4.1. Joint & Marginal distribution

"I confess that I have been blind as a mole, but it is better to learn wisdom late than never to learn it at all. Sherlock Holmes, The Man With the Twisted Lip."

• Instead of a univariate X, we will consider a random vector $\mathbf{X} = (X_1, \cdots, X_n)'$.

Definition

n-dimensional random vector $\mathbf{X} = (X_1, \cdots, X_n)'$ is a function from the sample space to n dimensional Euclidean space.

$$\mathcal{S} \xrightarrow{\mathbf{X}} \mathfrak{R}^n$$

The *joint cdf* of $(X_1, \cdots, X_n)'$ is

$$F(x_1, \cdots, x_n) = Pr[X_1 \leq x_1, \cdots, X_n \leq x_n]$$

4.1. Joint & Marginal distribution

- Properties of the joint cdf

1.
$$F \text{ is right-continuous in each } x_j$$

2.
$$\lim_{x_j \to -\infty} F(x_1, \cdots, x_n) = 0, \quad j = 1, \cdots, n$$

3.
$$\lim_{x_j \to \infty \text{ for all } j} F(x_1, \cdots, x_n) = 1$$

4.
$$\begin{aligned} F_{X_j}(x_j) =& F(\infty, \cdots, \infty, x_j, \infty, \cdots, \infty) \\ =& Pr[X_1 \leq \infty, \cdots, X_j \leq x_j, \cdots, X_n \leq \infty] \end{aligned}$$

4.1. Joint & Marginal distribution

Definition (Discrete random vector)

F is *discrete* [or $\mathbf{X}$ is a *discrete random vector*] if there exists a nonnegative function $f(x_1, \cdots, x_n)$ that is zero except on countable set $\mathrm{S}(X_1, \cdots, X_n)$[support of distribution] and is such that

$$F(x_1, \cdots, x_n) = \sum_{x_1', \cdots, x_n' \in A} f(x_1', \cdots, x_n'),$$

where $A = \{(-\infty, x_1] \times \cdots \times (-\infty, x_n]\} \cap \mathrm{S}(X_1, \cdots, X_n)$.

Then, the nonnegative function $f(x_1, \cdots, x_n)$ is called *(multivariate) pmf* of $X_1, \cdots X_n$ and

$$\sum_{x_1} \cdots \sum_{x_n} f(x_1, \cdots, x_n) = 1$$

4.1. Joint & Marginal distribution

Definition (Continuous random vector)

F is *(absolute) continuous* [or **X** is a *(absolute) continuous random vector] if there exists a nonnegative function* $f(x_1, \cdots, x_n)$ *such that*

$$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(x_1', \cdots, x_n') dx_n' \cdots dx_1'.$$

Then, the nonnegative function $f(x_1, \cdots, x_n)$ *is called* (multivariate) pdf *of* $X_1, \cdots X_n$ *and*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_n \cdots dx_1 = 1,$$

$$\frac{\partial^n}{\partial x_1, \cdots \partial x_n} F(x_1, \cdots, x_n) = f(x_1, \cdots, x_n).$$

4.1. Joint & Marginal distribution

Definition (Probability, Expectation)

$$\bullet P[(X_1, \cdots, X_n) \in \mathbf{A}]$$
$$= \sum \cdots \sum_{\mathbf{A}} f(x_1, \cdots, x_n), \text{discrete},$$
$$= \int \cdots \int_{\mathbf{A}} f(x_1, \cdots, x_n) dx_1 \cdots dx_n, \text{continuous}.$$

$$\bullet E[g(X_1, \cdots, X_n)]$$
$$= \sum \cdots \sum g(x_1, \cdots, x_n) f(x_1, \cdots, x_n), \text{discrete},$$
$$= \int \cdots \int g(x_1, \cdots, x_n) f(x_1, \cdots, x_n) dx_1 \cdots dx_n, \text{continuous}.$$

4.1. Joint & Marginal distribution

Definition (Marginal pmf, pdf)

- $f_{X_j}(x_j)$

$$= \sum \cdots \sum_{x'_1,\cdots,x'_{j-1},x'_{j+1},\cdots,x'_n} f(x'_1, \cdots, x'_n), \text{discrete},$$

$$= \underset{x'_1,\cdots,x'_{j-1},x'_{j+1},\cdots,x'_n}{\int \cdots \int} f(x'_1, \cdots, x'_n) dx'_1 \cdots dx'_n, \text{continuous}.$$

- $f_{X_h X_j}(x_h, x_j)$

$$= \sum \cdots \sum_{?} f(x'_1, \cdots, x'_n), \text{discrete},$$

$$= \int \cdots_{?} \int f(x'_1, \cdots, x'_n) dx'_1 \cdots dx'_n, \text{continuous}.$$

4.1. Joint & Marginal distribution

▷ Example

		$X = x$			$f_Y(y)$
		1	2	3	
	1	1/6	1/12	1/12	
$Y = y$	2	1/12	1/6	1/12	
	3	1/12	1/12	1/6	
	$f_X(x)$				

$$f(x,y) = \begin{cases} 1/6, & \text{if } x = y, \quad x, y = 1, 2, 3 \\ 1/12, & \text{if } x \neq y, \quad x, y = 1, 2, 3 \end{cases}$$

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

$P[X \geq Y] =$, $E[XY] =$

4.1. Joint & Marginal distribution

▷ Example

$$f(x,y) = \frac{1}{x},\ 0 < y < x < 1.$$

- Support:
- Is this a joint pdf ?

$$f_X(x) = \int_0^x f(x,y)dy, \qquad f_Y(y) = \int_y^1 f(x,y)dx$$

$P[X + Y \leq 1] =$

$E[XY] =$

▷ Example

$$f(x, y) = e^{-y}, \ 0 < x < y < \infty.$$

- Support:
- Is this a joint pdf ?

$$f_X(x) = \qquad\qquad\qquad , f_Y(y) =$$

$P[X + Y \leq 1] =$

$E[XY] =$

4.1. Joint & Marginal distribution

▷ Example

$$f(x,y) = \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda}\lambda^y}{y!}, \; x = 0, 1, \cdots, y, \; y = 0, 1, \cdots.$$

- Support:

$f_X(x) =$

$f_Y(y) =$

$E[XY] =$

4.2. Independence & conditional distribution

Definition

$X_1, \cdots, X_n$ are *mutually independent random variables* if and only if for any Borel sets $A_1, \cdots, A_n$ in $\mathfrak{R}$, the events $\{X_1 \in A_1\}, \cdots,$ $\{X_n \in A_n\}$ are mutually independent.

- if and only if

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

  for all $(x_1, \cdots, x_n) \in \mathfrak{R}^n$.

- if and only if

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

  for all $(x_1, \cdots, x_n) \in \mathfrak{R}^n$.

4.2. Independence & conditional distribution

Theorem

If $X_1, \cdots, X_n$ are mutually independent random variables, then

$$E[h_1(X_1)h_2(X_2)\cdots h_n(X_n)] = \prod_{i=1}^{n} E[h_i(X_i)] \ .$$

Proof.

4.2. Independence & conditional distribution

Lemma

X and Y are independent if and only if there exist nonnegative functions g and h of x and y only, respectively such that

$$f(x,y) = g(x)h(y), \quad \textit{for all} \;\; x, y$$

▷ Example:

$$f(x,y) = e^{-4}\frac{2^{x+y}}{x!y!}, \; x = 0, 1, \cdots, \; y = 0, 1, \cdots.$$

Are X and Y independent ?

4.2. Independence & conditional distribution

▷ Example:

$$f(x, y) = x + y, \quad 0 < x < 1, \quad 0 < y < 1.$$

Are X and Y independent ?

▷ Example:

$$f(x, y) = 8xy, \quad 0 < x < y < 1.$$

Are X and Y independent ?

4.2. Independence & conditional distribution

Lemma

If X and Y are independent, so also are $U = g(X)$ and $V = h(Y)$.

Definition

For discrete random variables X and Y with the joint pmf $f(x, y)$ for each x with $f_X(x) > 0$, the *conditional pmf of Y given $X = x$* is

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

Similarly

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

4.2. Independence & conditional distribution

Definition

For continuous random variables X and Y with the joint pdf $f(x, y)$ for each x with $f_X(x) > 0$, the *conditional pdf of* Y *given* $X = x$ is

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

Similarly

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

4.2. Independence & conditional distribution

To define the conditional pdf of the continuous r.v., conditional cdf is used. The conditional cdf is

$$
\begin{aligned}
&\lim_{h\to 0} P[X \le x | y \le Y \le y+h] \\
=&\lim_{h\to 0} \frac{(1/h)\int_y^{y+h}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy}{(1/h)\int_y^{y+h} f_Y(y)dy} \\
=&\frac{\int_{-\infty}^{x} f_{X,Y}(x,y)dx}{f_Y(y)} \\
=&F(x|y)
\end{aligned}
$$

and thus

$$
f(x|y) = \frac{\partial}{\partial x}F(x|y) \ .
$$

4.2. Independence & conditional distribution

▷ Example **[D1]**

	$f_{X,Y}(x,y)$	$X=x$: 1	2	3	$f_Y(y)$
$Y=y$	1	1/6	1/12	1/12	1/3
	2	1/12	1/6	1/12	1/3
	3	1/12	1/12	1/6	1/3
	$f_X(x)$	1/3	1/3	1/3	1

- Conditional pmf of X given $Y=y$.

$X=x$	$f(x\|Y=1)$	$f(x\|Y=2)$	$f(x\|Y=3)$
1	(1/6)/(1/3) =1/2	1/4	1/4
2	(1/12)/(1/3)=1/4	1/2	1/4
3	(1/12)/(1/3)=1/4	1/4	1/2

4.2. Independence & conditional distribution

$\triangleright$ Example **[D2]**

$$f_{X,Y}(x,y) = \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda}\lambda^y}{y!}, \; x = 0, 1, \cdots, y, \; y = 0, 1, \cdots .$$

$$Y \sim \text{Poisson}(\lambda), \quad X \sim \text{Poisson}(\lambda p)$$

$f_{Y|X}(y|x) =$ (Result: $[Y|X] \sim \text{Poisson}(\lambda(1-p))$)

$f_{X|Y}(x|y) =$ (Result: $[X|Y] \sim \text{Binomial}(y, p)$)

4.2. Independence & conditional distribution

▷ Example **[C1]**

$$f(x,y) = \frac{1}{x},\ 0 < y < x < 1.$$

$$f_X(x) = 1,\ 0 < x < 1, \quad f_Y(y) = -\ln y,\ 0 < y < 1.$$

$f_{Y|X}(y|x) = \frac{1}{x} I(0 < y < x < 1)$

$f_{X|Y}(x|y) = -\frac{1}{x \log(y)} I(0 < y < x < 1)$

$P[X < 3/4 | Y = 1/2] = \frac{\log(3/4) + \log(2)}{\log(2)}$

4.2. Independence & conditional distribution

▷ Example **[C2]**

$$f(x,y) = e^{-y},\ 0 < x < y < \infty.$$

$$f_X(x) = e^{-x},\ 0 < x < \infty, \quad f_Y(y) = ye^{-y}, 0 < y < \infty.$$

$f_{Y|X}(y|x) =$

$f_{X|Y}(x|y) =$

4.2. Independence & conditional distribution

Definition

Let $f_{Y|X}(y|x)$ be the conditional pmf or pdf of Y given $X = x$. Then, for the function $h(x, y)$, provided

$$\sum_y |h(x,y)| f_{Y|X}(y|x) < \infty \text{ or } \int_y |h(x,y)| f_{Y|X}(y|x) dy < \infty,$$

$$E\left[h(X,Y)|X=x\right] = \begin{cases} \sum_y h(x,y) f_{Y|X}(y|x), & \text{discrete} \\ \int_y h(x,y) f_{Y|X}(y|x) dy, & \text{continuous} \end{cases}$$

This is of course a parallel definition for X conditioning on Y.

4.2. Independence & conditional distribution

- With $h(x,y)=y$, $E[Y|X=x]$: Conditional Mean
- Conditional Variance

$$\begin{aligned} Var(Y|X=x) =& E\left\{[Y-E(Y|X=x)]^2\,|X=x\right\} \\ =& E\left[Y^2|X=x\right]-[E(Y|X=x)]^2 \end{aligned}$$

▷ Example: D1

$$E(Y|X=1)= \qquad , Var(Y|X=1)=$$

$$E(Y|X=2)= \qquad , Var(Y|X=2)=$$

$$E(Y|X=3)= \qquad , Var(Y|X=3)=$$

4.2. Independence & conditional distribution

- Example: **[D2]**

$$([Y|X] \sim \text{Poisson}(\lambda(1-p)), \quad [X|Y] \sim \text{Binomial}(y, p))$$

$$E(X|Y=y) =$$

$$Var(X|Y=y) =$$

$$E(Y|X=x) =$$

$$Var(Y|X=x) =$$

4.2. Independence & conditional distribution

▷ Example: **[C1]**

- $f_{Y|X}(y|x) = \frac{1}{x} I(0 < y < x < 1)$

$$E(Y|X = x) = \int_0^x \frac{y}{x} dy = \frac{1}{x}\frac{x^2}{2} = \frac{x}{2}$$

$$E(Y^2|X = x) = \int_0^x \frac{y^2}{x} dy = \frac{x^3}{6}$$

$$Var(Y|X = x) = \frac{x^2(2x - 3)}{12}$$

4.2. Independence & conditional distribution

Theorem

Assume all conditional means below are well defined and $f_X(x) > 0$. Then

1. $E[ah(X,Y) + b|X = x] = aE[h(X,Y)|X = x] + b$
2. $E[h(X,Y) + g(X,Y)|X = x] =$
   $$E[h(X,Y)|X = x] + E[g(X,Y)|X = x]$$
3. $E[g(X)h(X,Y)|X = x] = g(X)E[h(X,Y)|X = x]$

4.2. Independence & conditional distribution

Lemma

Provided all conditional means (given $X = x$) are well defined for a set of x with probability 1. Then

1. $E[ah(X, Y) + b|X] = aE[h(X, Y)|X] + b$

2. $E[h(X, Y) + g(X, Y)|X] =$
$$E[h(X, Y)|X] + E[g(X, Y)|X]$$

3. $E[g(X)h(X, Y)|X] = g(X)E[h(X, Y)|X]$

4.2. Independence & conditional distribution

Theorem

1. *If* $E[|Y|] < \infty$, *then* $E[Y] = E[E(Y|X)]$.
2. *If* $E\left[Y^2\right] < \infty$, *then*

$$Var[Y] = E[Var(Y|X)] + Var[E(Y|X)].$$

▷ Example: C1

$$X \sim \text{Uniform}(0,1), \quad Y|X \sim \text{Uniform}(0,X)$$

$E[Y] =$
$Var[Y] =$

4.2. Independence & conditional distribution

Theorem

Let X and Y be independent random variables with mgf $M_X(t)$ and $M_Y(t)$. Then the mgf of the random variable $Z = X + Y$ is given by

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

Proof:

▷ Example:

$$X \sim N(\mu, \sigma^2), \quad Y \sim N(\gamma, \tau^2) \longrightarrow Z = X + Y \sim$$

4.2. Independence & conditional distribution

Theorem

Let X_i, $i = 1, \cdots, n$ be independent random variables with $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$. Then

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mu_i, \quad Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \sigma_i^2$$

Proof:

4.3. Bivariate Transformation

- Goal: To find the distribution of $U = g(X, Y)$ or the joint distribution of $U = g_1(X, Y)$ and $V = g_2(X, Y)$

- Discrete random variables
- Continuous random variables
  - Using change of variable technique (Jacobian)
  - Using CDF
- Using MGF (discrete, continuous)

4.3. Bivariate Transformation

Discrete random variable

- Given the joint pmf $f_{X,Y}(x,y) = P[X = x, Y = y]$, derive the joint or marginal pmf of $U = g_1(X,Y)$ and $V = g_2(X,Y)$.

▷ Example: $U = X + Y$, $V = Y$. Then
$f_{U,V}(u,v) = f_{X,Y}(u-v,v)$

$f_U(u) = \sum_v f_{U,V}(u,v)$
$f_V(v) = \sum_u f_{U,V}(u,v)$

▷ Example: $X \sim$ Poisson (λ_1), $Y \sim$ Poisson (λ_2), where X and Y are independent. $U = X + Y$, $V = Y$. Then

$$f_{U,V}(u, v) = f_X(u - v) f_Y(v) = \frac{e^{-\lambda_1} \lambda_1^{(u-v)}}{(u-v)!} \frac{e^{-\lambda_2} \lambda_2^{v}}{v!},$$

where $u = v, v + 1, v + 2, ...,$ and $v = 0, 1, 2, ...$

Then, $f_U(u) = \sum_{v=0}^{\infty} f_{U,V}(u, v) = \sum_{v=0}^{u} f_{U,V}(u, v)$
(since $f_{U,V}(u, v) > 0$ for $v \leq u$, and 0 otherwise.)
Thus, $f_U(u) =$

4.3. Bivariate Transformation

Discrete random variable

- General Rule:
  Given $(X, Y) \sim f_{X,Y}(x, y)$, $U = g_1(X, Y)$, $V = g_2(X, Y)$.
  Assume there exist inverse mapping, $X = h_1(U, V)$ and $Y = h_2(U, V)$. Then

$$\begin{aligned} f_{U,V}(u, v) &= P[U = u, V = v] \\ &= P[g_1(X, Y) = u, g_2(X, Y) = v] \\ &= P[X = h_1(u, v), Y = h_2(u, v)] \\ &= f_{X,Y}[h_1(u, v), h_2(u, v)]. \end{aligned}$$

Let (X, Y) be a continuous random vector with joint pdf $f_{X,Y}(x, y)$. Assume the function $\mathbf{g}(x, y) = [g_1(x, y), g_2(x, y)]$ is an *one to one function from* $\mathcal{A}$ *onto* $\mathcal{B}$ where
$\mathcal{A} = \{(x, y) : f(x, y) > 0\}$ and
$\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y), (x, y) \in \mathcal{A}\}$. Denote the inverse transformation and jacobian as $x = h_1(u, v)$, $y = h_2(u, v)$ and

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

Assume $J \neq 0$. Then

$$f_{U,V}(u, v) = f_{X,Y}[h_1(u, v), h_2(u, v)]|J|, \quad (u, v) \in \mathcal{B}$$

▷ Example: $X \sim N(0,1)$, $Y \sim N(0,1)$. X and Y are independent.

$$U = g_1(X,Y) = \frac{X+Y}{\sqrt{2}}, \quad V = g_2(X,Y) = \frac{X-Y}{\sqrt{2}}$$

Find the joint and marginal distribution of U and V.

▷ Example: Convolution formula

X and Y are independent, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

$$W = g_1(X,Y) = X + Y, \quad V = g_2(X,Y) = Y$$

Find the joint distribution of W and V and marginal distribution of W.

4.3. Bivariate Transformation

Continuous random variable-Jacobian

▷ Example: $X \sim \text{Gamma}(\alpha, 1)$, $Y \sim \text{Gamma}(\beta, 1)$. X and Y are independent.

$$U = g_1(X, Y) = \frac{X}{X+Y}, \quad V = g_2(X, Y) = X + Y$$

Find the marginal distribution of U.

4.3. Bivariate Transformation

[Example 4.3.6.]

▷ Example: $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent.

$$U = \frac{X}{Y}, \quad V = |Y|$$

$$F_{U,V}(u,v) = P[U \leq u, V \leq v]$$

4.3. Bivariate Transformation

Using MGF-discrete or continuous

▷ Example: Independent X and Y. $X \sim$ Poisson (λ_1), $Y \sim$ Poisson (λ_2). $X + Y \sim ?$
[Note 1: For mgf, $E(e^{tx}) = \exp(\lambda_1(e^t - 1))$.]
[Note 2: For transformation, let $U = X + Y, V = Y$.]

4.3. Bivariate Transformation

Using MGF-discrete or continuous

$\triangleright$ Results: X and Y are independent.

$X \sim$ Poisson (λ_1), $Y \sim$ Poisson (λ_2). $X + Y \sim$

$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$. $X + Y \sim$

$X \sim$ Gamma (α_1, β), $Y \sim$ Gamma (α_2, β). $X + Y \sim$

$X \sim$ Binomial (n_1, p), $Y \sim$ Binomial (n_2, p). $X + Y \sim$

$X \sim$ Negative Binomial (r_1, p), $Y \sim$ Negative Binomial (r_2, p).
$X + Y \sim$

4.4. Hierarchical and mixture model

- Main point: By the definition of a conditional distribution

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

$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x)$. Suggests a fruitful way to dream up models, namely in an hierarchical fashion.
$\triangleright$ Example: $X \sim$Uniform(0,1), $Y|X \sim$ Uniform(0,X). The joint pdf of (X, Y) is defined by the marginal and condition al distributions.
$\triangleright$ Example: (Mixture distribution) $P \sim$Uniform(0,1), $X|P \sim$Binomial$(2, P)$. For $0 < p < 1$, $x = 0, 1, 2$,

$$f(x,p) = f_{X|P}(x|p) f_P(p) = \binom{2}{x} p^x (1-p)^{2-x}.$$

4.5. Covariance and correlation

- Measures of the strength of a linear relationship between two random variables

Definition

Provided $E|(X-EX)(Y-EY)|<\infty$, the *covariance* between X and Y is defined as

$$\begin{aligned}\sigma_{XY}=Cov(X,Y)&=E[(X-\mu_X)(Y-\mu_Y)]\\&=EXY-\mu_X\mu_Y,\end{aligned}$$

where $\mu_X=EX$ and $\mu_Y=EY$.

Note:

- $Cov(Y,X)=Cov(X,Y)$
- $Cov(X,X)=Var(X)$

4.5. Covariance and correlation

Definition

The *correlation* between X and Y is defined as

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

where σ_X^2 and σ_Y^2 are the variances of X and Y.

Theorem

$$Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}.$$

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_i a_i^2 \sigma_{X_i}^2 + 2\sum_{i,j}\sum_{i<j} a_i a_j \sigma_{X_i}\sigma_{X_j}.$$

4.5. Covariance and correlation

Theorem

The correlation of X and Y satisfies

$$|\rho_{XY}| \leq 1.$$

The equality holds ($\rho_{XY}^2 = 1$) if and only if $Y(X)$ is a linear function of $X(Y)$.

Proof: [See Theorem 4.5.7]

Definition ($(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$)

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, μ_Y, *variances* σ_X, σ_Y *and correlation* ρ is given by, for $-\infty < x < \infty$, $-\infty < y < \infty$,

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{q}{2(1-\rho^2)}\right],$$

where

$$q = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2.$$

Theorem

$$(X, Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$$

1. $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$
2. $Y|X = x \sim N[\mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X), \sigma_Y^2(1 - \rho^2)]$
   $X|Y = y \sim N[\mu_X + \rho(\sigma_X/\sigma_Y)(y - \mu_Y), \sigma_X^2(1 - \rho^2)]$
3. ρ *is the correlation between* X *and* Y, ρ_{XY}
4. $aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY})$.
5. X *and* Y *are independent if and only if* $\sigma_{XY} = \rho_{XY} = 0$.

4.5. Covariance and correlation

[Exercise 4.47]

$\triangleright$ Example: Marginal normal $\rightarrow$ Bivariate arrow ?
Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ be independent. Define

$$Z = \begin{cases} X, & XY > 0 \\ -X, & XY < 0. \end{cases}$$

What is the marginal distribution of Z and is the joint distribution of (Y, Z) ?

4.6. Multivariate Distribution

Moment Generating Function

- MGF of $\mathbf{X} = (X_1, \cdots, X_n)'$

$$M_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = M_{\mathbf{X}}(\mathbf{t}) = E\left(e^{t_1X_1+\cdots t_nX_n}\right) = E\left(e^{\mathbf{t}'\mathbf{X}}\right),$$

where $\mathbf{t} = (t_1, \cdots, t_n)'$.

- Marginal MGF

$$M_{X_i}(t) = M_{X_1,\cdots,X_n}(0, \cdots, 0, t, 0, \cdots 0).$$

- Moments

$$\frac{\partial^r M_{\mathbf{X}}(\mathbf{t})}{\partial t_1^{r_1} \cdots \partial t_n^{r_n}} = E\left(X_1^{r_1} \cdots X_n^{r_n}\right),$$

where $r = r_1 + \cdots + r_n$.

Let n and m be positive integers and let $p_1, \cdots, p_n$ be number satisfying $0 \leq p_i \leq 1$, $i = 1, \cdots, n$ and $\sum_i p_i = 1$. Then random vector $\mathbf{X} = (X_1, \cdots, X_n)$ has a multinomial distribution with m trials and cell probabilities $p_1, \cdots, p_n$ if the joint pmf of $\mathbf{X}$ is given

$$f(x_1, \cdots, x_n) = \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} = m! \prod_{i=1}^{n} \frac{p_i^{x_i}}{x_i!}$$

on the set of $(x_1, \cdots, x_n)$ such that each x_i is a nonnegative integers and $\sum_i x_i = m$.

4.6. Multivariate Distribution

Multinomial Distribution

- Distribution to model the result of statistical experiment with
- m independent and identical trials
- each trial results in one of n outcomes
- an outcome of type i has probability p_i

Theorem

$$(p_1 + \cdots + p_n)^m = \sum_{\mathbf{x} \in A} \frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n},$$

where

$$A = \{(x_1, \cdots, x_n) : \ x_i = 0, 1, \cdots, m, \ x_1 + \cdots x_n = m\}$$

.

4.6. Multivariate Distribution

Multinomial Distribution

- Other results for multinomial distribution

$$(X_1, \cdots, X_n) \sim \text{Multinomial}(m, p_1, \cdots, p_n)$$

- $X_i \sim \text{Binomial}(m, p_i)$
- $(Y_1, Y_2, Y_3) \sim \text{Trinomial}(m, p_1, p_2, p_3 + \cdots + p_n)$
  where $Y_1 = X_1$, $Y_2 = X_2$, $Y_3 = X_3 + \cdots + X_n$.

▷ Find the conditional distribution of $(X_1, \cdots, X_{n-1})$ conditioning on $X_n = x_n$.

4.7. Inequalities

- Hölder's Inequality
Let $X$and Y be random variables and let p and q be constants such that $1/p + 1/q = 1$. Then

$$EXY \le |EXY| \le E|XY| \le \left[E\left(|X|^p\right)\right]^{1/p}\left[E\left(|X|^q\right)\right]^{1/q}.$$

- Cauchy-Schwarz Inequality [Hölder's Inequality with $p = q = 2$]

$$EXY \le |EXY| \le E|XY| \le \left[E\left(|X|^2\right)\right]^{1/2}\left[E\left(|X|^2\right)\right]^{1/2}.$$

- $|Cov(X, Y)| \le \sqrt{Var(X)Var(Y)}$.

4.7. Inequalities

Definition

A function $g(x)$ is called *convex* if

$$g[\lambda x + (1-\lambda)y] \leq \lambda g(x) + (1-\lambda)g(y),$$

for all x, y and $\lambda \in [0,1]$. If the strict inequality holds, then the function is called *strictly convex*.

If $g(x)$ is concave then $-g(x)$ is convex.

Lemma

If $g(x)$ is differentiable and 2^{nd} derivative of $g(x)$ is nonnegative for all x, then g is convex.

4.7. Inequalities

- Jensen's Inequality

If $g(x)$ is a convex function then

$$E[g(X)] \geq g[E(X)].$$

$\triangleright$ Examples: $g(x) = x^2,\ 1/x,\ e^{tx}$