Chapter 4: Multiple Random Variables

MATH 667-01
Statistical Inference
University of Louisville
Textbook: Statistical Inference by Casella and Berger

September 10, 2015
Last corrected: 9/21/15

- *Definition*: An n*-dimensional random vector* is a function from a sample space S into $\mathbb{R}^n$, n-dimensional Euclidean space.
- *Definition*: Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ defined by $f(x, y) = P(X = x, Y = y)$ is called the *joint probability mass function* or *joint pmf* of (X, Y). Sometimes, if necessary, the notation $f_{X,Y}(x, y)$ will be used.
- Let $g(x, y)$ be a real-valued function defined for all possible values (x, y) of the discrete random vector (X, Y). Then $g(X, Y)$ is a random variable with expected value

$$\mathsf{E}g(X, Y) = \sum_{(x,y)\in\mathbb{R}^2} g(x, y) f(x, y).$$

- *Theorem*: Let (X, Y) be a discrete bivariate random vector with joint pmf $f_{X,Y}(x, y)$. Then the marginal pmfs of X and Y are given by

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \text{ and } f_Y(y) = \sum_{x \in \mathbb{R}} f_{X,Y}(x, y).$$

- *Example*: Suppose that X and Y are discrete random variables with pmf

$$f(x, y) = \frac{xy}{11} I_{\{2,3\}}(x) I_{\{1,2\}}(y) I_{(0,x)}(y).$$

Compute $\mathrm{E}XY$ and $\mathrm{E}X$. (*Answers*: $\frac{49}{11}$ and $\frac{31}{11}$)

- *Definition*: A function $f(x,y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ is called a *joint probability density function* or *joint pdf of the continuous bivariate random vector* (X,Y) if, for every $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \int\int_{A} f(x,y)\ dx\ dy.$$

- If $g(x,y)$ be a real-valued function, then the *expected value of* $g(X,Y)$ is

$$\mathsf{E}g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\ dx\ dy.$$

- *Example*: Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{45}{16}xy(x-y)(2-x) & \text{if } 0 < y < x < 2 \\ 0 & \text{otherwise} \end{cases} .$$

- The density function is shown in Figure 1.

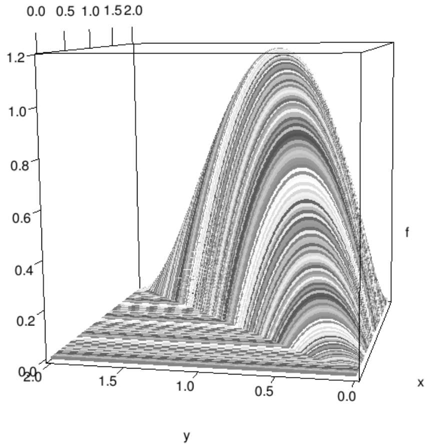

Figure 1: Density function for the joint distribution of X and Y.

Example continued:
Now suppose we want to compute the probability that $X > Y^2$.
To compute this probability, we can integrate with respect to y first and consider the region of integration shown in Figure 2; the infinitesimal slices shown in Figure 2 correspond to the density shown in Figure 3.

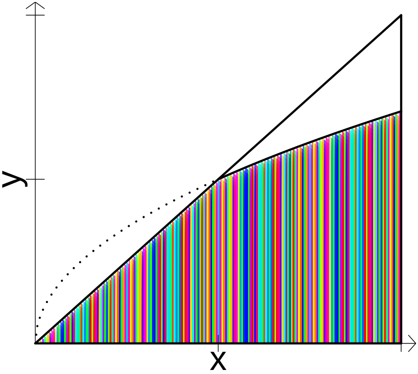

Figure 2: Region of integration that can be used to integrate with respect to y first.

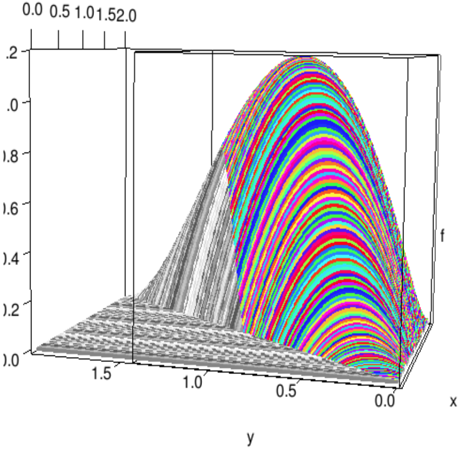

Figure 3: Density function for the joint distribution of X and Y, shown with the region of integration shaded with infinitesimal slices that can be used to integrate with respect to y first.

Example continued:
Then the probability can be computed as follows.

$$
\begin{aligned}
P(X > Y^2) &= \int_0^1 \int_0^x \frac{45}{16} xy(x-y)(2-x)\, dy\, dx + \int_1^2 \int_0^{\sqrt{x}} \frac{45}{16} xy(x-y)(2-x)\, dy\, dx \\
&= \frac{45}{16} \left\{ \int_0^1 \int_0^x (2x^2y - x^3y - 2xy^2 + x^2y^2)\, dy\, dx \right. \\
&\quad \left. + \int_1^2 \int_0^{\sqrt{x}} (2x^2y - x^3y - 2xy^2 + x^2y^2)\, dy\, dx \right\} \\
&= \frac{45}{16} \left\{ \int_0^1 \left[x^2y^2 - \frac{1}{2}x^3y^2 - \frac{2}{3}xy^3 + \frac{1}{3}x^2y^3 \right]_0^x dx \right. \\
&\quad \left. + \int_1^2 \left[x^2y^2 - \frac{1}{2}x^3y^2 - \frac{2}{3}xy^3 + \frac{1}{3}x^2y^3 \right]_0^{\sqrt{x}} dx \right\} \\
&= \frac{45}{16} \left\{ \int_0^1 \left(\frac{1}{3}x^4 - \frac{1}{6}x^6 \right) dx + \int_1^2 \left(-\frac{2}{3}x^{5/2} + x^3 + \frac{1}{3}x^{7/2} - \frac{1}{2}x^4 \right) dx \right\} \\
&= \frac{45}{16} \left\{ \left[\frac{1}{15}x^5 - \frac{1}{36}x^6 \right]_0^1 + \left[-\frac{4}{21}x^{7/2} + \frac{1}{4}x^4 + \frac{2}{27}x^{9/2} - \frac{1}{10}x^5 \right]_1^2 \right\} \\
&= \frac{45}{16} \left\{ \frac{7}{180} + \left(-\frac{64}{189}\sqrt{2} + \frac{2897}{3780} \right) \right\} \approx 0.918.
\end{aligned}
$$

Alternately, we can compute the probability by integrating with respect to x first and consider the infinitesimal slices and region of integration shown in Figures 4 and 5.

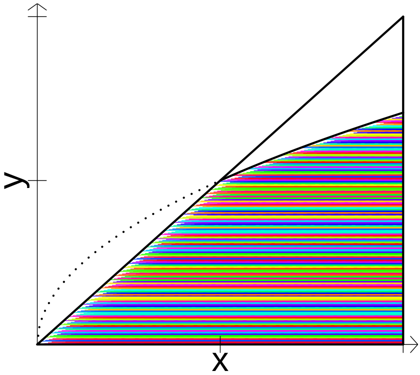

Figure 4: Region of integration that can be used to integrate with respect to x first.

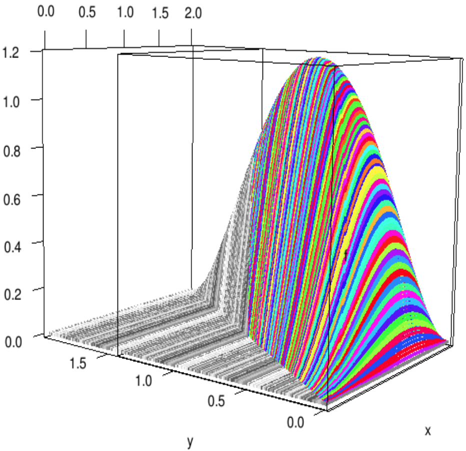

Figure 5: Density function for the joint distribution of X and Y, shown with the region of integration shaded with infinitesimal slices that can be used to integrate with respect to x first.

Example continued:
This is computed by calculating

$$P(X > Y^2) = \int_0^1 \int_y^2 \frac{45}{16} xy(x-y)(2-x)\; dx\; dy$$

$$+ \int_1^{\sqrt{2}} \int_{y^2}^2 \frac{45}{16} xy(x-y)(2-x)\; dx\; dy \approx 0.918.$$

Alternately, we could compute this probability as follows.

$$P(X > Y^2) = 1 - P(X \le Y^2) = 1 - \int_1^2 \int_{\sqrt{x}}^x \frac{45}{16} xy(x-y)(2-x)\; dy\; dx \approx 0.918.$$

- The marginal pdfs of X and Y are defined by

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

and

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

Example continued:

Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{45}{16}xy(x-y)(2-x) & \text{if } 0<y<x<2 \\ 0 & \text{otherwise} \end{cases}.$$

When $0 < x < 2$, the marginal density of X is

$$\begin{aligned} f_X(x) &= \int_0^x \frac{45}{16}(2x^2y - x^3y - 2xy^2 + x^2y^2)\,dy \\ &= \frac{45}{16}\left[x^2y^2 - \frac{1}{2}x^3y^2 - \frac{2}{3}xy^3 + \frac{1}{3}x^2y^3\right]_0^x \\ &= \frac{45}{16}\left(\frac{1}{3}x^4 - \frac{1}{6}x^5\right) \\ &= \frac{15}{32}x^4(2-x) \end{aligned}$$

Example continued:

so that

$$f_X(x) = \begin{cases} \frac{15}{32}x^4(2-x) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases} .$$

The marginal density of X is shown in Figure 6.

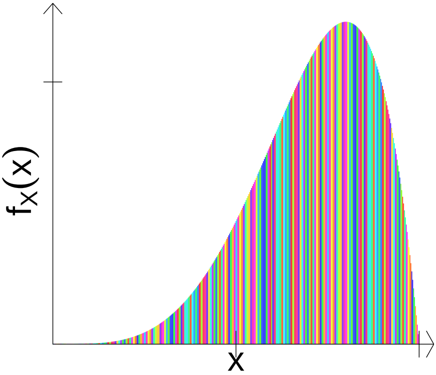

Figure 6: Density function for the marginal distribution of X.

- The *joint cdf* (cumultive distribution function) is the function $F(x, y)$ defined by

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

for all $(x, y) \in \mathbb{R}^2$.

- For a continuous bivariate random vector, $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, t)\ dt\ ds$. From the bivariate Fundamental Theorem of Calculus, this implies that

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

at the continuity points of $f(x, y)$.

- The *joint moment generating function* of X and Y is

$$M_{X,Y}(t_1, t_2) = \mathsf{E}e^{t_1 X + t_2 Y}.$$

- The moments of X and Y can be obtained in a manner analogous to the univariate case:

$$\mathsf{E}(X^n Y^m) = \frac{\partial^{n+m}}{\partial^n \partial^m} M_{X,Y}(t_1, t_2)\,|_{t_1=t_2=0}\ .$$

- *Definition*: Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the *conditional pmf of Y given that $X = x$* is the function of y defined by

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}.$$

For any y such that $P(Y = y) = f_Y(y) > 0$, the *conditional pmf of X given that $Y = y$* is the function of x defined by

$$f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}.$$

- If $g(Y)$ is a function of a discrete random variable Y, then the *conditional expected value of $g(Y)$ given that $X = x$* is

$$\mathsf{E}(g(Y)|x) = \sum_y g(y) f(y|x).$$

- *Example*: Suppose that X and Y are discrete random variables with pmf

$$f(x, y) = \frac{xy}{11} I_{\{2,3\}}(x) I_{\{1,2\}}(y) I_{(0,x)}(y).$$

Then the conditional pmf of X given $Y = 1$ is

$$f(x|1) = \frac{x}{5} I_{\{2,3\}}(x)$$

and the conditional pmf of X given $Y = 2$ is

$$f(x|2) = I_{\{3\}}(x).$$

- *Definition*: Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the *conditional pmf of* Y *given that* $X = x$ is the function of y defined by

$$f(y|x) = \frac{f(x, y)}{f_X(x)}.$$

  For any y such that $f_Y(y) > 0$, the *conditional pmf of* X *given that* $Y = y$ is the function of x defined by

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

- If $g(Y)$ is a function of a continuous random variable Y, then the *conditional expected value of* $g(Y)$ *given that* $X = x$ is

$$\mathsf{E}(g(Y)|x) = \int_{-\infty}^{\infty} g(y) f(y|x)\, dy.$$

- *Example*: Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y)=\begin{cases}\frac{45}{16}xy(x-y)(2-x) & \text{if } 0<y<x<2\\ 0 & \text{otherwise}\end{cases}.$$

When $0<y<x$, the conditional density function of Y given $X=x$ is

$$f_{Y|X=x}(y)=\frac{\frac{45}{16}xy(x-y)(2-x)}{\frac{15}{32}x^4(2-x)}=\frac{6y(x-y)}{x^3}$$

so that

$$f_{Y|X=x}(y)=\begin{cases}6x^{-3}y(x-y) & \text{if } 0<y<x\\ 0 & \text{otherwise}\end{cases}.$$

The conditional density of Y given $X=1$ is shown in Figures 8 and 9.

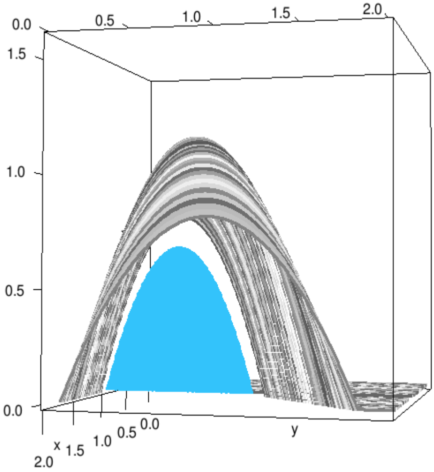

Figure 8: Conditional density function for distribution of Y given $X = 1$, shown with respect to the original joint density.

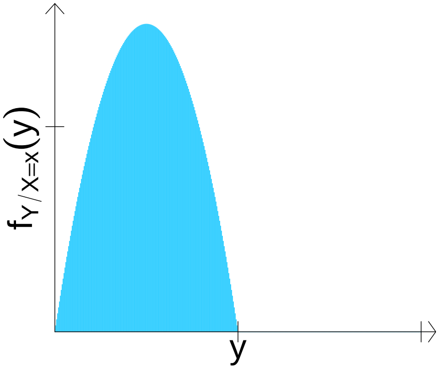

Figure 9: Conditional density function for distribution of Y given $X = 1$.

Example continued:
The moments of this conditional distribution can be computed as follows.

$$\begin{aligned} \mathsf{E}[Y^n|X=x] &= \int_0^x 6x^{-3}y^{1+n}(x-y)\ dy \\ &= 6x^{-3}\int_0^x (xy^{1+n}-y^{2+n})\ dy \\ &= 6x^{-3}\left[\frac{xy^{2+n}}{2+n}-\frac{y^{3+n}}{3+n}\right]_0^x \\ &= \frac{6x^n}{(2+n)(3+n)}. \end{aligned}$$

Thus, it follows that $\mathsf{E}[Y|X=x]=\dfrac{x}{2}$, $\mathsf{E}[Y^2|X=x]=\dfrac{3x^2}{10}$, and $\mathsf{Var}\,[Y|X=x]=\dfrac{3x^2}{10}-\left(\dfrac{x}{2}\right)^2=\dfrac{x^2}{20}$.

- *Definition*: Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called *independent random variables* if, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,

$$f(x, y) = f_X(x) f_Y(y).$$

- *Lemma*: Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$, $f(x, y) = g(x)h(y)$.
- *Theorem*: Let X and Y be independent random variables.
  a. For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$, that is the events $\{X \in A\}$ and $\{Y \in B\}$ are independent events.
  b. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then

$$\mathsf{E}(g(X)h(Y)) = (\mathsf{E}g(X))(\mathsf{E}h(Y)).$$

- *Theorem*: Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of $Z = X + Y$ is given by

$$M_Z(t) = M_X(t)M_Y(t).$$

- *Theorem*: Let $X \sim \mathsf{n}(\mu, \sigma^2)$ and $Y \sim \mathsf{n}(\gamma, \tau^2)$ be independent normal random variables. Then the random variable $Z = X + Y$ has a $\mathsf{n}(\mu + \gamma, \sigma^2 + \tau^2)$ distribution.

- *Convolution method*: If X_1 and X_2 are discrete non-negative integer-valued random variables with probability function $f(x_1, x_2)$, then, for an integer k,

$$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f(x_1, k - x_1).$$

- If X_1 and X_2 are continuous random variables with joint density $f(x_1, x_2)$, then $Y = X_1 + X_2$ has density

$$f_Y(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1)\ dx_1.$$

- *Example*: Suppose X follows a Binomial(n, p) distribution, Y follows a Bernoulli(p) distribution, and X and Y are independent. Then we have

$$
\begin{aligned}
P(X+Y=k) &= P(X=k-1, Y=1) + P(X=k, Y=0) \\
&= P(X=k-1)P(Y=1) + P(X=k)P(Y=0) \\
&= \left(\binom{n}{k-1} p^{k-1}(1-p)^{n-k+1}\right) p + \left(\binom{n}{k} p^{k}(1-p)^{n-k}\right)(1-p) \\
&= \left(\binom{n}{k-1} + \binom{n}{k}\right) p^{k}(1-p)^{n-k+1} \\
&= \binom{n+1}{k} p^{k}(1-p)^{n-k+1}.
\end{aligned}
$$

So, $X+Y$ follows a Binomial($n+1$, p) distribution.

- Bivariate continuous case: Suppose (X, Y) has density $f(x, y)$ and $U = u(X, Y)$ and $V = v(X, Y)$ where u and v have inverses such that there is an h_1 and h_2 where $x = h_1(u(x, y), v(x, y))$ and $y = h_2(u(x, y), v(x, y))$. Then the joint density of U and V is

$$g(u, v) = f(h_1(u, v), h_2(u, v))\,|J|$$

where

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

- *Example*: Suppose that X and Y are independent Exponential random variables each with mean 3 so that their joint density function is

$$f(x, y) = \left(\frac{1}{3}e^{-x/3}\right)\left(\frac{1}{3}e^{-y/3}\right) = \frac{1}{9}e^{-(x+y)/3}$$

for $x > 0$ and $y > 0$. Consider the bivariate transformation $U = X + Y$ and $V = Y$. Then $u = x + y$ and $v = y$ imply that $x = u - v$ and $y = v$ so that

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} = (1)(1) - (-1)(0) = 1.$$

Example continued:

Thus, it follows that

$$\begin{aligned} f(x,y)\ dx\ dy &= f(u-v,v)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\ du\ dv \\ &= \frac{1}{9}e^{-u/3}|1|\ du\ dv = \frac{1}{9}e^{-u/3}\ du\ dv \end{aligned}$$

so that

$$g(u,v) = \frac{1}{9}e^{-u/3}$$

for $u > v > 0$ ($x = u - v > 0$ and $y = v > 0$). Then suppose we want the marginal density for the sum $U = X + Y$. We obtain this by integrating out the v and get

$$g_U(u) = \int_0^u g(u,v)\ dv = \frac{1}{9}ue^{-u/3}$$

for $u > 0$. Thus, U follows a Gamma($\alpha = 2$, $\beta = \frac{1}{3}$) distribution.

- *Theorem*: Let X and Y be independent random variables. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.

- $\mathsf{E}X = \mu_X$, $\mathsf{E}Y = \mu_Y$, Var $X = \sigma_X^2$, Var $Y = \sigma_Y^2$
- Assume $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$
- The *covariance of* X and Y is the number defined by

$$\mathsf{Cov}(X, Y) = \mathsf{E}((X - \mu_X)(Y - \mu_Y)).$$

- *Definition*: The *correlation of* X *and* Y is the number defined by

$$\rho_{XY} = \frac{\mathsf{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

The value ρ_{XY} is also called the *correlation coefficient*.

- *Theorem*: For any random variables X and Y,

$$\mathsf{Cov}(X,Y) = \mathsf{E}XY - \mu_X \mu_Y.$$

- *Theorem*: If X and Y are independent random variables, then $\mathsf{Cov}(X,Y) = 0$ and $\rho_{XY} = 0$.
- *Theorem*: If X and Y are any two random variables, and a and b are any two constants, then

$$\mathsf{Var}\ (aX + bY) = a^2 \mathsf{Var}\ X + b^2 \mathsf{Var}\ Y + 2ab\mathsf{Cov}(X,Y).$$

  If X and Y are independent random variables, then

$$\mathsf{Var}\ (aX + bY) = a^2 \mathsf{Var}\ X + b^2 \mathsf{Var}\ Y.$$

- *Theorem*: For any random variables X and Y,
  a. $-1 \leq \rho_{XY} \leq 1$.
  b. $|\rho_{XY}| = 1$ if and only if there exists numbers $a \neq 0$ and b such that $P(Y = aX + b) = 1$. If $\rho_{XY} = 1$ then $a > 0$, and if $\rho_{XY} = -1$ then $a < 0$.

- *Example*: Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{45}{16}xy(x-y)(2-x) & \text{if } 0 < y < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

Then, we have

$$\begin{aligned}
\mathsf{E}[X^m Y^n] &= \int_0^2 \int_0^x x^m y^n \frac{45}{16} xy(x-y)(2-x)\ dy\ dx \\
&= \frac{45}{16} \int_0^2 \int_0^x \left((2x^{2+m} - x^{3+m})y^{1+n} - (2x^{1+m} - x^{2+m})y^{2+n} \right)\ dy\ dx \\
&= \frac{45}{16} \int_0^2 \left[\frac{2x^{2+m} - x^{3+m}}{2+n} y^{2+n} - \frac{2x^{1+m} - x^{2+m}}{3+n} y^{3+n} \right]_0^x\ dx \\
&= \frac{45}{16(2+n)(3+n)} \int_0^2 (2x^{4+m+n} - x^{5+m+n})\ dx \\
&= \frac{45}{16(2+n)(3+n)} \left[\frac{2}{5+m+n} x^{5+m+n} - \frac{1}{6+m+n} x^{6+m+n} \right]_0^2 \\
&= \frac{45}{16(2+n)(3+n)} \frac{2^{6+m+n}}{(5+m+n)(6+m+n)} \\
&= \frac{180}{(2+n)(3+n)(5+m+n)(6+m+n)} 2^{m+n}.
\end{aligned}$$

Example continued:

Thus, it follows that

$$\mathsf{E}X = \frac{10}{7}, \mathsf{E}Y = \frac{5}{7}, \mathsf{E}[XY] = \frac{15}{14}, \mathsf{E}[X^2] = \frac{15}{7}, \mathsf{E}[Y^2] = \frac{9}{14}.$$

Then we can compute

$$\mathsf{Cov}[X, Y] = \frac{15}{14} - \left(\frac{10}{7}\right)\left(\frac{5}{7}\right) = \frac{5}{98},$$

$$\mathsf{Var}\ X = \frac{15}{7} - \left(\frac{10}{7}\right)^2 = \frac{5}{49},$$

$$\mathsf{Var}\ Y = \frac{9}{14} - \left(\frac{5}{7}\right)^2 = \frac{13}{98},$$

and

$$\rho_{X,Y} = \frac{5/98}{\sqrt{10/98}\sqrt{13/98}} = \frac{5}{\sqrt{130}} \approx 0.4385.$$

- *Definition*: Let $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $0 < \sigma_X$, $0 < \sigma_Y$, and $-1 < \rho < 1$ be five real numbers. The *bivariate normal pdf with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ* is the bivariate pdf given by

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]}$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

- *Definition*: A random variable X is said to have a *mixture distribution* if the distribution of X depends on a quantity which also has a distribution.
- *Theorem*: If X and Y are any two random variables, then

$$\mathsf{E}X = \mathsf{E}(\mathsf{E}(X|Y)),$$

provided that the expectations exist.
- *Theorem*: For any two random variables X and Y,

$$\mathsf{Var}\ X = \mathsf{E}(\mathsf{Var}(X|Y)) + \mathsf{Var}(\mathsf{E}(X|Y))$$

provided that the expectations exist.

- *Example*: Suppose X and Y are continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{45}{16}xy(x-y)(2-x) & \text{if } 0<y<x<2 \\ 0 & \text{otherwise} \end{cases}.$$

Then we have

$$\mathsf{E}Y = \mathsf{E}[\mathsf{E}[Y|X]] = \mathsf{E}\left[\frac{X}{2}\right] = \frac{1}{2}\mathsf{E}X = \frac{1}{2}\left(\frac{10}{7}\right) = \frac{5}{7}$$

and

$$\text{Var } Y = \text{Var } [\mathsf{E}[Y|X]] + \mathsf{E}[\text{Var } [Y|X]] = \text{Var } \left[\frac{X}{2}\right] + \mathsf{E}\left[\frac{X^2}{20}\right]$$

$$= \frac{1}{4}\text{Var } X + \frac{1}{20}\mathsf{E}[X^2] = \frac{1}{4}\left(\frac{5}{49}\right) + \frac{1}{20}\left(\frac{15}{7}\right) = \frac{13}{98}.$$

- *Definition*: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors with point pdf or pmf $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$ denote the marginal pdf or pmf of $\boldsymbol{X}_i$. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are called *mutually independent random vectors* if, for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \bullet \cdots \bullet f_{\boldsymbol{X}_n}(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\boldsymbol{X}_i}(\boldsymbol{x}_i).$$

  If the X_is are all one-dimensional, then $X_1, \ldots, X_n$ are called *mutually independent random variables*.

- *Theorem*: Let $X_1, \ldots, X_n$ be mutually independent random variables. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(x_i)$ is a function only of $x_i, i = 1, \ldots, n$. Then

$$\mathsf{E}(g_1(X_1) \bullet \cdots \bullet g_n(X_n)) = (\mathsf{E} g_1(X_1)) \bullet \cdots \bullet (\mathsf{E} g_n(X_n)).$$

- *Theorem*: Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $Z = X_1 + \ldots + X_n$. Then the mgf of Z is

$$M_Z(t) = M_{X_1}(t) \bullet \cdots \bullet M_{X_n}(t).$$

In particular, if $X_1, \ldots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = (M_X(t))^n.$$

- *Corollary*: Let $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_1}(t), \ldots, M_{X_n}(t)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Let $Z = (a_1 X_1 + b_1) + \cdots + (a_n X_n + b_n)$. Then the mgf of Z is

$$M_Z(t) = (e^{t(\sum b_i)}) M_{X_1}(a_1 t) \bullet \cdots \bullet M_{X_n}(a_n t).$$

- *Corollary*: Let $X_1, \ldots, X_n$ be mutually independent random variables with $X_i \sim \mathsf{n}(\mu_i, \sigma_i^2)$. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be fixed constants. Then

$$Z = \sum_{i=1}^{n} (a_i X_i + b_i) \sim \mathsf{n}\left(\sum_{i=1}^{n} (a_i \mu_i + b_i), \sum_{i=1}^{n} a_i^2 \sigma_i^2 \right).$$

- *Theorem*: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are mutually independent random vectors if and only if there exist functions $g_i(x_i), i = 1, \ldots, n$, such that the joint pdf or pmf of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ can be written as

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = g_1(\boldsymbol{x}_1) \bullet \cdots \bullet g_n(\boldsymbol{x}_n).$$

- *Theorem*: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random vectors. Let $g_i(\boldsymbol{x}_i)$ be a function only of $\boldsymbol{x}_i, i = 1, \ldots, n$. Then the random variables $U_i = g_i(\boldsymbol{X}_i), i = 1, \ldots, n$, are mutually independent.