Lab 1: Probability Theory

W203: Statistics for Data Science

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 $P(T|H_1)$?

> Let T and $!T$ denote the event that coin is trick coin and not a trick coin (i.e. a fair coin), respectively.
>
> From the problem statement we know that

$$\begin{aligned} P(T) &= 0.01 \\ P(!T) &= 1 - P(T) = 0.99 \\ P(H_1|T) &= 1 \\ P(H_1|!T) &= 0.5 \end{aligned}$$

> First let's calculate the probability of heads

$$\begin{aligned} P(H_1) &= P(H_1|T)\,P(T) + P(H_1|!T)\,P(!T) \\ &= (1)\,(0.01) + (0.5)\,(0.99) \\ &= 0.505 \end{aligned}$$

> Now, using Bayes Theorem let's calculate

$$\begin{aligned} P(T|H_1) &= \frac{P(H_1|T)\,P(T)}{P(H_1)} \\ &= \frac{(1)\,(0.01)}{0.505} \\ &= 0.01980 \end{aligned}$$

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 $P(T|H_k)$.

> The probability of obtaining H_K for each coin type can be expressed as

$$\begin{aligned} P(H_k|T) &= 1 \\ P(H_k|!T) &= P(H_1|!T)^k \end{aligned}$$

> Using this, we can expand the full equation for the conditional probability of H_k given T

$$\begin{aligned} P(T|H_k) &= \frac{P(H_k|T)\,P(T)}{P(H_k)} \\ &= \frac{P(H_k|T)\,P(T)}{P(H_k|T)\,P(T) + P(H_k|!T)\,P(!T)} \\ &= \frac{P(H_k|T)\,P(T)}{P(H_k|T)\,P(T) + P(H_1|!T)^k\,P(!T)} \\ &= \frac{(1)\,(0.01)}{(1)\,(0.01) + (0.5^k)\,(0.99)} \\ &= \frac{0.01}{0.01 + (0.5^k)\,(0.99)} \end{aligned}$$

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%?

$$\begin{aligned} P(T|H_k) &= \frac{P(H_k|T)\,P(T)}{P(H_k|T)\,P(T) + P(H_1|!T)^k\,P(!T)} \\ &= \frac{(1)\,(0.01)}{(1)\,(0.01) + (0.5^{14})\,(0.99)} \\ &= 0.99399 \end{aligned}$$

> The coin would have to come up heads 14 times consecutively in order for the conditional probability to be higher than 99%.

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.

> The probability mass function (pmf) can be expressed as

$$
\begin{aligned}
p(0) &= p(0) = P(X = 0) = 1 - p = 1 - 3/4 = 1/4 \\
p(1) &= p(0) = P(X = 1) = p = 3/4
\end{aligned}
$$

> The entire pmf

$$
p(x) = \begin{cases} 1/4, & x = 0 \\ 3/4, & x = 1 \end{cases}
$$

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

> The cumulative distribution function (cdf) can be calculated as

$$
\begin{aligned}
p(0) &= p(X = 0) = P(FF) = (1 - p)^2 \\
p(1) &= p(X = 1) = P(FS) + P(SF) = 2(1 - p)\,p \\
p(2) &= p(X = 2) = P(SS) = p^2
\end{aligned}
$$

$$
\begin{aligned}
p(0) &= (1 - p)^2 = (1 - 3/4)^2 = 1/16 \\
p(1) &= 2(1 - p)\,p = 2(1 - 3/4)\,3/4 = 6/16 \\
p(2) &= p^2 = 3/4^2 = 9/16
\end{aligned}
$$

> The entire cdf

$$
F(x) = \begin{cases} 1/16, & x \le 0 \\ 7/16, & 0 < x < 2 \\ 1, & 2 \le x \end{cases}
$$

c) Compute $E(X)$.

$$\begin{aligned} E(X) &= \sum x \cdot p(x) \\ &= 0 * 1/16 + 1 * 6/16 + 2 * 9/16 \\ &= 0 + 6/16 + 18/16 \\ &= 1.5 \end{aligned}$$

d) Compute $var(X)$.

First let's compute $E(X^2)$

$$\begin{aligned} E(X^2) &= \sum x \cdot p(x) \\ &= 0^2 * 1/16 + 1^2 * 6/16 + 2^2 * 9/16 \\ &= 0 + 6/16 + 36/16 \\ &= 2.625 \end{aligned}$$

Now let's calculate the variance of X

$$\begin{aligned} V(X) &= \sum x^2 \cdot p(x) - \mu^2 \\ &= E(X^2) - E(X)^2 \\ &= 2.625 - (1.5)^2 \\ &= 0.375 \end{aligned}$$

3. Relating Min and Max

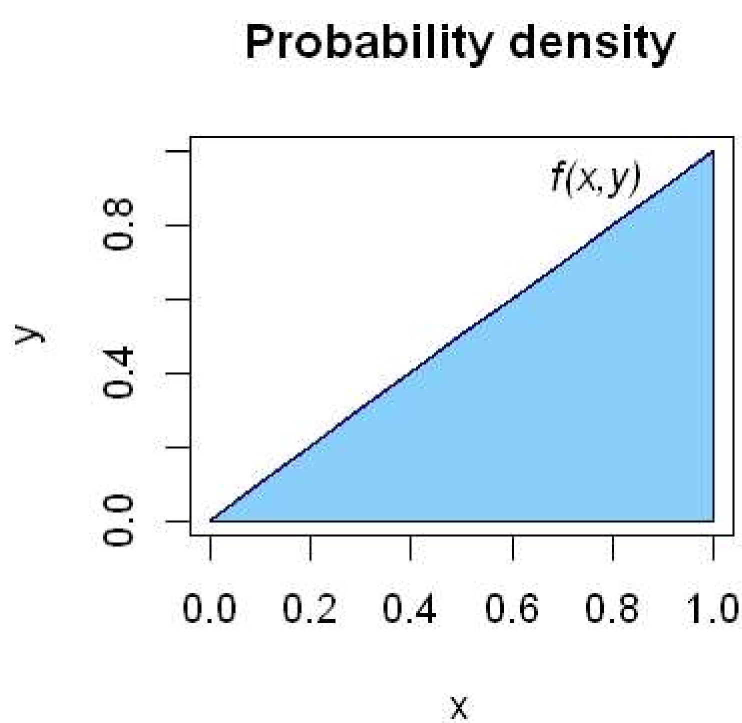
Continuous random variables X and Y have a joint distribution with probability density function,

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

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.

In [23]:
```
# Resize the plots
ratio = 1
width = 3.5
options(repr.plot.width=width, repr.plot.height=width*ratio)

# Options
step_size <- 0.1

# Generate intervals
x <- seq(0, 1, step_size)
y <- x

# Create plot
plot(x, y, main = 'Probability density',
     col='blue',
     lwd=2, type="l",
     xlab='x', ylab='y',
     xlim=c(0, max(x)))

# Shaded area
polygon(cbind(c(min(x), x, max(x)), c(min(y), y, min(y))), col="lightskyblue")

# Legend
legend('topright',
       inset=0.05,
       cex = 1,
       c('f(x,y)', '               '),
       horiz=TRUE, bty='n',
       col=c('blue'),
       text.font=3)
```



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 (pdf) of x can be calculated as

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x,y)\,dy \\ &= \int_0^x f(x,y)\,dy \\ &= \int_0^x 2\,dy \\ &= 2y\Big|_0^x \\ &= 2x - 2(0) \\ &= 2x \end{aligned}$$

> The entire marginal pdf of x

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

c) Derive the unconditional expectation of X.

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x)\,dx \\ &= \int_0^1 x \cdot 2x\,dx \\ &= \int_0^1 2x^2\,dx \\ &= \frac{2x^3}{3}\Big|_0^1 \\ &= \frac{2(1^3)}{3} - \frac{2(0^3)}{3}\Big|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

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

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f(x,y)}{f(x)} \\
&= \frac{2}{2x} \\
&= \frac{1}{x}
\end{aligned}
$$

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

$$
\begin{aligned}
E(Y|X) &= \int_{-\infty}^{+\infty} y \cdot f_{Y|X}(y|x)\, dy \\
&= \int_0^x y \cdot \frac{1}{x}\, dy \\
&= \frac{y^2}{2} \cdot \frac{1}{x} \Big|_0^x \\
&= \frac{y^2}{2x} \Big|_0^x \\
&= \frac{x^2}{2x} - \frac{0^2}{2x} \\
&= \frac{x}{2}
\end{aligned}
$$

f) Derive $E(XY)$. Hint: if you take an expectation conditional on X, X is just a constant inside the expectation. This means that $E(XY|X) = XE(Y|X)$.

$$
\begin{aligned}
E(XY) &= \int_0^1 E(Y|X) \cdot x \cdot f_X(x)\, dx \\
&= \int_0^1 \frac{x}{2} \cdot x \cdot 2x\, dx \\
&= \int_0^1 \frac{2x^3}{2}\, dx \\
&= \int_0^1 x^3\, dx \\
&= \frac{x^4}{4} \Big|_0^1 \\
&= \frac{1^4}{4} - \frac{0^4}{4} \\
&= \frac{1}{4}
\end{aligned}
$$

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

> First, let's calculate $E(Y)$ by multiplying $E(Y|X)$ with the marginal pdf of X

$$
\begin{aligned}
E(Y) &= \int_0^1 E(Y|X) \cdot f_X(x)\,dx \\
&= \int_0^1 \frac{x}{2} \cdot 2x\,dx \\
&= \int_0^1 x^2\,dx \\
&= \frac{x^3}{3}\Big|_0^1 \\
&= \frac{1^3}{3} - \frac{0^3}{3} \\
&= \frac{1}{3}
\end{aligned}
$$

> Now we can calculate the covariance of variables X and Y

$$
\begin{aligned}
cov(X, Y) &= E(XY) - E(X)E(Y) \\
&= \frac{1}{4} - \left(\frac{2}{3} \cdot \frac{1}{3}\right) \\
&= \frac{1}{4} - \frac{2}{9} \\
&= \frac{1}{36}
\end{aligned}
$$

4. Circles, Random Samples, and the Central Limit Theorem

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.

If variable D is a circle on the interval $[-1, 1]$ I believe we can calculate it's expected value as a ratio of the probability of D, with respect to the entire sample space. My rational for this approach stems from a continuation of how we'd calculate $E(D)$ in the discrete case.

Let $D = 1$ denote success, and $D = 0$ denote failure.

$$\begin{aligned}\text{Probability of success} &= \frac{\text{area of circle D}}{\text{area of sample space}} \\ &= \frac{\pi\, r^2}{s^2} \\ &= \frac{\pi\, 1^2}{2^2} \\ &= \frac{\pi}{4}\end{aligned}$$

The probability mass function (pmf) can be expressed as

$$\begin{aligned}p(1) &= p(D = 1) = p = \frac{\pi}{4} \\ p(0) &= p(D = 0) = 1 - p = 1 - \frac{\pi}{4}\end{aligned}$$

The expected value of D

$$\begin{aligned}E(D) &= \sum d \cdot p(d) \\ E(D) &= 1 * \frac{\pi}{4} + 0 * \left(1 - \frac{\pi}{4}\right) \\ &= \frac{\pi}{4}\end{aligned}$$

b) Compute the standard deviation of D.

First we must calculate $E(D^2)$

$$\begin{aligned} E(D^2) &= \sum d^2 \cdot p(d) \\ &= 1^2 * \frac{\pi}{4} + 0^2 * \left(1 - \frac{\pi}{4}\right) \\ &= \frac{\pi}{4} \end{aligned}$$

We can now calculate the variance of D

$$\begin{aligned} V(D) &= E(D^2) - E(D)^2 \\ &= \frac{\pi}{4} - \left(\frac{\pi}{4}\right)^2 \\ &= 0.16855 \end{aligned}$$

The standard deviation of D

$$\begin{aligned} \sigma &= \sqrt{V(D)} \\ &= \sqrt{0.16855} \\ &= 0.41054 \end{aligned}$$

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 [10]:
```
# Interval
int = seq(from = -1, to = 1, by = 0.001)

# Draws of x and y
x_draw = sample(int, 1)
y_draw = sample(int, 1)

# Calculate D
if(x_draw^2 + y_draw^2 < 1) {D = 1} else {D = 0}

# Print x, y and D
paste('x =', x_draw)
paste('y =', y_draw)
paste('D =', D)
```

'x = 0.579'

'y = -0.028'

'D = 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]:
```
# Function of variable D
d <- function(x, y) {
    if(x^2 + y^2 < 1) {1} else {0}
}

# Experiment simulated 1000 times
sim <- replicate(1000, d(sample(int, 1),sample(int, 1)))

paste('Mean', mean(sim))
paste('Standard Deviation', sd(sim))
```

'Mean 0.78'

'Standard Deviation 0.414453582167861'

The resulting mean and standard deviation from the simulation of the experiment are very similar to the true values in parts a and b. As the number of samples increases, the experimental mean and standard deviation approach the population mean and standard deviation, allowing us to begin making inferences about the wider population from our sample (providing n is sufficiently large).