# STA 256: Statistics and Probability I

Al Nosedal.
University of Toronto.

Fall 2017

My momma always said: "Life was like a box of chocolates. You never know what you're gonna get."

Forrest Gump.

# Definition 3.1

A random variable $Y$ is said to be **discrete** if it can assume only a finite or countably infinite number of distinct values.

Note. Recall that a set of elements is countably infinite if the elements in the set can be put into one-to-one correspondence with the positive integers.

# Definition 3.2

The probability that $Y$ takes on the value $y$, $P(Y = y)$, is defined as the sum of probabilities of all sample point in $S$ that are assigned the value $y$. We will sometimes denote $P(Y = y)$ by $p(y)$.

# Definition 3.3

The **probability distribution** for a discrete variable $Y$ can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all $y$.

# Theorem 3.1

For any discrete probability distribution, the following must be true:
1. $0 \leq p(y) \leq 1$ for all $y$.
2. $\sum_y p(y) = 1$, where the summation is over all values of $y$ with nonzero probability.

# Example

A couple plans to have three children. There are 8 possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All 8 arrangements are (approximately) equally likely.

a. Write down all 8 arrangements of the sexes of three children. What is the probability of any one of these arrangements?

b. Let $X$ be the number of girls the couple has. What is the probability that $X = 2$?

c. Starting from your work in (a), find the distribution of $X$. That is, what values can $X$ take, and what are the probabilities for each value?

# Solution

a. Sample space = S = {(G,G,G), (G,G,B), (G,B,G), (G,B,B), (B,G,G), (B,G,B), (B,B,G), (B,B,B)}.
Probability of any of these arrangements = $\frac{1}{8}$.

b. $P(X = 2) = P(G, G, B) + P(G, B, G) + P(B, G, G) = \frac{3}{8}$.

c. $P(X = 0) = \frac{1}{8}$, $P(X = 1) = \frac{3}{8}$,
$P(X = 2) = \frac{3}{8}$, $P(X = 3) = \frac{1}{8}$

A group of four components is known to contain two defectives. An inspector tests the components one at a time until the two defectives are located. Once she locates the two defectives, she stops testing, but the second defective is tested to ensure accuracy. Let $Y$ denote the number of the test on which the second defective is found. Find the probability distribution for $Y$.

# Solution

Let $D_i = \{$ the i-th component tested is defective $\}$ and
$N_i = \{$ the i-th is Nondefective$\}$.
$P(Y = 2) = P(D_1 \cap D_2) = P(D_2|D_1)P(D_1) = (1/3)(1/2) = 1/6.$

$$P(Y = 3) = P(D_1 \cap N_2 \cap D_3) + P(N_1 \cap D_2 \cap D_3)$$
$$= P(D_3|D_1 \cap N_2)P(D_1 \cap N_2) + P(D_3|N_1 \cap D_2)P(N_1 \cap D_2)$$
$$= P(D_3|D_1 \cap N_2)P(N_2|D_1)P(D_1) + P(D_3|N_1 \cap D_2)P(D_2|N_1)P(N_1)$$
$$= (1/2)(2/3)(1/2) + (1/2)(2/3)(1/2) = (1/6) + (1/6) = 2/6.$$

# Solution

$P(Y = 4) =$
$P(N_1 \cap N_2 \cap D_3 \cap D_4) + P(N_1 \cap D_2 \cap N_3 \cap D_4) + P(D_1 \cap N_2 \cap N_3 \cap D_4)$
Let's find $P(N_1 \cap N_2 \cap D_3 \cap D_4)$.
$P(N_1 \cap N_2 \cap D_3 \cap D_4) =$
$P(D_4|N_1 \cap N_2 \cap D_3)P(D_3|N_1 \cap N_2)P(N_2|N_1)P(N_1)$.
From the last equation, we have that
$P(N_1 \cap N_2 \cap D_3 \cap D_4) = (1)(1)(1/3)(1/2) = 1/6$
Doing something similar with $P(N_1 \cap D_2 \cap N_3 \cap D_4)$ and
$P(D_1 \cap N_2 \cap N_3 \cap D_4)$, we have that:

$$P(Y = 4) = 3/6.$$

In order to verify the accuracy of their financial accounts, companies use auditors on a regular basis to verify accounting entries. The company's employees make erroneous entries 5% of the time. Suppose that an auditor randomly checks three entries. (Assume independence)

a. Find the probability distribution for $Y$, the number of errors detected by the auditor.

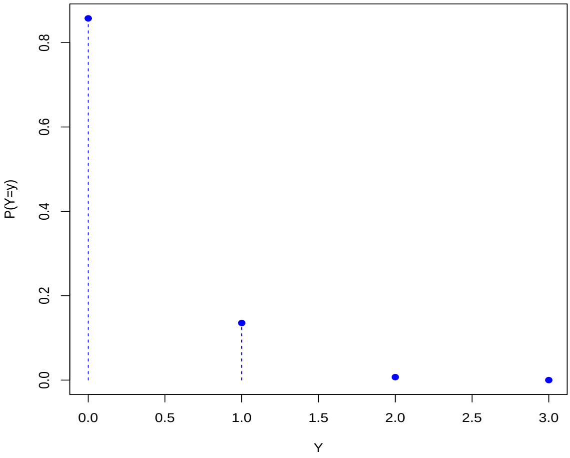
b. Construct a graph for $p(y)$.

c. Find the probability that the auditor will detect more than one error.

# Solution

Let $E$ denote an error on a single entry and let $N$ denote no error. There are 8 sample points: {(E,E,E), (E,E,N), (E,N,E), (E,N,N), (N,E,E), (N,E,N), (N,N,E), (N,N,N)}.
We also have that $P(E) = 0.05$ and $P(N) = 0.95$. Then
$P(Y = 0) = (0.95)^3 = 0.857375$
$P(Y = 1) = 3(0.05)(0.95)^2 = 0.135375$
$P(Y = 2) = 3(0.05)^2(0.95) = 0.007125$
$P(Y = 3) = (0.05)^3 = 0.000125.$

# Solution b)



# Solution c)

$$P(Y > 1) = P(Y = 2) + P(Y = 3)$$

$$= 0.007125 + 0.000125 = 0.00725.$$

# Definition 3.4

Let $Y$ be a discrete random variable with the probability function $p(y)$. Then the **expected value** of $Y$, $E(Y)$, is defined to be

$$E(Y) = \sum_{\text{all values of y}} yp(y).$$

# Exercise 3.22*

A single fair die is tossed once. Let $Y$ be the number facing up. Find

a) the expected value of $Y$,

b) the expected value of $W = 3Y$,

# Solution a)

By definition,

$$E(Y) = \sum_{y=1}^{6} yP(Y = y)$$

$$= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

$$E(Y) = \left(\frac{1}{6}\right)(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$$

# Solution b)

By definition,

$E(W) = \sum_{all\ ws} wP(W = w)$

$$= 3\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) + 12\left(\frac{1}{6}\right) + 15\left(\frac{1}{6}\right) + 18\left(\frac{1}{6}\right)$$

$$= 3[1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)]$$

$E(W) = 3[3.5] = 3E(Y)$

# Example

The probability distribution for a random variable $Y$ is given below. Find the expected value of $Y$ and $W = Y^2$.

| y | p(y) |
|---|---|
| -2 | 1/5 |
| -1 | 1/5 |
| 0 | 1/5 |
| 1 | 1/5 |
| 2 | 1/5 |

# Solution

By definition
$E(Y) = (-2)p(-2) + (-1)p(-1) + (0)p(0) + (1)p(1) + 2p(2)$

$$= \left(\frac{1}{5}\right)(-2 - 1 + 0 + 1 + 2) = 0$$

# Solution

Note that the probability distribution of $W$ is given by:

| w | p(w) |
|---|---|
| 0 | $1/5$ |
| 1 | $1/5 + 1/5 = 2/5$ |
| 4 | $1/5 + 1/5 = 2/5$ |

# Solution

By definition

$$E(W) = (0)P(W = 0) + (1)P(W = 1) + (4)P(W = 4)$$

$$= 0 + 2/5 + 8/5 = 10/5 = 2$$

# Solution

Now, let us rewrite this in terms of $Y$

$$E(W) = (0)[P(Y=0)] + (1)[P(Y=-1) + P(Y=1)] + (4)[P(Y=-2) + P(Y=2)]$$

$$E(W) = (0)P(Y=0) + (1)P(Y=-1) + (1)P(Y=1) + (4)P(Y=-2) + (4)P(Y=2)$$

$$E(W) = (0)P(Y=0) + (-1)^2 P(Y=-1) + (1)^2 P(Y=1) + (-2)^2 P(Y=-2) + (2)^2 P(Y=2)$$

$$E(W) = \sum_{y=-2}^{2} y^2 P(Y=y)$$

$$E(Y^2) = \sum_{y=-2}^{2} y^2 P(Y=y)$$

# Theorem 3.2

Let $Y$ be a discrete random variable with probability function $p(y)$ and $g(Y)$ be a real-valued function of $Y$. Then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \sum_{\text{all y}} g(y)p(y).$$

# Proof

We prove the result in the case where the random variable $Y$ takes on the finite number of values $y_1, y_2, ..., y_n$. Because the function $g(y)$ may not be one-to-one, suppose that $g(Y)$ takes on values $g_1, g_2, ..., g_m$ (where $m \leq n$). It follows that $g(Y)$ is a random variable such that for $i = 1, 2, ..., m$,

$$P[g(Y) = g_i] = \sum_{g(y_j)=g_i} p(y_j) = p^*(g_i).$$

# Proof

Thus, by Definition 3.4,

$$E[g(Y)] = \sum_{i=1}^{m} g_i p^*(g_i)$$
$$= \sum_{i=1}^{m} g_i \left\{ \sum_{g(y_j)=g_i} p(y_j) \right\}$$
$$= \sum_{i=1}^{m} \sum_{g(y_j)=g_i} g_i p(y_j)$$
$$= \sum_{j=1}^{n} g(y_j) p(y_j).$$

# Definition 3.5

If $Y$ is a random variable with mean $E(Y) = \mu$, the **variance** of a random variable $Y$ is defined to be the expected value of $(Y-\mu)^2$. That is,

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

The **standard deviation** of $Y$ is the positive square root of $V(Y)$.

# Theorem 3.3

Let $Y$ be a discrete random variable with probability function $p(y)$ and $c$ be a constant. Then $E(c) = c$.

Proof.
Easy! Let $g(Y) = c$ and apply Theorem 3.2.

# Theorem 3.4

Let $Y$ be a discrete random variable with probability function $p(y)$, $g(Y)$ be a function of $Y$, and $c$ be a constant. Then

$$E[cg(Y)] = cE[g(Y)].$$

Proof.
Easy! Apply Theorem 3.2.

# Theorem 3.5

Let $Y$ be a discrete random variable with probability function $p(y)$ and $g_1(Y)$, $g_2(Y)$, . . . , $g_k(Y)$ be $k$ functions of $Y$. Then

$$E[g_1(Y)+g_2(Y)+...+g_k(Y)] = E[g_1(Y)]+E[g_2(Y)]+...+E[g_k(Y)].$$

# Proof

We will demonstrate the proof only for the case $k = 2$, but analogous steps will hold for any $k$. By Theorem 3.2

$$\begin{aligned} E[g_1(Y) + g_2(Y)] &= \sum_y [g_1(y) + g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(y)] + E[g_2(y)]. \end{aligned}$$

# Theorem 3.6

Let $Y$ be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

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

or

$$V(Y) = E(Y^2) - [E(Y)]^2.$$

# Proof

$$V(Y) = E[(Y-\mu)^2] = E[Y^2 - 2Y\mu + \mu^2]$$

$$V(Y) = E(Y^2) - E(2Y\mu) + E(\mu^2) \text{ (by Theorem 3.5)}$$

$$V(Y) = E(Y^2) - 2\mu E(Y) + \mu^2 \text{ (Noting that } \mu \text{ is a constant)}$$

$$V(Y) = E(Y^2) - 2\mu^2 + \mu^2$$

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

# Example

The probability distribution for a random variable $Y$ is given below. Find $V(Y)$.

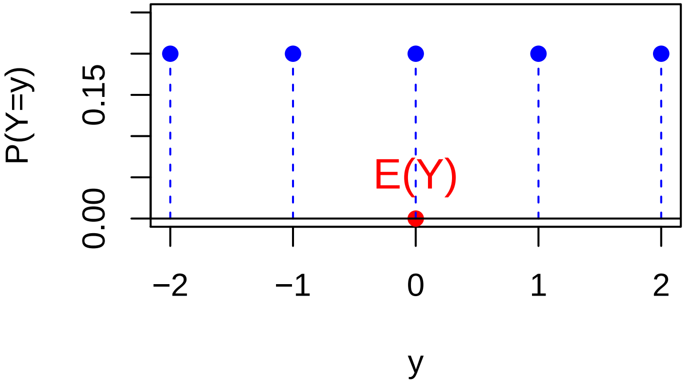
| y | p(y) |
|---|---|
| -2 | 1/5 |
| -1 | 1/5 |
| 0 | 1/5 |
| 1 | 1/5 |
| 2 | 1/5 |

# Solution

From what we did above, we know that $E(Y) = 0$ and $E(Y^2) = 2$. Now, recalling that $V(Y) = E(Y^2) - [E(Y)]^2$, we have that:

$$V(Y) = 2 - 0^2 = 2.$$

# Graph



# Example

The probability distribution for a random variable $W$ is given below. Find $V(W)$.

| w | p(w) |
|---|---|
| -3 | 1/7 |
| -2 | 1/7 |
| -1 | 1/7 |
| 0 | 1/7 |
| 1 | 1/7 |
| 2 | 1/7 |
| 3 | 1/7 |

# Solution

By definition,

$$E(W) = (-3)\left(\frac{1}{7}\right) + ... + 0\left(\frac{1}{7}\right) + ... + 3\left(\frac{1}{7}\right) = 0.$$

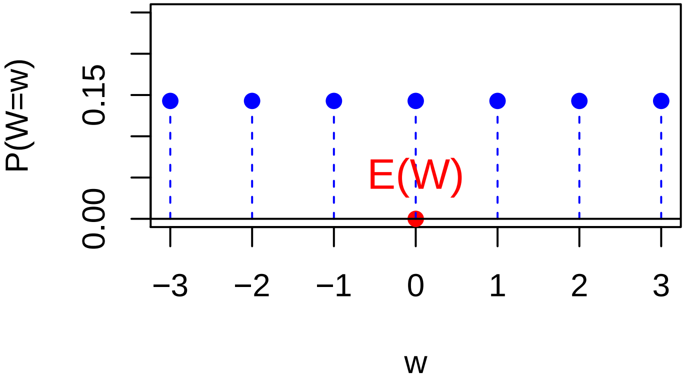
By Theorem 3.2,

$$E(W^2) = (-3)^2\left(\frac{1}{7}\right) + ... + 0^2\left(\frac{1}{7}\right) + ... + 3^2\left(\frac{1}{7}\right)$$

$$E(W^2) = \frac{1}{7}(9 + 4 + 1 + 0 + 1 + 4 + 9) = \frac{28}{7} = 4.$$

$$V(W) = 4 - (0)^2 = 4.$$

# Graph



# Variance and Standard Deviation

Consider two users. One receives either 8 or 12 e-mail messages per day, with a 50-50% chance of each. The other receives either 0 or 20 e-mails, also with a 50-50% chance. Find the mean, variance, and standard deviation of these two distributions.

# Solution

Y = number of emails received by user one.
$E(Y) = 8\left(\frac{1}{2}\right) + 12\left(\frac{1}{2}\right) = \frac{20}{2} = 10.$
$E(Y^2) = 8^2\left(\frac{1}{2}\right) + 12^2\left(\frac{1}{2}\right) = \frac{64+144}{2} = \frac{208}{2} = 104.$
$V(Y) = E(Y^2) - [E(Y)]^2 = 104 - 100 = 4.$
$\sigma_Y = \sqrt{V(Y)} = \sqrt{4} = 2.$

# Solution

X = number of emails received by user two.
$E(X) = 0\left(\frac{1}{2}\right) + 20\left(\frac{1}{2}\right) = \frac{20}{2} = 10.$
$E(X^2) = 0^2\left(\frac{1}{2}\right) + 20^2\left(\frac{1}{2}\right) = \frac{400}{2} = 200.$
$V(X) = E(X^2) - [E(X)]^2 = 200 - 100 = 100.$
$\sigma_X = \sqrt{V(X)} = \sqrt{100} = 10.$

# Exercise 3.27

A potential customer for an \$85,000 fire insurance policy possesses a home in an area that, according to experience, may sustain a total loss in a given year with probability of 0.001 and a 50% loss with probability 0.01. Ignoring all other partial losses, what premium should the insurance company charge for a yearly policy in order to break even on all \$85,000 policies in this area?

# Solution

Let $Y$ = Payout of an individual policy. First, we will find the probability distribution of $Y$.

$P(Y = 85,000) = 0.001$
$P(Y = 42,500) = 0.01$
$P(Y = 0) = 1 - (0.001 + 0.01) = 0.989$

# Solution (cont.)

By definition,

$$E(Y) = 85,000P(Y = 85,000) + 42,500P(Y = 42,500) + 0P(Y = 0)$$

$$E(Y) = 85,000(0.001) + 42,500(0.01) + 0(0.989) = 85 + 425 = 510.$$

# Exercise 3.31

Suppose that $Y$ is a discrete random variable with mean $\mu$ and variance $\sigma^2$ and let $W = 2Y$.
i) Find $E(W)$.
ii) Find $V(W)$.

# Solution

i) $E(W) = E(2Y) = 2E(Y) = 2\mu$ (By Theorem 3.4)
ii) $V(W) = E\{(W - \mu_W)^2\}$ (by definition)
$= E\{[2Y - 2\mu]^2\}$ (from part i) )
$= E\{2^2[Y - \mu]^2\}$
$= 2^2 E\{[Y - \mu]^2\}$ (by Theorem 3.4)
$= 4V(Y) = 4\sigma^2$ (by definition of $V(Y)$ ).

# Exercise 3.33

Let $Y$ be a discrete random variable with mean $\mu$ and variance $\sigma^2$.
If $a$ and $b$ are constants, prove that
a. $E(aY + b) = aE(Y) + b$.
b. $V(aY + b) = a^2 V(Y) = a^2\sigma^2$.

# Solution

a) $E(aY + b) = E[aY] + E[b]$ (by Theorem 3.5)
$= aE[Y] + b$ (by Theorem 3.4 and Theorem 3.3)

b) $V(aY + b) = E\{[aY + b - (aE(Y) + b)]^2\}$ (by definition of variance and from part a) )
$= E\{[aY - aE(Y)]^2\}$
$= E\{a^2[Y - E(Y)]^2\}$
$= a^2E\{[Y - E(Y)]^2\}$ (by Theorem 3.4)
$= a^2V(Y)$ (by definition of $V(Y)$).

# Definition 3.6

A **binomial experiment** possesses the following properties:

- The experiment consists of a fixed number, $n$, of identical trials.
- Each trial results in one of two outcomes: success, S, or failure, F.
- The probability of success on a single trial is equal to some value $p$ and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
- The trials are **independent**.
- The random variable of interest is $Y$, the number of successes observed during the $n$ trials.

# Definition 3.7

A random variable $Y$ is said to have a **binomial distribution** based on $n$ trials with success probability $p$ if and only if

$$p(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}, \ \ y = 0, 1, 2, ..., n \text{ and } 0 \leq p \leq 1.$$

or

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y}, \ \ y = 0, 1, 2, ..., n \text{ and } 0 \leq p \leq 1.$$

# Binomial Theorem

$$(a+b)^2 = a^2 + 2ab + b^2$$

or

$$(a+b)^2 = \binom{2}{2} a^2 b^0 + \binom{2}{1} a^1 b^1 + \binom{2}{0} a^0 b^2$$

# Binomial Theorem (again)

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

or

$$(a+b)^3 = \binom{3}{3}a^3b^0 + \binom{3}{2}a^2b^1 + \binom{3}{1}a^1b^2 + \binom{3}{0}a^0b^3$$

or

$$(a+b)^3 = \sum_{y=0}^{3}\binom{3}{y}a^y b^{3-y}$$

# Binomial Theorem (again...)

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

# Theorem 3.7

Let $Y$ be a binomial random variable based on $n$ trials and success probability $p$. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Proof.
Check page 107 of our textbook (there you will find something similar to what we did in class).

# Exercise 3.45

A fire-detection device utilizes three temperature-sensitive cells acting independently of each other in such a manner that any one or more may activate the alarm. Each cell possesses a probability of $p = 0.8$ of activating the alarm when the temperature reaches 100 degrees Celsius or more. Let $Y$ equal the number of cells activating the alarm when the temperature reaches 100 degrees.

a. Find the probability distribution for $Y$.

b. Find the probability that the alarm will function when the temperature reaches 100 degrees.

# Solution

a) Let $Y$ = number of cells activating the alarm. $Y$ has a Binomial distribution with parameters $n = 3$ and $p = 0.8$

b)
$$
\begin{aligned}
P[\text{alarm will function}] &= P[Y \geq 1] \\
&= P[Y \geq 1] \\
&= 1 - P[Y = 0] \\
&= 1 - 0.008 = 0.992
\end{aligned}
$$

# Exercise 3.39

A complex electronic system is built with a certain number of backup components in its subsystems. One subsystem has four identical components, each with a probability of 0.2 of failing in less than 1000 hours. The subsystem will operate in any two of the four components are operating. Assume that the components operate independently. Find the probability that

a. exactly two of the four components last longer than 1000 hours.

b. the subsystem operates longer than 1000 hours.

# Solution

Let $Y$ be the number of components failing in less than 1000 hours. $Y$ has a Binomial distribution with parameters $n = 4$ and $p = 0.2$.

a) $P(Y = 2) = \binom{4}{2}(0.2)^2(0.8)^2$
$= 6(4/100)(64/100) = 1536/10000 = 0.1536$

b) The system will operate if 0, 1, or 2 components fail in less than 1000 hours.

$P(\text{system operates}) = P(Y = 0) + P(Y = 1) + P(Y = 2)$
$= 0.4096 + 0.4096 + 0.1536 = 0.9728$

# Definition 3.8

A random variable $Y$ is said to have a **geometric probability** distribution if and only if

$$p(y) = q^{y-1}p, \; y = 1, 2, 3, ..., \; 0 \leq p \leq 1.$$

# Theorem 3.8

If $Y$ is a random variable with a geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \qquad \sigma^2 = V(Y) = \frac{1-p}{p^2}.$$

Proof.
Check page 117 of our textbook (there you will find something similar to what we did in class).

# Exercise 3.81

How many times would you expect to toss a balanced coin in order to obtain the first head?

# Solution

Let $Y$= number of the toss on which the first head is observed. $Y$ has a Geometric distribution with parameter $p = P(heads) = 1/2$. Then,

$$E(Y) = \frac{1}{p} = \frac{1}{1/2} = 2.$$

So, two times is our answer.

# Exercise 3.79

In responding to a survey question on a sensitive topic (such as "Have you ever tried marijuana?"), many people prefer not to respond in the affirmative. Suppose that 80% of the population have not tried marijuana and all of those individuals will truthfully answer no to your question. the remaining 20% of the population have tried marijuana and 70% of those individuals will lie. Derive the probability distribution of $Y$, the number of people you would need to question in order to obtain a single affirmative response.

# Solution

Let $Y$ = number of people questioned until a "yes" answer is given. Let $M^c$ be the event {randomly chosen individual has not tried marijuana } and $A$ be the event { randomly chosen individual answer "yes" (Affirmative) to question }.

Then, $Y$ has a Geometric distribution with parameter $p$. Let's find $p$

$$p = P(A) = P(A \cap M) + P(A \cap M^c) =$$
$$P(A|M)P(M) + P(A|M^c)P(M^c)$$
$$= (0.3)(0.2) + (0)(0.8) + = 0.06.$$

# Solution

Thus,

$$P(Y = y) = (0.94)^{y-1}(0.06), \ \ y = 1, \ 2, \ 3, \ldots$$

# Example

Observation of a waiting line at a medical clinic indicates the probability that a new arrival will be an emergency case is $p = 1/6$. Assume that conditions of arriving patients represent independent events, and let the random variable $Y$ denote the number of arrivals until the **fourth** emergency case. Find the probability distribution of $Y$.

# Solution

Let $A_i$ = the ith arrival is **not** an emergency. Define $Y$ as the number of arrivals until our **fourth** emergency case. From its definition, $Y$ could take on values $\{4, 5, 6, \ldots\}$.

$$
\begin{aligned}
P(Y = 4) &= P(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c) \\
&= P(A_1^c)P(A_2^c)P(A_3^c)P(A_4^c) \text{ (independence)} \\
&= \left(\tfrac{1}{6}\right)^4
\end{aligned}
$$

# Solution

$$
\begin{aligned}
P(Y=5) = &\ P(A_1 \cap A_2^c \cap A_3^c \cap A_4^c \cap A_5^c) \\
&+ P(A_1^c \cap A_2 \cap A_3^c \cap A_4^c \cap A_5^c) \\
&+ P(A_1^c \cap A_2^c \cap A_3 \cap A_4^c \cap A_5^c) \\
&+ P(A_1^c \cap A_2^c \cap A_3^c \cap A_4 \cap A_5^c)
\end{aligned}
$$

$P(Y=5) = 4\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^4$ or $P(Y=5) = \binom{4}{3}\left(\frac{5}{6}\right)^{(5-4)}\left(\frac{1}{6}\right)^4$

# Solution

If $k$ represents an integer such that $k \geq 4$, proceeding in a similar way we have
$P(Y = k) = \binom{k-1}{3} \left(\frac{5}{6}\right)^{(k-4)} \left(\frac{1}{6}\right)^{4}$

(This is the probability distribution of $Y$).

# Definition 3.9

A random variable $Y$ is said to have a **negative binomial** probability distribution if and only if

$$p(y) = \binom{y-1}{r-1}(1-p)^{y-r}p^r, \ \ y = r, r+1, r+2, ..., \ \ 0 \leq p \leq 1.$$

# Theorem 3.9

If $Y$ is a random variable with a negative binomial distribution,

$$\mu = E(Y) = \frac{r}{p} \qquad \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}.$$

The employees of a firm that manufactures insulation are being tested for indications of asbestos in their lungs. The firm is requested to send three employees who have positive indications of asbestos on to a medical centre for further testing. If 40% of the employees have positive indications of asbestos in their lungs and each test costs $20, find the expected value and variance of the total cost of conducting the tests necessary to locate the three positives.

# Solution

$Y$ = number of tests necessary to locate our first three positives.
$p$ = Probability of positive indications of asbestos = 0.40.
From the conditions described above, $Y$ has a Negative Binomial distribution with parameters $p = 0.40$ and $r = 3$.

$E(Y) = \frac{r}{p} = \frac{3}{0.4} = \frac{30}{4} = 7.5$
$V(Y) = \frac{r(1-p)}{p^2} = \frac{3(0.6)}{(0.4)^2} = 11.25$

# Solution

Let $T$ = Total cost. Clearly, $T = 20Y$.

$E(T) = E(20Y) = 20E(Y) = 20(7.5) = 150$ dollars.
$V(T) = V(20Y) = 20^2 V(Y) = 400(11.25) = 4500$ dollars$^2$.

# Example

Two additional jurors are needed to complete a jury for a criminal trial. There are six prospective jurors, two women and four men. Two jurors are randomly selected from the six available. Let $Y$ represent the number of women in this group of two additional jurors.

a) Find the probability distribution of $Y$.

b) $E(Y)$.

c) $V(Y)$.

# Solution

Let $Y$ represent the number of female jurors in a sample of size 2.
From Exercise 2.29 (check our notes from chapter 2), we have that:
$P(Y = 0) = \frac{6}{15}$
$P(Y = 1) = \frac{8}{15}$
$P(Y = 2) = \frac{1}{15}$

# Solution

$$P(Y=0)=\frac{\binom{2}{0}\binom{4}{2}}{\binom{6}{2}}=\frac{6}{15}$$
$$P(Y=1)=\frac{\binom{2}{1}\binom{4}{1}}{\binom{6}{2}}=\frac{8}{15}$$
$$P(Y=2)=\frac{\binom{2}{2}\binom{4}{0}}{\binom{6}{2}}=\frac{1}{15}$$

# Definition 3.10

A random variable $Y$ is said to have a **hypergeometric probability distribution** if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where $y$ is an integer $0,\ 1,\ 2, ..., n$, subject to the restrictions $y \leq r$ and $n - y \leq N - r$.

# Theorem 3.10

If $Y$ is a random variable with a hypergeometric distribution,

$$\mu = E(Y) = \frac{nr}{N} \text{ and } \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

# Exercise 3.103

A warehouse contains ten printing machines, four of which are defective. A company selects five of the machines at random, thinking all are in working condition. What is the probability that all five of the machines are nondefective?

# Solution

$N = 10.$
$r = 6$ (nondefective machines).
$N - r = 4$ (defective machines).
$n = 5$ (sample size).

$$P(Y = 5) = \frac{\binom{6}{5}\binom{4}{0}}{\binom{10}{5}} = \frac{(6)(1)}{252} = \frac{1}{42}.$$

# Relationship to the Binomial Distribution

There is an interesting relationship between the hypergeometric and the binomial distribution. As one might expect, if $n$ is small compared to $N$, the nature of the $N$ items changes very little in each draw. In fact, as a rule of thumb the approximation is good when $\frac{n}{N} \leq 0.05$.
Thus the quantity $\frac{r}{N}$ plays the role of the binomial parameter $p$. As a result, the binomial distribution may be viewed as a large population edition of the hypergeometric distributions. The mean and variance then come from the formulas

$$\mu = np = \frac{nr}{N}, \quad \sigma^2 = npq = n\frac{r}{N}\left(1 - \frac{r}{N}\right).$$

Comparing these formulas with those of Theorem 3.10, we see that the mean is the same whereas the variance differs by a correction factor of $\frac{(N-n)}{(N-1)}$, which is negligible when $n$ is small relative to $N$.

# Properties of Poisson Process

Experiments yielding numerical values of a random variable $X$, the number of outcomes occurring during a given time interval or in a specified region, are called **Poisson experiments**. The given time interval may be of any length, such as a minute, a day, a week, a month, or even a year. Hence a Poisson experiment can generate observations for the random variable $X$ representing the number of telephone calls per hour received by an office, the number of days school is closed due to snow during the winter, or the number of postponed games due to rain during a baseball season. The specified region could be a line segment, an area, a volume, or perhaps a piece of material. In such instances $X$ might represent the number of field mice per acre, the number of bacteria in a given culture, or the number of typing errors per page.

# Properties of Poisson Process

A Poisson experiment is derived from the **Poisson process** and possesses the following properties:

1. The number of outcomes occurring in one time interval or specified region is independent of the number that occurs in any other disjoint time interval or region of space.
2. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
3. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number $X$ of outcomes occurring during a Poisson experiment is called a **Poisson random variable**, and its probability distribution is called the **Poisson distribution**.

# Definition 3.11

A random variable $Y$ is said to have a **Poisson probability distribution** if and only if

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}, \; y = 0, \; 1, \; 2, ..., \; \lambda > 0.$$

# The Poisson Distribution as a limiting form of the Binomial

Although the Poisson usually finds applications in space and time problems, it can be viewed as a limiting form of the Binomial distribution. In the case of the Binomial, if $n$ is quite large and $p$ is small, the conditions begin to simulate the continuous space or time region implications of the Poisson process. The independence among Bernoulli trials in the Binomial case is consistent with property 2 of the Poisson process. Allowing the parameter $p$ to be close to zero relates to property 3. Indeed, we shall now derive the Poisson distribution as a limiting form of the Binomial distribution when $n \to \infty$, $p \to 0$, and $np$ remains constant. Hence, if $n$ is large and $p$ is close to 0, the Poisson distribution can be used, with $\mu = np$, to approximate Binomial probabilities.

# Derivation of Poisson distribution

Let $Y$ be a Binomial distribution with parameters $n$ and $p$. We are letting $\lambda = np$ and taking the limit of the Binomial probability $p(y)$ as $n \to \infty$.

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$= \frac{(n)(n-1)(n-2)...[n-(y-1)](n-y)!}{y!(n-y)!} \frac{\lambda^y}{n^y} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y}$$

$$= \frac{(n)(n-1)(n-2)...[n-(y-1)]}{(n)(n)(n)...(n)} \frac{\lambda^y}{y!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y}$$

$$= (1)\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) .... \left(1 - \frac{(y-1)}{n}\right) \frac{\lambda^y}{y!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y}$$

Recalling that $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$. We have that

$$\lim_{n\to\infty} P(Y=y) = \frac{\lambda^y}{y!}e^{-\lambda}$$

$y = 0,\ 1,\ 2,\ 3, \ldots$

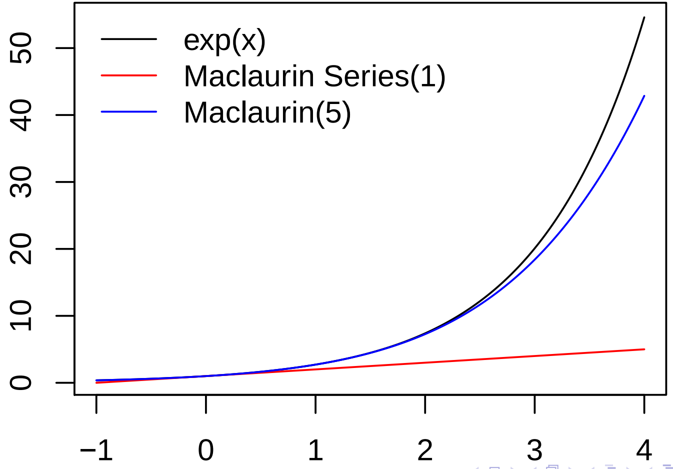
# Maclaurin Series

A Maclaurin series is a Taylor series expansion of a function about 0,

$$f(x) = f(0) + f^{'}(0)x + \frac{f^{''}(0)}{2!}x^2 + \frac{f^{3}(0)}{3!}x^3 + ... + \frac{f^{n}(0)}{n!}x^n + ...$$

(Maclaurin series are named after the Scottish mathematician Colin Maclaurin.)

# Example



# Theorem 3.11

If $Y$ is a random variable possessing a Poisson distribution with parameter $\lambda$, then

$$\mu = E(Y) = \lambda \text{ and } \sigma^2 = V(Y) = \lambda.$$

# Exercise 3.127

The number of typing errors made by a typist has a Poisson distribution with an average of four errors per page. If more than four errors appear on a given page, the typist must retype the whole page. What is the probability that a randomly selected page does not need to be retyped?

# Solution

$Y$ = number of typing errors made by a typist.
$\lambda = 4$

$$P(Y \leq 4) = \frac{4^4}{4!}e^{-4} + \frac{4^3}{3!}e^{-4} + \frac{4^2}{2!}e^{-4} + \frac{4^1}{1!}e^{-4} + \frac{4^0}{0!}e^{-4}$$

$$P(Y \leq 4) = 0.6288$$

# Exercise 3.135

A salesperson has found that the probability of a sale on a single contact is approximately 0.03. If the salesperson contacts 100 prospects, what is the approximate probability of making at least one sale?

# Solution

$X$ = number of sales made by salesperson out of 100 prospects.
$X$ has a Binomial distribution with parameters $n = 100$ and $p = 0.03$.
$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - 0.04755 = 0.95245$.

# Solution (approximation)

$n = 100$ and $p = 0.03$.

$\lambda = np = (100)(0.03) = 3$

Let $Y$ = number of sales.

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y < 1) \\ &= 1 - P(Y = 0) \\ &= 1 - \frac{3^0}{0!}e^{-3} = 0.9502 \end{aligned}$$

# Definition 3.12

The kth moment of a random variable $Y$ taken **about the origin** is defined to be $\mathbf{E(Y^k)}$ and is denoted by $\mu_k^{'}$.

# Definition 3.13

The kth moment of a random variable $Y$ taken **about its mean**, or the kth central moment of $Y$, is defined to be $\mathbf{E}[(\mathbf{Y}-\mu)^{\mathbf{k}}]$ and is denoted by $\mu_k$.

# Definition 3.14

The **moment-generating function** $M_Y(t)$ for a random variable $Y$ is defined to be $M_Y(t) = E(e^{tY})$. We say that a moment-generating function for $Y$ exists if there exists a positive constant $b$ such that $M_Y(t)$ is finite for $|t| \leq b$.

# Theorem 3.12

If $M_Y(t)$ exists, then for any positive integer $k$,

$$\frac{d^k M_Y(t)}{dt^k}]_{t=0} = M_Y^{'}(0) = \mu_k^{'}.$$

In other words, if you find the kth derivative of $M_Y(t)$ with respect to $t$ and set $t = 0$, the result will be $\mu_k^{'}$.

# Note

$$e^{tY} = 1 + tY + \frac{t^2Y^2}{2!} + \frac{t^3Y^3}{3!} + ...$$
$$E[e^{tY}] = E[1 + tY + \frac{t^2Y^2}{2!} + \frac{t^3Y^3}{3!} + ...]$$
$$E[e^{tY}] = E[1] + tE(Y) + \frac{t^2}{2!}E(Y^2) + \frac{t^3}{3!}E(Y^3) + ...$$
$$\frac{d}{dt}E[e^{tY}] = E(Y) + \frac{2t}{2!}E(Y^2) + \frac{3t^2}{3!}E(Y^3) + ...$$

Therefore, $M^{'}(0) = E(Y)$.

# Exercise 3.145

If $Y$ has a Binomial distribution with $n$ trials and probability of success $p$, show that the moment-generating function for $Y$ is

$$M_Y(t) = (pe^t + q)^n, \text{ where } q = 1 - p.$$

# Proof

$$M_Y(t) = E[e^{tY}] = \sum_{y=0}^{n} e^{ty}\binom{n}{y}p^y(1-p)^{n-y}$$

$$= \sum_{y=0}^{n}\binom{n}{y}(pe^t)^y(1-p)^{n-y}$$

Note that the last equation can be rewritten as:
$M_Y(t) = [pe^t + (1-p)]^n$.

# Exercise 3.147

If $Y$ has a geometric distribution with probability of success $p$, show that the moment-generating function for $Y$ is

$$M_Y(t) = \frac{pe^t}{1 - qe^t}, \text{ where } q = 1 - p.$$

# Proof

Before we prove this result, we will find a formula that will be useful to us.

$S = 1 + r + r^2 + ... + r^{n-1}$

Now, we multiply the last equation by $r$.

$rS = r + r^2 + r^3 + ... + r^n$

Then, we compute the difference between these two equations.

$S - rS = 1 - r^n$

Solving for $S$, we have

$S = \frac{1-r^n}{1-r}$

Assuming that $|r| < 1$ and letting $n$ go to infinity

$$\lim_{n\to\infty} S = \sum_{n=0}^{\infty} r^n = \frac{1}{1-r}.$$

# Proof

$$M_Y(t) = E[e^{tY}] = \sum_{y=1}^{\infty} e^{ty}(1-p)^{y-1}p$$
$$= \sum_{y=1}^{\infty} e^{ty}e^{t}e^{-t}(1-p)^{y-1}p$$
$$= pe^t \sum_{y=1}^{\infty} e^{t(y-1)}(1-p)^{y-1}$$
$$= pe^t \sum_{y=1}^{\infty} [(1-p)e^t]^{y-1}$$

(let $n = y - 1$)

$$= pe^t \sum_{n=0}^{\infty} [(1-p)e^t]^n$$
$$= pe^t \sum_{n=0}^{\infty} [qe^t]^n$$

Using our result from the previous slide,
$M_Y(t) = \frac{pe^t}{1-qe^t}$.

# Exercise 3.155

Let $M_Y(t) = \frac{1}{6}e^t + \frac{2}{6}e^{2t} + \frac{3}{6}e^{3t}$. Find the following:

a. The distribution of $Y$.

b. $E(Y)$.

c. $V(Y)$.

# HW?

Suppose that $Y$ is a Binomial random variable based on $n$ trials with success probability $p$ and let $W = n - Y$.

a. Find the moment-generating function of $W$, $M_W(t)$.

b. Based on your answer to part a), what is the distribution of $W$?

# A few comments about MGFs

The moment-generating function possesses two important applications. First, if we can find $E(e^{tY})$, we can find any of the moments for $Y$.

The second application of a moment-generating function is to prove that a random variable possesses a particular probability distribution $p(y)$. If $M_Y(t)$ exists for a probability distribution $p(y)$, it is unique. Also, if the moment-generating functions, for two random variables $Y$ and $Z$ are equal, then $Y$ and $Z$ must have the same probability distribution. It follows that, if we can recognize the moment-generating function of a random variable $Y$ to be one associated with a specific distribution, then $Y$ must have that distribution.