统计学方法及其应用

Statistical Methods with Applications

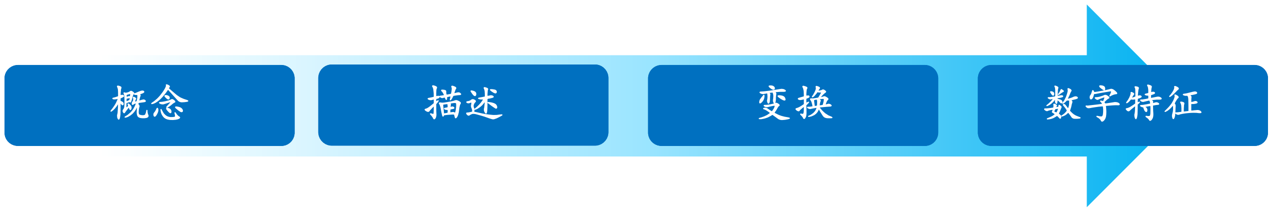
Rui Jiang, PhD

Associate Professor

Ministry of Education Key Laboratory of Bioinformatics
Bioinformatics Division, TNLIST/Department of Automation
Tsinghua University, Beijing 100084, China

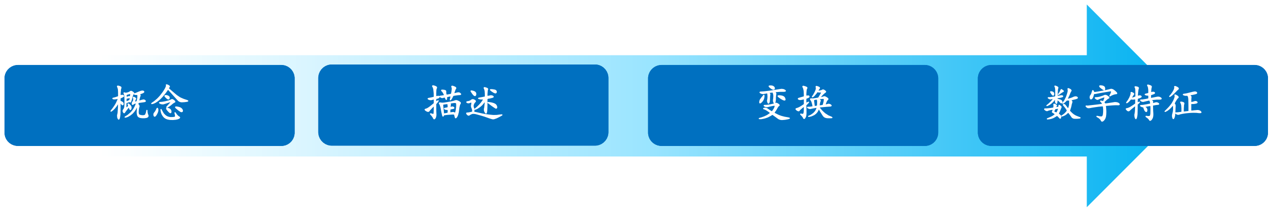
Random variables

- 随机变量



The need for multiple random variables

- 单个随机变量



- 多个随机变量
  - 感兴趣的是多个数字特征，例如，身高、血压、体温，每一个建模为一个随机变量
  - 研究的是同一个数字特征的多个观测，此时，每一个观测建模为一个随机变量

Bivariate Random Vectors

统计学方法及其应用 统计学基础 随机变量的函数

"A random variable is a quantity whose values are random and to which a probability distribution is assigned."

Tossing two fair dice

Tossing two fair dice, the sample space is the Cartesian product of two sets {1,2,3,4,5,6}

$$
\begin{aligned}
S &= \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \\
&= \{ \\
&\qquad (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\
&\qquad (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\
&\qquad (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\
&\qquad (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\
&\qquad (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\
&\qquad (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \\
&\quad \}
\end{aligned}
$$

Univariate random variable

$$X : S \to R$$

X = Sum

Second die						
6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	**2**	**3**	**4**	**5**	**6**

First die

$$H_X(1,1) = 2$$
$$\dots$$
$$H_X(3,3) = 6$$
$$\dots$$
$$H_X(6,6) = 12$$

$$P(X = 2) = 1 / 36$$
$$\dots$$
$$P(X = 6) = 5 / 36$$
$$\dots$$
$$P(X = 12) = 1 / 36$$

Univariate random variable

$$Y : S \to R$$

Y=|First - Second|

Second die						
6	5	4	3	2	1	0
5	4	3	2	1	0	1
4	3	2	1	0	1	2
3	2	1	0	1	2	3
2	1	0	1	2	3	4
1	0	1	2	3	4	5
	1	**2**	**3**	**4**	**5**	**6**

First die

$$H_Y(1,1) = 0$$
$$\dots$$
$$H_Y(3,3) = 0$$
$$\dots$$
$$H_Y(6,6) = 0$$

$$P(Y = 0) = 6 / 36$$
$$\dots$$
$$P(Y = 2) = 8 / 36$$
$$\dots$$
$$P(Y = 5) = 2 / 36$$

Bivariate random vector

***X*=Sum**

Second die						
6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	**2**	**3**	**4**	**5**	**6**

First die

***Y*=|First - Second|**

Second die						
6	5	4	3	2	1	0
5	4	3	2	1	0	1
4	3	2	1	0	1	2
3	2	1	0	1	2	3
2	1	0	1	2	3	4
1	0	1	2	3	4	5
	1	**2**	**3**	**4**	**5**	**6**

First die

$$(X,Y):S\to R^2$$

$$H_{XY}(1,1)=(2,0),\ldots,H_{XY}(3,3)=(6,0),\ldots,H_{XY}(6,6)=(12,0)$$

Bivariate random vectors

Bivariate random vectors

A **bivariate random vector** is a function from a sample space S to $\Re^2$, the 2-dimensional Euclidean space.

Define probability functions

In the original sample space S (domain of the random vector), a probability function can be defined, e.g.

$$P(\{(1,1)\}) = 1/36$$
$$P(\{(1,4),(4,1)\}) = 2/36$$
$$\ldots$$

In the space R^2 (range of the random vector), a probability function can be induced, e.g.,

$$P(X = 2 \text{ and } Y = 0) = 1/36$$
$$P(X = 5 \text{ and } Y = 3) = 2/36$$
$$\ldots$$

Probability distribution

Y&X	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	2/36	0	0	0	0	0
4	0	0	0	0	2/36	0	2/36	0	0	0	0
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36

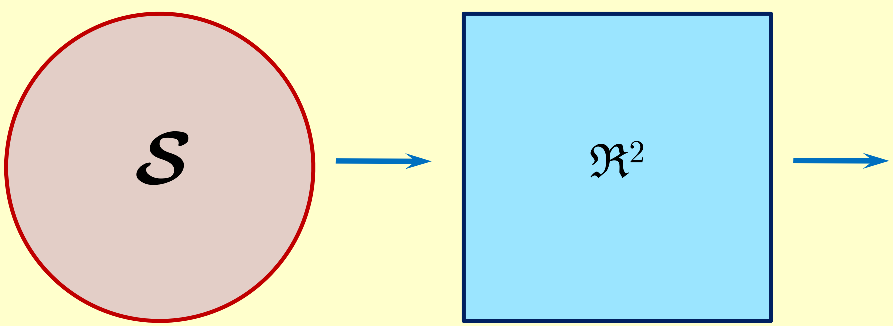
Joint probability mass functions

Joint pmf

Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\Re^2$ into $\Re$ defined by

$$f(x, y) = P(X = x, Y = y)$$

is called the **joint probability mass function** or **joint pmf** of (X, Y), denoted by $f_{X,Y}(x, y)$ for emphasizing the random vector (X, Y).



Y&X	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	2/36	0	0	0	0	0
4	0	0	0	0	2/36	0	2/36	0	0	0	0
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36

Marginal distribution

Y&X	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	2/36	0	0	0	0	0
4	0	0	0	0	2/36	0	2/36	0	0	0	0
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36

Y&X	2	3	4	5	6	7	8	9	10	11	12	Y
5	0	0	0	0	0	2/36	0	0	0	0	0	2/36
4	0	0	0	0	2/36	0	2/36	0	0	0	0	4/36
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0	6/36
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0	8/36
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0	10/36
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	6/36
X	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	

Marginal probability mass functions

Marginal pmf

Let (X,Y) be a discrete bivariate random vector with joint pmf $f_{X,Y}(x,y)$. Then the **marginal pmf** of X, $f_X(x) = P(X = x)$, is

$$f_X(x) = \sum_{y \in \Re} f_{X,Y}(x,y),$$

and the marginal pmf of Y, $f_Y(y) = P(Y = y)$, is

$$f_Y(y) = \sum_{x \in \Re} f_{X,Y}(x,y)$$

Conditional distribution

Y&X	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	2/36	0	0	0	0	0
4	0	0	0	0	2/36	0	2/36	0	0	0	0
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36
X	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Y|X	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	1/3	0	0	0	0	0
4	0	0	0	0	2/5	0	2/5	0	0	0	0
3	0	0	0	1/2	0	1/3	0	1/2	0	0	0
2	0	0	2/3	0	2/5	0	2/5	0	2/3	0	0
1	0	1/1	0	1/2	0	1/3	0	1/2	0	1/1	0
0	1/1	0	1/3	0	1/5	0	1/5	0	1/3	0	1/1

Conditional distribution

Y&X	2	3	4	5	6	7	8	9	10	11	12	Y
5	0	0	0	0	0	2/36	0	0	0	0	0	2/36
4	0	0	0	0	2/36	0	2/36	0	0	0	0	4/36
3	0	0	0	2/36	0	2/36	0	2/36	0	0	0	6/36
2	0	0	2/36	0	2/36	0	2/36	0	2/36	0	0	8/36
1	0	2/36	0	2/36	0	2/36	0	2/36	0	2/36	0	10/36
0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	0	1/36	6/36

X|Y	2	3	4	5	6	7	8	9	10	11	12
5	0	0	0	0	0	1/1	0	0	0	0	0
4	0	0	0	0	1/2	0	1/2	0	0	0	0
3	0	0	0	1/3	0	1/3	0	1/3	0	0	0
2	0	0	1/4	0	1/4	0	1/4	0	1/4	0	0
1	0	1/5	0	1/5	0	1/5	0	1/5	0	1/5	0
0	1/6	0	1/6	0	1/6	0	1/6	0	1/6	0	1/6

Conditional probability mass functions

Conditional pmf

Let (X, Y) be a discrete bivariate random vactor 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 denoted by $f(y \mid x)$ and defined by

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

Joint probability density functions

Joint pdf

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

$$P((X,Y)\in A) = \iint_A f(x,y)dxdy.$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

Marginal probability density functions

Marginal pdf

Let (X, Y) be a continuous bivariate random vector with joint pdf $f_{X,Y}(x, y)$. Then the **marginal pdfs** of X and Y, $f_X(x)$ and $f_Y(y)$, are given by

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

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx.$$

Conditional probability density functions

Conditional pdf

Let (X, Y) be a continuous bivariate random vactor 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 pdf** of Y given that $X = x$ is the function of y denoted by $f(y \mid x)$ and defined by

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

Joint cumulative distribution functions

Joint cdf

The joint cdf of the continuous bivariate random vector (X, Y) is the funciton $F(x, y)$ defined by

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

for all $(x, y) \in \Re^2$.

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv \\ f(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} \end{aligned}$$

Bivariate standard normal distribution

A vector (X, Y) is said to have a **bivariate standard normal distribution** if the joint pdf of (X, Y) is

$$f(x, y \mid \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right],$$

where $x, y \in (-\infty, \infty), \rho \in (-1, 1)$.

Now, what are marginal distributions of X and Y?

what is the conditional distribution of X given $Y = y$?

Marginal distribution

If (X, Y) has a bivariate standard normal distribution, that is,

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right].$$

Then

$$\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right] dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{y^2 - 2\rho xy + (\rho x)^2 + x^2 - (\rho x)^2}{2(1-\rho^2)}\right] dy \\
&= \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right] dy \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)
\end{aligned}$$

$X \sim N(0,1)$

Conditional distribution

If (X,Y) has a bivariate standard normal distribution, that is,

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right].$$

Then

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).$$

Therefore

$$\begin{aligned} f(x \mid y) &= \frac{f(x,y)}{f(y)} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right]. \end{aligned}$$

$$X \mid Y \sim N(\rho y, 1-\rho^2).$$

Relation

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right]; \\
f(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right); \\
f(y) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).
\end{aligned}
$$

$$\text{If } \rho = 0, \text{ then } f(x,y) = f(x)f(y)$$

$$\text{If } \rho \neq 0, \text{ then } f(x,y) \neq f(x)f(y)$$

In some cases, the joint distribution is equal to the product of marginal distributions; in some other cases, they are not equal. When are they equal?

Relation

$$
\begin{aligned}
f(x) \quad &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right); \\
f(x \mid y) &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left[-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right].
\end{aligned}
$$

$$
\text{If } \rho = 0, \ \text{then } f(x \mid y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = f(x)
$$

$$
\text{If } \rho \neq 0, \ \text{then } f(x \mid y) \neq f(x)
$$

In some cases, the conditional distribution is equal to the marginal distribution; in some other cases, they are not equal. When are they equal?

Independence

Independence

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 $x \in \Re$ and $y \in \Re$,

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

Independence

If X and Y are independent, the conditional pdf of Y given X is

$$\begin{aligned} f(y \mid x) &= \frac{f(x, y)}{f_X(x)} \\ &= \frac{f_X(x) f_Y(y)}{f_X(x)} \\ &= f_Y(y) \end{aligned}$$

That is, the conditional pdf is the same as the marginal pdf, regardless of the value of x. The knowledge of $X = x$ gives us **no** additional information about Y.

Independence

Sufficient and necessary

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 $x \in \Re$ and $y \in \Re$,

$$f(x, y) = g(x)h(y).$$

Sufficiency

Because $f(x, y)$ is the joint pdf, we have that

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)\int_{-\infty}^{\infty} h(y)dy;$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)\int_{-\infty}^{\infty} g(x)dx.$$

Therefore

$$\begin{aligned}
f_X(x)f_Y(y) &= \left[g(x)\int_{-\infty}^{\infty} h(y)dy\right]\left[h(y)\int_{-\infty}^{\infty} g(x)dx\right] \\
&= g(x)h(y)\left[\int_{-\infty}^{\infty} g(x)dx \int_{-\infty}^{\infty} h(y)dy\right] \\
&= f(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)dxdy \\
&= f(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy \\
&= f(x,y)
\end{aligned}$$

Independence

Let X and Y be independent random variables. For any $A \subset \Re$ and $B \subset \Re$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

In other words,

events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

$$\begin{aligned} P(X \in A, Y \in B) &= \int_A \int_B f_{X,Y}(x,y)dxdy \\ &= \int_A \int_B f_X(x) f_Y(y)dxdy \\ &= \int_A f_X(x)dx \int_B f_Y(y)dy \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

Independence

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 indepdendent.

For any $u \in R$ and $v \in R$, let

$$A_u = \{x : g(x) \le u\} \text{ and } B_v = \{y : h(y) \le v\}$$

Then

$$\begin{aligned} F_{U,V}(u,v) &= P(U \le u, V \le v) \\ &= P(X \in A_u, Y \in B_v) \\ &= P(X \in A_u) P(Y \in B_v) \end{aligned}$$

Therefore

$$\begin{aligned} f_{U,V}(u,v) &= \frac{\partial^2}{\partial u \partial v} F_{U,V}(u,v) \\ &= \left[\frac{d}{du} P(X \in A_u)\right]\left[\frac{d}{dv} P(Y \in B_v)\right] \end{aligned}$$

Expectations of random vectors

Expectations

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

$$\mathrm{E}g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

if (X,Y) is a continuous random vector, and

$$\mathrm{E}g(X,Y) = \sum_{(x,y)\in\Re^2} g(x,y)f(x,y)$$

if (X,Y) is a discrete random vector.

Independence

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,

$$\mathrm{E}\Big[g(X)h(Y)\Big] = \mathrm{E}\Big[g(X)\Big]\mathrm{E}\Big[h(Y)\Big].$$

$$\begin{aligned}
\mathrm{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)\varphi(x)\psi(y)dxdy \\
&= \left[\int_{-\infty}^{\infty} g(x)\varphi(x)dx\right]\left[\int_{-\infty}^{\infty} h(y)\psi(y)dy\right] \\
&= \mathrm{E}[g(X)]\mathrm{E}[h(Y)]
\end{aligned}$$

Moment generating function

For two independent random variables X and Y,

let $g(x) = e^{tx}$ and $h(y) = e^{ty}$, then

$$\mathrm{E}[g(X)] = \mathrm{E}[e^{tX}] = M_X(t)$$

$$\mathrm{E}[h(Y)] = \mathrm{E}[e^{tY}] = M_Y(t)$$

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[e^{t(X+Y)}] = M_{X+Y}(t)$$

Because of the independence,

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$$

We thus have

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Summation of two random variables

Let X and Y be two independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable

$$Z = X + Y$$

is given by

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

Summation of two Normal's

Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ be two independent normal random variables, then

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

$$M_Y(t) = \exp\left(\nu t + \frac{1}{2}\tau^2 t^2\right)$$

$$M_Z(t) = M_X(t) M_Y(t) = \exp\left((\mu + \nu)t + \frac{1}{2}(\sigma^2 + \tau^2)t^2\right)$$

Therefore,

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

Conditional expectations

Conditional Expectations

If $g(Y)$ is a real-valued function of Y, then the conditional expected value of $g(Y)$ given that $X = x$ is denoted by $\mathrm{E}\left[g(Y) \mid x\right]$ and is defined to be

$$\mathrm{E}\left[g(Y) \mid x\right] = \int_{-\infty}^{\infty} g(y) f(y \mid x) dy$$

in the continuous case and

$$\mathrm{E}\left[g(Y) \mid x\right] = \sum_{y \in \Re} g(y) f(y \mid x)$$

in the discrete case.

Conditional expectations

Conditional Expectations

If X and Y are two random variables, then

$$\mathrm{E}X = \mathrm{E}(\mathrm{E}(X \mid Y))$$

provided that the expectations exists.

$$\begin{aligned}
\mathrm{E}X &= \int\int xf(x,y)dxdy \\
&= \int\int xf(x \mid y)f(y)dxdy \\
&= \int\left[\int xf(x \mid y)dx\right]f(y)dy \\
&= \int \mathrm{E}(X \mid y)f(y)dy \\
&= \mathrm{E}(\mathrm{E}(X \mid Y))
\end{aligned}$$

Univariate transformations of pmfs

Let X be a random variable with range $\mathcal{X}$. For transformation $Y = g(X)$, the sampel space is $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$.

If $y \in \mathcal{Y}$, then

$$\begin{aligned} f_Y(y) &= P(Y = y) \\ &= \sum_{x \in g^{-1}(y)} P(X = x) \\ &= \sum_{x \in g^{-1}(y)} f_X(x). \end{aligned}$$

If $y \notin \mathcal{Y}$, then

$$f_Y(y) = 0.$$

Bivariate transformations of pmfs

Let (X, Y) be a discrete bivariate random vector with a known joint pmf $f_{X,Y}(x,\ y)$. Let (U, V) be a bivariate random vector defined by $U = g(X, Y)$ and $V = h(X, Y)$.

Define

$$\mathcal{A} = \{(x, y) : f_{X,Y}(x,\ y) > 0\},$$
$$\mathcal{B} = \{(u, v) : u = g(x, y) \text{ and } v = h(x, y) \text{ for some } (x, y) \in \mathcal{A}\},$$
$$\mathcal{C} = \{(x, y) : g(x, y) = u \text{ and } h(x, y) = v \text{ for any } (u, v) \in \mathcal{B}\}.$$

Then,

$$f_{U,V}(u,\ v) = \sum_{(x,y)\in\mathcal{C}} f_{X,Y}(x,\ y).$$

Poisson distribution

- A random variable X is said to have a $Poisson(\lambda)$ distribution if

$$P(X = x \mid \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x\,!}, x = 0, 1, 2, \ldots$$

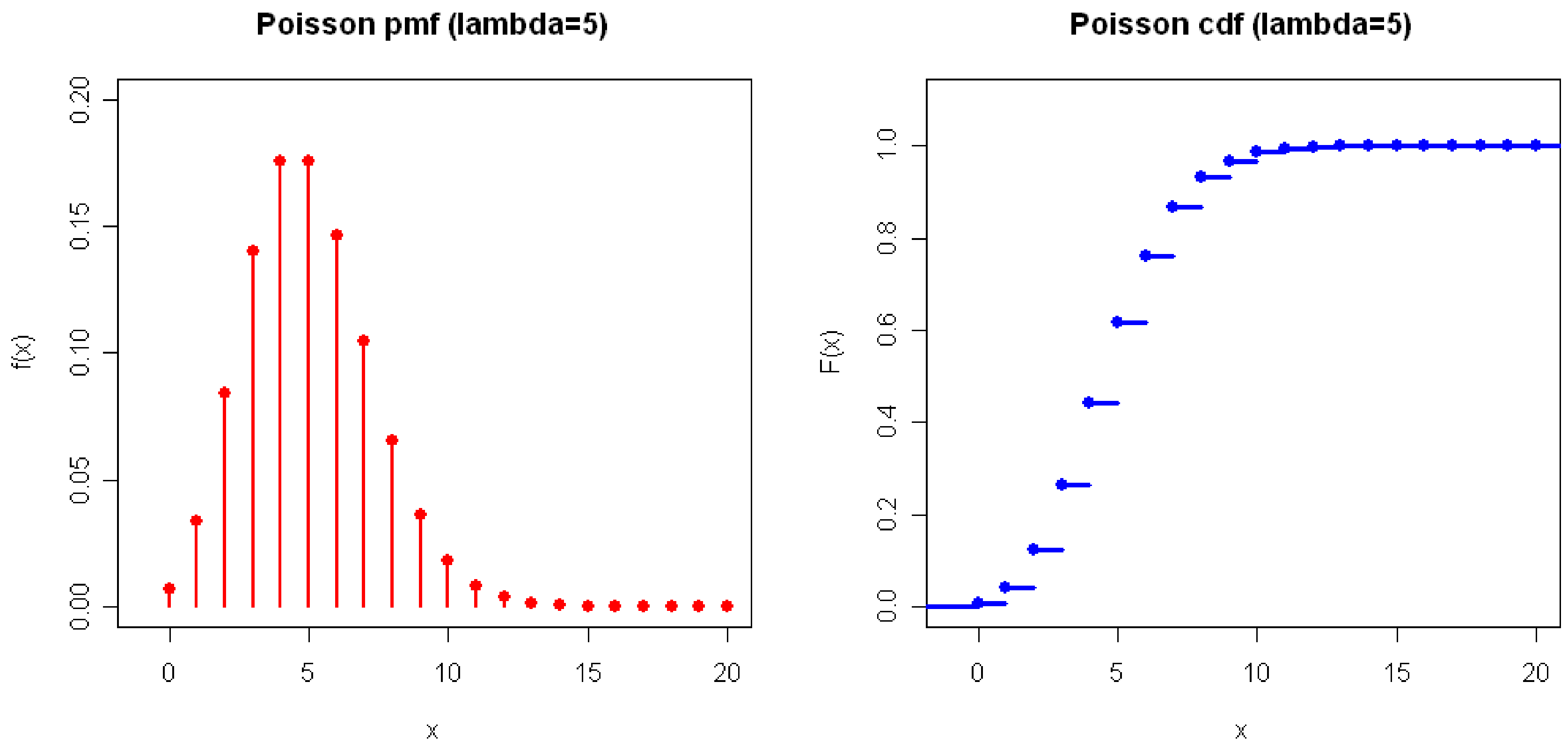
  where λ is the **intensity parameter**
- The probability of a number of events occurring in a fixed period of time if these events occur with a known average rate (intensity) and independently of the time since the last event
- Mean

$$\mathrm{E}X = \lambda$$ **(why?)**

- Variance

$$\mathrm{Var}\,X = \lambda$$ **(why?)**

pmf and cdf



Poisson pmf (lambda=5)
f(x)
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
x
Poisson cdf (lambda=5)
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
x

Poisson distribution in R

- pmf

  `dpois(x, lambda)`

- cdf

  `ppois(q, lambda)`

- Quantile function

  `qpois(p, lambda)`

- Random numbers

  `rpois(n, lambda)`

Summation of two independent Poisson's

Let X and Y be two independent Poisson random variables with intensities θ and λ, respectively. The joint pmf of (X,Y) is then

$$f_{X,Y}(x,y) = \frac{e^{-\theta}\theta^x}{x!}\frac{e^{-\lambda}\lambda^y}{y!}, x = 0,1,2,\ldots; y = 0,1,2,\ldots$$

Define transformation $U = X+Y, V = Y$

$\mathcal{A} = \{(x,y) : x = 0,1,2,\ldots; y = 0,1,2,\ldots\}$

$\mathcal{B} = \{(u,v) : v = 0,1,2,\ldots; u = v+0, v+1, v+2,\ldots\}$

$\mathcal{C} = \{(x,y) : y = v = 0,1,2,\ldots; x = u-y = u-v\}$

$$f_{U,V}(u,v) = \frac{e^{-\theta}\theta^{u-v}}{(u-v)!}\frac{e^{-\lambda}\lambda^v}{v!}, v = 0,1,2,\ldots; u = v, v+1, v+2,\ldots$$

$$\begin{aligned} f_U(u) &= \sum_{v=0}^{u}\frac{e^{-\theta}\theta^{u-v}}{(u-v)!}\frac{e^{-\lambda}\lambda^v}{v!}, u = 0,1,2,\ldots \\ &= \frac{e^{-(\theta+\lambda)}}{u!}\sum_{v=0}^{u}\frac{u!}{(u-v)!v!}\theta^{u-v}\lambda^v, u = 0,1,2,\ldots \\ &= \frac{e^{-(\theta+\lambda)}}{u!}(\theta+\lambda)^u, 0,1,2,\ldots \end{aligned}$$

$X+Y \sim Poisson(\theta+\lambda)$

Univariate transformations of pdfs

Let X have pdf $f_X(x)$, let $Y = g(x)$, where g is a **monotone** function. Suppose $f_X(x)$ is continuous on $\mathcal{X} = \{x : f_X(x) > 0\}$ and $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$.

Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \dfrac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Bivariate transformations of pdfs

Let (X,Y) be a continuous bivariate random vector with joint pdf $f_{X,Y}(x,\ y)$. Let (U,V) be a bivariate random vector defined by $U = g(X,Y)$ and $V = h(X,Y)$. Let

$$\mathcal{A} = \{(x,y) : f_{X,Y}(x,y) > 0\} \text{ and}$$
$$\mathcal{B} = \{(u,v) : u = g(x,y) \text{ and } v = h(x,y) \text{ for some } (x,y) \in \mathcal{A}\}.$$

If the transformation is a **one - to - one** transformation of $\mathcal{A}$ **onto** $\mathcal{B}$, then

$$f_{U,V}(u,\ v) = f_{X,Y}(\varphi(U,V),\ \psi(U,V))\,|\,J\,|,$$

where

$$X = \varphi(U,V),\ \ Y = \psi(U,V),$$
$$J = \begin{vmatrix} \dfrac{\partial \varphi}{\partial u} & \dfrac{\partial \varphi}{\partial v} \\ \dfrac{\partial \psi}{\partial u} & \dfrac{\partial \psi}{\partial v} \end{vmatrix} = \frac{\partial \varphi}{\partial u}\frac{\partial \psi}{\partial v} - \frac{\partial \psi}{\partial u}\frac{\partial \varphi}{\partial v}.$$

Summation of two random variables

Convolution formula

Let (X,Y) be two independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw.$$

Let the transformation be $W = X$ and $Z = X + Y$.
Define $X = \varphi(W,Z) = W$ and $Y = \psi(W,Z) = Z - W$.
Then the Jacobian is 1, and the joint pdf is $f_{W,Z}(w,z) = f_X(w) f_Y(z-w)$.
The marginal pdf $f_Z(z)$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_{W,Z}(w,z)\,dw = \int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw.$$

Summation of two Normal's

Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ be two independent normal random variables, then

$$Z = X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2),$$
$$Z = X - Y \sim N(\mu - \nu, \sigma^2 + \tau^2).$$

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(w) f_Y(z-w)\,dw &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(z-w-\nu)^2}{2\tau^2}\right) dw \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2+\tau^2}} \exp\left(-\frac{(z-\mu-\nu)^2}{2(\sigma^2+\tau^2)}\right) \times \\
&\quad \underbrace{\frac{1}{\sqrt{2\pi}} \sqrt{\frac{\sigma^2+\tau^2}{\sigma^2\tau^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\frac{\sigma^2+\tau^2}{\sigma^2\tau^2}\left(w - \frac{\mu\tau^2 + (z-\nu)\sigma^2}{\sigma^2+\tau^2}\right)^2\right] dw}_{1}
\end{aligned}$$

Gamma function

- Gamma function

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

- Properties
  - $\Gamma(\alpha+1) \quad = \alpha\,\Gamma(\alpha), \qquad \alpha > 0$
  - $\Gamma(1) \quad = 1$
  - $\Gamma(½) \quad = \sqrt{\pi}$
  - $\Gamma(n) \quad = (n-1)!$
- Gamma function in R
  - **gamma(x)**
  - **lgamma(x)**

$$\Gamma(\alpha+n)\,/\,\Gamma(\alpha) = \alpha(\alpha+1)\cdots(\alpha+n-1)$$

$$\begin{aligned}
&\Gamma(\alpha+1)\,/\,\Gamma(\alpha) &&= \alpha \\
&\Gamma(\alpha+2)\,/\,\Gamma(\alpha+1) &&= \alpha+1 \\
&\cdots \\
&\Gamma(\alpha+n)\,/\,\Gamma(\alpha+n-1) &&= \alpha+n-1
\end{aligned}$$

Gamma distribution

- pdf

$$f(x \mid \text{shape}{=}\alpha, \text{scale}{=}\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, 0 \le x < \infty, \alpha > 0, \theta > 0$$

- Mean

$$\mathrm{E}X = \alpha\theta$$

- Variance

$$\mathrm{Var}X = \alpha\theta^2$$

Gamma distribution

- pdf

$$f(x \mid \text{shape}{=}\alpha, \text{rate}{=}\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, 0 \leq x < \infty, \alpha > 0, \beta > 0$$

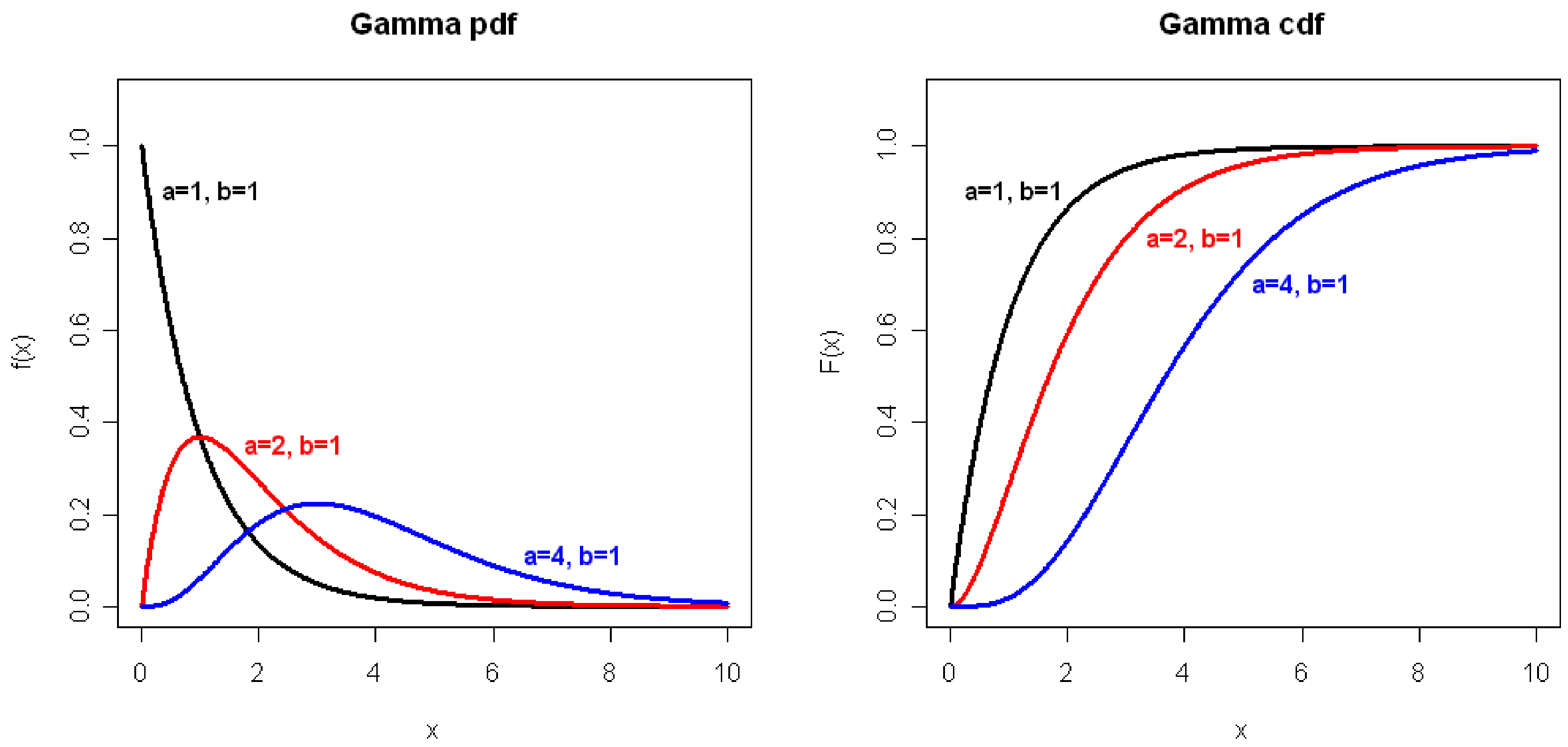
- Mean

$$\mathrm{E}X = \frac{\alpha}{\beta}$$

- Variance

$$\mathrm{Var}X \quad = \frac{\alpha}{\beta^2}$$

pdf and cdf


Gamma pdf
a=1, b=1
a=2, b=1
a=4, b=1
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
Gamma cdf
a=1, b=1
a=2, b=1
a=4, b=1
F(x)
x

Gamma distribution in R

- pmf

  ```
  dgamma(x, shape, rate=1, scale=1/rate)
  ```

- cdf

  ```
  pgamma(q, shape, rate=1, scale=1/rate)
  ```

- Quantile function

  ```
  qgamma(p, shape, rate=1, scale=1/rate)
  ```

- Random numbers

  ```
  rgamma(n, shape, rate=1, scale=1/rate)
  ```

Gamma(shape=1, scale=λ) $\rightarrow$ exponential

- Gamma pdf

$$f(x \mid \text{shape}=\alpha, \text{scale}=\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, 0 \leq x < \infty, \alpha > 0, \theta > 0$$

- Gamma(shape=1, scale=λ) pdf

$$f(x \mid \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, 0 \leq x < \infty, \lambda > 0$$

Gamma mgf

$$f(x \mid \text{shape}=\alpha, \text{scale}=\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, 0 \le x < \infty, \alpha > 0, \theta > 0$$

$$\begin{aligned}
M(t) &= \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{\infty} e^{tx} x^{\alpha-1} e^{-x/\theta} dx \\
&= \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-x(1/\theta - t)} dx \\
&= \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-\frac{x}{\theta/(1-\theta t)}} dx \\
&= \left[\frac{1}{\Gamma(\alpha)\theta^{\alpha}}\right] \left[\frac{1}{\Gamma(\alpha)(\theta \,/\, (1-\theta t))^{\alpha}}\right]^{-1} \\
&= \left(\frac{1}{1-\theta t}\right)^{\alpha}
\end{aligned}$$

Gamma moments

$$M(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha}$$

$$\frac{d}{dx}M(t) = \alpha\theta(1-\theta t)^{-(\alpha+1)}$$

$$\Rightarrow \mu_1' = \alpha\theta \Rightarrow \mu = \alpha\theta$$

$$\frac{d^2}{dx^2}M(t) = \alpha(\alpha+1)\theta^2(1-\theta t)^{-(\alpha+2)}$$

$$\Rightarrow \mu_2' = \alpha(\alpha+1)\theta^2 \Rightarrow \sigma^2 = \alpha\theta^2$$

$$\frac{d^3}{dx^3}M(t) = \alpha(\alpha+1)(\alpha+2)\theta^3(1-\theta t)^{-(\alpha+3)}$$

$$\Rightarrow \mu_3' = \alpha(\alpha+1)(\alpha+2)\theta^3 \Rightarrow \beta_s = \frac{2}{\sqrt{\alpha}}$$

$$\frac{d^4}{dx^4}M(t) = \alpha(\alpha+1)(\alpha+2)(\alpha+4)\theta^4(1-\theta t)^{-(\alpha+4)}$$

$$\Rightarrow \mu_4' = \alpha(\alpha+1)(\alpha+2)(\alpha+4)\theta^4 \Rightarrow \beta_k - 3 = \frac{6}{\alpha}$$

Summation of two Gamma's

Let $X \sim Gamma\left(\alpha, \text{scale} = \theta\right)$ and $Y \sim Gamma\left(\beta, \text{scale} = \theta\right)$ be two independent Gamma random variables, then

$$M_X(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha}$$

$$M_Y(t) = \left(\frac{1}{1-\theta t}\right)^{\beta}$$

$$M_Z(t) = M_X(t) M_Y(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha+\beta}$$

Therefore,

$$Z = X + Y \sim Gamma(\alpha + \beta, \text{scale} = \theta)$$

Beta distribution

Let $U \sim Gamma(\alpha, \text{scale} = \theta)$ and $V \sim Gamma(\beta, \text{scale} = \theta)$ be two independent Gamma random variables, Consider the transform

$$X = U \,/\, (U+V), \;\; Y = U + V$$

Clearly, $U = XY, \;\; V = Y(1-X)$

The Jacobian is therefore

$$J = \begin{vmatrix} y & x \\ -y & 1-x \end{vmatrix} = y(1-x) + xy = y$$

Because $p(u,v) = \left[\frac{1}{\Gamma(\alpha)\theta^{\alpha}} u^{\alpha-1} e^{-u/\theta}\right]\left[\frac{}{\Gamma(\beta)\theta^{\beta}} v^{\beta-1} e^{-v/\theta}\right]$

$$p(x,y) = \left[\frac{1}{\Gamma(\alpha)\theta^{\alpha}} (xy)^{\alpha-1} e^{-xy/\theta}\right]\left[\frac{1}{\Gamma(\beta)\theta^{\beta}} \left[y(1-x)\right]^{\beta-1} e^{-y(1-x)/\theta}\right] y$$

$$= \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right]\left[\frac{1}{\Gamma(\alpha+\beta)\theta^{\alpha+\beta}} y^{(\alpha+\beta)-1} e^{-y/\theta}\right]$$

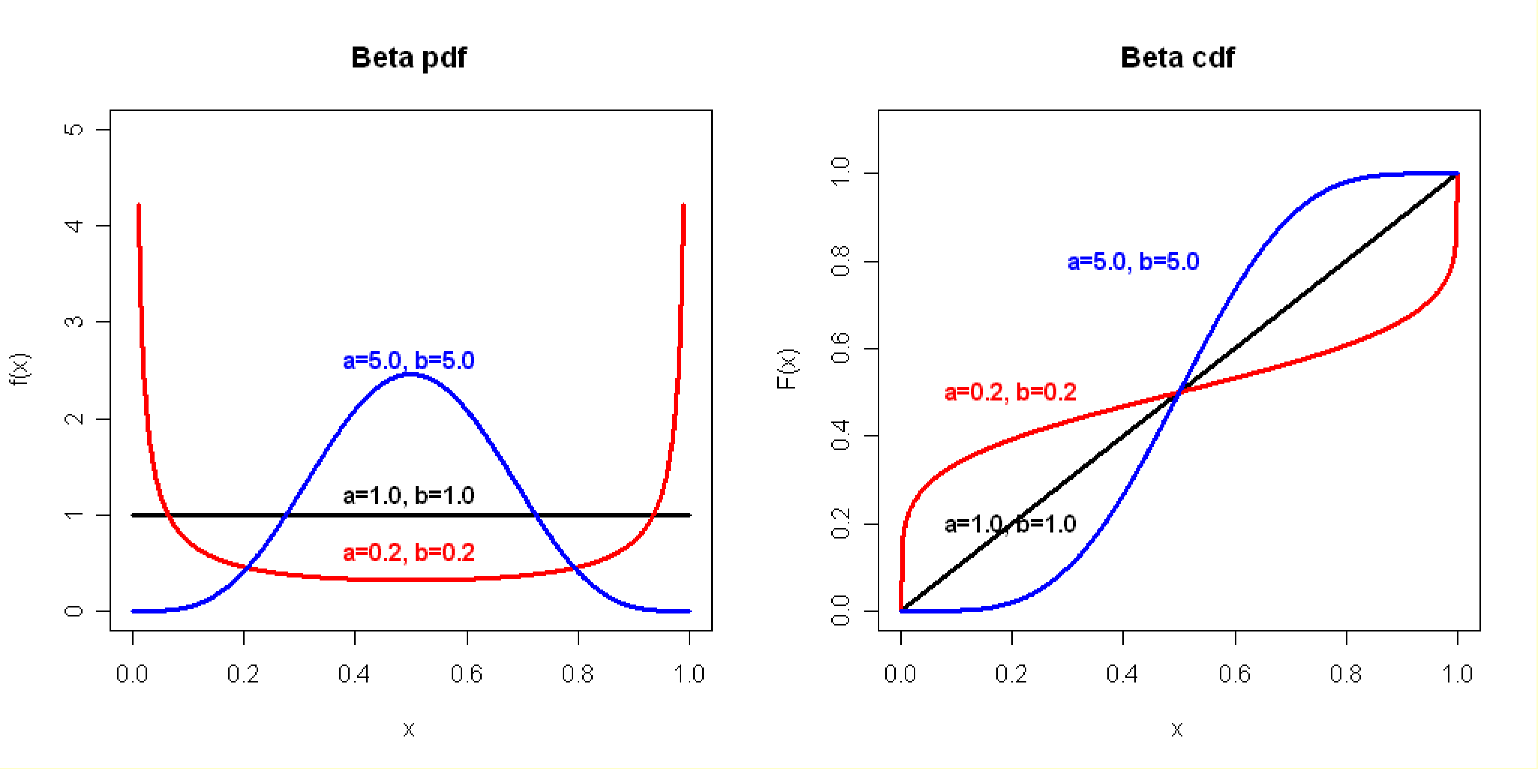
Therefore

$$X \sim Beta(\alpha, \beta); Y \sim Gamma(\alpha+\beta, \text{scale}{=}\theta)$$

Beta distribution

A random variable is said to have a $Beta(\alpha, \beta)$ distribution if the pdf is

$$f(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1, \alpha > 0, \beta > 0$$



Beta distribution in R

- pmf

  `dbeta(x, shape1, shape2)`

- cdf

  `pbeta(q, shape1, shape2)`

- Quantile function

  `qbeta(p, shape1, shape2)`

- Random numbers

  `rbeta(n, shape1, shape2)`

Integral

Since

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

is a pdf, we have

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx = 1$$

In other words,

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]^{-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

This is called a **beta function**

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Covariance and correlation

Covariance

The **covariance** of X and Y is the number defined by

$$\operatorname{Cov}(X,Y) = \mathrm{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big].$$

Correlation coefficient

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

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

Covariance

For any random variables X and Y,

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

If X and Y are independent random variables,

$$\mathrm{Cov}(X,Y)=\mathrm{E}XY-\mu_X\mu_Y=\mu_X\mu_Y-\mu_X\mu_Y=0,$$

and

$$\rho_{\mathrm{XY}}=0.$$

Variance

If X and Y are any random variables and a and b are any two constants, then

$$\begin{aligned}
\mathrm{Var}(aX+bY) &= \mathrm{E}\left[(aX+bY)-\mathrm{E}(aX+bY)\right]^2 \\
&= \mathrm{E}\left[(aX-\mathrm{E}(aX))+(bY-\mathrm{E}(bY))\right]^2 \\
&= \mathrm{E}\left[a(X-\mathrm{E}X)+b(Y-\mathrm{E}Y)\right]^2 \\
&= \mathrm{E}\left[a^2(X-\mathrm{E}X)^2\right]+\mathrm{E}\left[b^2(Y-\mathrm{E}Y)^2\right]+\mathrm{E}\left[2ab(X-\mathrm{E}X)(Y-\mathrm{E}Y)\right] \\
&= a^2\mathrm{E}(X-\mathrm{E}X)^2+b^2\mathrm{E}(Y-\mathrm{E}Y)^2+2ab\mathrm{E}(X-\mathrm{E}X)(Y-\mathrm{E}Y) \\
&= a^2\mathrm{Var}X+b^2\mathrm{Var}Y+2ab\mathrm{Cov}(X,Y).
\end{aligned}$$

If X and Y are independent random variables, then

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

Linear relationship

Linear relationship

For any random variables X and Y,

1. $-1 \le \rho_{XY} \le 1$.
2. $|\rho_{XY}| = 1$ if and only if there exist numbers $a \ne 0$ and b such that $P(Y = aX + b) = 1$. If $\rho_{XY} = 1$, then $a > 0$, and if $\rho_{XY} = -1$, then $a < 0$.

Consider the function

$$\begin{aligned} h(t) &= \mathrm{E}\left((X-\mu_X)t + (Y-\mu_Y)\right)^2 \\ &= \sigma_X^2 t^2 + 2\mathrm{Cov}(X,Y)t + \sigma_Y^2 \end{aligned}$$

and its discriminant

$$\left(2\mathrm{Cov}(X,Y)\right)^2 - 4\sigma_X^2\sigma_Y^2 \le 0.$$

Bivariate normal distribution

A random vector (X,Y) is said to has a bivariate normal distribution if their joint pdf is

$$f(x,y \mid \mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}.$$

Bivariate normal distribution

Let $\mathbf{x} = (x, y)$;

$$\boldsymbol{\mu} = (\mu_X, \mu_Y),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix},$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\left(1-\rho^2\right)\sigma_X^2\sigma_Y^2}\begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}.$$

Bivariate normal distribution becomes

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi) \mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Bivariate normal distributions

If $(X, Y) \sim$ bivariate normal $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, the marginal distribution of X is $N(\mu_X, \sigma_X^2)$, the marginal distribution of Y is $N(\mu_Y, \sigma_Y^2)$.

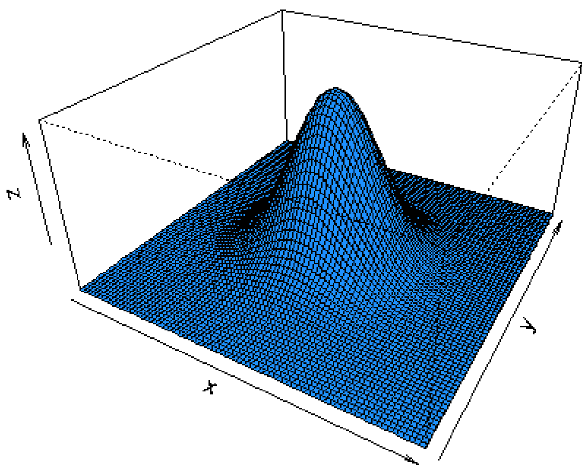
$W = X - Y$ has a $N(\mu_W, \sigma_W^2)$ distribution, where

$$\mu_W = \mu_X - \mu_Y,$$
$$\sigma_W^2 = \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2.$$

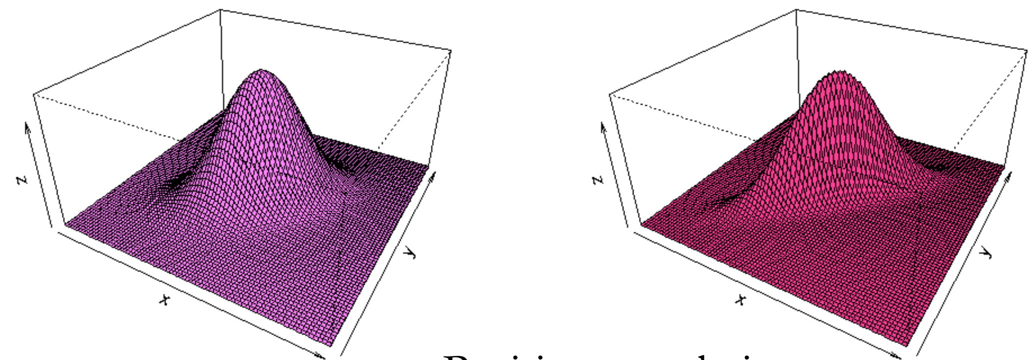
In general, $Z = aX + bY$ has a $N(\mu_Z, \sigma_Z^2)$ distribution, where

$$\mu_Z = a\mu_X + b\mu_Y,$$
$$\sigma_Z^2 = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2.$$

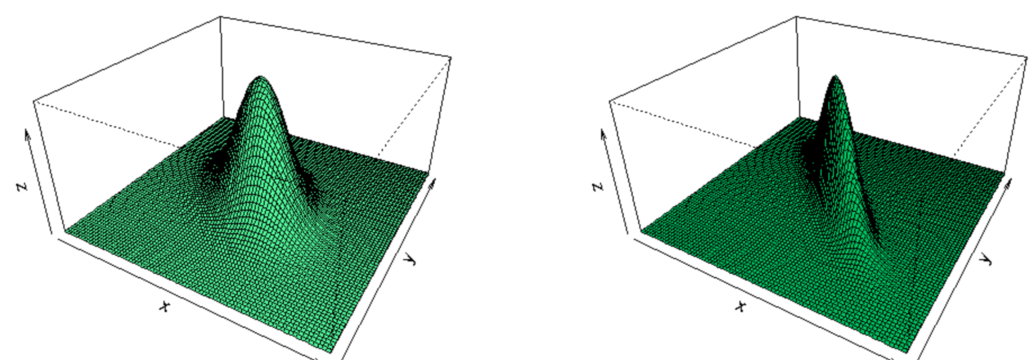
Bivariate normal density



No correlation



Positive correlation



Negative correlation

Hierarchical models

Hierarchical models

The distribution of a random variable depends on a quantity that also has a distribution.

Beta-Binomial

Let $X \sim Binomial(n,\theta),\ \theta \sim Beta(\alpha,\beta)$. Then

$$p(x \mid \theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Therefore

$$\begin{aligned} p(x) &= \int p(\theta)p(x \mid \theta)d\theta = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{x}\int \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}d\theta \\ &= \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]\left[\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\right]^{-1}\binom{n}{x} \end{aligned}$$

According to Baye's rule,

$$p(\theta \mid x) = \frac{p(\theta)p(x \mid \theta)}{p(x)} = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$$

Hence,

$$\theta \mid x \sim Beta(\alpha+x, \beta+n-x)$$

Laplace distribution

If $X \sim Exponential(\lambda)$, that is $f(x) = \frac{1}{\lambda} e^{-x/\lambda}, x \geq 0, \lambda > 0$, then the transformation $Y = -X$ yields $f(y) = \frac{1}{\lambda} e^{y/\lambda}, y \leq 0, \lambda > 0$. In other words,

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} I_{[0,\infty)}(x) \text{ and } f(y) = \frac{1}{\lambda} e^{-y/\lambda} I_{(-\infty,0]}(y)$$

Now, consider the hierarchical model

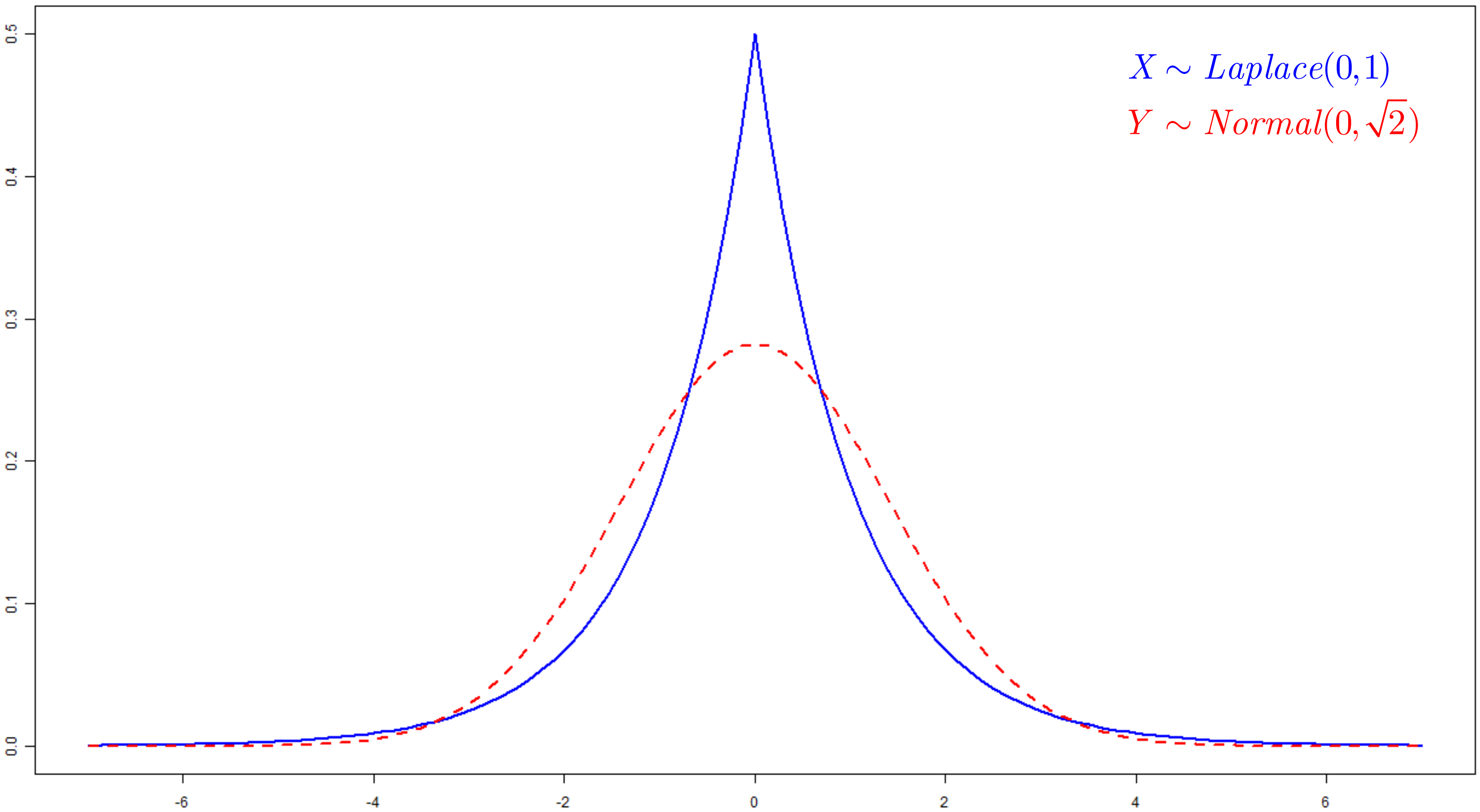
$$f(z \mid b = 1) = \frac{1}{\lambda} e^{-z/\lambda} I_{[0,\infty)}(z)$$
$$f(z \mid b = 0) = \frac{1}{\lambda} e^{z/\lambda} I_{(-\infty,0]}(z)$$
$$B \sim Bernoulli(1/2)$$

We have

$$\begin{aligned} f(z) &= f(z \mid b = 1)P(b = 1) + f(z \mid b = 0)P(b = 0) \\ &= \frac{1}{2\lambda} e^{-z/\lambda} I_{[0,\infty)}(z) + \frac{1}{2\lambda} e^{z/\lambda} I_{(-\infty,0]}(z) \\ &= \frac{1}{2\lambda} e^{-|z|/\lambda} \end{aligned}$$

This pdf defines a **Laplace distribution**, aka **double exponential distribution**.

Laplace versus normal (equal variances)



Multivariate Random Vectors

统计学方法及其应用 | 统计学基础 | 随机变量的函数

"A random variable is a quantity whose values are random and to which a probability distribution is assigned."

Multivariate random vectors

Random vectors

An n-dimensional random vector is a function from a sample space S to $\Re^n$, the n-dimensional Euclidean space.

$$\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$$

$$\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$$

Multivariate random vectors

	Univariate	Bivariate	Multivariate
概念	$S \longrightarrow \mathfrak{R}$	$S \longrightarrow \mathfrak{R}^2$	$S \longrightarrow \mathfrak{R}^n$
描述	pmf or pdf	joint, marginal, and conditional pmf or pdf	joint, marginal, and conditional pmf or pdf
变换	Derivative(Jacobian)	Jacobian	Jacobian
特征	Moments, expectations, mean, variance	Independence, conditional expectations, covariance, correlation	Mutual independence, Pair-wise correlation

Joint distributions

Discrete case

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

$$F(\mathbf{x}) = F(x_1, \ldots, x_n) = \sum_{t_1 \le x_1, \ldots, t_n \le x_n} f(t_1, \ldots, t_n)$$

$$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x})$$

Continuous case

$$f(\mathbf{x}) = f(x_1, \ldots, x_n)$$

$$F(\mathbf{x}) = F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

$$P(\mathbf{X} \in A) = \int \cdots \int_A f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

Marginal distributions

Discrete case

$$f(x_1,\ldots,x_k) = \sum_{(x_{k+1},\ldots,x_n)\in\Re^{n-k}} f(x_1,\ldots,x_k,x_{k+1},\ldots,x_n)$$

Continuous case

$$f(x_1,\ldots,x_k) = \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_{k+1}} f(x_1,\ldots,x_k,x_{k+1},\ldots,x_n)dx_{k+1}\cdots dx_n$$

Conditional distributions

Conditional distributions

$$f(x_{k+1},\ldots,x_n \mid x_1,\ldots,x_k) = \frac{f(x_1,\ldots,x_k,x_{k+1},\ldots,x_n)}{f(x_1,\ldots,x_k)}$$

Transformations

Transformations

$$Y_i = g_i(\mathbf{X}) = g_i(X_1, \ldots, X_n)$$

$$\Rightarrow \qquad X_i = h_i(\mathbf{Y}) = h_i(Y_1, \ldots, Y_n), i = 1, \ldots, n$$

$$\Rightarrow \qquad J = \begin{vmatrix} \dfrac{\partial h_1}{\partial y_1} & \cdots & \dfrac{\partial h_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial h_n}{\partial y_1} & \cdots & \dfrac{\partial h_n}{\partial y_n} \end{vmatrix}$$

$$f_{\mathbf{X}}(\mathbf{x}) = f(x_1, \ldots, x_n)$$

$$\Rightarrow \qquad f_Y(\mathbf{y}) = f(h_1, \ldots, h_n) \mid J \mid$$

Expectations

Expectations

$$\mathrm{E}g(\mathbf{X}) = \sum_{\mathbf{x}\in\Re^n} g(\mathbf{x})f(\mathbf{x})$$

$$\mathrm{E}g(\mathbf{X}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\mathbf{x})f(\mathbf{x})\; d\mathbf{x}$$

$$\mathrm{E}_{f(x_1,\ldots,x_k)}g(x_1,\ldots,x_k) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_k) \underbrace{f(x_1,\ldots,x_k)}_{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_k,x_{k+1},\ldots,x_k)dx_{k+1}\cdots dx_n} dx_1\cdots dx_k$$

$$\mathrm{E}_{f(x_1,\ldots,x_k|x_{k+1},\ldots,x_k)}g(x_1,\ldots,x_k) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_k)\; \underbrace{f(x_1,\ldots,x_k \mid x_{k+1},\ldots,x_k)}_{\frac{f(x_1,\ldots,x_n)}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_k,x_{k+1},\ldots,x_k)dx_1\cdots dx_k}}\; dx_1\cdots dx_k$$

Mutually independent random vectors

Mutually independent

Let $X_1, \dots, X_n$ be random variables with joint pdf or pmf $f(x_1, \dots, x_n)$. Let $f_{X_i}(x_i)$ denote the marginal pdf or pmf of X_i. Then $X_1, \dots, X_n$ are called **mutually independent** random variables if, for every $(x_1, \dots, x_n)$,

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Mutually independent $\neq$ pairwise independent

mgf

mgf

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 ony of x_i, $i=1,\ldots,n$. Then

$$\mathrm{E}\left[\prod_{i=1}^{n} g_i(x_i)\right] = \prod_{i=1}^{n} \mathrm{E}\left[g_i(x_i)\right].$$

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

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

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

$$M_Z(t) = \left[M_X(t)\right]^n.$$

Summation of multiple Gamma's

Let $X_i \sim Gamma(\alpha_i, \text{scale} = \theta), i = 1, \ldots, n,$ be n independent Gamma random variables, then

$$M_{X_i}(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha_i}$$

Let $Z = X_1 + \cdots + X_n$

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left(\frac{1}{1-\theta t}\right)^{\sum_{i=1}^{n} \alpha_i}$$

Therefore,

$$Z = \sum_{i=1}^{n} X_i \sim Gamma(\sum_{i=1}^{n} \alpha_i, \text{scale} = \theta)$$

Bernoulli trial

A Bernoulli trial is an experiment has two possible outcomes, represented by 0 and 1.

A *Bernoulli*(p) random variable has pmf

$$p(x = 1) = p$$

$$p(x = 0) = 1 - p$$

Let

$$I(x = k) = \begin{cases} 1 & x = k \\ 0 & \text{otherwise} \end{cases}$$

Then

$$p(x) = p^{I(x=1)}(1 - p)^{I(x=0)}$$

Multinomial trial

A Multinomial trial is an experiment has m possible outcomes, indexed by $1,\ldots,m$. Let X be the index of the outcome

Suppose in the experiment

$$p(x=1)=p_1$$
$$\ldots$$
$$p(x=m)=p_m$$

Let $\mathbf{p}=(p_1,\ldots,p_m)$ with $\sum_{k=1}^{m}p_k=1$

Then

$$p(x)=\prod_{k=1}^{m}p_k^{I(x=k)}$$

This is the pmf of a multinomial trial$(m,\mathbf{p})$ random variable.

Binomial distribution

Repeat a Bernoulli trial a number of n times, Let

$$X = \#\{1 \text{ in the experiments}\}$$

Then X has a *binomial*(n, p) distribution with pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

Multinomial distribution

Repeat a multinomial trial $(m, \mathbf{p})$ a number of n times, Let

$$X_k = \#\{k \text{ in the experiments}\}$$

Let

$$\mathbf{X} = (X_1, \ldots, X_m)$$

Then $\mathbf{X}$ has a $multinomial(n, m, \mathbf{p})$ distribution with pmf

$$p(\mathbf{x} \mid n, m, \mathbf{p}) = \frac{n\,!}{x_1\,! \times \cdots \times x_m\,!}\, p_1^{x_1} \cdots \cdot p_m^{x_m} = \frac{(\sum_{k=1}^m x_k)!}{\prod_{k=1}^m x_k\,!} \prod_{k=1}^m p_k^{x_k}$$

Binomial theorem

Binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

Multinomial theorem

Multinomial theorem

Let n and m be positive integers. Let $\mathcal{A}$ be the set of vectors $\mathbf{x} = (x_1, \ldots, x_m)$ such that each x_i is a nonnegative integer and $\sum_{k=1}^{m} x_k = n$. Then, for any real number $p_1, \ldots, p_m$,

$$(p_1 + \cdots + p_m)^n = \sum_{\mathbf{x} \in \mathcal{A}} \frac{n!}{x_1! \times \cdots \times x_m!} p_1^{x_1} \cdot \cdots \cdot p_m^{x_m}.$$

Particularly, if $\sum_{k=1}^{m} p_k = 1$, then

$$\sum_{\mathbf{x} \in \mathcal{A}} \frac{n!}{x_1! \times \cdots \times x_m!} p_1^{x_1} \cdot \cdots \cdot p_m^{x_m} = (p_1 + \cdots + p_m)^n = 1$$

Multinomial distribution

Multinomial distribution

Let n and m be positive integers and let $p_1,\ldots,p_m$ be numbers satisfying $0 \le p_k \le 1, k=1,\ldots,m,$ and $\sum_{k=1}^m p_k = 1$. Then the random vector $\mathbf{X}=(X_1,\ldots,X_m)$ has a **multinomial distribution** with n trials and cell probabilities $\mathbf{p}=(p_1,\ldots,p_m)$. The joint pmf of $\mathbf{X}$ is

$$f(\mathbf{x} \mid n, m, \mathbf{p}) = \frac{n!}{x_1! \times \cdots \times x_m!} p_1^{x_1} \cdot \cdots \cdot p_m^{x_m} = \frac{\left(\sum_{k=1}^m x_k\right)!}{\prod_{k=1}^m x_k!} \prod_{k=1}^m p_k^{x_k}$$

on the set of $\mathbf{x}=(x_1,\ldots,x_m)$ such that each x_k is a nonnegative integer and $\sum_{k=1}^m x_k = n$.

Marginal distribution

Consider a single count x_k

$$p(x_k) = \sum_{\sum_{i\neq k} x_i = n - x_k} \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m}$$

$$= \sum_{\sum_{i\neq k} x_i = n - x_k} \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m} \frac{(n-x_k)!(1-p_k)^{n-x_k}}{(n-x_k)!(1-p_k)^{n-x_k}}$$

$$= \frac{n!}{x_k!(n-x_k)!} p_k^{x_k} (1-p)_k^{n-x_k} \underbrace{\sum_{\sum_{i\neq k} x_i = n - x_k} \frac{(n-x_k)!}{\prod_{i\neq k} x_i!} \prod_{i\neq k} \left(\frac{p_i}{1-p_k} \right)^{x_i}}_{1}$$

$$= \frac{n!}{x_k!(n-x_k)!} p_k^{x_k} (1-p_k)^{n-x_k}$$

Therefore

$$x_k \sim Binomial(n, p_k)$$

Conditional distribution

Exclude a single count x_k

$$
\begin{aligned}
&p(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_m \mid x_k) \\
&\qquad = \frac{p(x_1,\ldots,x_m)}{p(x_k)} \\
&\qquad = \frac{(n-x_k)!}{x_1!\cdots x_{k-1}!\,x_{k+1}!\cdots x_m!}\prod_{i\neq k}\left(\frac{p_i}{1-p_k}\right)^{x_i}
\end{aligned}
$$

Therefore

$$x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_m \mid x_k \sim Multinomial(\tilde{n},\tilde{m},\tilde{\mathbf{p}}),$$

where

$$\tilde{n} = n - x_k;$$

$$\tilde{m} = m - 1;$$

$$\tilde{\mathbf{p}} = \left(\frac{p_1}{1-p_k},\ldots,\frac{p_{k-1}}{1-p_k},\frac{p_{k+1}}{1-p_k},\ldots,\frac{p_m}{1-p_k}\right)$$

Thank you very much