Joint Continuous Distributions

Statistics 110

Summer 2006

Joint Continuous Distributions

Not surprisingly we can look at the joint distribution of 2 or more continuous RVs. For example, we could look at the amount of time it takes to get to the Science Center from home each morning for the remaining days this week ($X =$ Thursday travel time and $Y =$ Friday's travel time).

Probabilities are based on the joint PDF $f_{X,Y}(x, y)$. The probability of being in the event A is given by

$$P[(X, Y) \in A] = \int_A f_{X,Y}(x, y) dx dy$$

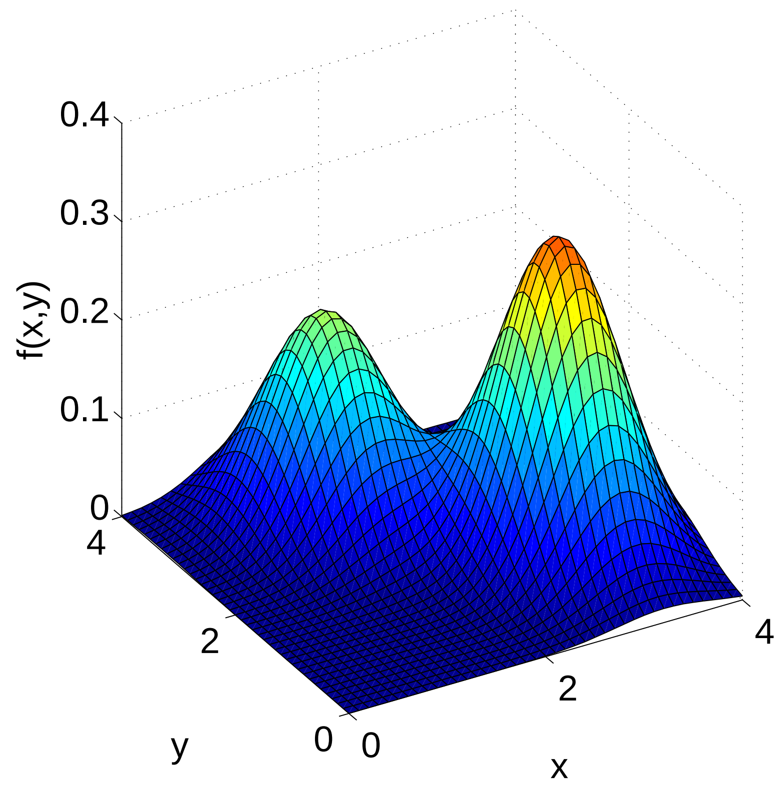
for any $A \subset \mathbb{R}^2$.

Note that a joint density must satisfy

$$f(x,y) \geq 0$$

$$\int_{\Omega} f(x,y) dx dy = 1$$

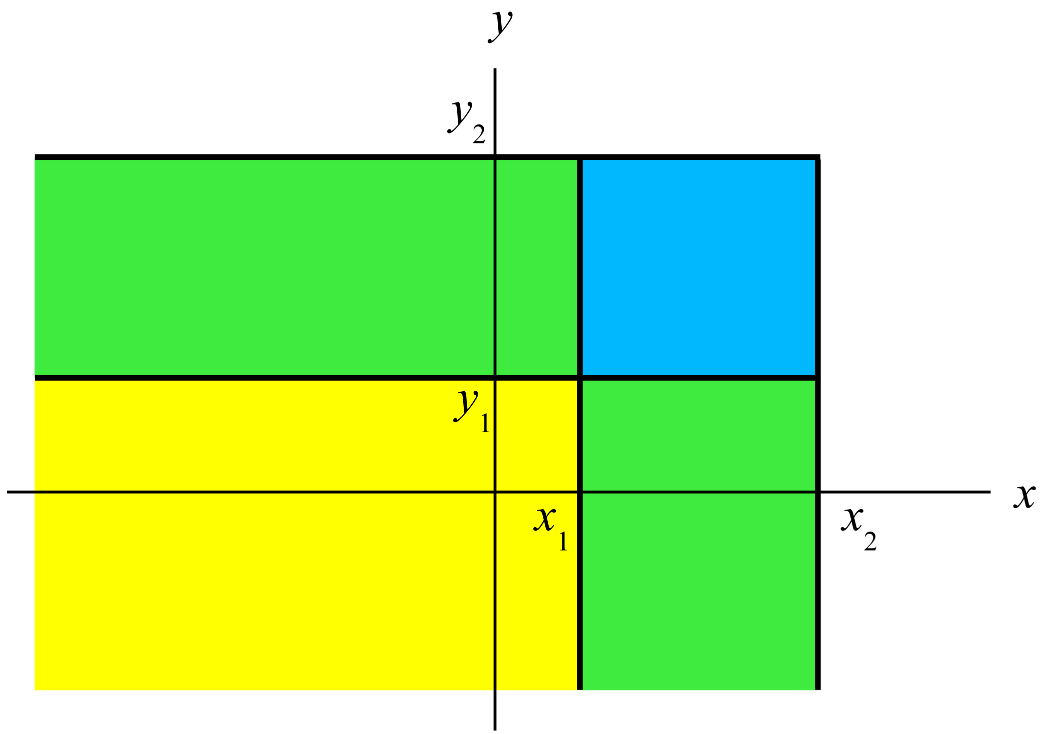
where Ω is the sample space for the combination of RVs.



For rectangular regions, the joint CDF is useful for calculating probabilities.

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y]$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y)dydx$$

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



As with the univariate case, the joint PDF is given by

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

wherever the partial derivative is defined.

For small Δ_x and Δ_y, if f is continuous at (x,y),

$$P[x \leq X \leq x + \Delta_x, y \leq Y \leq y + \Delta_y] \approx f_{X,Y}(x,y)\Delta_x \Delta_y$$

so the probability of getting in a small region around (x,y) is proportional to $f_{X,Y}(x,y)$ so the density is giving information about how likely and observation at the point is.

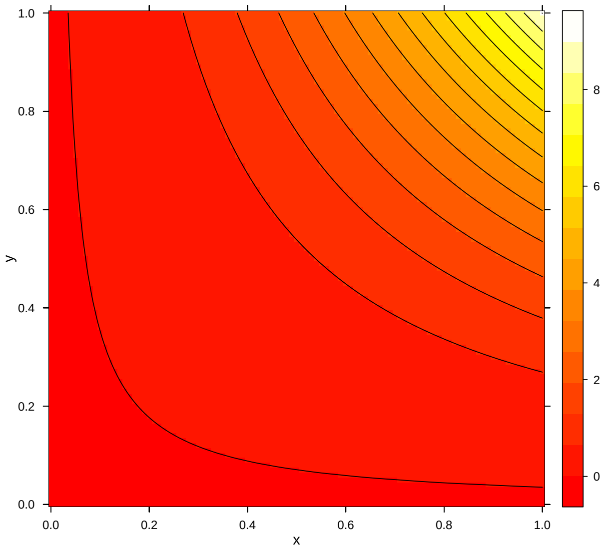
The marginal distribution of each component is “easily” determined from the joint density as

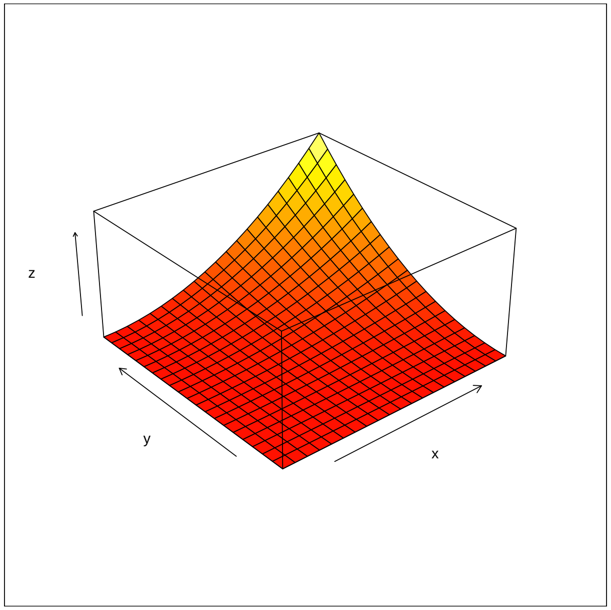
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

This is the analogue to the discrete case, where we are integrating out y instead of adding over it.

For example, let

$$f_{X,Y}(x,y) = 9x^2y^2; \quad 0 \le x \le 1, 0 \le y \le 1$$





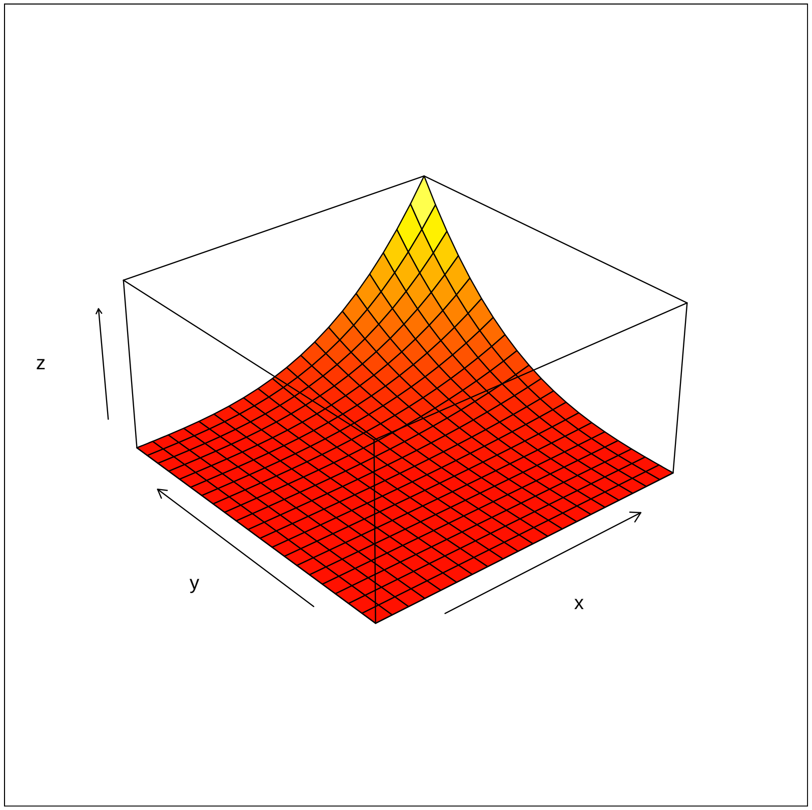
Then

$$f_X(x) = \int_0^1 9x^2y^2dy = 3x^2; \quad 0 \le x \le 1$$

Similarly $f_Y(y) = 3y^2$.

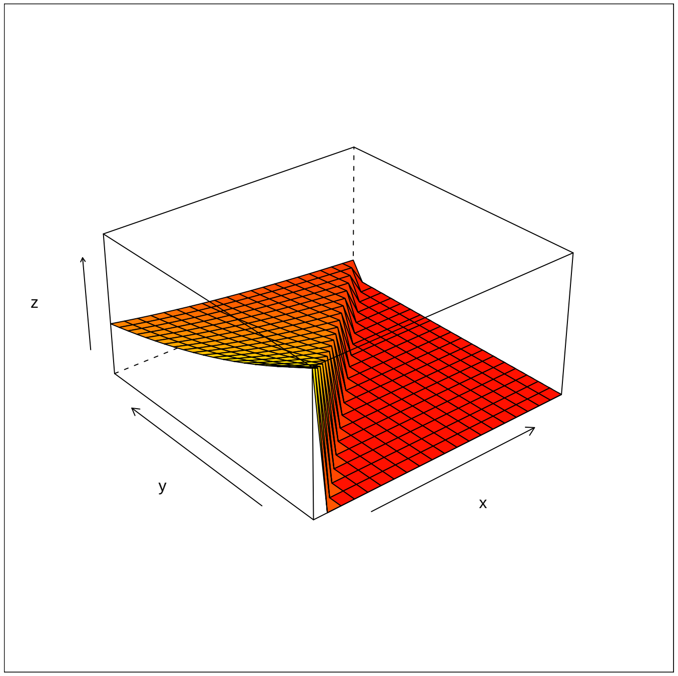
Also

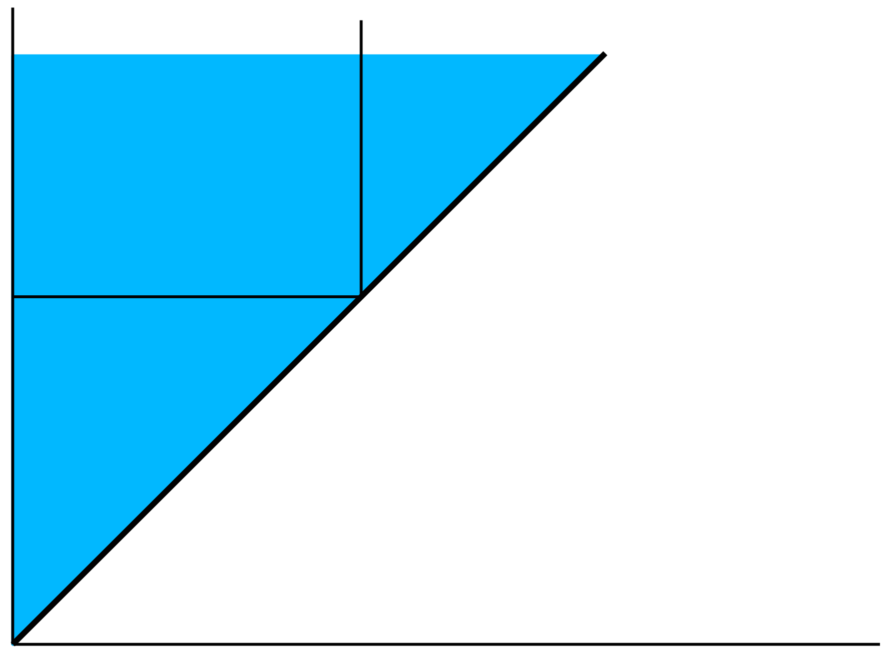
$$F_{X,Y}(x,y) = x^3y^3; \quad 0 \leq x \leq 1, 0 \leq y \leq 1$$



$$F_{X,Y}(x,y) = \int_0^x \int_0^y 9x^2y^2 dxdy; \quad 0 \leq x \leq 1, 0 \leq y \leq 1$$

Not all joint RVs are defined on nice rectangles. For example

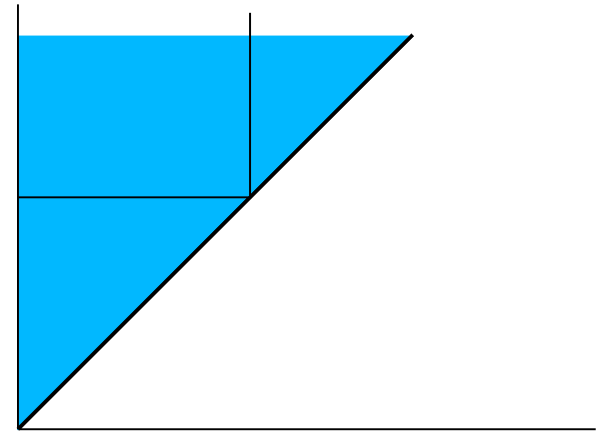




$$f_{X,Y}(x,y) = 2e^{-(x+y)}; \quad 0 \leq x \leq y, y \geq 0$$

is defined on an infinite triangle. You need to be careful in determining the marginal distributions.

$$f_X(x) = \int_x^{\infty} 2e^{-x}e^{-y}dy = 2e^{-2x}; \quad x \geq 0$$



and

$$f_Y(y) = \int_0^{y} 2e^{-y}e^{-x}dx = 2e^{-y}(1-e^{-y}); \quad y \geq 0$$

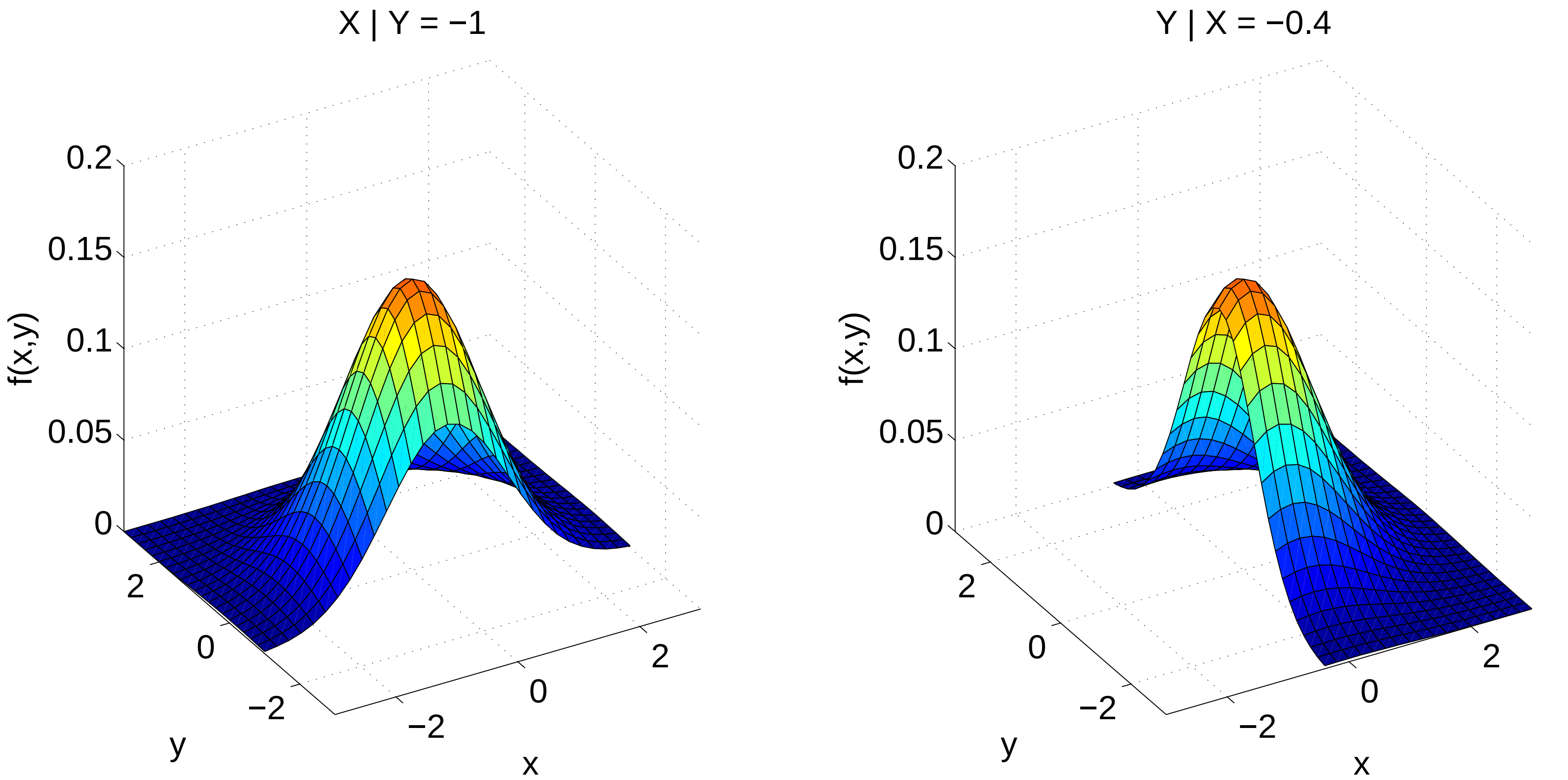
The concept of conditional distribution is a bit more complex in the continuous case.

$$
\begin{aligned}
& P[x \le X \le x + \Delta_x | y \le Y \le y + \Delta_y] \\
& \qquad = \frac{P[x \le X \le x + \Delta_x, y \le Y \le y + \Delta_y]}{P[y \le Y \le y + \Delta_y]} \\
& \qquad \approx \frac{f_{X,Y}(x,y)\Delta_x\Delta_y}{f_Y(y)\Delta_y} \\
& \qquad = \left\{ \frac{f_{X,Y}(x,y)}{f_Y(y)} \right\} \Delta_x
\end{aligned}
$$

So conditional on $Y \in [y, y + \Delta_y]$, X has, approximately, a density given by the expression in $\{\ \}$. Note that this density does not depend on Δ_y as long as it is small.

Thus we write the conditional PDF of $X|Y = y$ as

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



For the examples seen so far

- $f_{X,Y}(x,y) = 9x^2y^2$

$$f_{X|Y}(x|y) = \frac{9x^2y^2}{3y^2} = 3x^2; \quad 0 \leq x \leq 1$$

$$f_{Y|X}(y|x) = \frac{9x^2y^2}{3x^2} = 3y^2; \quad 0 \leq y \leq 1$$

- $f_{X,Y}(x,y) = 2e^{-(x+y)}$

$$f_{X|Y}(x|y) = \frac{2e^{-(x+y)}}{2e^{-y}(1-e^{-y})} = \frac{e^{-x}}{1-e^{-y}}; \quad 0 \le x \le y$$

This is an example of a truncated distribution. X has an exponential distribution except that values larger than y are removed.

$$f_{Y|X}(y|x) = \frac{2e^{-(x+y)}}{2e^{-2x}} = e^{x-y}; \quad y \ge x$$

This is an example an example of a shifted exponential. Y is exponential on the interval $[x, \infty)$.

Independent Continuous Random Variables

A set of continuous RVs $X_1, X_2, \ldots, X_n$ are independent if and only if

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

for all $x_1, x_2, \ldots, x_n$.

Note that the text defines independence in terms of CDFs instead of the densities. These are equivalent definitions since

$$\frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_n}(x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n)$$

and

$$\int_{\infty}^{x_1}\int_{\infty}^{x_2}\dots\int_{\infty}^{x_n} f_{X_1}(u_1)f_{X_2}(u_2)\dots f_{X_n}(u_n)du_1du_2\dots du_n$$

$$= \prod_{i=1}^{n}\left[\int_{\infty}^{x_i} f_{X_i}(u_i)du_i\right]$$

$$= F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n)$$

Both of these are equivalent to requiring

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

for all sets A and B (with the obvious extension to 3 or more RVs)

The proof of this is similar to that for the discrete case. Replace the sums by integrals.

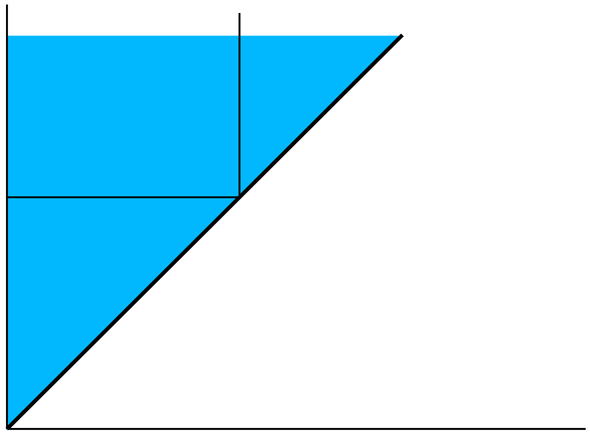
This factorization of densities (or CDFs) gives an easy way to check whether RVs are independent. If the joint density can be written as (using 2 RVs for example where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$)

$$f_{\mathrm{X,Y}}(x, y) = g(x)h(y)\text{for all } x \in \mathcal{X}, y \in \mathcal{Y}$$

with $g(x) \geq 0$ and $h(y) \geq 0$, X and Y are independent. Note that in this factorization g and h don't have to be densities (they will be proportional to the marginal densities).

For example, with $f(x, y) = 9x^2y^2$, this can be decomposed with $g(x) = 9x^2$ and $h(y) = y^2$.

However an example discussed in section 3.3, $f(x, y) = 2x + 2y - 4xy$, X and Y are not independent since there is no decomposition of the valid form.



The condition for all $x \in \mathcal{X}, y \in \mathcal{Y}$ is important. The example where the sample space was defined on the triangle

$$f_{X,Y}(x,y) = 2e^{-(x+y)}; \quad 0 \leq x \leq y, y \geq 0$$

appears that it can be factored in the desired form ($2e^{-(x+y)} = 2e^{-x} \times e^{-y}$). However it doesn't account for the region with 0 probability.

This result isn't usually used to show dependence. To do this, usually you will show that the joint density is not the product of the marginals or in terms of conditional distributions.

There is the similar result with the CDF.

$$F(x,y) = G(y)H(y)$$

with G and H both nondecreasing, non-negative functions.

Two continuous RVs are independent iff

$$f_{Y|X}(y|x) = f_Y(y) \text{ for all } y$$

or

$$f_{X|Y}(x|y) = f_X(x) \text{ for all } x$$

Actually if one holds, the other has to as well.

Technical point: Actually needs to be for all but a countable number of points.

Dependent Continuous Random Variables

As with the discrete case, joint distributions can be built up with the use of conditional distributions.

The joint density of two RVs can be written as

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x) f_{Y|X}(y|x) \\ &= f_Y(y) f_{X|Y}(x|y) \end{aligned}$$

There is the obvious extension to three variables of

$$f_{X,Y,Z}(x,y,z) = f_X(x) f_{Y|X}(y|x) f_{Z|X,Y}(z|x,y)$$

Of course there are versions with all 6 possible orderings of X, Y, and Z.

Example: A model for SAT like scores

Let X be the results of test 1 (e.g. math) and Y be the results of test 2 (e.g. English). A possible model for this this is

$$X \sim N(\mu_X, \sigma_X^2)$$
$$Y|X = x \sim N(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2)$$

where $-1 \leq \rho \leq 1$. If $\rho > 0$ this model suggests that if X is bigger than its mean, Y tends to be bigger than its mean.

For those of you that have seen linear regression

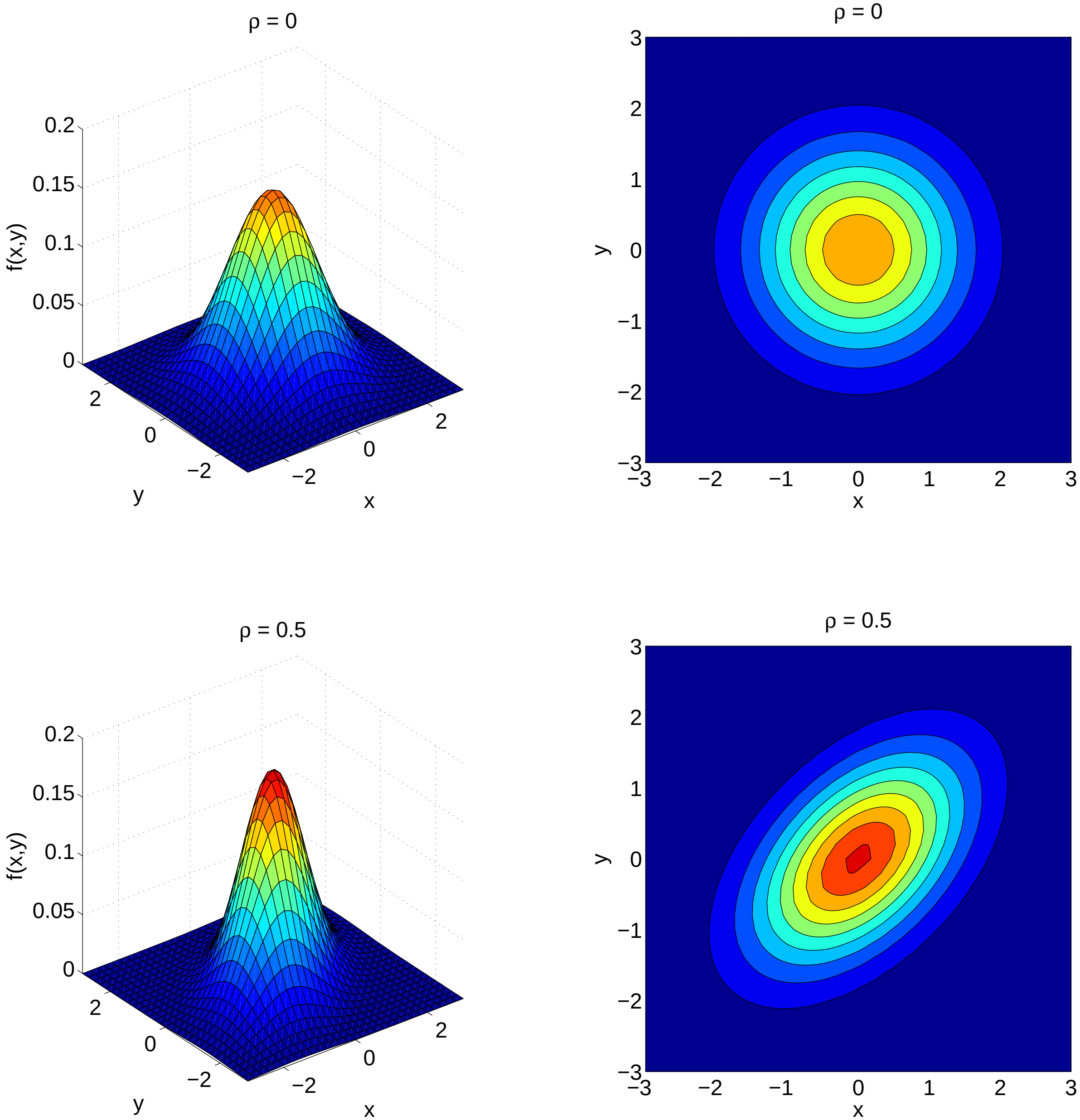
$$E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

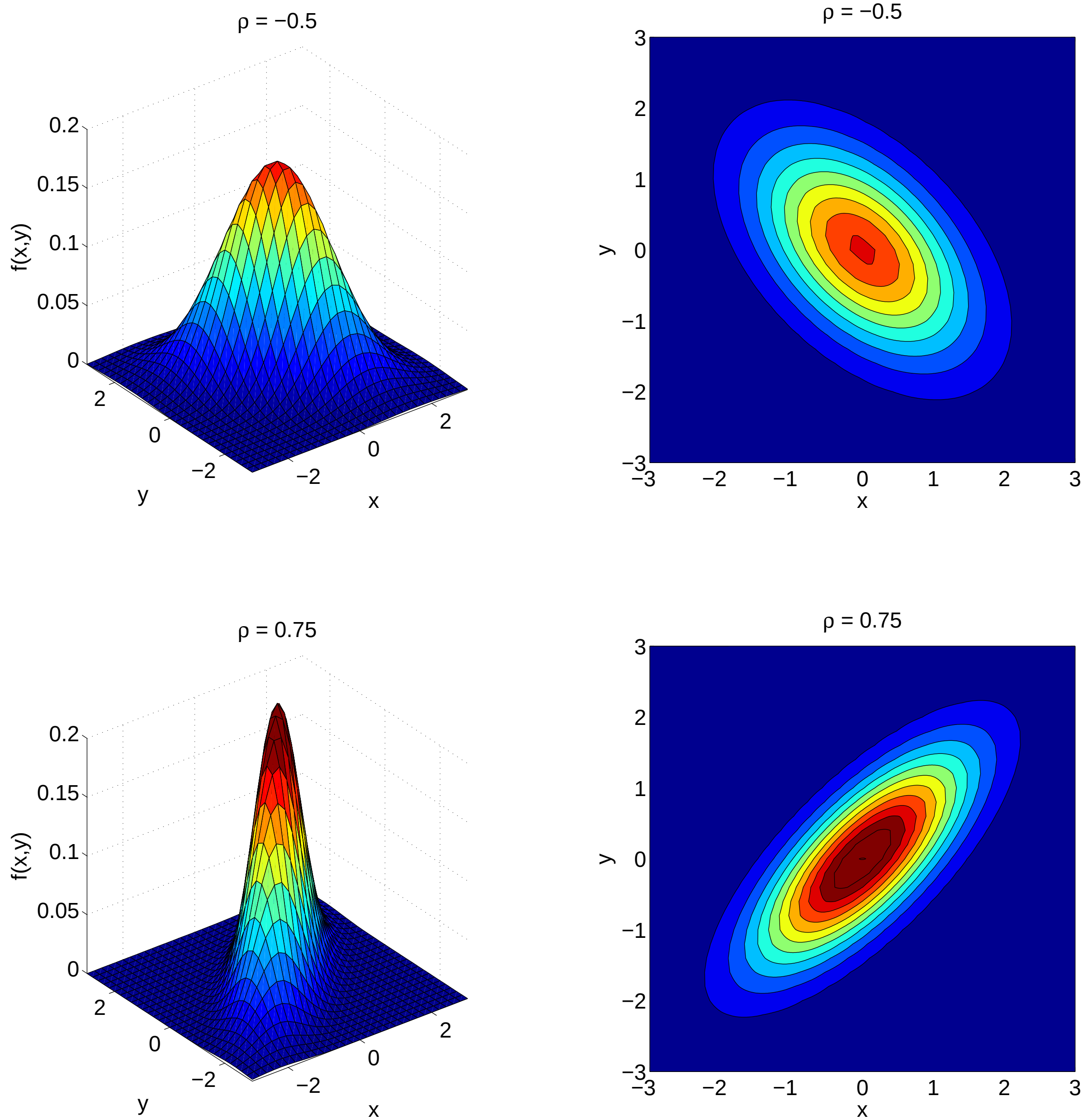
is the population regression line and ρ is the population correlation.

The joint density of X and Y is

$$
\begin{aligned}
f_{X,Y}(x,y) = &\frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right) \\
&\times \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2}\frac{\left(y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)^2}{\sigma_Y^2(1-\rho^2)}\right) \\
= &\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}\right.\right. \\
&\left.\left.+\frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right)
\end{aligned}
$$

This is known as the bivariate normal density.


ρ = 0
f(x,y)
0.2
0.15
0.1
0.05
0
2
0
−2
y
−2
0
2
x
ρ = 0
3
2
1
0
−1
−2
−3
y
−3
−2
−1
0
1
2
3
x
ρ = 0.5
f(x,y)
0.2
0.15
0.1
0.05
0
2
0
−2
y
−2
0
2
x
ρ = 0.5
3
2
1
0
−1
−2
−3
y
−3
−2
−1
0
1
2
3
x


ρ = −0.5
f(x,y)
0.2
0.15
0.1
0.05
0
2
0
−2
y
−2
0
2
x
ρ = −0.5
3
2
1
0
−1
−2
−3
y
−3
−2
−1
0
1
2
3
x
ρ = 0.75
f(x,y)
0.2
0.15
0.1
0.05
0
2
0
−2
y
−2
0
2
x
ρ = 0.75
3
2
1
0
−1
−2
−3
y
−3
−2
−1
0
1
2
3
x

Useful properties of the bivariate normal

- Marginals are univariate normal

$$X \sim N(\mu_X, \sigma_X^2) \text{ and } Y \sim N(\mu_Y, \sigma_Y^2)$$

- Conditional distributions are univariate normal

$$Y|X = x \sim N(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2)$$
$$X|Y = y \sim N(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), (1 - \rho^2)\sigma_X^2)$$

- Sums of normals are normal. If $Z = X + Y$, then

$$Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y)$$

Note that for all the examples presented $\mu_X = \mu_Y = 0$ and $\sigma_X^2 = \sigma_Y^2 = 1$, so the pairs of marginal distributions is the same in all 4 cases. However the joint distributions, and thus the conditional distributions, are all different.

To prove that a sum of normals is normal, you can use the following lemma.

Lemma. *If X and Y are two continuous RVs and $Z = X + Y$, then the PDF of Z is*

$$
\begin{aligned}
f_Z(z) &= \int_{\mathcal{X}} f_{X,Y}(x, z - x)dx \\
&= \int_{\mathcal{Y}} f_{X,Y}(z - y, y)dy
\end{aligned}
$$

Example: Let $X_1, X_2, \ldots X_n$ be independent $Exp(\lambda)$ RVs. Then $S_n = X_1 + X_2 + \ldots + X_n \sim Gamma(n, \lambda)$.

For $S_2 = X_1 + X_2$

$$\begin{aligned} f_{S_2}(s) &= \int_0^\infty f_{X_1}(x) f_{X_2}(s-x) dx \\ &= \int_0^s \left(\lambda e^{-\lambda x}\right)\left(\lambda e^{-\lambda(s-x)}\right) dx \\ &= \lambda^2 e^{-\lambda s} \int_0^s dx = \lambda^2 s e^{-\lambda s} \end{aligned}$$

Then for $S_3 = S_2 + X_3$

$$\begin{aligned} f_{S_3}(s) &= \int_0^\infty f_{S_2}(x) f_{X_3}(s-x) dx = \int_0^s \left(\lambda^2 x e^{-\lambda x}\right)\left(\lambda e^{-\lambda(s-x)}\right) dx \\ &= \lambda^3 e^{-\lambda s} \int_0^s x dx = \frac{\lambda^3 s^2}{2} e^{-\lambda s} \end{aligned}$$

which is a $Gamma(3, \lambda)$ density. The general case follows by induction.