A. **Conditional Probability Density.** Let X and Y be **jointly continuous** random variables. Let's say we want a **conditional probability density** function, say of X given $Y = y$. That is we roughly want:

$$\frac{\mathbb{P}\left(X \in \qquad\qquad \mid Y \in \qquad\qquad\right)}{dx} = \frac{\mathbb{P}\left((X, Y) \in \qquad\qquad\right)}{dx \cdot \mathbb{P}\left(Y \in \qquad\qquad\right)} \approx$$

Conditional Probability Density. Let X and Y be **jointly continuous** random variables. Then the **conditional pdf** of X given $Y = y$ is:

$f(x \mid y) =$

If $f_Y(y) = 0$ this is undefined, but we adopt the same convention that it is finite but undefined, so multiplying it by 0 always yields 0. Thus:

$f(x, y) \overset{\text{a.e.}}{=}$

The conditional pdf let's us **define** the conditional probability that $X \in D$ given $Y = y$, where D is a "measurable" subset of the x–axis, as:

$\mathbb{P}(X \in D \mid Y = y) =$

Recall "a.e." stands for the technical term "almost everywhere".

Why do we say "define"? Have we not already defined it. Actually, no we have not. Prior to today, we would have tried to define this conditional probabilities by:

$$\mathbb{P}(X \in D \mid Y = y) \overset{?}{=} \frac{\mathbb{P}(X \in D, Y = y)}{\mathbb{P}(Y = y)}$$

but this is **not** possible, because the probability that $Y = y$ is zero. This is why we have this alternate, you could say "extended", definition.

There is a simple connection to independence, that perhaps is expected, since $\mathbb{P}(E \mid F) = \mathbb{P}(E)$ for independent events. Let X and Y be jointly continuous.

Then: X and Y are independent $\iff f(x, y) \overset{\text{a.e.}}{=} f_X(x) f_Y(y)$

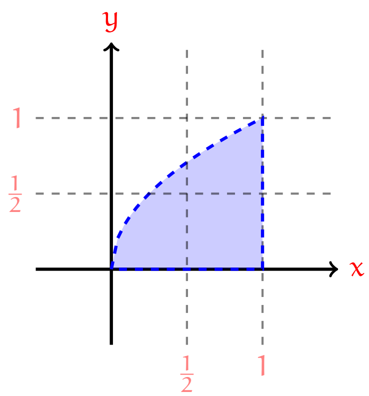
$\iff$ for a.e. y with $f_Y(y) > 0$

Independence and Conditional Density. Two jointly continuous random variables X and Y are **independent** if and only if:

$f_X(x) \overset{\text{a.e.}}{=}$ for a.e. y with $f_Y(y) > 0$

Example 1. Suppose X and Y are jointly continuous, with joint density function:

$$f(x,y) = \begin{cases} 8x^2y & \text{if } 0 < x < 1 \text{ and } 0 < y < \sqrt{x} \\ 0 & \text{otherwise} \end{cases}$$



(a) Find the conditional density of $f(x \mid y)$.

(b) How does part a confirm that X and Y are **not** independent?

(c) Find: $\mathbb{P}\left(X < \frac{1}{2} \mid Y = \frac{1}{2}\right)$

B. **Jdfs and Jacobians.** Let X and Y be **jointly continuous** random variables. And suppose you know how X and Y depend on each other, in the sense that your know their joint density function:

$f_{X,Y}(x,y) \leftarrow$ jdf of X and Y

Next define two new random variables in terms of X and Y. For example:

$U = X + Y$

$V = X - Y$

I want to determine how U and V depend on each other, in other words I want:

$f_{U,V}(u,v) \leftarrow$ jdf of U and V

In what proceeds, it is important that $\vec{g}$ be **invertible**, meaning it is possible to solve for X and Y **uniquely** in terms of U and V.

$X =$

$Y =$

If $\vec{g}$ is **differentiable**, then the way **infinitesimal areas** transform under $\vec{g}$ is determined by the **absoute value** of the **Jacobian determinant**:

$$\frac{dudv}{dxdy} = \left|\frac{\partial(u,v)}{\partial(x,y)}\right| \text{ where } \frac{\partial(u,v)}{\partial(x,y)} =$$

The determinant of a matrix is:

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

That infinitesimal areas transform the way described to the left way is tied to a combination of **linear algebra** (the determinant of a matrix tells you how area scales under a linear transformation) and **multivariable calculus** (which tells you that, at the infinitesimal level, a differentiable function can be treated as a linear transformation).

In our running example:

$$\frac{dudv}{dxdy} =$$

If further $\vec{g}$ is **nonsingular**, meaning the Jacobian determinant is nonzero:

$$\frac{dxdy}{dudv} =$$

In our running example:

$$\frac{dxdy}{dudv} =$$

C. **Continued: Jdfs and Jacobians.** Now, let's return to probability.

$$\mathbb{P}\left((X, Y) \in \qquad\qquad\qquad\qquad\right) = \mathbb{P}\left((U, V) \in \qquad\qquad\qquad\qquad\right)$$

In our running example, where $U = X + Y$ and $V = X - Y$, we had:

$f_{U,V}(u, v) =$

Let's summarize the key aspects of the approach just outlined.

Jdfs and Jacobians. Let X and Y be jointly continuous random variables, and suppose $\langle U, V \rangle = \vec{g}(X, Y)$, where $\vec{g}$ has an "almost everywhere" **inverse** and is **nonsingular** "almost everywhere".

Then letting $\langle u, v \rangle = \vec{g}(x, y)$, the jdfs of X, Y and U, V are related according to:

$f_{U,V}(u, v) =$

Here, having an "almost everywhere" inverse means almost every (u, v) can be **uniquely** solved for in terms of (x, y), and "almost every" (x, y) can be **uniquely** solved for in terms of (u, v). Nonsingular means the Jacobian determinant is **nonzero**.