ORF 245 Fundamentals of Statistics
Chapter 3
Joint Distributions

Robert Vanderbei

Fall 2014

Slides last edited on October 1, 2014

Introduction

Joint Cumulative Distribution Function (cdf):

$$F(x, y) = P(X \leq x,\ Y \leq y).$$

Probability that (X, Y) belongs to a given rectangle:

$$P(x_1 < X \leq x_2,\ y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

Probability that (X, Y) belongs to an infinitesimal rectangle:

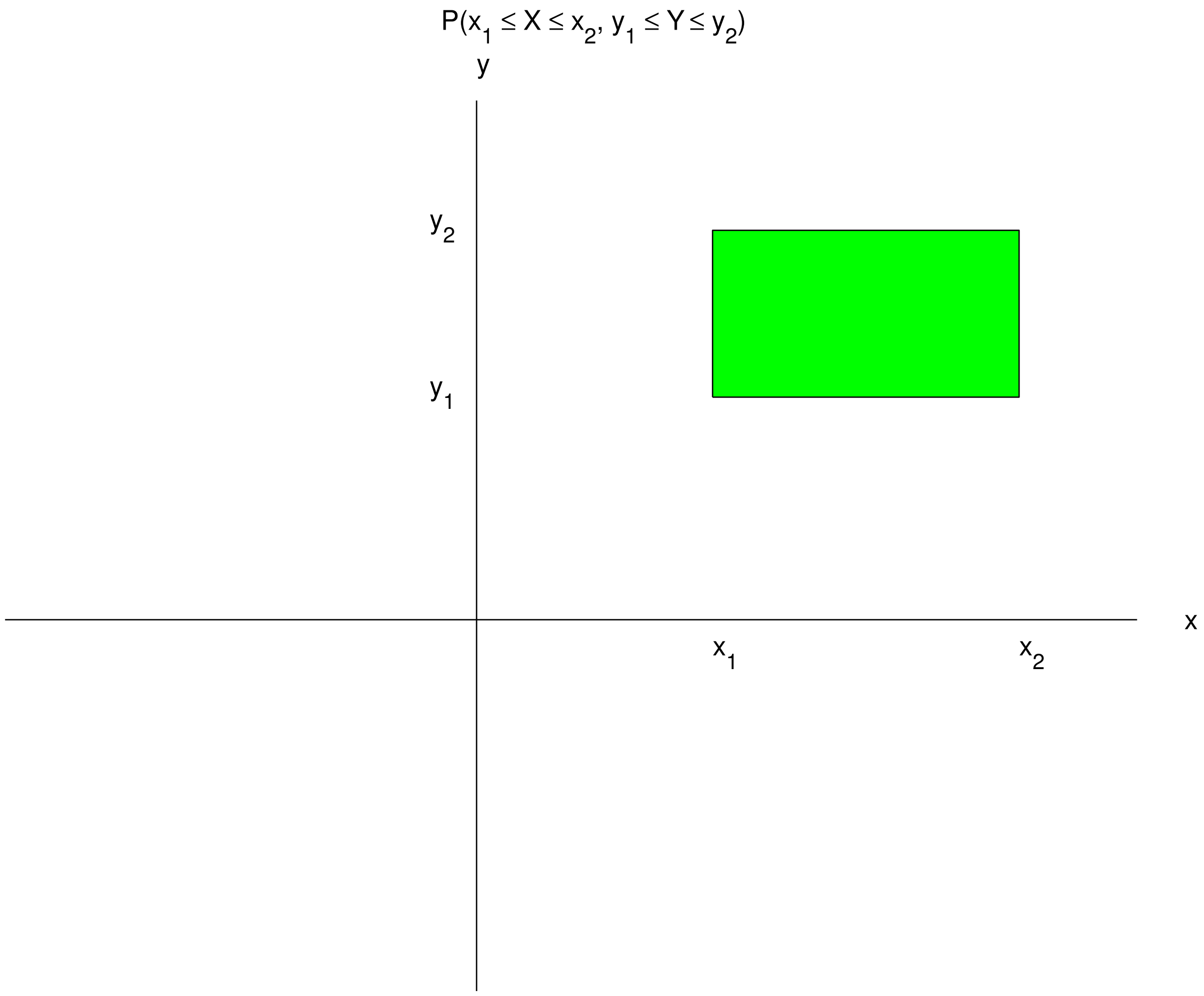
$$\begin{aligned} P(x < X &\leq x + dx,\ y < Y \leq y + dy) \\ &= F(x + dx, y + dy) - F(x, y + dy) - F(x + dx, y) + F(x, y) \\ &\approx \frac{\partial^2 F}{\partial x \partial y}(x, y)\ dxdy \end{aligned}$$

Joint Probability Density Function:

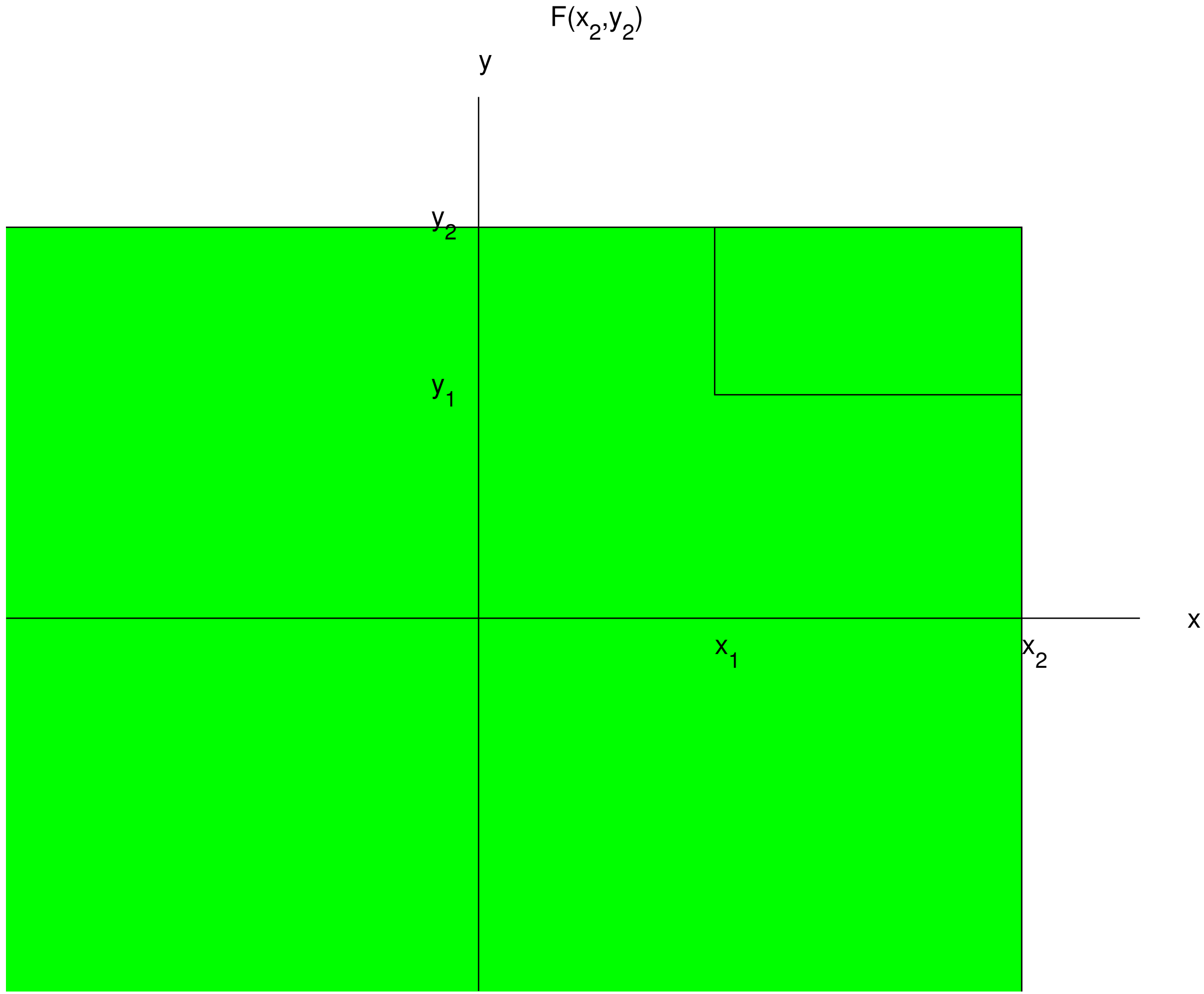
$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y)$$

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

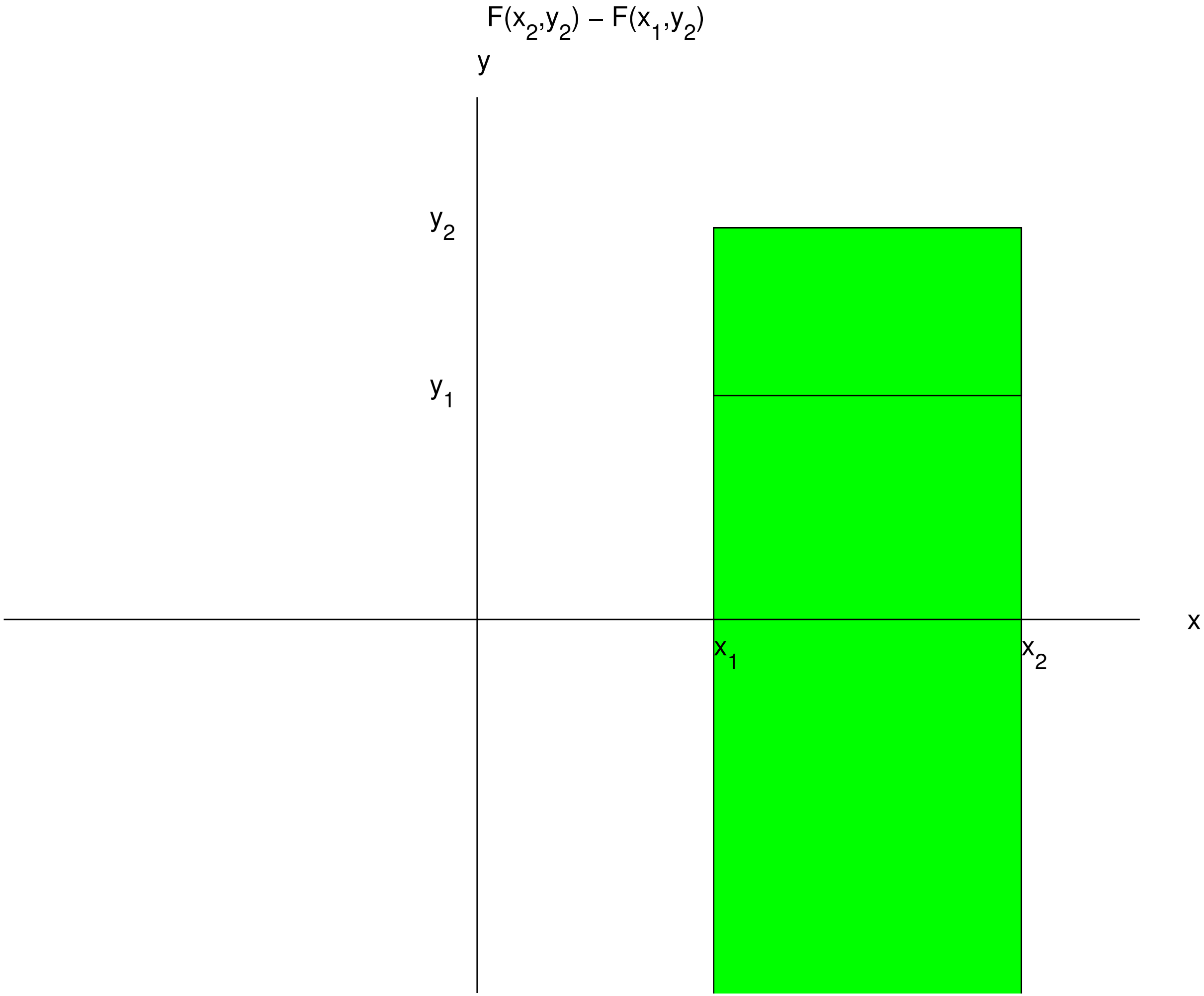
Probability of (X, Y) Falling in a Rectangle



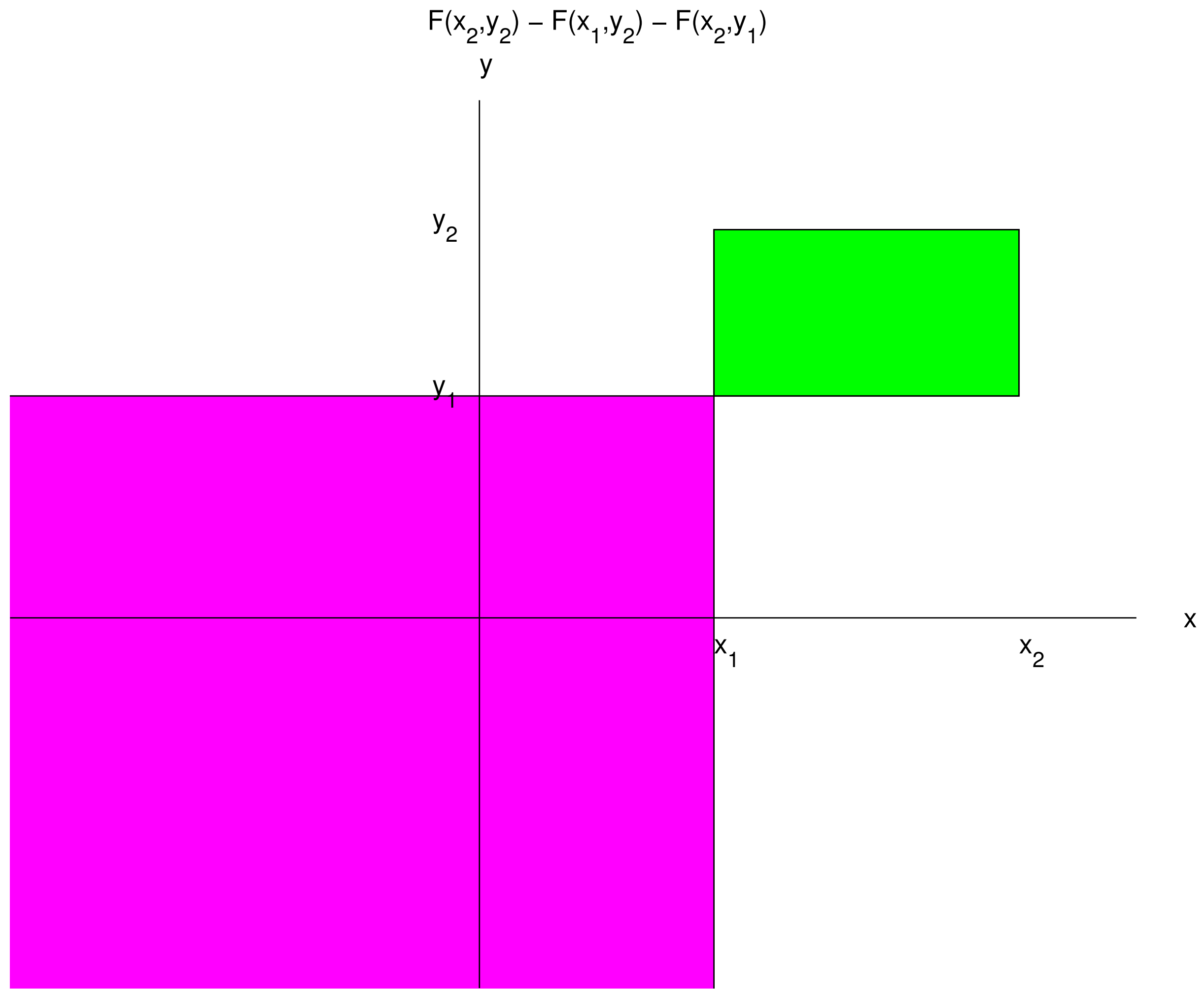
Probability of (X, Y) Falling in a Rectangle



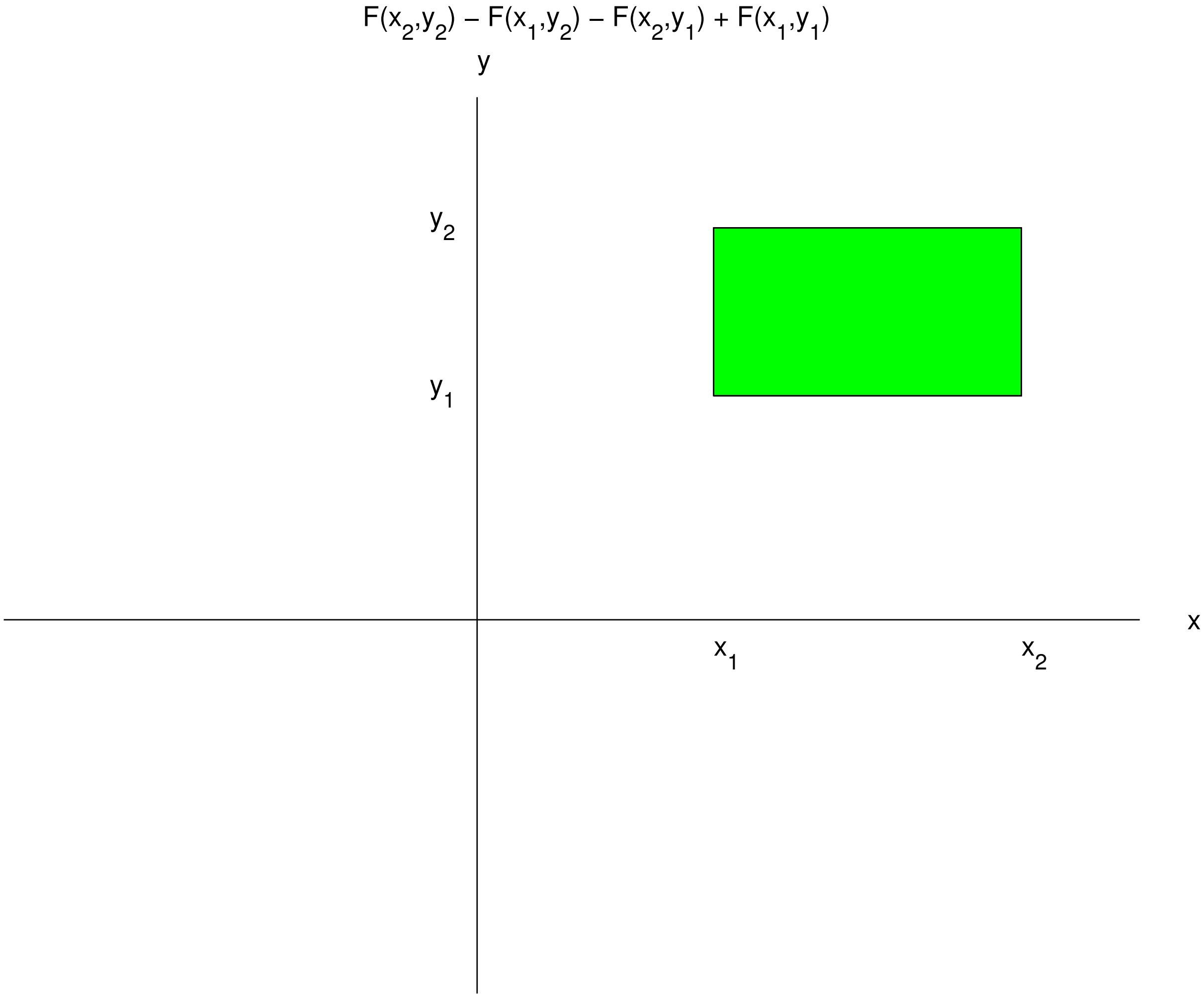
Probability of (X, Y) Falling in a Rectangle



Probability of (X, Y) Falling in a Rectangle



Probability of (X, Y) Falling in a Rectangle



Discrete Case

If X and Y are "independent", then

RV			Y				
	value		y_1	y_2	y_3	$\cdots$	y_n
		prob	q_1	q_2	q_3	$\cdots$	q_n
X	x_1	p_1	p_1q_1	p_1q_2	p_1q_3	$\cdots$	p_1q_n
	x_2	p_2	p_2q_1	p_2q_2	p_2q_3	$\cdots$	p_2q_n
	x_3	p_3	p_3q_1	p_3q_2	p_3q_3	$\cdots$	p_3q_n
			$\vdots$	$\vdots$	$\vdots$		$\vdots$
	x_m	p_m	p_mq_1	p_mq_2	p_mq_3	$\cdots$	p_mq_n

If they are not independent, then the situation can be more complicated.
For example, we can shuffle things a little bit...

RV			Y				
	value		y_1	y_2	y_3	$\cdots$	y_n
		prob	q_1	q_2	q_3	$\cdots$	q_n
X	x_1	p_1	$p_1q_1+\varepsilon$	$p_1q_2-\varepsilon$	p_1q_3	$\cdots$	p_1q_n
	x_2	p_2	p_2q_1	p_2q_2	p_2q_3	$\cdots$	p_2q_n
	x_3	p_3	$p_3q_1-\varepsilon$	$p_3q_2+\varepsilon$	p_3q_3	$\cdots$	p_3q_n
			$\vdots$	$\vdots$	$\vdots$		$\vdots$
	x_m	p_m	p_mq_1	p_mq_2	p_mq_3	$\cdots$	p_mq_n

Marginals

Given a joint cdf, $F(x, y)$, for a pair of random variables X and Y, the distribution of X is easy to find:

$$F_X(x) = P(X \leq x) = P(X \leq x,\ Y < \infty) = F(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y) dy du$$

And, the density function for X is then found by differentiating:

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

In a similar way, we can find the density $f_Y(y)$ associated with random variable Y.

These univariate densities are called *marginal densities*.

In the discrete case, they can be found by summing the entries in a column (or row):

$$p_X(x_i) = \sum_j p(x_i, y_j)$$

Example 3.2 E

A point is chosen at random from the disk of radius 1 centered at the origin of a coordinate system.
Let (X, Y) denote the rectangular coordinates of this random point.
Since the area of the disk is π, the density function must be

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2 + y^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Let $R = \sqrt{X^2 + Y^2}$ denote the radial coordinate of the random point.
It is easy to compute the cdf for R:

$$F_R(r) = P(R \leq r) = \frac{\pi r^2}{\pi} = r^2, \qquad 0 \leq r \leq 1$$

The density function for R is therefore

$$f_R(r) = 2r, \qquad 0 \leq r \leq 1$$

The marginal density of X is also easy to compute:

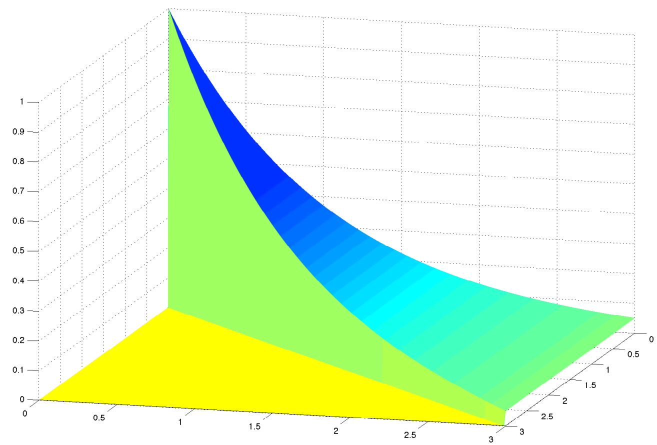
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi}\sqrt{1-x^2}, \qquad -1 \leq x \leq 1$$

Example 3.3 D

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y}, & 0 \le x \le y \\ 0, & \text{elsewhere} \end{cases}$$

$$f_X(x) = \int_x^{\infty} \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x}, \qquad x \ge 0 \qquad \Longleftarrow \text{Exponential}$$

$$f_Y(y) = \int_0^{y} \lambda^2 e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y}, \qquad y \ge 0 \qquad \Longleftarrow \text{Gamma}$$



Independence

Random variables X_1, X_2, ..., X_n are *independent* if

$$P(X_1 \in A_1,\ X_2 \in A_2,\ \ldots\ X_n \in A_n) = P(X_1 \in A_1)\ P(X_2 \in A_2)\ \cdots\ P(X_n \in A_n).$$

An equivalent definition is that the cdf factors into a product:

$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)\ F_{X_2}(x_2)\ \cdots\ F_{X_n}(x_n)$$

If each of the random variables has a density function, then it is also equivalent to say that the joint density factors into a product:

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\ f_{X_2}(x_2)\ \cdots\ f_{X_n}(x_n)$$

If X and Y are independent, then $h(X)$ and $g(Y)$ are independent for any pair of functions $h(\cdot)$ and $g(\cdot)$.

Conditional Distribution

Discrete Case: Let X and Y be a pair of discrete random variables.

The *conditional probability* that $X = x$ given that $Y = y$ is

$$p_{X|Y}(x|y) \;=\; P(X = x \mid Y = y) \;=\; \frac{P(X = x,\; Y = y)}{P(Y = y)} \;=\; \frac{p_{XY}(x,y)}{p_Y(y)}$$

provided, of course, that $p_Y(y) > 0$. This probability is best left undefined if $p_Y(y) = 0$. Holding y to be fixed and viewing $p_{X|Y}(x|y)$ as a function of x, this function is a probability mass function since it is nonnegative and sums to one.
If X and Y are independent, then $p_{X|Y}(x|y) = p_X(x)$.
A useful formula:

$$p_X(x) = \sum_y p_{XY}(x,y) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

Continuous Case: The conditional density is given by:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Usefull formula:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy$$

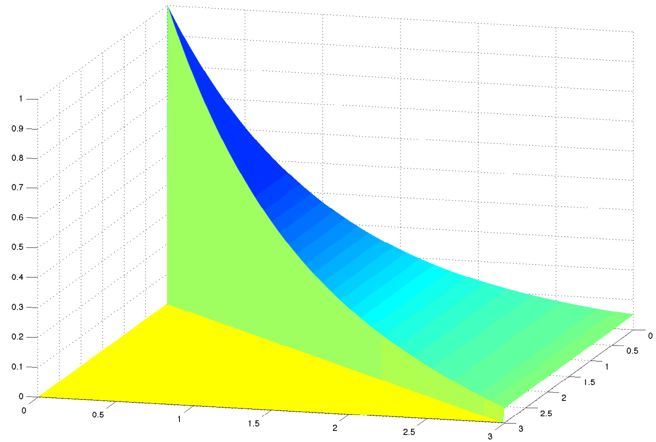
Example 3.5.2 A

Continuation of earlier example:

$$f(x,y) = \lambda^2 e^{-\lambda y}, \qquad 0 \le x \le y$$

$$f_X(x) = \int_x^\infty \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x}, \qquad x \ge 0$$

$$f_Y(y) = \int_0^y \lambda^2 e^{-\lambda y} dy = \lambda^2 y e^{-\lambda y}, \qquad y \ge 0$$



$$f_{Y|X}(y|x) = \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)}, \qquad y \ge x$$

$$f_{X|Y}(x|y) = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \frac{1}{y}, \qquad 0 \le x \le y$$

Bayesian Inference (Example 3.5.2 E)

Consider a coin tossing experiment. Suppose that we do not know the probability of tossing a heads. Let Θ be a random variable representing this probability. The *Bayesian philosophy* is to assume a *prior* distribution for the unknown parameter and then update the prior based on observations. If we assume that we know nothing about Θ (not a reasonable assumption!), then we might take as our prior the uniform distribution on $[0, 1]$:

$$f_\Theta(\theta) = 1, \qquad 0 \le \theta \le 1$$

Now, suppose we toss the coin n times and let X denote the number of heads. The conditional distribution for X given that $\Theta = \theta$ is binomial with parameters n and θ:

$$f_{X|\Theta}(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n$$

The joint distribution for X and Θ is gotten by simply multiplying:

$$f_{X,\Theta}(x, \theta) = f_{X|\Theta}(x|\theta) f_\Theta(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \qquad 0 \le \theta \le 1, \quad x = 0, 1, \ldots, n$$

Bayesian Inference (Example 3.5.2 E) – Continued

From this joint distribution, we can compute the marginal distribution for X by integrating:

$$f_X(x) \;=\; \int_0^1 f_{X,\Theta}(x,\theta)d\theta \;=\; \binom{n}{x} \int_0^1 \theta^x (1-\theta)^{n-x} d\theta$$

This last integral can be computed using integration-by-parts (over and over).
Here's what one gets

$$\int_0^1 \theta^x (1-\theta)^{n-x} d\theta = \frac{x!(n-x)!}{(n+1)!}$$

Hence, the marginal distribution for X reduces to a very simple formula:

$$f_X(x) \;=\; \frac{1}{n+1}, \qquad\qquad x = 0, 1, \ldots, n$$

From the joint distribution and the marginal distribution, we can compute the conditional distribution of Θ given that $X = x$:

$$f_{\Theta|X}(\theta|x) \;=\; \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} \;=\; (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}, \qquad 0 \le \theta \le 1$$

The conditional density

$$f_{\Theta|X}(\theta|x) \;=\; (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}, \qquad\qquad 0 \le \theta \le 1$$

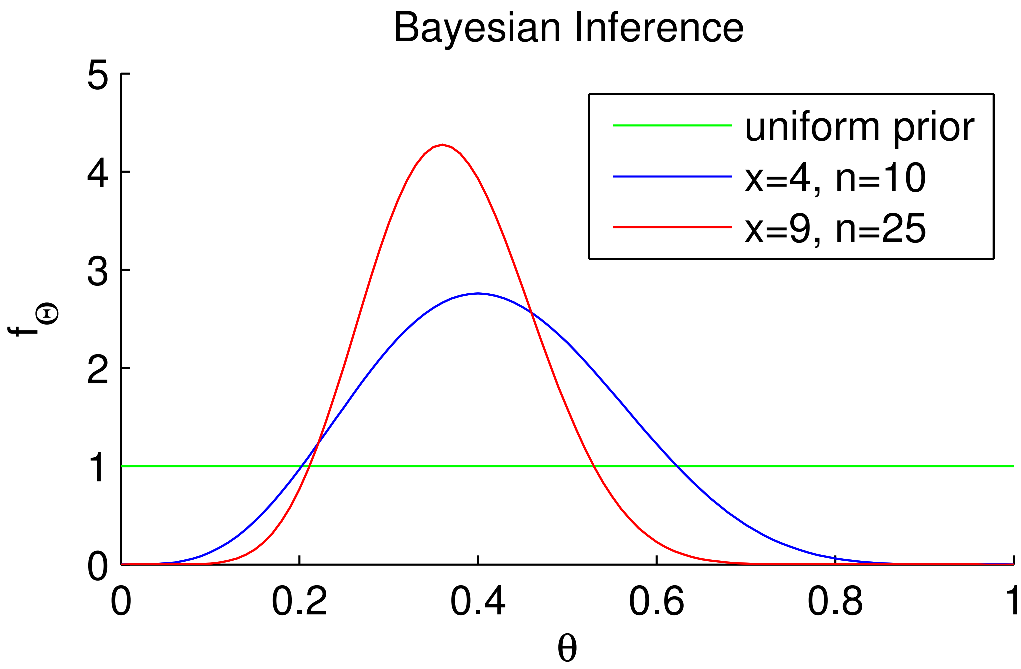
is a bit hard to visualize. Some plots are shown on the next slide. But, it's interesting first to figure out where this function has its maximum (as that's an indicator of where this density is concentrated). So, we differentiate and set the derivative to zero:

$$\begin{aligned}\frac{d}{d\theta}f_{\Theta|X}(\theta|x) \;&=\; (n+1)\binom{n}{x}\frac{d}{d\theta}\theta^x(1-\theta)^{n-x} \\ &=\; (n+1)\binom{n}{x}\left(x\theta^{x-1}(1-\theta)^{n-x} - \theta^x(n-x)(1-\theta)^{n-x-1}\right) \\ &=\; 0\end{aligned}$$

Dividing both sides by $n+1$ and by $\binom{n}{x}$ and by θ^{x-1} and by $(1-\theta)^{n-x-1}$, we get

$$x(1-\theta) - \theta(n-x) = 0 \qquad \Longrightarrow \qquad \theta = x/n$$

Bayesian Inference (Example 3.5.2 E) – Continued



Distribution of a Sum: $Z = X + Y$

Discrete Case: Let X and Y be a pair of discrete random variables *taking integer values*.

Using the law of total probability, we get that

$$\begin{aligned} P(Z = z) &= \sum_{x=-\infty}^{\infty} P(X = x,\ Z = z) = \sum_{x=-\infty}^{\infty} P(X = x,\ X + Y = z) \\ &= \sum_{x=-\infty}^{\infty} P(X = x,\ x + Y = z) = \sum_{x=-\infty}^{\infty} P(X = x,\ Y = z - x) \end{aligned}$$

Hence, the distribution of Z is

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p(x, z - x)$$

If X and Y are *independent*, then

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z - x)$$

A sum of this form is called a *convolution*.

Distribution of a Sum: $Z = X + Y$

Continuous Case: We start by computing the cdf:

$$F_Z(z) = \iint_{\{(x,y):x+y\leq z\}} f(x,y)dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx$$

In the inner integral, make a change of variable from y to $v = x + y$ and then reverse the order of integration:

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,v-x)dvdx = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x,v-x)dxdv$$

Finally, differentiate to find the density:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)dx$$

Again, if X and Y are *independent*, the result is a *convolution*:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

Example 3.6.1 A

Let X and Y be independent exponential random variables with the same parameter λ.

Find the distribution of their sum: $Z = X + Y$.

The distribution of the sum is given by the convolution:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \\
&= \int_0^z \lambda e^{-\lambda x} \, \lambda e^{-\lambda(z-x)} \, dx \\
&= \lambda^2 \int_0^z e^{-\lambda z} dx \\
&= \lambda^2 z e^{-\lambda z} \qquad \Longleftarrow \text{Gamma}
\end{aligned}
$$

NOTE: The sum of n independent exponential random variables with parameter λ is a random variable with a Gamma distribution with parameters n and λ.

Example 3.6.1 B

Let X and Y be independent standard normal random variables. That is, $N(0,1)$.

Find the distribution of the ratio: $Z = Y/X$.

$$
\begin{aligned}
F_Z(z) &= P(Y/X \le z) = P(Y \le zX,\ X \ge 0) + P(Y \ge zX,\ X < 0) \\
&= P(Y \le zX,\ X \ge 0) + P(-Y \le z(-X),\ -X \ge 0) \\
&= 2P(Y \le zX,\ X \ge 0) \\
&= 2\int_0^{\infty}\int_{-\infty}^{xz} f_X(x) f_Y(y)\,dy\,dx = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{xz} e^{-x^2/2} e^{-y^2/2}\,dy\,dx
\end{aligned}
$$

Differentiating, we get

$$
\begin{aligned}
f_Z(z) &= \frac{1}{\pi}\int_0^{\infty} e^{-x^2/2}\, e^{-(xz)^2/2} x\,dx = \frac{1}{\pi}\int_0^{\infty} e^{-x^2(1+z^2)/2}\, x\,dx \\
&= \frac{1}{\pi}\int_0^{\infty} e^{-u(1+z^2)}\,du = \frac{1}{\pi(1+z^2)} \qquad \Longleftarrow \textit{Cauchy}
\end{aligned}
$$

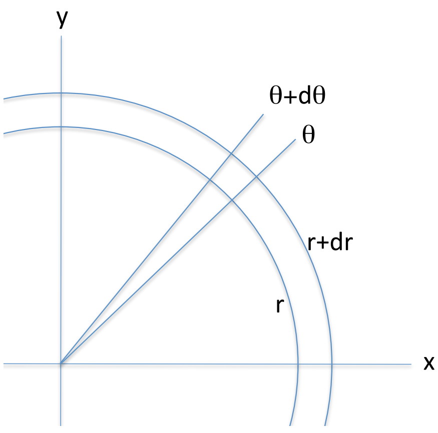
Example 3.6.2 A

Let X and Y be a pair of independent standard normal random variables.

Find the joint distribution of the polar coordinates: $R = \sqrt{X^2 + Y^2}$ and $\Theta = \text{atan2}(Y, X)$.

Let's work infinitesmally:

$$\begin{aligned} f_{R\Theta}(r,\theta)drd\theta &= P(r \leq R \leq r + dr,\ \theta \leq \Theta \leq \theta + d\theta) \\ &= P((X,Y) \in A) \\ &= f_{XY}(r\cos\theta,\ r\sin\theta)rdrd\theta \\ &= f_X(r\cos\theta)\ f_Y(r\sin\theta)\ rdrd\theta \\ &= \frac{1}{\sqrt{2\pi}}e^{-r^2\cos^2(\theta)/2}\ \frac{1}{\sqrt{2\pi}}e^{-r^2\sin^2(\theta)/2}\ rdrd\theta \\ &= \frac{1}{2\pi}e^{-r^2(\cos^2\theta+\sin^2\theta)/2}\ rdrd\theta \end{aligned}$$



Hence,

$$f_{R\Theta}(r,\theta) = \frac{1}{2\pi}\ re^{-r^2/2}$$

Example 3.6.2 A – Continued

The joint density can be factored into a product of densities:

$$f_{R\Theta}(r,\theta) = \left(\frac{1}{2\pi}\right)\,\left(re^{-r^2/2}\right) = f_\Theta(\theta)\; f_R(r), \qquad -\pi \le \theta \le \pi,\; r \ge 0$$

Hence, *R and Θ are independent* random variables.

And, Θ is uniformly distributed on $[-\pi, \pi]$.

The distribution of R is called the *Rayleigh distribution*.

Remark: A nonnegative random variable R has a Rayleigh distribution if and only if R^2 has an exponential distribution.

The Rayleigh Distribution

The extra r factor in the Rayleigh density function

$$f_R(r) = re^{-r^2/2} \qquad r \geq 0$$

allows us to do the integral to find the cdf for R

$$F_R(r) = \int_0^r f_R(u)du = \int_0^r ue^{-u^2/2}du$$

by making the obvious change of variables $v = u^2/2$

$$F_R(r) = \int_0^{r^2/2} e^{-u}du = \left[-e^{-u}\right]_0^{r^2/2} = 1 - e^{-r^2/2}$$

Simulating Standard Normal Random Variables

The cdf for the Rayleigh distribution is easy to "invert":

$$F_R^{-1}(u) = \sqrt{-2\ln(1-u)}$$

Hence, if U is a uniform random variable on $[0, 1]$, then

$$R = \sqrt{-2\ln(1-U)}$$

has a Rayleigh distribution. And, if V is a second uniform random variable on $[-\pi, \pi]$, then

$$X = R\cos(V) \quad \text{and} \quad Y = R\sin(V)$$

are a pair of independent standard normal random variables.

Matlab

```
n = 500;
U = rand([n 1]);  % doc rand
V = 2*pi*rand([n 1]) - pi;
R = sqrt(-2*log(1-U));
X = R.*cos(V);
Y = R.*sin(V);

figure(1);   % plot (X,Y)
plot(X,Y,'b+');
axis equal;
xlim([-4 4]);
ylim([-4 4]);
xlabel('x');
ylabel('y');
title('Standard Normals in the Plane');

figure(2);   % plot empirical cdf next to true cdf
Xsort = sort(X);
x = (-400:400)/100;
y = cdf('norm',x);
plot(Xsort,(1:n)/n,'r-');
hold on;
plot(x,y,'k-');
hold off;
xlabel('x');
ylabel('F(x)');
legend('Empirical CDF','\Phi(x)','Location','Northwest');
title('Cumulative Distribution Function');
```