Expected Value and Variance: Chapter 7

Chapter Outcomes:

Summarizing data on random variables

Expectation of a random variable

Some applications of expectation

Means and variances of special discrete probability distributions

Motivation:

Oftentimes, it is more helpful (and quicker) to provide a person with “summary statistics” rather than giving full details of every outcome in an experiment (i.e. random phenomenon).

This way, one can immediately get a sense for the data just by knowing, for example, the *mean* and/or *median* (which we will discuss more fully in this chapter).

A useful technique to present data is to express it in terms of the *frequency distribution*.

Example:

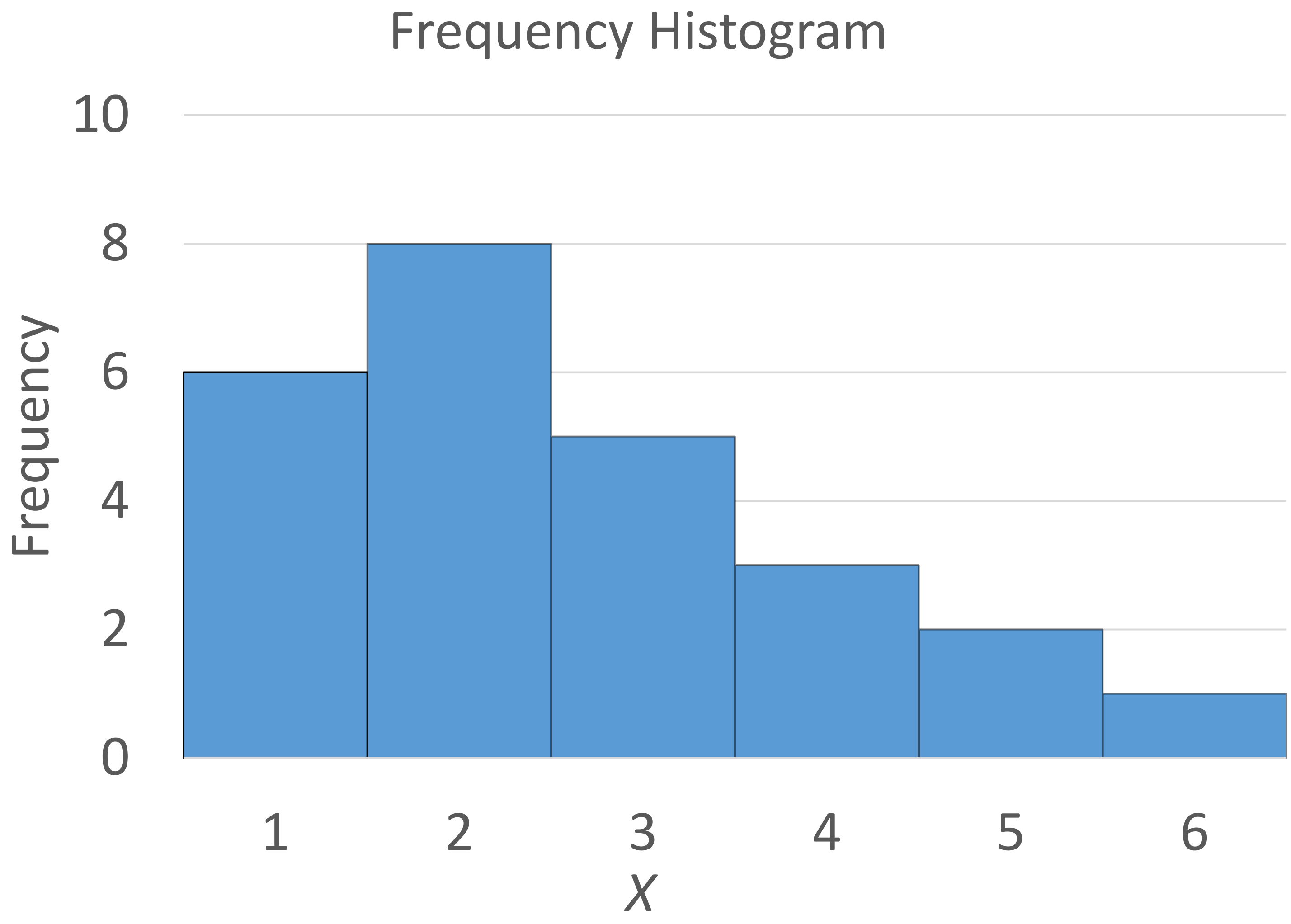
Suppose we are observing the number of people in cars crossing a toll bridge. Let X represent the number of people in a given car.

Clearly, X is a **discrete** random variable.

If we observed a total of 25 cars, we could then present our data as follows:

X	**Frequency**	**Relative Frequency**	**Percent Relative Frequency**
1	6	6/25=0.24	24
2	8	8/25=0.32	32
3	5	5/25=0.20	20
4	3	3/25=0.12	12
5	2	2/25=0.08	8
6	1	1/25=0.04	4
Total	25	1	100

We could also represent the above data using a *Frequency Histogram*.



Frequency Histogram
10
8
6
4
2
0
Frequency
1
2
3
4
5
6
X

Of course, sometimes we might prefer a *single-number* summary.

The most common single-number summary is the **sample mean (average)** or **arithmetic mean** of the outcomes.

If we have n outcomes $x_1, x_2, \ldots, x_n$ for a r.v. *X*, then the arithmetic mean is given by:

$$\boldsymbol{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

In our earlier example, we had:

X	**1**	**2**	**3**	**4**	**5**	**6**
Frequency	6	8	5	3	2	1

We will calculate the sample mean, $\bar{x}$.

Note: This is an **estimate** of the true mean, μ.

Remarks:

A set of observed outcomes $x_1, x_2, \ldots, x_n$ for a r.v. X is called a **sample**.

The average (or mean) we calculate is the average of **that particular sample**, termed the *sample mean,* and is denoted by $\bar{x}$.

If we **repeatedly did this experiment**, we should expect that each sample would produce a **different sample mean**. This is a very important concept, and will be revisited in STAT 231!

Note: $\bar{x}$ does not have to be an integer, and likely won't be, even if the r.v. X only takes on integer values. Remember that $\bar{x}$ is an **average of the sample of observations**.

Two other common summary statistics are:

1. Median

The value such that, when the results are placed in order, from lowest to highest, half the results lie below it and the other half lie above it.

Example:

Determine the median of the following sample data:

14, 15, 13, 21, 15, 15, 26, 16, 20, 13

First of all, we need to **order** the n =10 outcomes, from lowest to highest, as follows:

13, 13, 14, 15, 15, 15, 16, 20, 21, 26

Rules: There are two cases to consider:

1. If n is **odd**, then the median is given by the $\left(\frac{n+1}{2}\right)^{th}$ ordered observation.
2. If n is **even**, then the median is the average of the 2 middle values.

In this particular example, n = 10 is **even**, so the sample median is given by **the average of the 5th and 6th ordered observations**, leading to a sample median of: (15+15)/2 = 15.

2. Mode

The value which occurs **most often**.

Note: It is possible we could observe more than one mode, when two or more values occur just as often. It's also possible to have no mode.

Example:

Determine the mode given the following data:

14, 15, 13, 21, 15, 15, 26, 16, 20, 13

We see that the value 15 occurs 3 times → **sample mode is 15**

Note: Although the mean, median, and mode are referred to as **measures of central tendency**, they do not always result in the same value.

You Try:

Calculate the median and mode of the 25 sample of data points from the toll bridge example.

Solution:

Mode: We see that 2 is the most common value. This is the sample mode.

Median: We have an odd number of observations, so the median is given by the $\frac{(25+1)}{2}th$ ordered observation, or the 13^{th} ordered observation. (There were 6 1's and 8 2's).

So, the 13^{th} ordered observations is 2 and thus, the sample median is 2.

Expectation of a Random Variable (Section 7.2):

In our earlier example where X represented the number of people in cars crossing a certain toll bridge, we looked at finding summary statistics (like the mean) for a sample of **observed** x-values.

The same ideas can be used to summarize the probability distribution of a r.v. X.

Recall that in our toll bridge example,

$$\bar{x} = \frac{(6\times1)+(8\times2)+(5\times3)+(3\times4)+(2\times5)+(1\times6)}{25}$$

Now, suppose we knew that the probability function of X was actually given by:

x	1	2	3	4	5	6
$f(x) = P(X = x)$	0.30	0.25	0.20	0.15	0.09	0.01

This means that if we were to observe **a very large number of cars (over a long period of time)**, the proportion of time we would observe X=1 would be 0.30. For X=2 the proportion would equal to 0.25. For X=3, the proportion would be 0.20, etc.

Hence, in the **long-run**, if we use the p.f.

x	1	2	3	4	5	6
$f(x)$	0.30	0.25	0.20	0.15	0.09	0.01

we would observe an **expected value** of

$$(0.30) \times 1 + (0.25) \times 2 + (0.20) \times 3 + (0.15) \times 4 + (0.09) \times 5 + (0.01) \times 6$$

$$= \mathbf{2.51}$$

Note: Here is a case where we have a discrete r.v. and the expected value is not an integer. Remember that the expected value represents a **long-run average**.

Therefore, the mean we just calculated is essentially the "theoretical mean" and is generally denoted by μ or $E(X)$.

As we just saw, this mean requires that we know the probability distribution of the r.v. X.

Definition: The **expected value** (also called the **mean** or **expectation**) of a discrete r.v. X with p.f. $f(x)$ is

$$\mu = E(X) = \sum_{x \in A} x f(x).$$

You Try:

You are in charge of the maintenance of a particular machine and notice that its breakdowns are due to operator misuse, the mechanical failure of some component, or electrical failure within the machine.

For each of these failures, there is an associated cost modelled as follows:

Breakdown problem	**Misuse**	**Mechanical**	**Electrical**
Likelihood of breakdown	0.3	0.5	0.2
Repair cost	$350	$200	$50

Solve for the expected repair cost.

Solution: Based on the given information:

the expected repair cost = $350*0.3 + $200*0.5 + $50*0.2 = $215

Sometimes, we are interested in finding the average value of **some function of X**.

Let's revisit the toll gate example:

Say a toll of $2 is paid per car and 50 cents per occupant. Find the average long run toll payment using the probability distribution given earlier, namely:

x	1	2	3	4	5	6
$f(x)$	0.30	0.25	0.20	0.15	0.09	0.01

That is, we are interested in the mean of the r.v.

$$Y = 0.5X + 2$$

Here we see that Y is a function of X.

Theorem: If the discrete r.v. X has p.f. $f(x)$, then the expected value of some function $g(X)$ of X is given by

$$E[g(X)] = \sum_{x \in A} g(x) f(x).$$

In the previous example, recall that we are interested in the mean of the r.v. $\mathbf{Y = 0.5X + 2}$.

Using the previous theorem, we obtain:

Example:

Consider the r.v. X with p.f. of the form

$$f_X(x) = \frac{1}{3} \text{ for } x = -1, 0, 1.$$

We define the r.v. $Y = X^2$

Now, calculate $E(Y)$.

Remarks:

1. As we saw in calculating the expected value of X, the expected value of $g(X)$ may be a value that $g(X)$ can never actually take.
2. Notice that when $g(x) = x$, we simply obtain the expected value of X just like before.
3. Both μ and $E(X)$ refer to the same quantity (i.e. both represent the **expected value / average / mean** of a r.v.). However, there is a small advantage to using the greek letter μ, in that it makes it visually clearer that the expected value is **NOT** a r.v. like X but **it is a constant**.

4. Always look at the value of your expected value and make sure it makes sense in the context of the problem.

For example, if $0 < X < 5$ and you get $E(X)$=12, then something has gone WRONG! Here, your expected value should also be between 0 and 5.

5. The expected value has a physical interpretation. In physical terms, $E(X)$ is the **balance point** for the probability distribution of $f(x)$.

Linearity Properties of Expectation:

- The expectation $E(\cdot)$ is a **linear operator**, satisfying the properties:

1. For constants a and b,

$$E[ag(X) + b] = aE[g(X)] + b.$$

Note: Property 1 tells us that:

$$E(aX + b) = aE(X) + b,$$

implying that if $g(X) = aX + b$, then

$$E[g(X)] = g(E(X)).$$

However, if $g(X)$ is a **non-linear function** of X, it is **NOT** generally true that $E[g(X)] = g(E(X))$, ***so be careful*!**

2. Property 1 can be easily expanded:

For constants a and b and two functions g_1 and g_2:

$$E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)].$$

Example: Suppose that X has p.f. of the form

$$\mathrm{f}(x) = \frac{x}{10} \text{ for } x = 1, 2, 3, 4.$$

Find $E[X(5 - X)]$.

Example:

A local television station sells only 15 second, 30 second, and 60 second advertising spots. Let X denote the length of a randomly selected commercial appearing on this station, and suppose that the probability distribution of X is given by:

x	15	30	60
$f(x)$	0.1	0.3	0.6

a) Find $E(X)$.

b) If a 15 second spot sells for \$500, a 30 second spot for \$800, and a 60 second spot for \$1000, find the average amount paid for a commercial appearing on this station.

You Try:

Let X have the following probability function:

x	1	2	3	4
$f(x)$	0.4	0.3	0.1	0.2

a) Calculate $E(X)$.

b) Calculate $E\left(\frac{1}{X}\right)$. (Be careful! Use first principles.)

Solution:

a) With the given information:

E(X) = 1*0.4 + 2*0.3 + 3*0.1 + 4*0.2 = 2.1

b) We can organize our preliminary calculations for this expected value, as E(1/X) = E[g(X)]

So, g(1) = 1, g(2) = 1/2, g(3) = 1/3, g(4) = 1/4.

So, E(1/X) = (1)*0.4 + (1/2)*0.3 + (1/3)*0.1 + (1/4)*0.2

E(1/X) = 19/30 = 0.63333

c) In a win-win game, a player will win a monetary prize, but has to decide between the **fixed price** of $\frac{\$1000}{E(X)}$ and the **random price** of $\frac{\$1000}{X}$, where the r.v. X has the p.f. defined earlier. Which choice would you recommend?

Solution:

Fixed price: We want to calculate:

1,000/E(X) = 1,000/2.1 = 476.19

Random price: We want to calculate:

E(1000/X) = 1,000(0.4) + (1,000/2)(0.3) + (1000/3)(0.1) + (1,000/4)0.2 = 633.3333

So, the random price is better!

Means of Special Distributions (Section 7.4)

Now, we look at finding the mean, $\mu = E(X)$, for each of the probability models derived in Chapter 5.

We begin with the **binomial distribution**:

Let X~Binomial(n, p), then $\mu = E(X) = np$

Next, let us consider the **Poisson distribution**:

Let X~Pois(λt), then $\mu = E(X) = \lambda t$

Proceeding in a similar fashion for the other discrete probability distributions, we find:

Discrete Uniform: $X \sim DU(a,b): \mu = E(X) = \frac{a+b}{2}$

Hypergeometric: $X \sim HG(N,r,n): \mu = E(X) = \frac{nr}{N}$

Negative Binomial: $X \sim NB(k,p): \mu = E(X) = \frac{k(1-p)}{p}$

Geometric: $X \sim Geo(p): \mu = E(X) = \frac{1-p}{p}$

Note: The Negative Binomial and Geometric distributions are defined in terms of the number of failures before the kth and first success respectively.

Variability

The expected value tells us where the distribution is *on average* and is a useful summary statistic, but it tells us nothing about the **variability within the set of observations**.

How do we measure variability (or dispersion)?

Does it make sense to look at the **average difference** between X and μ?

$$E(X-\mu) = E(X) - \mu = \mu - \mu = 0.$$

Based on the above observation, we see that this is not really informative.

What qualifies as a "good" or "appropriate" measure of variability?

One idea is to consider the **mean absolute deviation** of a r.v. X, namely:

$$E(|X-\mu|)$$

While this is certainly a worthy idea, one disadvantage to this measure is that, in general, absolute values do not possess "nice" mathematical properties.

An alternative but related formulation first **squares** the distance between X and μ and then takes its expected value.

This formulation leads to the **variance**.

Definition: The **variance** of a r.v. X is

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

From this definition, we see that the variance is the average squared distance the r.v. X is **away from its mean**.

We also talk about the positive **square root** of the variance, because if our r.v. X is associated with a particular unit of measurement, then the variance is expressed in (units)^2 which may sometimes be inconvenient.

By taking the positive square root of the variance, we regain the original units.

Definition: The **standard deviation** of a r.v. X is

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

Based on these definitions, note that variance and standard deviation can **NEVER** be **negative**.

The smallest these values can be is 0.

Although the definitions of the variance and standard deviation allow for an easier understanding of what the measurement of variability is doing, these definitions are often awkward to use computationally.

Instead, we will typically use:

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

   ** tends to be used in most situations*

2. $Var(X) = E[X(X-1)] + \mu - \mu^2$

   ** most often used when there is an $x!$ in the denominator of the p.f. $f(x)$*

Example: Suppose X is a r.v. with p.f. given by

x	1	2	3	4	5	6	7	8	9
$f(x)$	0.07	0.1	0.12	0.13	0.16	0.13	0.12	0.1	0.07

Find:

a. $\mathrm{E}(\mathrm{X})$
b. $\mathrm{Var}(\mathrm{X})$
c. $\mathrm{sd}(\mathrm{X})$

Illustrations of Variance

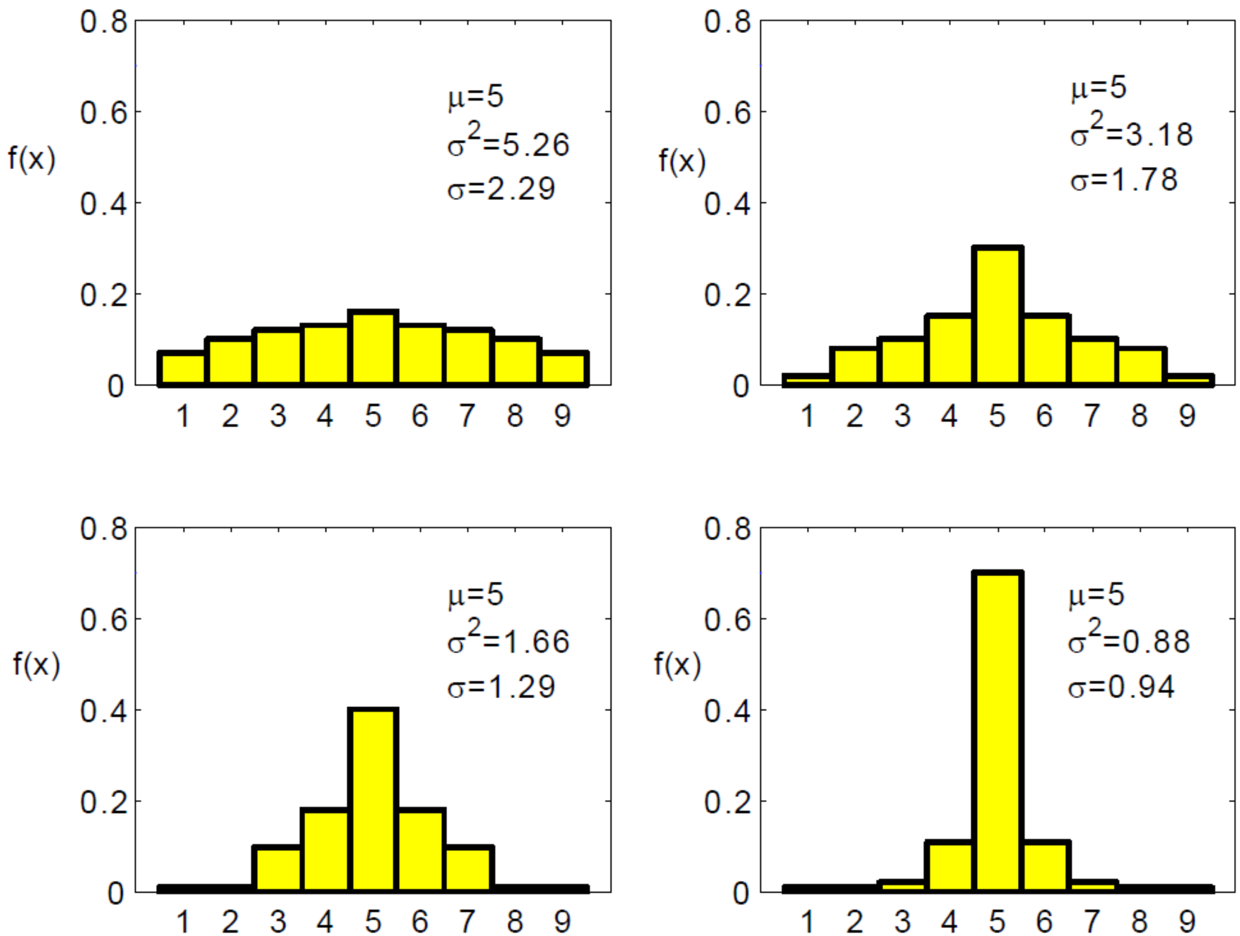


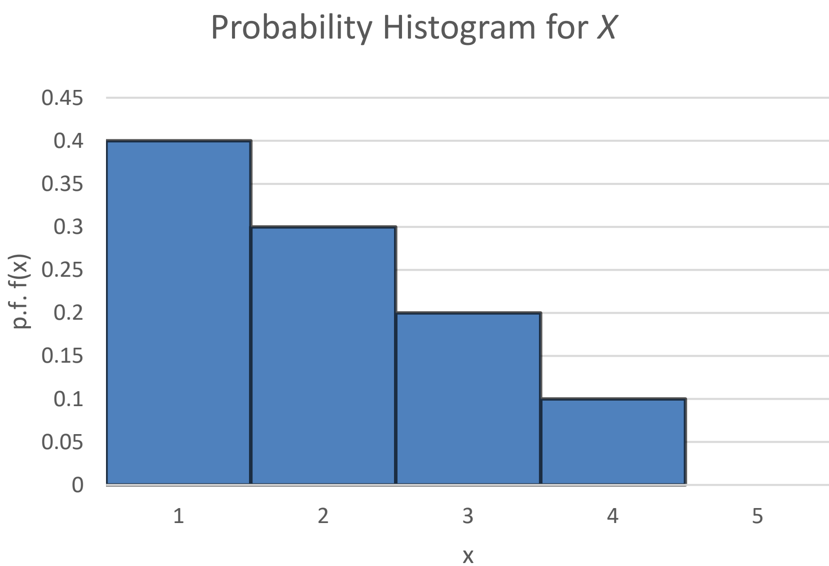
Figure 7.3: How $Var(X)$ or $sd(X)$ reflects the spread of a probability histogram

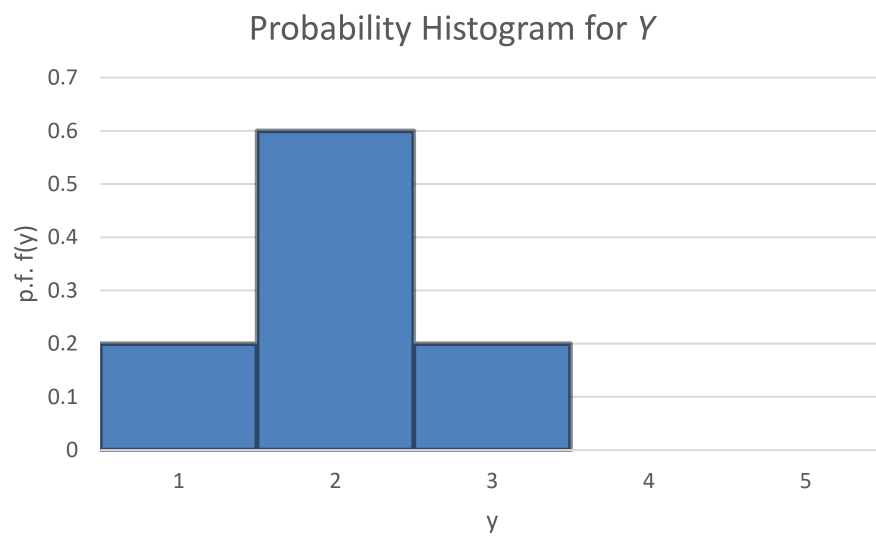
Example: A television manufacturer receives certain components in lots of four from two different suppliers. Let X and Y denote the number of defective components in randomly selected lots from the first and second suppliers, respectively.

The probability distributions and associated probability histograms are given on the next slide.

x	0	1	2	3
$f_X(x)$	0.4	0.3	0.2	0.1

y	0	1	2
$f_Y(y)$	0.2	0.6	0.2





Solve for the expected value and variance of X and Y and compare their distributions.

Variances of Special Distributions

Just as we did for the mean, we next look at finding the variance, $\sigma^2 = Var(X)$, for each of the probability models derived in Chapter 5.

We again begin with the **binomial distribution**:

If X~Bin(n, p), then:

Var(X) = $\sigma^2 =$ n*p*(1-p)

Sd(X) = $\sqrt{Var(X)} = \sqrt{n * p * (1 - p)}$

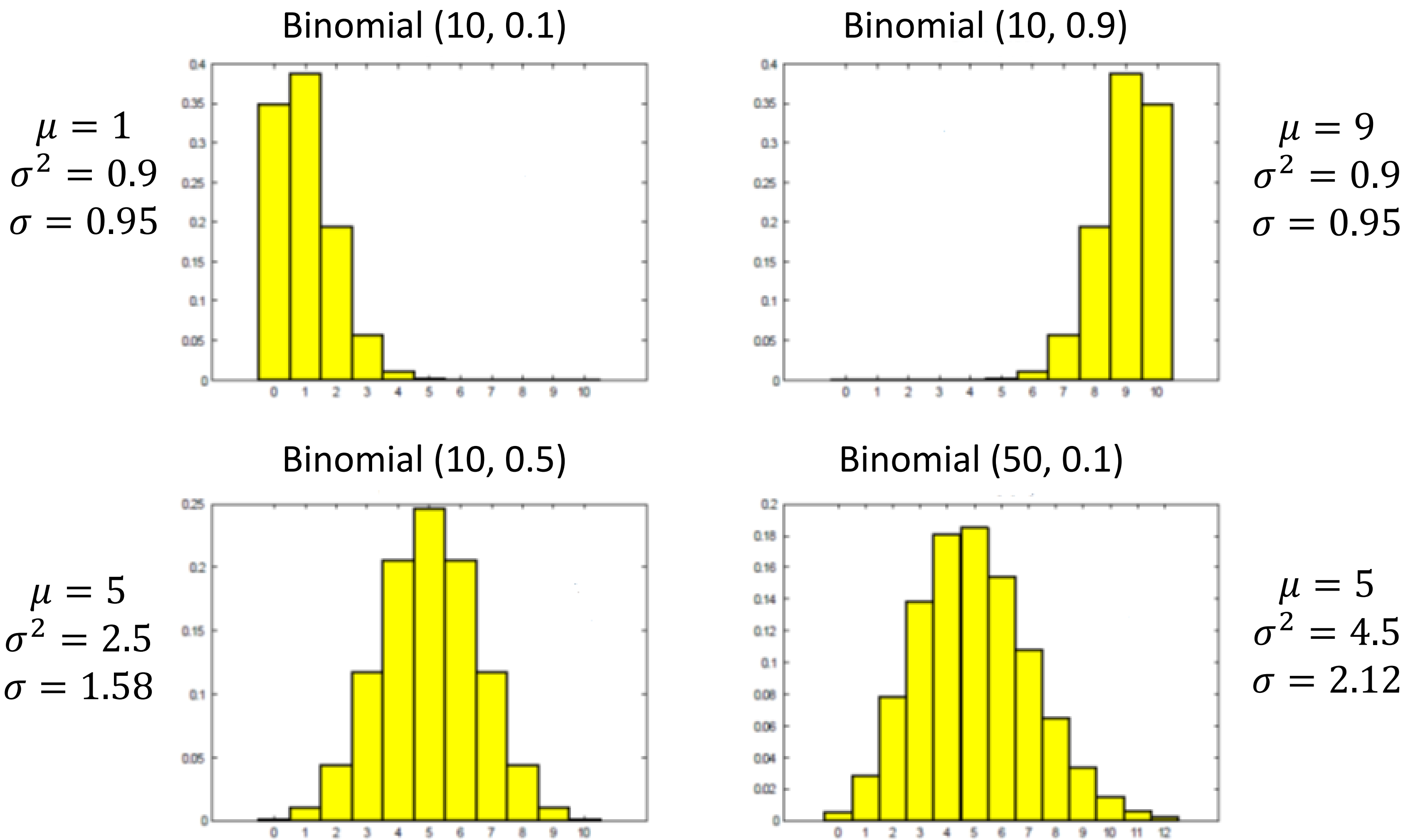


Figure 7.4: Probability histograms, means and variances for various Binomial(n, p) random variables

Next, let us consider the **Poisson distribution:**

One of the unique features of a Poisson r.v. is that the mean and variance are equal!

That is, if $X \sim Pois(\mu = \lambda t)$, then:

Recall: $E(X) = \lambda t$

$Var(X) = \lambda t$

$Sd(X) = \sqrt{\lambda t}$

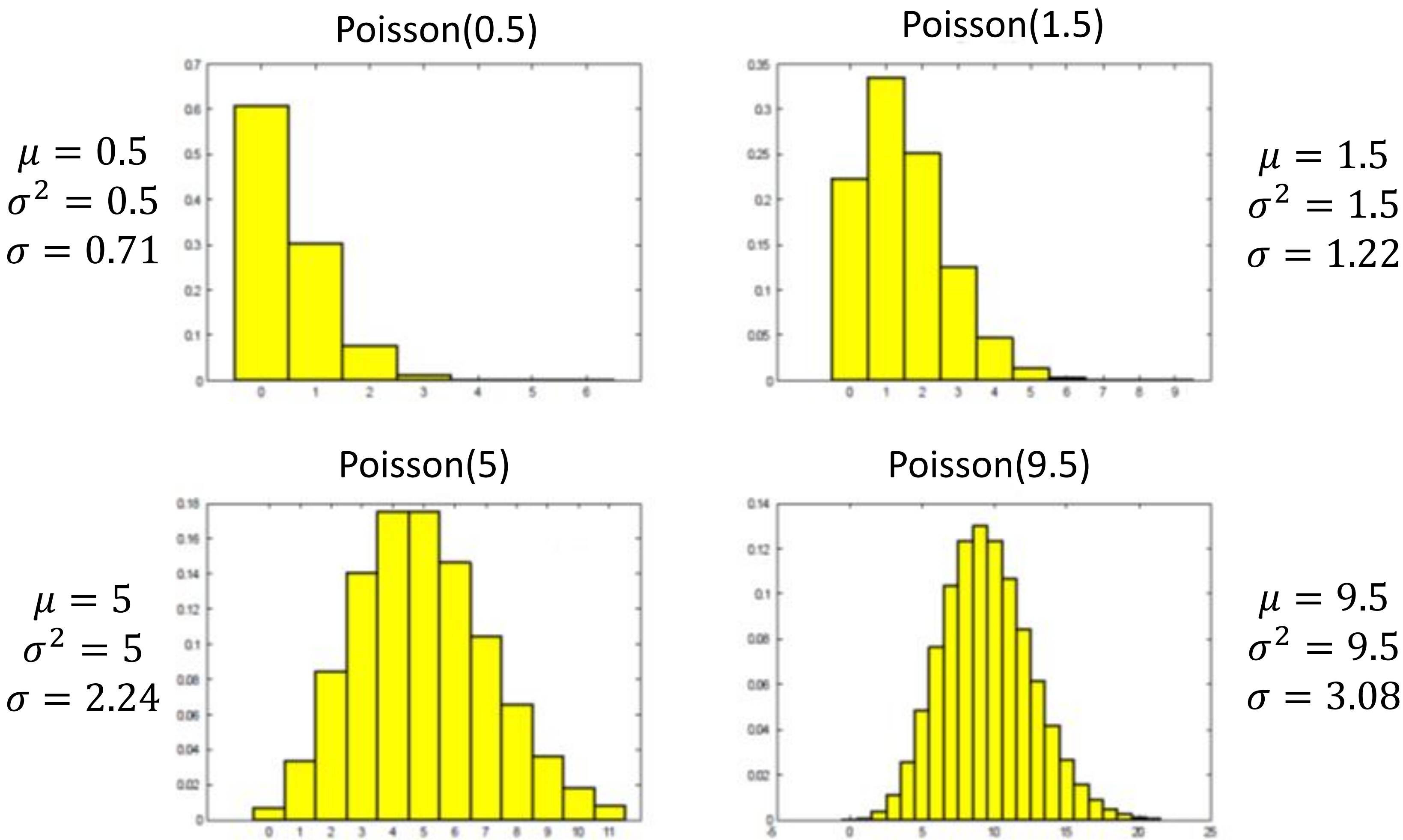


Figure 7.5: Probability histograms, means and variances for various Poisson(μ) random variables

Proceeding in a similar fashion for the other discrete probability distributions, we find:

$$X \sim DU(a,b): Var(X) = \frac{(b-a+1)^2-1}{12}$$

$$X \sim HG(N,r,n): Var(X) = \frac{nr}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$X \sim NB(k,p): Var(X) = \frac{k(1-p)}{p^2}$$

$$X \sim Geo(p): Var(X) = \frac{1-p}{p^2}$$

Note: In the case of the NB and Geo distributions, we are considering the number of failures before the kth and 1st successes, respectively.

Important Properties:

If a and b are constants and $Y = aX + b$, then:

1. $\mu_Y = E(Y) = aE(X) + b = a\mu_X + b$
2. $\sigma_Y^2 = Var(Y) = a^2 Var(X) = a^2 \sigma_X^2$
3. $\sigma_Y = sd(Y) = |a|\, sd(X) = |a|\sigma_X$

For 3.: Note that *a* could be negative, so the absolute value sign guards against a negative standard deviation, which we know is not possible.