Lab 1: Probability Theory

W203, Section 4: Statistics for Data Science

Brad Andersen - February 20, 2019

1. Meanwhile, at the Unfair Coin Factory...

You are given a bucket that contains 100 coins. 99 of these are fair coins, but one of them is a trick coin that always comes up heads. You select one coin from this bucket at random. Let T be the event that you select the trick coin. This means that $P(T) = 0.01$.

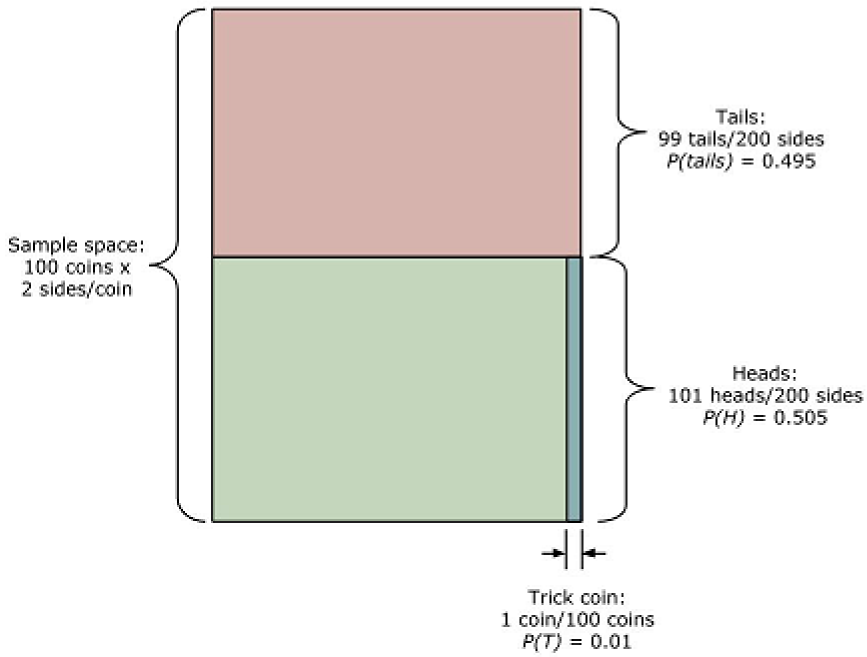
> a. Suppose you flip the coin once and it comes up heads. Call this event H_1. If this event occurs, what is the conditional probability that you have the trick coin? In other words, what is $P(T|H_1)$?

Given 100 coins whose individual events (*heads and tails* or, in the case of the trick coin, *heads and heads*) are mutually exclusive, there exist 200 individual outcomes from selecting one coin at random and flipping it. Knowing that the trick coin *always* reveals heads, of these 200 outcomes, there exist 99 tail events and 101 head events. The probability of one outcome being heads is as follows:

$$P(H_1) = \frac{101}{200}$$

$$= 0.505$$

An area diagram documenting the percentages of events is as follows:



Note that the area representing the intersection of heads events and the trick coin selection event is comprised of no tails events; the trick coin produces only heads events. Rather, it is equivalent to the number of trick coin heads events (2) within the entire event sample space of 200 potential outcomes. Therefore,

$$P(T \cap H_1) = \frac{2}{200}$$

$$= 0.01$$

Calculating the probability of selecting the trick coin conditional upon one coin flip revealing heads is as follows:

$$P(T|H_1) = \frac{P(T \cap H_1)}{P(H_1)}$$

$$= \frac{0.01}{0.505}$$

$$= 0.0198$$

Therefore, $P(T|H_1) = 0.0198$.

> b. Suppose instead that you flip the coin k times. Let H_k be the event that the coin comes up heads all k times. If you see this occur, what is the conditional probability that you have the trick coin? In other words, what is $P(T|H_k)$.

Let:

- $P(F)$ represent the probability of selecting a fair coin, calculated as $P(F) = 1 - P(T) = 0.99$.
- H_F represent the event of revealing a head when flipping a fair coin once.
- H_1 represent the event of revealing a head when flipping the trick coin once (as defined with the problem, above).
- N_F represent the event of revealing a tail when flipping a fair coin once.
- H_{Fk} represent the event of revealing k consecutive heads using a fair coin.
- H_k represent the event of revealing k consecutive heads using the trick coin.

Assume:

- Flipping a fair coin reveals a head or a tail with equal probability (i.e. $P(H_F) = 0.5$ and $P(N_F) = 1 - P(H_F) = 0.5$).

Using a fair coin, the probability of revealing a head with each flip decreases exponentially due to the multiplication rule. For example, the probability of revealing two consecutive heads can be calculated as follows:

$$\begin{aligned} P(H_{F2}) &= P(H_F) \cdot P(H_F) \\ &= 0.5 \cdot 0.5 \\ &= 0.5^2 \\ &= 0.25 \end{aligned}$$

The same calculation holds true for the flip of the trick coin, but the probability of revealing a head is always 1:

$$\begin{aligned} P(H_2) &= P(H_1) \cdot P(H_1) \\ &= 1 \cdot 1 \\ &= 1^2 \\ &= 1 \end{aligned}$$

The probability of having selected the one trick coin from the collection of 100 coins remains constant at $P(T) = 0.01$. Therefore, because the probabilty of revealing a head with the trick coin is always 1, the probability of revealing k consecutive heads remains constant:

$$\begin{aligned} P(T \cap H_k) &= P(T) \cdot P(H_1)^k \\ &= 0.01 \cdot 1^k \\ &= 0.01 \end{aligned}$$

The conditional probability of having selected the trick coin given consecutive revealed heads increases, as the conditional probability of having selected a fair coin given consecutive revealed heads decreases. Calculating the conditional probability of having selected the trick coin given k consecutive revealed heads revealed can be performed as follows:

$$P(T|H_k) = \frac{P(T \cap H_k)}{(P(H_F)^k \cdot P(F)) + (P(H_1)^k \cdot P(T))}$$

$$= \frac{0.01}{(0.5^k \cdot 0.99) + (1^k \cdot 0.01)}$$

$$= \frac{0.01}{(0.5^k \cdot 0.99) + 0.01}$$

Therefore, $P(T|H_k)$ can be calculated using the following equation:

$$\frac{0.01}{(0.5^k \cdot 0.99) + 0.01}$$

> c. How many heads in a row would you need to observe in order for the conditional probability that you have the trick coin to be higher than 99%?

The following R snippet calculates the number of consecutive head events required for the conditional probability that the trick coin was selected is greater than 99%:

In [12]:

```r
# Probability definitions
p_trick <- 0.01
p_fair <- 0.99
p_fairhead <- 0.5
p_trickhead <- 1

# Function calculating P(T|H_k), derived from answer to problem 1b
p_conditional <- function(k) {
  p_fairhead_k <- p_fairhead ^ k
  return(p_trick / ((p_fair * p_fairhead_k) + p_trick))
}

# Iterate through coin tosses, calculating the probability accordingly.
# Stop and print the number of tosses performed with the probability
# exceeds 99.
tosses <- 0
p_incremental <- 0
while (p_incremental <= 0.99) {
  tosses <- tosses + 1
  p_incremental <- p_conditional(tosses)
}
print(paste("Consecutive heads required for 99% probability:", tosses))
```

```
[1] "Consecutive heads required for 99% probability: 14"
```

Therefore, the number of consecutive head events required for the conditional probability that the trick coin was selected is greater than 99% is 14.

2. Wise Investments

You invest in two startup companies focused on data science. Thanks to your growing expertise in this area, each company will reach unicorn status (valued at $1 billion) with probability 3/4, independent of the other company. Let random variable X be the total number of companies that reach unicorn status. X can take on the values 0, 1, and 2. Note: X is what we call a binomial random variable with parameters $n = 2$ and $p = 3/4$.

> a. Give a complete expression for the probability mass function of X.

Given binomial random variable X with the following properties:

- Values x can assume: 0, 1 and 2
- n (number of trials): 2
- p (probability): 0.75

Let:

- F represent trial failures
- S represent trial successes

To compute the probability mass function for binomial random variables X, we will use the following theorem:

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Knowing the properties of X and the means with which its pmf can be calculated, probabilities associated with the results of all possible trials are represented as follows:

Outcome	x	Probability Equation	Probability
FF	0	$(1-p)^2$	0.0625
FS	1	$p \cdot (1-p)$	0.1875
SF	1	$p \cdot (1-p)$	0.1875
SS	2	p^2	0.5625

The pmf for X is therefore represented as follows:

$$b(x; 2, 0.75) = \begin{cases} \binom{2}{x} 0.75^x \cdot (1-0.75)^{2-x} & x = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

> b. Give a complete expression for the cumulative probability function of X.

Substituting probability values calculated above, the cumulative probability function of X is therefore represented as follows:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.0625 & 0 \leq x < 1 \\ 0.4375 & 1 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

c. Compute $E(X)$.

Substituting probability values calculated above, the expected value of X can be calculated as follows:

$$\begin{aligned} E(X) &= \sum_{x \in [0,1,2]} x \cdot p(x) \\ &= 0 \cdot p(0.0625) + 1 \cdot p(0.375) + 2 \cdot p(0.5625) \\ &= 0 + 0.375 + 1.125 \\ &= 1.5 \end{aligned}$$

Results are the same when the expected value of X is calculated without summation:

$$\begin{aligned} E(X) &= n \cdot p \\ &= 2 \cdot 0.75 \\ &= 1.5 \end{aligned}$$

The expected value of X (i.e. $E(X)$) is 1.5.

d. Compute $var(X)$.

Substituting probability values calculated above, the variance of X can be calculated as follows:

$$\begin{aligned} var(X) &= \sum_{x \in [0,1,2]} (x - \mu)^2 \cdot p(x) \\ &= (0 - 1.5)^2 \cdot 0.0625 + (1 - 1.5)^2 \cdot 0.375 + (2 - 1.5)^2 \cdot 0.5625 \\ &= 2.25 \cdot 0.0625 + 0.25 \cdot 0.375 + 0.25 \cdot 0.5625 \\ &= 0.375 \end{aligned}$$

Results are the same when the variance of X is calculated without summation:

$$\begin{aligned} var(X) &= np(1-p) \\ &= 1.5 \cdot 0.25 \\ &= 0.375 \end{aligned}$$

The variance of X (i.e. $var(X)$) is 0.375.

3. A Really Bad Darts Player

Let X and Y be independent uniform random variables on the interval $[-1, 1]$. Let D be a random variable that indicates if (X, Y) falls within the unit circle centered at the origin. We can define D as follows:

$$D = \begin{cases} 1, & X^2 + Y^2 < 1 \\ 0, & otherwise \end{cases}$$

Note that D is a Bernoulli variable.

> a. Compute the expectation $E(D)$. Hint: it might help to remember why we use area diagrams to represent probabilites.

Probability density functions for random variables X and Y

Given that random variables X and Y are continuous and uniform on the interval $[-1, 1]$, their probability density functions (pdf) can be represented as follows:

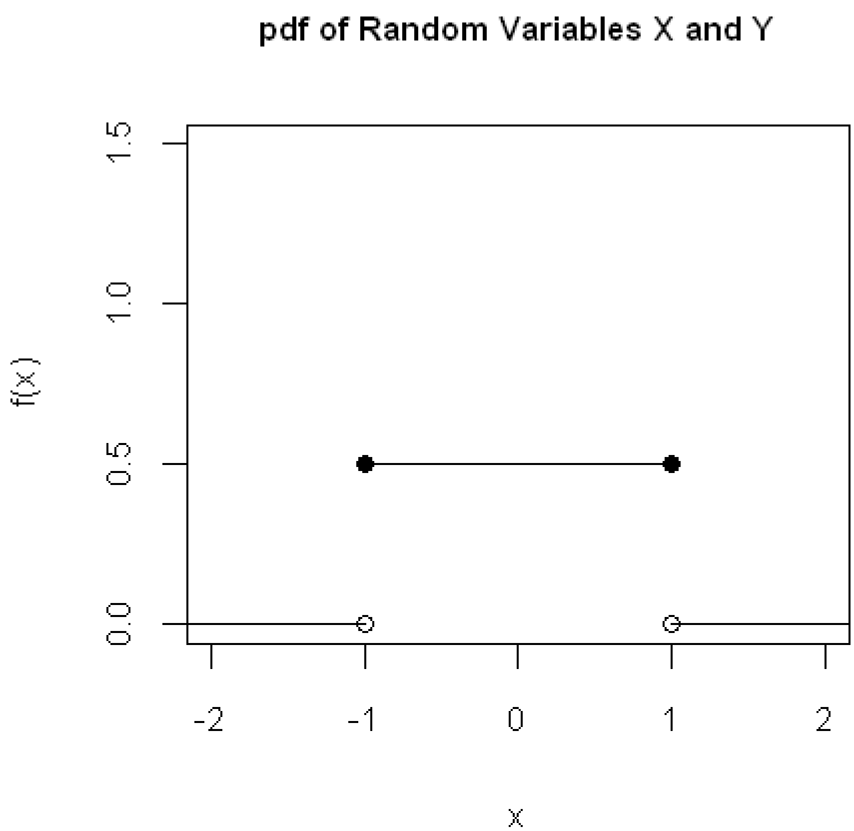
$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & otherwise \end{cases}$$

Substituting the bounds of the interval on which X and Y are defined for A and B, their pdf and graphical representation can be defined as follows (note that because both X and Y share the same properties, only the pdf for X is shown):

$$f(x; -1, 1) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

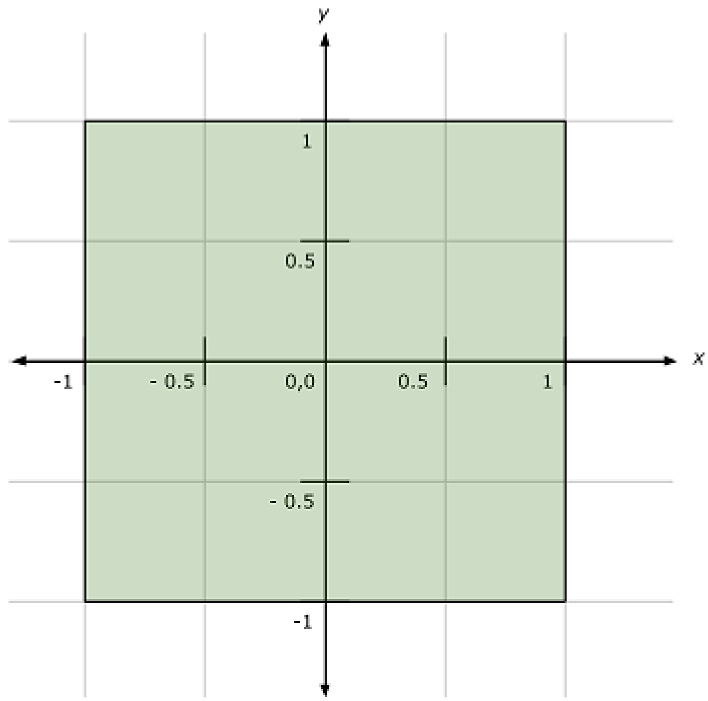
In [13]:

```
options(repr.plot.width=4, repr.plot.height=4)
par(ps = 10, cex = 1, cex.main = 1)
a <- -1; b <- 1; min_x <- a - 1; max_x <- b + 1; prob_x <- 1 / (b - a);
knots <- c(a, b); heights <- c(0, prob_x, 0);
plot(
    stepfun(knots, heights),
    verticals = FALSE,
    xlab="x",
    ylab="f(x)",
    xlim = c(-2, 2),
    ylim = c(0, 1.5),
    do.points = FALSE,
    main = "pdf of Random Variables X and Y"
)
points(c(a, b), c(0, 0), pch = 1)
points(c(a, b), c(prob_x, prob_x), pch = 19)
```



Sample space and "target" geometries

The intervals of random variables X and Y define a sample space that is square, has a width and length of 2, and is centered on the origin $(0, 0)$. On a Cartesian plane, this sample space is represented as follows:

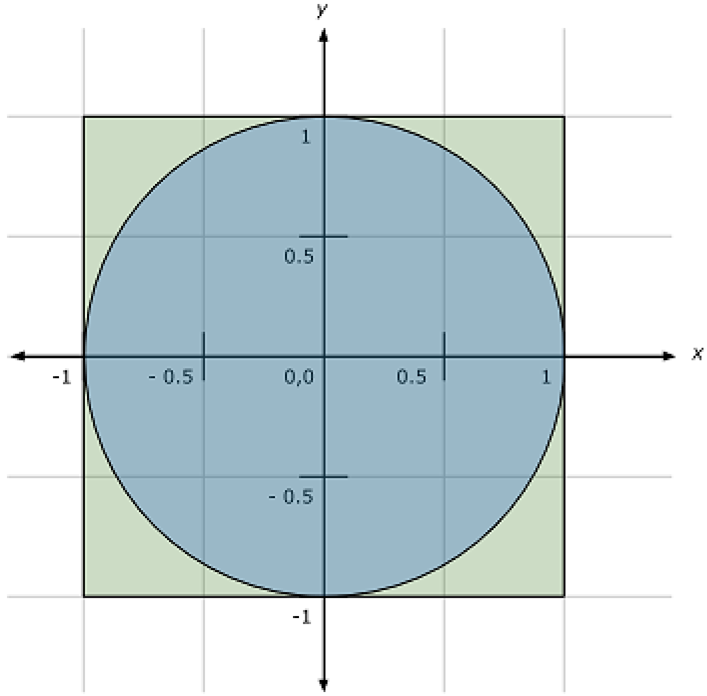


Values of these random variables can be represented as (x, y) coordinates on this plane.

The "target" area for which we are calculating probabilities is a unit circle, also centered on the origin $(0, 0)$ of a Cartesian plane. D, a random Bernoulli variable, is defined with its value being assigned 1 when $X^2 + Y^2 < 1$. Given the Pythagorean theorem, with adjacent and opposite lengths of a right triangle represented by a and b:

$$a^2 + b^2 = c^2$$

a likeness with the definition of D is apparent; random variables X and Y can be associated with the legs of a right triangle whose hypotenuse is less than 1. On the Cartesian plane, points where $x^2 + y^2 = 1$ can be visualized as a circle whose radius is the maximum length of the hypotenuse, or 1:



The areas of the square representing the sample space, and the circle representing the area of the sample space when $D = 1$ can be calculated as follows:

$$\begin{aligned}\text{Sample space area} &= width \times height \\ &= 2 \times 2 \\ &= 4 \text{ square units}\end{aligned}$$

$$\begin{aligned}\text{Target area} &= \pi \times radius^2 \\ &= \pi \times 1^2 \\ &= \pi \text{ square units}\end{aligned}$$

The ratio of target area to sample space area is $\frac{\pi}{4}$ or approximately 0.7854. This value represents the probability that the (x, y) coordinates will fall within the area of the target circle. It also represents the probability that the value of random variable D will equal 1.

The expected value of a Bernoulli random variable is equal to the probability that the random variable will assume the value of 1.

Therefore, the expected value of D is as follows:

$$\begin{aligned}E(D) &= p \\ &= 0.7854\end{aligned}$$

b. Compute the standard deviation of D.

The variance of D can be calculated as follows:

$$\sigma_D^2 = p \cdot (1 - p)$$
$$= 0.7854 \cdot (1 - 0.7854)$$
$$= 0.7854 \cdot 0.2146$$
$$= 0.1685$$

The standard deviation is the square root of the variance.

Therefore, the standard deviation of D is as follows:

$$\sigma_D = \sqrt{p \cdot (1 - p)}$$
$$= \sqrt{0.7854 \cdot (1 - 0.7854)}$$
$$= \sqrt{0.7854 \cdot 0.2146}$$
$$= \sqrt{0.1685}$$
$$= 0.4105$$

> c. Write an R function to compute the value of D, given a value for X and a value for Y. Use R to simulate a draw for X and a draw for Y, then compute the value of D.

In [14]:

```r
# Returns the value of random variable D given x and y coordinates
# in the range [-1:1].
# Note that no validation of input coordinate is performed.
compute_D <- function(coordinate) {
  if (coordinate[1] ^ 2 + coordinate[2] ^ 2 < 1) {
    return(1)
  }
  return(0)
}

# Compute the value of D for x and y values retrieved at random from
# a uniform distribution
print(compute_D(runif(2, -1, 1)))
```

```
[1] 1
```

> d. Use R to simulate the previous experiment 1000 times, resulting in 1000 samples for D. Compute the sample mean and sample standard deviation of your result, and compare them to the true values in parts a. and b.

In [15]:

```
# Define the expected value and standard deviation of random variable
# D from computations in Lab 1 (see question 4, a and b)
D_mean_lab <- 0.7854
D_stdv_lab <- 0.4105

# Invoke compute_D 1,000 times, and retain the values (i.e. 0 or 1)
ones <- 0; zeros <- 0;
for (i in c(1:1000)) {

  D_current <- compute_D((runif(2, -1, 1)))
  if (D_current == 0) {
    zeros <- zeros + 1
  }
  if (D_current == 1) {
    ones <- ones + 1
  }
}

# Compute the expected value and standard deviation from the 1,000
# invocations of the experiment
D_mean_computed <- ones * 0.001
D_stdv_computed <- sqrt(D_mean_computed * (1 - D_mean_computed))

# Print the laboratory and computed expected values and standard
# deviations of D
sprintf("rv D - Expected values: Lab 1: %1.4f, Computed: %1.4f", D_mean_lab, D
sprintf("rv D - Standard deviations: Lab 1: %1.4f, Computed: %1.4f", D_stdv_la
```

'rv D - Expected values: Lab 1: 0.7854, Computed: 0.7810'

'rv D - Standard deviations: Lab 1: 0.4105, Computed: 0.4136'

4. Relating Min and Max

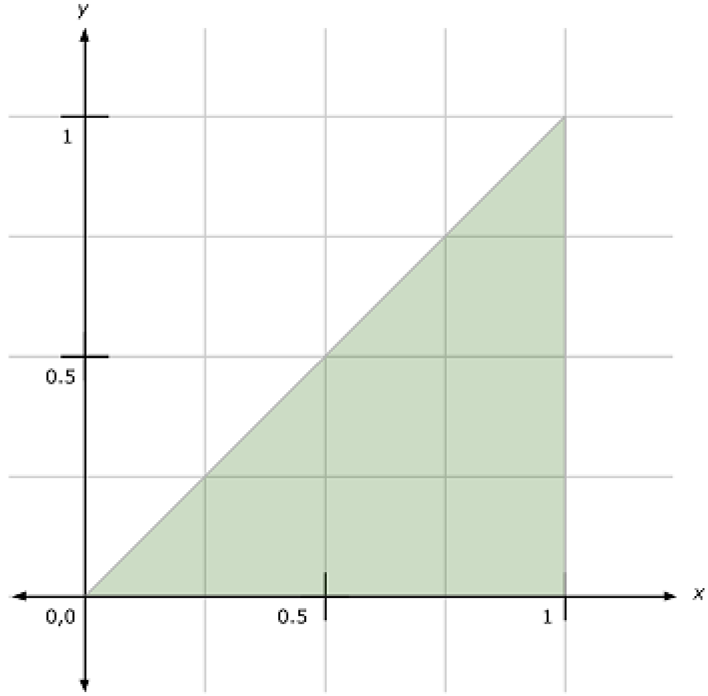
Continuous random variables X and Y have a joint distribution with probability density function,

$$f(x, y) = \begin{cases} 2, & 0 < y < x < 1 \\ 0, & otherwise. \end{cases}$$

You may wonder where you would find such a distribution. In fact, if A_1 and A_2 are independent random variables uniformly distributed on $[0, 1]$, and you define $X = max(A_1, A_2)$, $Y = min(A_1, A_2)$, then X and Y will have exactly the joint distribution defined above.

> a. Draw a graph of the region for which X and Y have positive probability density.

Continuous random variables X and Y have positive probability density as represented by the following graph:



where $0 < y < x < 1$.

b. Derive the marginal probability density function of $X, f_X(x)$. Make sure you write down a complete expression.

The marginal probability density function of $X, f_X(x)$ may be derived as follows:

$$
\begin{aligned}
g(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\
&= \int_{0}^{x} f(x, y)\, dy \\
&= \int_{0}^{x} 2\, dy \\
&= \left[\, 2y \,\right]_0^x \\
&= 2(x) - 2(0) \\
&= 2x
\end{aligned}
$$

The marginal probability density function of $X, f_X(x)$ ***is*** $2x$ ***when*** $0 < y < x < 1$.

c. Derive the unconditional expectation of X.

The unconditional expectation of X may be derived as follows:

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\, dx \\
&= \int_{0}^{1} x \cdot 2\, dx \\
&= \int_{0}^{1} 2x\, dx \\
&= \left[x^2 \right]_0^1 \\
&= 1^2 - 0^2 \\
&= 1
\end{aligned}
$$

The unconditional expectation of X is 1 when $0 < y < x < 1$.

> d. Derive the conditional probability density function of Y, conditional on X, $f_{Y|X}(y|x)$

Using values derived above, the conditional probability density function of Y, conditional on X, $f_{Y|X}(y|x)$ may be derived as follows:

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{2}{2x} \\
&= \frac{1}{x} \quad \text{when } 0 < y < x < 1
\end{aligned}
$$

The conditional probability density function of Y, conditional on X, $f_{Y|X}(y|x)$ is $\frac{1}{x}$ when $0 < y < x < 1$.

> e. Derive the conditional expectation of Y, conditional on X, $E(Y|X)$.

Using values derived above, the conditional expectation of Y, conditional on X, $E(Y|X)$ may be derived as follows:

$$
\begin{aligned}
E(Y|X) &= \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\;dy \\
&= \int_{0}^{x} y \cdot \frac{1}{x}\;dy \\
&= \int_{0}^{x} \frac{y}{x}\;dy \\
&= \left[\frac{y^2}{2x} \right]_0^x \\
&= \frac{x}{2} - 0 \\
&= \frac{x}{2}
\end{aligned}
$$

The conditional expectation of Y, conditional on X, $E(Y|X)$ is $\frac{x}{2}$ when $0 < y < x < 1$.

> f. Derive $E(XY)$. Hint 1: Use the law of iterated expectations. Hint 2: If you take an expectation conditional on X, X is just a constant inside the expectation. This means that $E(XY|X) = XE(Y|X)$.

```
$E(XY)$ can be derived by integrating $xy$ on probability density function $f(x, y)$ as follows:

\begin{equation}
\begin{split}
E(XY) & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy \cdot f(x, y)\;dx\;dy \\ \\
& = \int_{0}^{x}\int_{y}^{1}xy \cdot f(x, y)\;dx\;dy \\ \\
& = \int_{0}^{x}\int_{y}^{1}xy \cdot 2\;dx\;dy \\ \\
\end{split}
\end{equation}

Calculating the inner integral:

\begin{equation}
\begin{split}
& = \int_{y}^{1}2xy\;dx \\ \\
& = \begin{bmatrix}x^2y\end{bmatrix}_y^1 \\ \\
& = y - y^3 \\ \\
\end{split}
\end{equation}

Continuing, calculating the outer integral:
```

```

\begin{equation}
\begin{split}
E(XY) & = \int_{0}^{x}y - y^3\;dy \\ \\
& = \begin{bmatrix}\dfrac{y^2}{2} - \dfrac{y^4}{4}\end{bmatrix}_0^x \\ \\
& = \dfrac{x^2}{2} - \dfrac{x^4}{4}
\end{split}
\end{equation}

##### Expectation $E(XY)$ is $\dfrac{x^2}{2} - \dfrac{x^4}{4}$ when $0 < y < x < 1$.
```

> g. Using the previous parts, derive $cov(X, Y)$

Using values derived above in association with the simplified formula for calculating covariances, $cov(X, Y)$ can be derived as follows:

$$
\begin{split}
cov(X, Y) &= E(XY) - \mu_x \cdot \mu_y \\
&= \frac{x^2}{2} - \frac{x^4}{4} - 1 \cdot \frac{x}{2} \\
\\
&= \frac{x^2}{2} - \frac{x^4}{4} - \frac{x}{2} \\
\\
&= \frac{x^2 - x}{2} - \frac{x^4}{4}
\end{split}
$$

The covariance of random variables X and Y, $cov(X, Y)$ ***is*** $\dfrac{x^2 - x}{2} - \dfrac{x^4}{4}$.

In [ ]:

```

```