p).$$

*We say X is b(n,p).*

Here the probability that $X = x$ is given by the term $p^x(1-p)^{n-x}$ multiplied by the number of ways of selecting x positions for defectives in the n possible positions. This number is $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

# The Binomial Distribution

Recall the binomial expansion:

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}.$$

Observe, using the binomial expansion with $a = p$, $b = (1-p)$ that

$$\sum_x f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1.$$

# Summary, Binomial Distribution

A random variable X has a binomial distribution if:

- The experiment consists of n identical trials.
- Each trial has exactly two outcomes; a success or a failure, that is each trial is a Bernoulli trial.
- The probability of success, p, is constant from trial to trial.
- The trials are independent.
- X is the number of successes among the n trials.

# Examples.

- Number of defectives in a sample of n items.
- Number of heads in a sequence of n coin tosses.
- Number of people with type $O^+$ blood of the n people that enter a blood bank.
- Number of children of a couple who has a genetic disease.

# R-code

Let n be the number of trials.
Let $X = x$ be the number of successes among the n trials.
Let p be the probability for success.
Then in R,
$> f = dbinom(x, n, p)$ # probability of x successes among the n trials.
$> F = pbinom(x, n, p)$ # the probability of at most x successes among the n trials which is F(x)).

- The probability of 8 or fewer successes is

$$P(X \leq 8) = F(8) = \sum_{x=0}^{8} p(x).$$

- The probability of 8 or fewer successes is

$$P(X \leq 8) = F(8) = \sum_{x=0}^{8} p(x).$$

- The probability of more than 8 successes is

$$P(X > 8) = 1 - F(8) = \sum_{x=9}^{n} p(x)$$

- The probability of 8 or fewer successes is

$$P(X \leq 8) = F(8) = \sum_{x=0}^{8} p(x).$$

- The probability of more than 8 successes is

$$P(X > 8) = 1 - F(8) = \sum_{x=9}^{n} p(x)$$

- The probability of at least 8 successes is

$$P(X \geq 8) = 1 - F(7) = \sum_{x=8}^{n} p(x).$$

- The probability of more than 3 successes but at most 8 successes is

$$P(3 < X \leq 8) = F(8) - F(3) = \sum_{x=4}^{8} p(x).$$

- The probability of more than 3 successes but at most 8 successes is

$$P(3 < X \leq 8) = F(8) - F(3) = \sum_{x=4}^{8} p(x).$$

- The probability of more than 3 successes but fewer than 8 successes is

$$P(3 < X < 8) = F(7) - F(3) = \sum_{x=4}^{7} p(x).$$

## Example

Suppose a student take a final multiple choice exam with 20 questions. The student has not studied during the whole semester, so he plans to guess on each question. Each question has 5 choices.
(A) Determine the probability that the student guesses correctly on 7 of the questions.
(B) Determine the probability that the student guesses correctly on at least one question.

## Example

Suppose a student take a final multiple choice exam with 20 questions. The student has not studied during the whole semester, so he plans to guess on each question. Each question has 5 choices.
(A) Determine the probability that the student guesses correctly on 7 of the questions.
(B) Determine the probability that the student guesses correctly on at least one question.

## Solution

*There are* $n = 20$ *identical trials.*
*Each trial has two outcomes, right and wrong.*
*The probability of success is* $p = \frac{1}{5}$ *on each trial.*
*The guesses are independent.*
*Let* $X$ *be the number of questions the student guesses correctly.*

# Solution

(A) The probability that the student guesses correctly on 7 questions is

$$f(7) = \binom{20}{7}(0.20)^7(0.80)^{13} = 0.0545.$$

In R:
$> f = -dbinom(7, 20, 0.2)$
[1]0.0545
(B) The probability that the student guesses correctly on at least one question is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{20}{0}(0.20)^0(0.80)^{20} = 0.988.$$

# Binomial distributions

**Example**

Let X be $b(1, p)$ and let $Y$ be $b(2, p)$. If $P(X = 0) = \frac{2}{3}$, find $P(Y = 0)$.

# Binomial distributions

**Example**

Let X be $b(1, p)$ and let $Y$ be $b(2, p)$. If $P(X = 0) = \frac{2}{3}$, find $P(Y = 0)$.

**Solution**

- $f(x) = \binom{1}{x} p^x (1-p)^{1-x}$
- $g(y) = \binom{2}{y} p^y (1-p)^{2-y}$
- $P(X = 0) = p^0(1-p)^1 = 1 - p = \frac{2}{3}$, *so* $p = \frac{1}{3}$.
- *Hence* $P(Y = 0) = \left(\frac{1}{3}\right)^0 (1 - \frac{1}{3})^2 = \frac{4}{9}$.

# Unbiased estimator

## Definition

We say that $X$ is an unbiased estimator of a parameter $\theta$ if $E(X) = \theta$.

## Example

- Let $X$ be $b(n, p)$ and let $Y = \frac{X}{n}$.
- Then $E(Y) = \frac{1}{n}E(X) = p$
- Thus $Y = \frac{X}{n}$ is an unbiased estimator of $p$.
- We have $Var(Y) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$.

## Problem

*In any given year, a farmer tries to decide whether to spray his crop to control for insects. The probability that spraying is necessary is* $0.7$ *If spraying is necessary, the farmer sprays* $98\%$ *of the time. If spraying is not completely necessary, the farmer sprays* $35\%$ *of the time.*
*(A) What is the probability that the farmer sprays at least* $1$ *year in a* $4$ *year period?*
*(B) Each year the farmer hires a consultant to determine whether he needs spraying. To hire the consultants cost* $\$2000$. *If spraying is necessary it cost* $\$10,000$.
*Find the expected cost and variance of spraying in a* $4$ *year period.*

Solution given in class.