New notation:

The *product* of two sets $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ is
$A \times B = \{(x, y) : x \in A, y \in B\}$.

For example $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}$. This is a fancy way of writing the x-y plane.

The set $(1, 2) \times [3, 4]$ is the box $\{(x, y) : 1 < x < 2, 3 \leq y \leq 4\}$. See the next slide.

The set $\{1, 2\} \times \{3, 4\}$ only has four elements
$\{(1, 3), (1, 4), (2, 3), (2, 4)\}$.

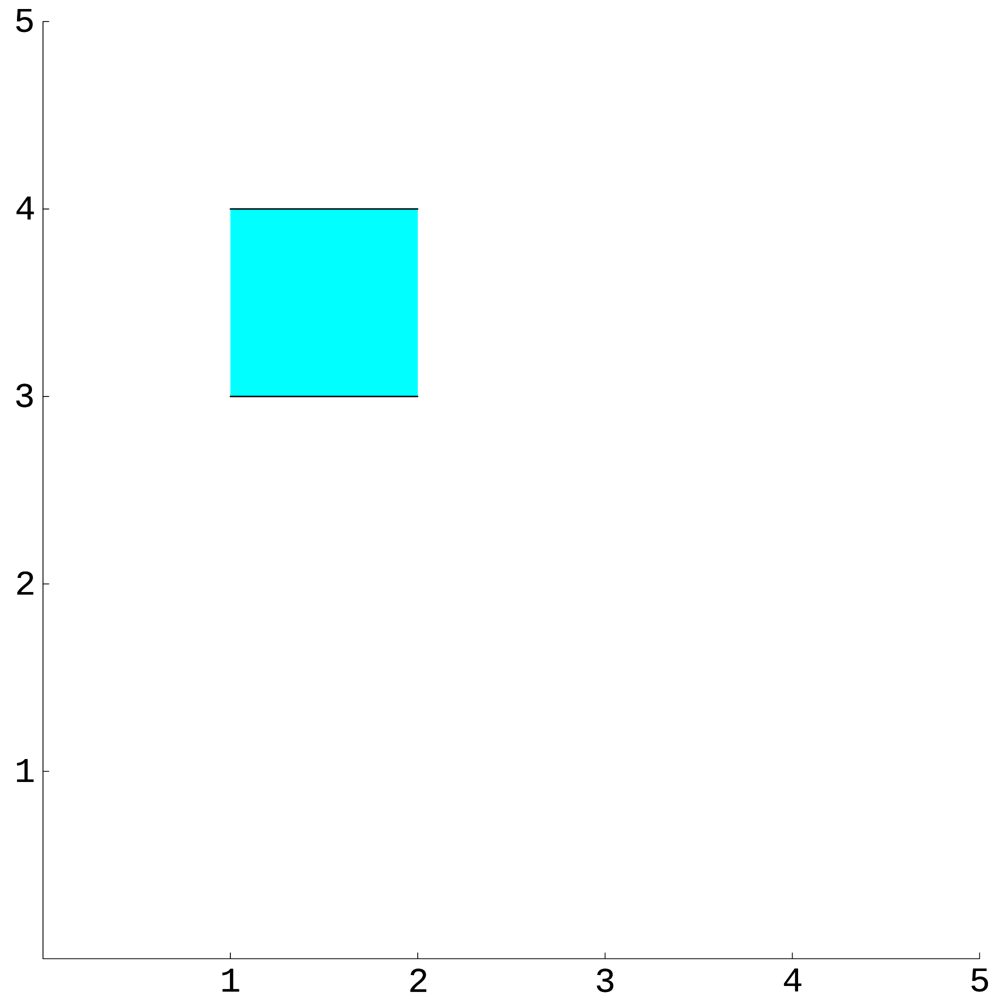

Figure 1: $(1, 2) \times [3, 4]$.

Chapter 3: Jointly Distributed Random Variables

We have considered a single random variable X, and possibly a function of this random variable, say $Y = g(X)$ so far.

If we *observe*, say, $X = 3.8$, then we know $Y = g(3.8) = 14.44$ if $g(x) = x^2$. In this case knowing X tells us everything about Y and vice-versa. We can think of the pair (X, Y) as two, perfectly related random variables.

Sometimes there is a relationship between X and Y that is not so deterministic. Perhaps X tends to be large when Y is large, but knowing X does not tell us *everything* about Y.

Randomly pick a U.S. adult. Let X be her years of post high school education and let Y be her salary. Knowing $X = 6$ years rather than $X = 1$ year tells us *something* about her salary Y, but not everything.

Sometimes there is *no* relationship between X and Y beyond knowing their probability distributions $f_X(x)$ and $f_Y(y)$. This idea we've already seen and termed *independence*.

Think of collecting data $X_1, X_2, \ldots, X_n$. Often we wish to collect data in such a way that we obtain a representative sample from the overall population of size n. We often assume data are collected *independently*.

If we further assume a probability model for the data X_i, say $X_i \sim \exp(\theta)$, then the object of *statistical* inference is to estimate θ using only the data $X_1, \ldots, X_n$ and then answer questions of interest involving θ, for example for any $X \sim \exp(\theta)$, what is $P(X > 100) = e^{-\theta 100}$?

3.2 Discrete jointly distributed random variables

We first start with two random variables X and Y that are both defined on the same sample space.

Example: Roll a 6-sided die. The sample space for the experiment is $S = \{1, 2, 3, 4, 5, 6\}$. Let $X = 0$ if a roll is an odd number and $X = 1$ if a roll is even. Let $Y = 1$ if a roll is a prime number (i.e. one of $\{2, 3, 5\}$) and $Y = 0$ if the roll is not a prime number.

Example: A subject from the population of all individuals exposed basement radon is randomly selected. Let X be the measured radon level in the individual's home in picoCuries per liter, let Y be the time exposed in months.

Example: Two people enter a gym independently and their times in minutes spent running on a treadmill X and Y are recorded.

def'n A random pair (X, Y) is discrete if there exists a probability mass function (pmf) $p(x, y) \geq 0$ that is positive on a finite or countable number of values $R_{X,Y} = \{(x, y) : p(x, y) > 0\}$ such that for any set $A \subset \mathbb{R}^2$,

$$P((X, Y) \in A) = \sum_{A \cap R_{X,Y}} p(x, y).$$

The joint pmf of (X, Y) gives probabilities paired outcomes (x, y):

$$p(x, y) = P(X = x, Y = y) = P(X = x \text{ and } Y = y).$$

The range of (X, Y) are those pairs of outcomes that are possible

$$R_{X,Y} = \{(x, y) : p(x, y) > 0\}.$$

For discrete (X, Y) with a finite number of outcomes, a table is often used to define $p(x, y)$.

Text example: A fair coin is flipped three times, let X be the number of heads on the first toss and Y the number of heads out of the three tosses. We can summarize all possible outcomes of the *experiment* and hence all possible outcomes of (X, Y) in a table:

Outcomes	x	y	$p(x, y)$
ttt	0	0	1/8
htt	1	1	1/8
hth, *hht*	1	2	2/8
tth, *tht*	0	1	2/8
thh	0	2	1/8
hhh	1	3	1/8

$R_{X,Y}$ is listed in the second column of (x, y) pairs:
$R_{X,Y} = \{(0,0), (1,1), (1,2), (0,1), (0,2), (1,3)\}$.

Question? Is $R_{X,Y} = R_X \times R_Y$?

The possible values that X can take on are $R_X = \{0, 1\}$, for Y we have $R_Y = \{0, 1, 2, 3\}$.

So $R_X \times R_Y = \{(0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 2), (1, 3)\}$.

Nope, here we have $R_{X,Y} \subset R_X \times R_Y$ but not $R_{X,Y} = R_X \times R_Y$.

We will see later that this means X and Y cannot be *independent*. That is, knowing X tells us something about Y and vice-versa.

Probabilities for events $(X, Y) \in A$, where A is any subset of paired numbers $A \subset \mathbb{R}^2$, are computed by simply finding those (x, y) that are in A that can actually happen (i.e. also in $R_{X,Y}$) and adding up the corresponding $p(x, y)$'s.

For example

$$P(X = Y) = p(0,0) + p(1,1) = 2/8.$$

$$P(Y = X + 1) = p(0,1) + p(1,2) = 4/8.$$

In the first case $X = Y$ if and only if $(X, Y) \in A = \{(x, y) : x = y\}$. The event A is a line in $\mathbb{R}^2$. But those elements in A that can actually happen are $A \cap R_{X,Y} = \{(0,0), (1,1)\}$. Draw a picture. Don't let the notation get in the way of what's really going on.

What if we only care about probabilities for either X or Y, not (X, Y)?

We can formulate events concerning X or Y *only* in terms of events involving (X, Y) and use $p(x, y)$. For example if we want $P(X = 0)$, this corresponds to the set $A = \{(x, y) : x = 0, -\infty < y < \infty\}$. We need to find those pairs (x, y) in $R_{X,Y}$ that satisfy $x = 0$.

$$P(X = 0) = p(0, 0) + p(0, 1) + p(0, 2) = 1/8 + 2/8 + 1/8 = 4/8.$$

Similarly,

$$\begin{aligned} P(1 \leq Y \leq 2) &= P(1 \leq Y \leq 2 \text{ and } -\infty < X < \infty) \\ &= p(0, 1) + p(1, 1) + p(0, 2) + p(1, 2) \\ &= 2/8 + 1/8 + 1/8 + 2/8 = 6/8. \end{aligned}$$

In fact, we can find the *marginal distributions* as

$$p_X(j) = P(X = j) = \sum_{(j,y)\in R_{X,Y}} p(j, y),$$

and

$$p_Y(j) = P(Y = j) = \sum_{(x,j)\in R_{X,Y}} p(x, j).$$

For the previous example, $p_X(0) = 1/8 + 2/8 + 1/8 = 4/8$ and $p_X(1) = 1/8 + 2/8 + 1/8 = 4/8$. Similarly, $p_Y(0) = 1/8$, $p_Y(1) = 3/8$, $p_Y(2) = 3/8$, $p_Y(3) = 1/8$. Note that marginally $X \sim \text{Bern}(0.5)$ and $Y \sim \text{bin}(3, 0.5)$.

This is really an application of the law of total probability

$$P(X = j) = \sum_{y\in R_Y} P(X = j \cap Y = y).$$

3.3 Continuous jointly distributed random variables

Instead of a finite or countable number of outcomes in $R_{X,Y}$, continuous random pairs (X, Y) take on a continuum of values in some subset $R_{X,Y} \subset \mathbb{R}^2$.

def'n A random pair (X, Y) is continuous if there exists a probability density function (pdf) $f(x, y) \geq 0$ such that for any set $A \subset \mathbb{R}^2$,

$$P((X, Y) \in A) = \iint_A f(x, y) dx dy.$$

Conversely, $f(x, y)$ is a pdf for some random pair (X, Y) if

1. $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$, and
2. $\iint_{\mathbb{R}^2} f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

When finding the probabilities of events $(X, Y) \in A$, we only need to consider those paired outcomes that can actually happen,

$$P((X, Y) \in A) = P((X, Y) \in A \cap R_{X,Y}).$$

This is useful when finding domains to integrate $f(x, y)$ to get a desired probability.

Example: Let $f(x, y) = e^{-x-y}$ on $R_{X,Y} = \{(x, y) : x > 0, y > 0\}$. Find $P(X + Y \leq 1)$.

This probability is the same as $P((X, Y) \in A)$ where A is the half plane $A = \{(x, y) : x + y \leq 1\}$.

We need to find $A \cap R_{X,Y}$ to evaluate $\iint_{A \cap R_{X,Y}} f(x, y) dx dy$.

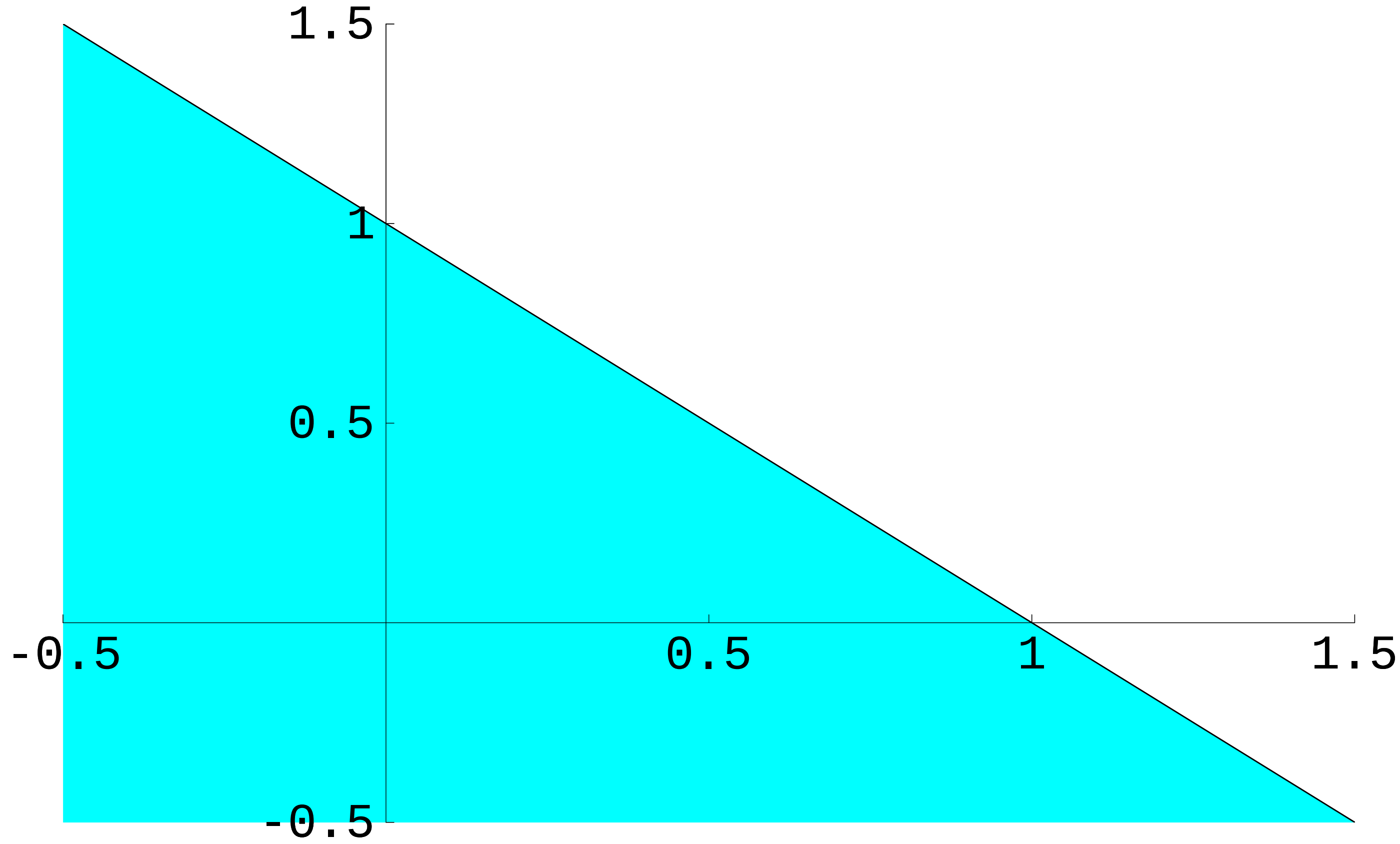

Figure 2: Set $A = \{(x, y) : x + y \leq 1\}$.

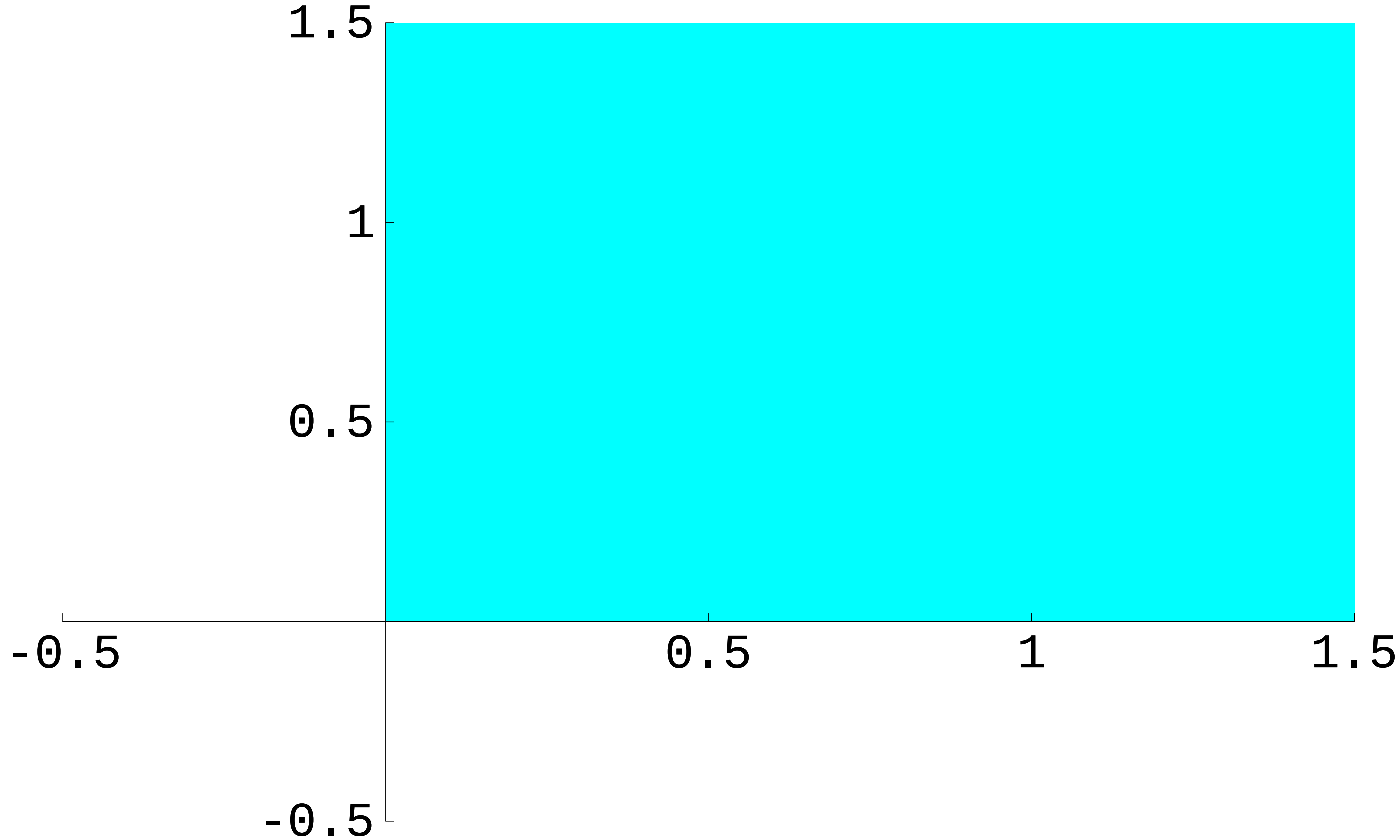

Figure 3: Set $R_{X,Y} = \{(x, y) : x > 0, y > 0\}$.

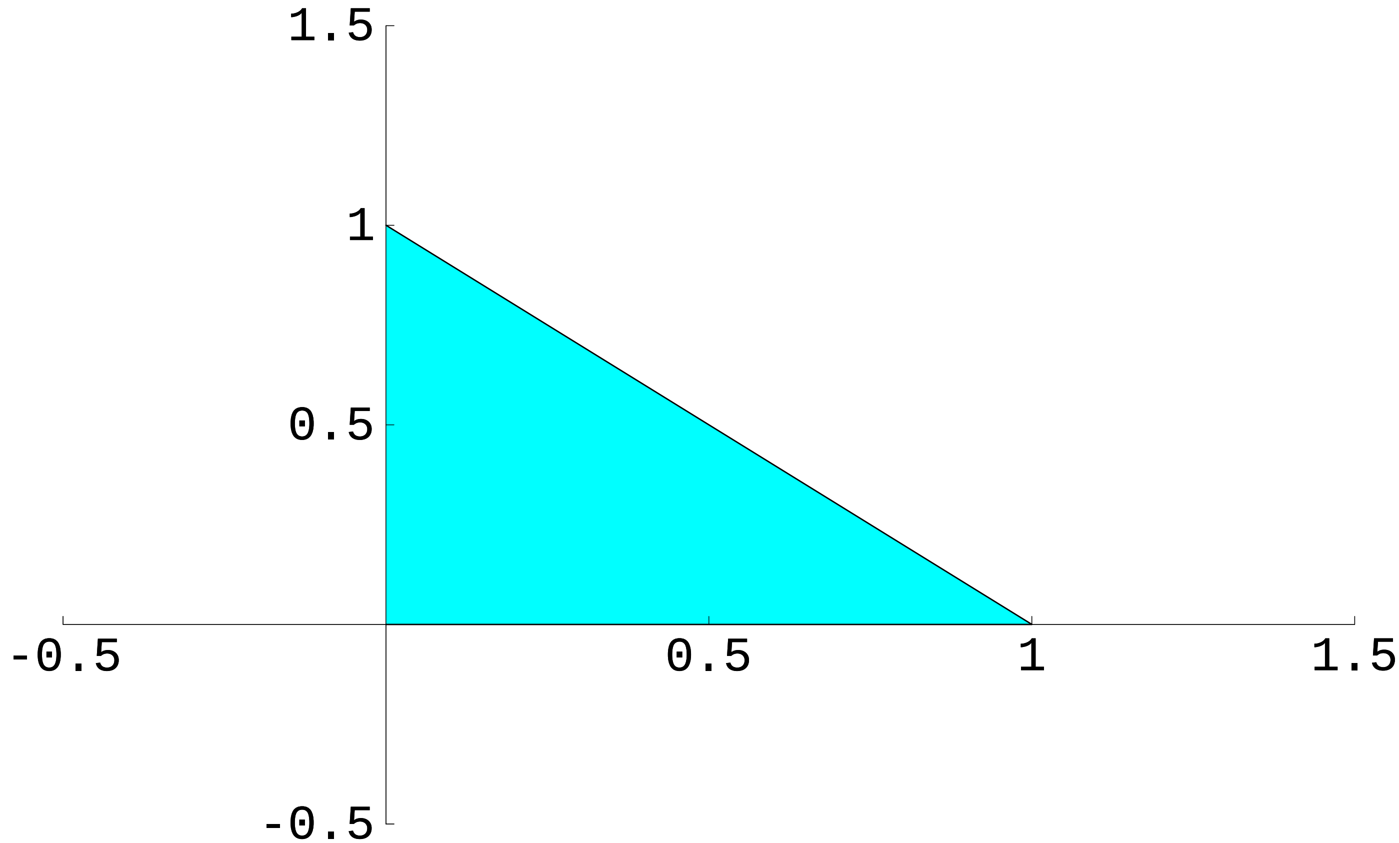

Figure 4: Set $A \cap R_{X,Y} = \{(x, y) : 0 < x < 1, 0 < y < 1 - x\}$.

$$
\begin{aligned}
P(X+Y \le 1) &= \int_0^1 \left[\int_0^{1-x} e^{-x-y} dy\right] dx \\
&= \int_0^1 e^{-x} \left[\int_0^{1-x} e^{-y} dy\right] dx \\
&= \int_0^1 e^{-x} \left[[-e^{-y}]_0^{1-x}\right] dx \\
&= \int_0^1 e^{-x}[1-e^{x-1}]dx \\
&= \int_0^1 [e^{-x}-e^{-1}]dx \\
&= [-e^{-x}-e^{-1}x]_0^1 \\
&= -e^{-1}-e^{-1}--e^0 \\
&= 1-\frac{2}{e} \approx 0.264.
\end{aligned}
$$

Let $R_{X,Y}$ be some region in $\mathbb{R}^2$ and let $|R_{X,Y}|$ denote the area of this region. The continuous random pair (X, Y) is said to be **uniform** over $R_{X,Y}$ if

$$f(x,y) = \left\{ \begin{array}{ll} \frac{1}{|R_{X,Y}|} & \text{when } (x,y) \in R_{X,Y} \\ 0 & \text{else} \end{array} \right\}.$$

Finding probabilities for unform random pairs reduces to finding areas. Let (X, Y) be uniform over $R_{X,Y}$. Then

$$P((X,Y) \in A) = \iint_{A \cap R_{X,Y}} f(x,y)dxdy = \frac{|A \cap R_{X,Y}|}{|R_{X,Y}|}.$$

Example: let (X, Y) be uniform over the unit disk $R_{X,Y} = \{(x,y) : x^2 + y^2 \leq 1\}$. Let's find $P(X > Y)$.

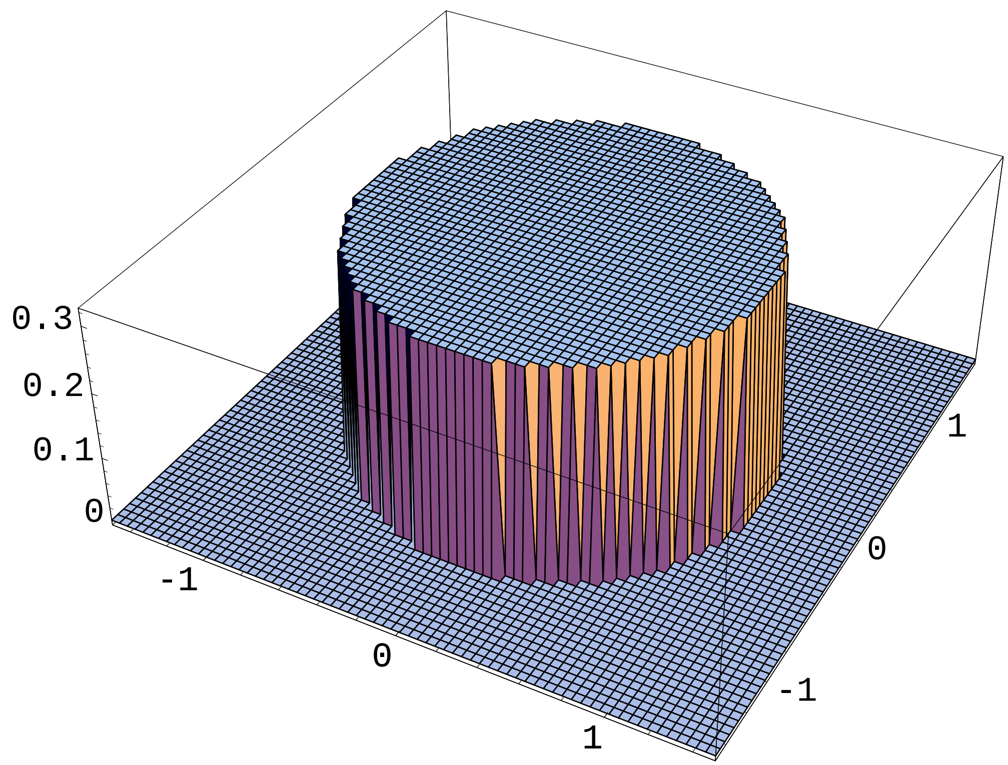

Figure 5: Density $f(x, y)$ for (X, Y) uniform over unit disk.

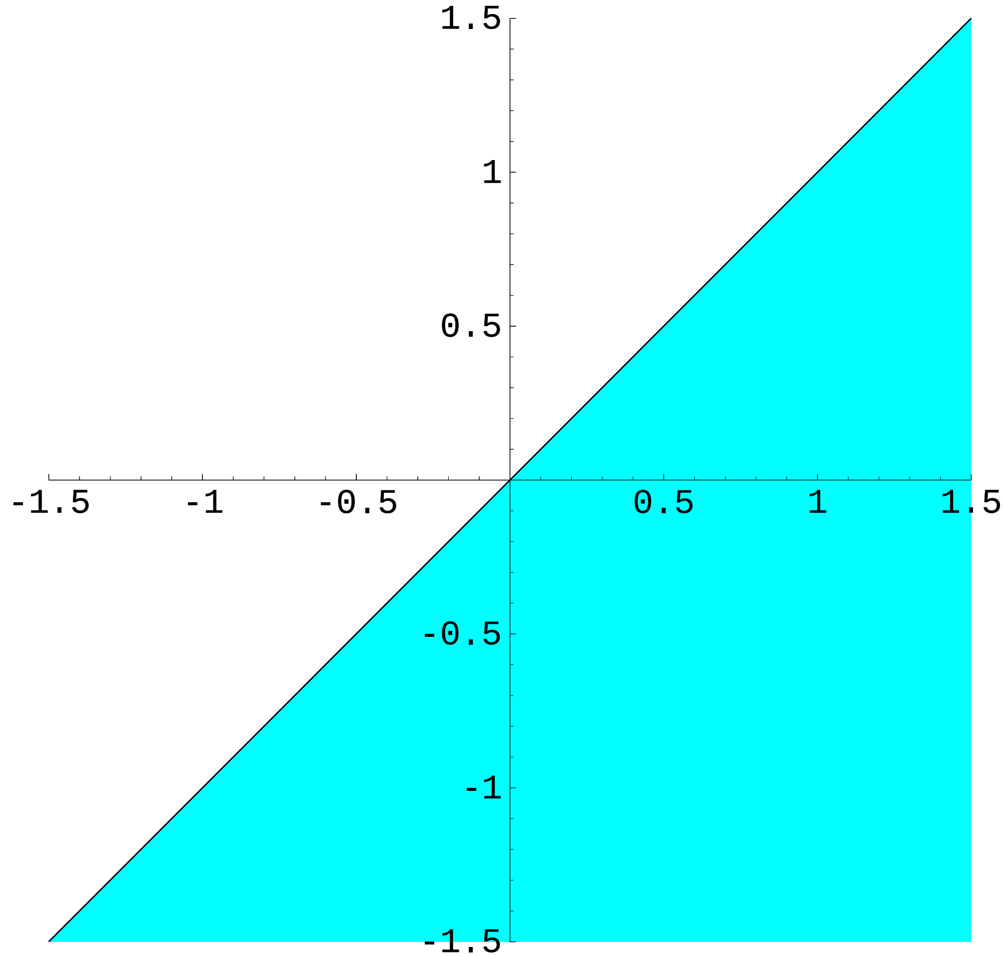

Figure 6: Set $A = \{(x, y) : x > y\}$.

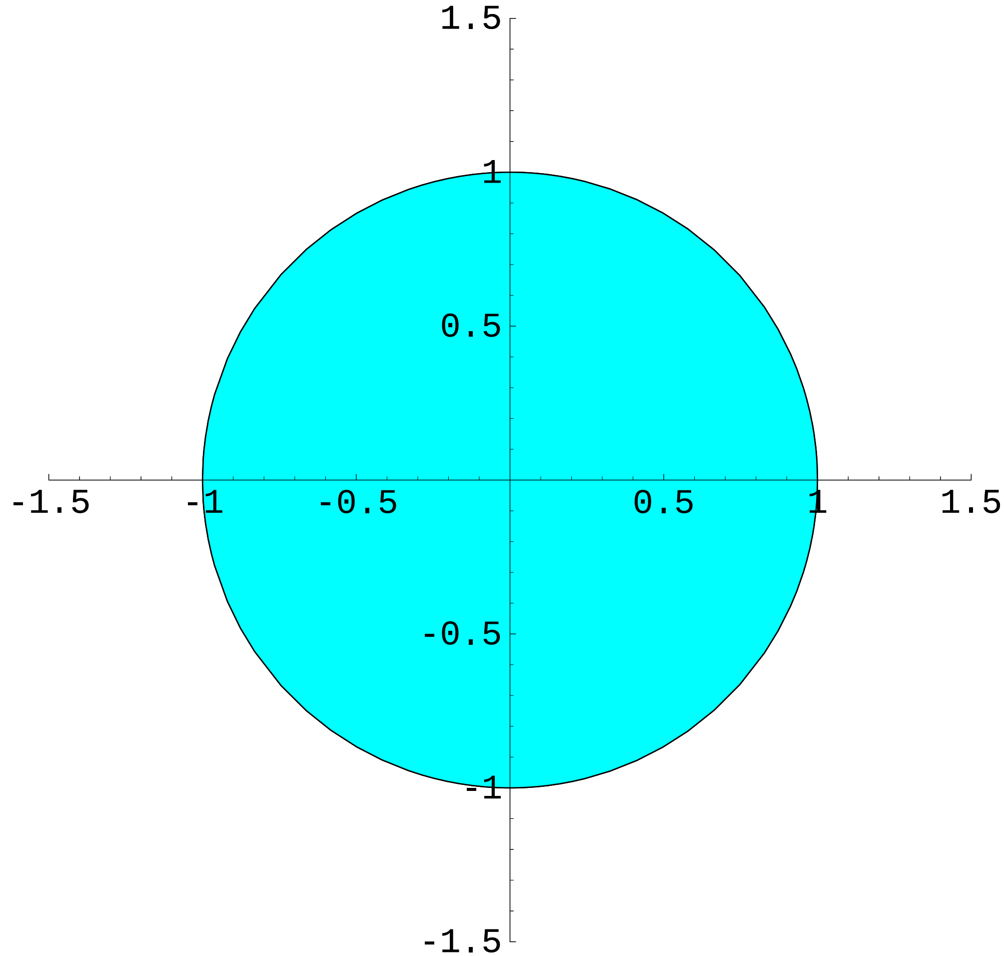

Figure 7: Set $R_{X,Y} = \{(x, y) : x^2 + y^2 \leq 1\}$.

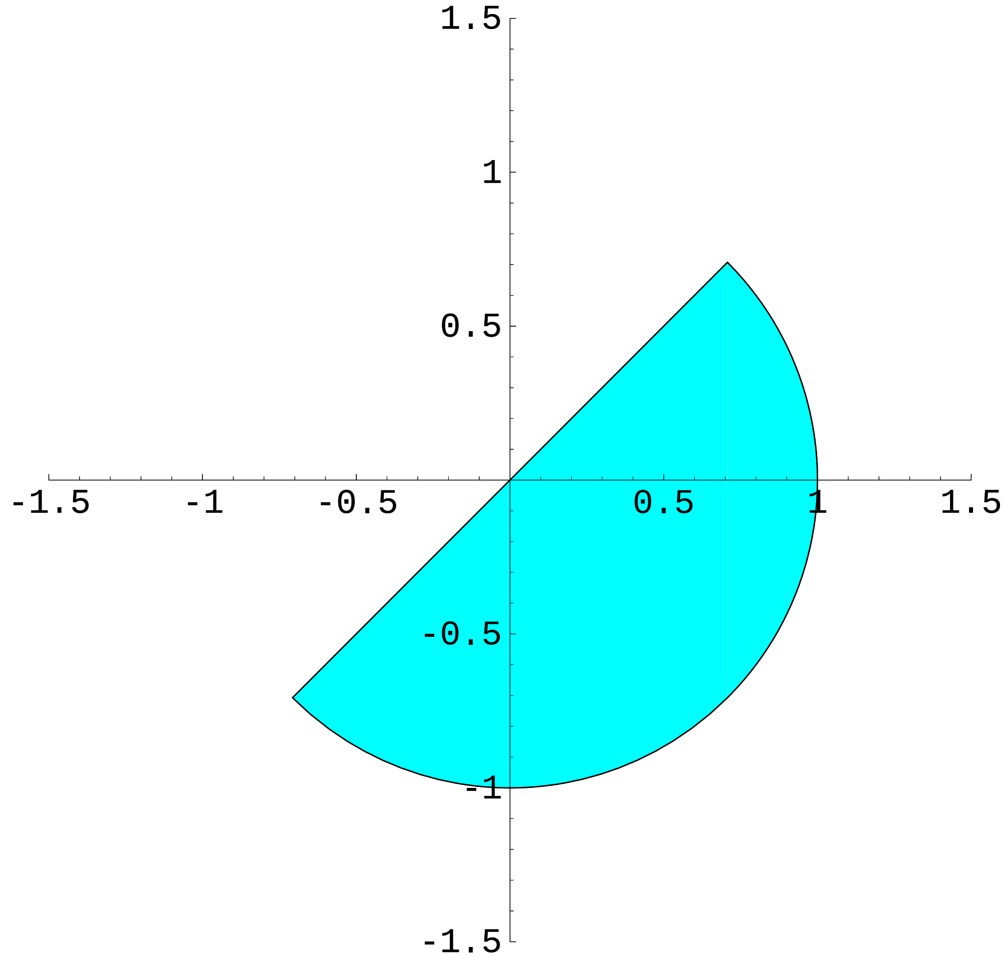

Figure 8: Set $A \cap R_{X,Y} = \{(x, y) : x > y, x^2 + y^2 \leq 1\}$.

Since $|A \cap R_{X,Y}| = \frac{1}{2}\pi 1^2$ and $|R_{X,Y}| = \pi 1^2$, $P(X > Y) = \frac{1}{2}$.

An important continuous bivariate distribution is the *bivariate normal* distribution (pp. 81-84). We'll explore the bivariate normal distribution in a bit more detail later on.

For a jointly distributed continuous random pair (X, Y) with pdf $f(x, y)$, the marginal densities $f_X(x)$ and $f_Y(y)$ are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

These are the continuous analogues of the law of total probability.

Let's find $f_X(x)$ for (X, Y) distributed uniform over the unit disk. For $-1 < x < 1$ we compute

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \pi^{-1} dy \\
&= \pi^{-1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} [y]' dy \\
&= \pi^{-1} [y]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \\
&= \pi^{-1} \left(\sqrt{1-x^2} - -\sqrt{1-x^2} \right) \\
&= 2\pi^{-1} \sqrt{1-x^2}.
\end{aligned}
$$

on $R_X = (-1, 1)$.

Random vectors

The idea of random pairs generalizes to *random vectors*. A vector is an ordered list of numbers; a vector is random if each of the ordered numbers is a random variable. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$. The *elements* of $\mathbf{Y}$ are $Y_1, Y_2, \ldots, Y_n$. The positive integer n is the *size* or *dimension* of the vector $\mathbf{Y}$.

A continuous random vector $\mathbf{Y}$ has a pdf $f(\mathbf{y}) = f(y_1, y_2, \ldots, y_n)$ defined on $\mathbb{R}^n$. A discrete $\mathbf{Y}$ has a pmf $p(\mathbf{y}) = p(y_1, y_2, \ldots, y_n)$. These guys are summed or integrated over sets $A \subset \mathbb{R}^n$ to get probabilities $P(\mathbf{Y} \in A)$.

There are not many "special" jointly defined random vectors (i.e. with their own names); the most widely used (and they're used a lot) are the *multivariate normal distribution* and the *multinomial distribution*.

3.4 Independent random variables

All random pairs have a joint cumulative distribution function given by

$$F(x, y) = P(X \le x, Y \le y) = P(X \le x \text{ and } Y \le y).$$

This generalizes to random vectors as

$$F(\mathbf{y}) = F(y_1, y_2, \ldots, y_n) = P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n).$$

The idea of independence of events from Chapter 1 extends to random vectors:

def'n Random variables $Y_1, Y_2, \ldots, Y_n$ are *independent* if for any sets $A_1, A_2, \ldots, A_n$

$$P(Y_1 \in A_1, Y_2 \in A_2, \ldots, Y_n \in A_n) =$$

$$P(Y_1 \in A_1) P(Y_2 \in A_2) \cdots P(Y_n \in A_n).$$

Proposition: $Y_1, \ldots, Y_n$ are independent iff

$$F(\mathbf{y}) = F(y_1, y_2, \ldots, y_n) = F_{Y_1}(y_1) F_{Y_2}(y_2) \cdots F_{Y_n}(y_n).$$

If $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is continuous this happens iff

$$f(\mathbf{y}) = f(y_1, y_2, \ldots, y_n) = f_{Y_1}(y_1) f_{Y_2}(y_2) \cdots f_{Y_n}(y_n).$$

If $\mathbf{Y}$ is discrete this happens iff

$$p(\mathbf{y}) = p(y_1, y_2, \ldots, y_n) = p_{Y_1}(y_1) p_{Y_2}(y_2) \cdots p_{Y_n}(y_n).$$

In other words, the elements of $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are independent iff the joint distribution factors into the product of the marginal cdf's, pdf's, or pmf's.

This is especially useful to obtain joint distributions from independent random variables.

Example: let $X \sim \exp(0.1)$ independent of $Y \sim N(0,1)$. The joint pdf for (X,Y) on $R_{X,Y} = R_X \times R_Y = (0,\infty) \times (-\infty,\infty)$ is

$$f(x,y) = f_X(x) f_Y(y) = 0.1 \exp(-0.1x)(2\pi)^{-0.5} \exp(-0.5y^2).$$

Note: in general, if X is independent of Y then
$R_{X,Y} = R_X \times R_Y = \{(x,y) : x \in R_X, y \in R_Y\}$.

In fact, if $R_{X,Y} \neq R_X \times R_Y$ then X and Y cannot be independent.

Example: let (X,Y) be uniformly distributed over the triangle with vertices $(0,0)$, $(0,1)$, and $(1,0)$. Clearly $R_X = (0,1)$, $R_Y = (0,1)$, but $R_{X,Y} \neq (0,1) \times (0,1)$, the unit square. So X and Y cannot be independent. This makes good intuitive sense. If knowing Y somehow restricts the range of X then Y is giving information about X.

Another useful consequence is the following:

Proposition: If the joint pdf $f(x, y)$ factors into functions of x only and y only, i.e. $f(x, y) = g(x)h(y)$ for some $g(x)$ and $h(y)$ such that they both integrate to one *and* $R_{X,Y}$ is the product of two sets, say $R_{X,Y} = A \times B$ then $f_X(x) = g(x)$ and $f_Y(y) = h(y)$, $R_X = A$, $R_Y = B$ and X and Y are independent.

This idea generalizes to random vectors with pdf $f(y_1, \ldots, y_n)$ and also discrete random variables (substitute "pmf" for "pdf").

Example: Let (X, Y) have pdf $f(x, y) = e^{-x-y}$ on $R_{X,Y} = \{(x, y) : x > 0, y > 0\}$. We can write $f(x, y) = e^{-x}e^{-y} = g(x)h(y)$ and $R_{X,Y} = (0, \infty) \times (0, \infty)$ so X and Y are independent and $f_X(x) = e^{-x}$ on $R_X = (0, \infty)$ and $f_Y(y) = e^{-y}$ on $R_Y = (0, \infty)$. $X \sim \exp(1)$ indep. of $Y \sim \exp(1)$.

Is X independent of Y for the uniform unit disk example?

Recall that $f_X(x) = 2\pi^{-1}\sqrt{1 - x^2}$ on $R_X = (-1, 1)$. Similarly, $f_Y(y) = 2\pi^{-1}\sqrt{1 - y^2}$ on $R_Y = (-1, 1)$. No, X and Y cannot be independent. You can check either of

$$R_X \times R_Y \neq R_{X,Y},$$

or

$$f(x, y) \neq f_X(x)f_Y(y).$$

Either one of these implies that X is not independent of Y.