Mathematical Statistics and Data Analysis

Lecture 5: Review of Probability - Part IV

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

October 11, 2019

Outlines

Reading Material

Textbook:

- Rice: Chapter 3;
- Mao: Chapter 3;

Multivariate

Definition

Suppose $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ are n random variables defined on a sample space Ω. Then

$$\mathbf{X}(\omega) = \{X_1(\omega), X_2(\omega), \cdots, X_n(\omega)\}$$

is said to be a n-dimensional multivariate or n-dimensional random vector.

Cases

- We are interested in the height X_1 and weight X_2 of a school-aged child. (X_1, X_2) is a bivariate.
- We want to study the household expenses. Suppose X_1, X_2, X_3, X_4 are respectively denoted as the clothing, food, shelter, and transportation. Then, (X_1, X_2, X_3, X_4) is a four-dimensional random vector.

Multivariate

Definition

For any n real numbers $x_1, x_2, \cdots, x_n$, the joint c.d.f. of a $n-$dimensional random vector is

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$$

which is the probability that n events $\{X_1 \leq x_1\}$, $\{X_2 \leq x_2\}$, $\cdots$, $\{X_n \leq x_n\}$ simultaneously occur.

Special Case

When $n = 2$, the two-dimensional random variable (X, Y) are considered. The joint c.d.f. is

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

which means these two events simultaneously occur.

Multivariate

Theorem

$F(x, y)$ is the joint c.d.f. of a bivariate in and only if the function $F(x, y)$ satisfies

- (Monotonicity) For either x or y, $F(x, y)$ is increasing, i.e.
  - If $x_1 < x_2$, then $F(x_1, y) \leq F(x_2, y)$;
  - If $y_1 < y_2$, then $F(x, y_1) \leq F(x, y_2)$;
- (Boundedness) $0 \leq F(x, y) \leq 1$ for every x and y and

$$\begin{aligned} F(-\infty, y) &= \lim_{x \to -\infty} F(x, y) = 0 \\ F(x, -\infty) &= \lim_{y \to -\infty} F(x, y) = 0 \\ F(\infty, \infty) &= \lim_{x, y \to \infty} F(x, y) = 0 \end{aligned}$$

Multivariate

Theorem (Con'd)

$F(x, y)$ is the joint c.d.f. of a bivariate in and only if the function $F(x, y)$ satisfies

- (Right-continuousness) Each variate is right-continuous, that is,

$$\begin{aligned} F(x+0, y) &= F(x, y) \\ F(x, y+0) &= F(x, y) \end{aligned}$$

- (Non-negativity) For any $a < b$ and $c < d$, then

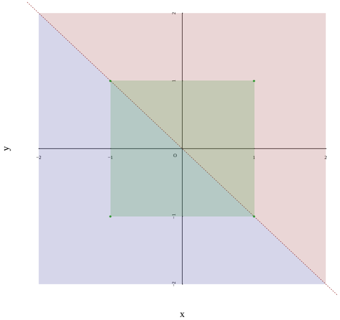
$$\begin{aligned} & F(a < X \leq b, c < Y \leq y) \\ = \; & F(b, d) - F(a, d) - F(b, c) + F(a, c) \end{aligned}$$

Why we need the non-negativity?

Example

A function is defined as follows:

$$G(x,y) = \begin{cases} 0, & x+y<0 \\ 1, & x+y \geq 0 \end{cases}$$



Multivariate

Definition

Suppose $F(x, y)$ is the joint c.d.f. of a random bivariate (X, Y).

- The **marginal c.d.f.** of X is

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \leq x, Y < \infty) \\ &= \lim_{y \to \infty} F(x, y) \overset{\text{def}}{=} F(x, \infty) \end{aligned}$$

- The **marginal c.d.f.** of Y is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X < \infty, Y \leq y) \\ &= \lim_{x \to \infty} F(x, y) \overset{\text{def}}{=} F(\infty, y) \end{aligned}$$

Discrete bivariate

Definition

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

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

is called the **joint probability mass function** or **joint p.m.f.** of (X, Y) If necessary, the notation $f_{X,Y}(x, y)$ will be also used.

Property

- **Non-negativity**: $f(x, y) \geq 0$ for any (x, y).
- **Normalization**: $\sum_{x \in \Re} \sum_{y \in \Re} f(x, y) = 1$.

Discrete bivariate

Suppose that

- X: $x_1, x_2, \cdots, x_n, \cdots$;
- Y: $y_1, y_2, \cdots, y_n, \cdots$;

Let $p_{ij} = f(x_i, y_j)$. The joint p.m.f. of (X, Y) is also presented as follows:

	y_1	y_2	$\cdots$	y_n	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nn}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

Discrete bivariate

Theorem

Let (X, Y) be a discrete bivariate random vector with joint p.m.f. $f_{X,Y}(x, y)$. Then the **marginal p.m.f.**s of X and Y, $f_X(x) = P(X = x)$ and $f_Y(y) = P(Y = y)$, are given by

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

Proof: We will prove the result for $f_X(x)$ and the proof for $f_Y(y)$ is similar. For any $x \in \Re$, let $A_x = \{(x, y) : -\infty < y < \infty\}$. That is, A_x is the line in the plane with first coordinate equal to x. Then for any $x \in \Re$,

$$\begin{aligned} f_X(x) &= P(X = x) = P(X = x, -\infty < y < \infty) = P((X, Y) \in A_x) \\ &= \sum_{(x,y) \in A_x} f_{X,Y} = \sum_{y \in \Re} f_{X,Y}(x, y) \end{aligned}$$

Example

Consider the experiment of tossing two fair dice. The samle space for this experiment has 36 equally likely points. Now, with each of these 36 points associate two numbers, X and Y. Let

$$X = \text{ sum of two dice and } Y = |\text{difference of the two dice}|$$

The values of the joint p.m.f. of (X, Y) are as follows:

		X										
		2	3	4	5	6	7	8	9	10	11	12
	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$	
Y	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		
	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			
	4					$\frac{1}{18}$		$\frac{1}{18}$				
	5						$\frac{1}{18}$					

Example(Con'd)

The marginal p.m.f.s can be calculated as follows:

		x											
		2	3	4	5	6	7	8	9	10	11	12	$f(y)$
y	0	$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$		$\frac{1}{36}$	$1/6$
	1		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$5/18$
	2			$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$			$2/9$
	3				$\frac{1}{18}$		$\frac{1}{18}$		$\frac{1}{18}$				$1/6$
	4					$\frac{1}{18}$		$\frac{1}{18}$					$1/9$
	5						$\frac{1}{18}$						$1/18$
$f_X(x)$		$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$	

Multinomial Distribution

Definition

Suppose that each of n independent trails can result in one of r types of outcomes and that on each trial the probabilities of the r outcomes are $p_1, p_2, \cdots, p_r$. Let X_i be the total number of outcomes of type i in the n trials, $i = 1, 2, \cdots, r$. $(X_1, X_2, \cdots, X_r)$ is said to be distributed as a **multinomial distribution** and the joint p.m.f. is

$$f(n_1, n_2, \cdots, n_r) = \binom{n}{n_1 n_2 \cdots n_r} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

where $n = n_1 + n_2 + \cdots + n_r$.

Multivariate Hypergeometric Distribution

Definition

Suppose that N balls are in a bag, where the number of ball No.i is N_i, $i = 1, 2, \cdots, r$ and $N = N_1 + N_2 + \cdots + N_r$. n balls are randomly taken out of the bag. Let X_i be the number of ball No.i among the n balls, $i = 1, 2, \cdots, r$. $(X_1, X_2, \cdots, X_r)$ is said to be distributed as a **multivariate hypergeometric distribution** and the joint p.m.f. is

$$f(n_1, n_2, \cdots, n_r) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_r}{n_r}}{\binom{N}{n}}$$

where $n_1 + n_2 + \cdots + n_r = n$.

Continuous bivariate

Definition

Suppose that $F(x, y)$ is a joint c.d.f. of a continuous bivariate random vector (X, Y). A function $f(x, y)$ from $\Re^2$ into $\Re$ is called a **joint probability density function** or **joint p.d.f.** of (X, Y) if, for any $x \in \Re$ and $y \in \Re$,

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y) \mathrm{d}x \mathrm{d}y$$

Property

- **Non-negativity**: $f(x, y) \geq 0$ for any (x, y);
- **Normalization**: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x \mathrm{d}y = 1$

Continuous bivariate

Definition

The **marginal probability density functions** or **joint p.d.f.** of X and Y are also defined by

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y, -\infty < x < \infty \\
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x, -\infty < x < \infty
\end{aligned}
$$

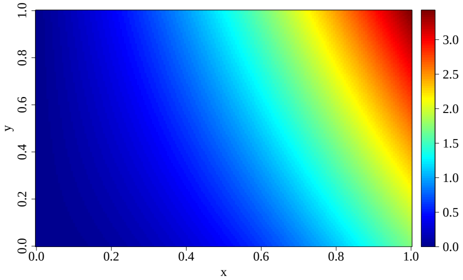
Continuous bivariate

Example

Consider the bivariate density function

$$f(x, y) = \frac{12}{7}(x^2 + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$$

which is plotted as follows.

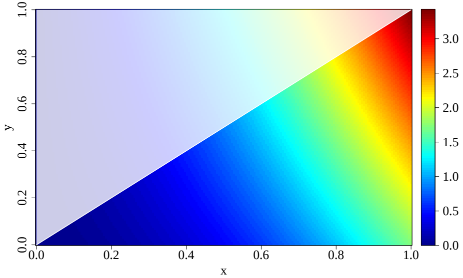


Continuous bivariate

Example (Con'd)

$P(X > Y)$ can be found by intergating f over the set

$$\{(x, y) : 0 \leq y \leq x \leq 1\}$$



$$P(X > Y) = \int_0^1 \int_0^x \frac{12}{7}(x^2 + xy)\mathrm{d}y\mathrm{d}x = \frac{9}{14}$$

Continuous bivariate

Example (Con'd)

The marginal p.d.f. of X is

$$f_X(x) = \int_0^1 \frac{12}{7}(x^2 + xy)\mathrm{d}y = \frac{12}{7}\left(x^2 + \frac{x}{2}\right).$$

The marginal p.d.f. of Y is

$$f_Y(y) = \int_0^1 \frac{12}{7}(x^2 + xy)\mathrm{d}x = \frac{12}{7}\left(\frac{1}{3} + \frac{y}{2}\right).$$

Multivariate uniform distribution

Definition

Suppose $D \subset R^n$ is a bounded region and the area of D is S_D. A multivariate random vector $(X_1, X_2, \cdots, X_n)$ is said to be a **multivariate uniform distribution**. The joint p.d.f. of $(X_1, X_2, \cdots, X_n)$ is

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{S_D} I\left\{(x_1, x_2, \cdots, x_n) \in D\right\}.$$

Bivariate Normal Distribution

Definition

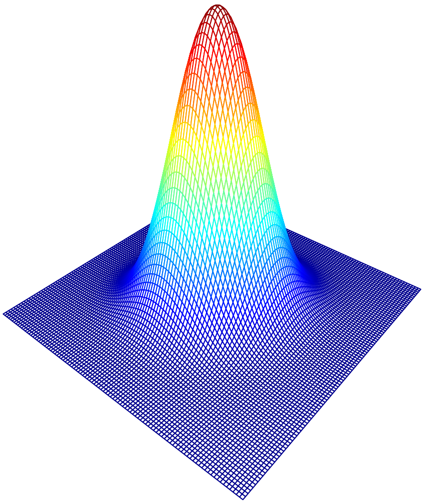
A random bivariate (X, Y) is said to be distributed as a **bivariate normal distribution**. The joint p.d.f. of (X, Y) is

$$
\begin{aligned}
f(x, y) \quad &= \quad \frac{1}{2\pi\sqrt{\sigma_X^2\sigma_Y^2(1-\rho^2)}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}\right.\right. \\
&\left.\left. + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right\}
\end{aligned}
$$

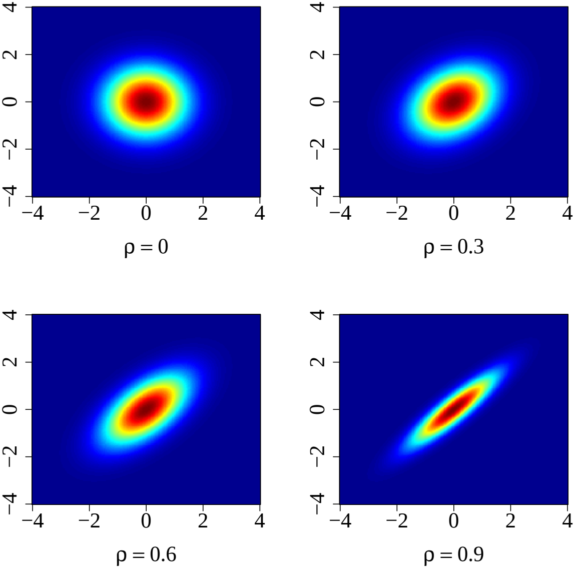
- (X, Y) is denoted as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$.
- Five parameters:

$$
-\infty < \mu_X, \mu_Y < \infty \quad \sigma_X, \sigma_Y > 0 \quad -1 < \rho < 1
$$

Bivariate Normal Distribution



Bivariate Normal Distribution



Bivariate Normal Distribution

Property

The marginal distributions of X and Y are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively.

Proof: Let $u = (x - \mu_X)/\sigma_X$ and $v = (y - \mu_Y)/\sigma_Y$. The joint p.d.f. of X is

$$f_X(x) = \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\rho uv)\right\} \mathrm{d}v$$

Using the identity

$$u^2 + v^2 - 2\rho uv = (v - \rho u)^2 + u^2(1-\rho^2)$$

we have

$$f_X(x) = \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} e^{-u^2/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}(v - \rho u)^2\right\} \mathrm{d}v$$

Bivariate Normal Distribution

Property

The marginal distributions of X and Y are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively.

Proof (Con'd): Finally, recognizing the integral as that of a normal density with mean ρv and variance $(1-\rho^2)$, we obtain

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}$$

which is a normal density, as was to be shown.

Independence

Definition

Random variables $X_1, X_2, \cdots, X_n$ are said to be **(mutually) independent** if their joint c.d.f. factors into the product of their marginal c.d.f.'s

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

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

- If $X_1, X_2, \cdots, X_n$ are discrete r.v.s and $f(x_i)$ is the p.m.f. of X_i. Then, $f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$;
- If $X_1, X_2, \cdots, X_n$ are continuous r.v.s and $f(x_i)$ is the p.d.f. of X_i. Then, $f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$;

Sum of Poisson random variables

Example

Suppose that $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$ and X and Y are independent. Then

$$Z = X + Y \sim P(\lambda_1 + \lambda_2).$$

Solution: The possible values of Z are non-negative integers such as $0, 1, 2, \cdots$. The event $\{Z = z\}$ is equivalent to the union of such disjoint events as

$$\{X = x, Y = z - x\}, i = 0, 1, \cdots, k.$$

For any non-negative integer k,

$$P(Z = z) = \sum_{i=0}^{z} P(X = x, Y = z - x)$$

Sum of Poisson random variables

Solution (Con'd): Since X and Y are independent, the marginal p.m.f. of Z is

$$
\begin{aligned}
f_Z(z) &= P(Z=z) = \sum_{x=0}^{z} \left(\frac{\lambda_1^x}{x!} e^{-\lambda_1} \right) \left(\frac{\lambda_2^{(z-x)}}{(z-x)!} e^{-\lambda_2} \right) \\
&= \frac{(\lambda_1+\lambda_2)^z}{z!} e^{-(\lambda_1+\lambda_2)} \cdot \\
&\quad \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \cdot \left(\frac{\lambda_1}{\lambda_1+\lambda_2} \right)^x \cdot \left(\frac{\lambda_2}{\lambda_1+\lambda_2} \right)^{z-x} \\
&= \frac{(\lambda_1+\lambda_2)^z}{z!} e^{-(\lambda_1+\lambda_2)}
\end{aligned}
$$

Sum of Poisson random variables

Remark

- The **convolution** of two Poisson distribution functions is still a Poisson distribution function, that is,

$$P(\lambda_1) * P(\lambda_2) = P(\lambda_1 + \lambda_2)$$

- Generalization:

$$P(\lambda_1) * P(\lambda_2) * \cdots * P(\lambda_n) = P(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$

  Particularly, when $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$,

$$P(\lambda) * P(\lambda) * \cdots * P(\lambda) = P(n\lambda)$$

Sum of continuous variables

Theorem

Suppose X and Y are two independent continuous r.v.s and the p.d.fs are respectively $p_X(x)$ and $p^Y(y)$. Then the p.d.f. of $Z = X + Y$ is

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z-y)p_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)\mathrm{d}x$$

Proof: The c.d.f. of Z is

$$\begin{aligned} F_Z(z) &= P(X+Y \leq z) = \iint\limits_{x+y\leq z} p_X(x)p_Y(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} p_X(x)\mathrm{d}x\right)p_Y(y)\mathrm{d}y \end{aligned}$$

Sum of continuous variables

Proof (Con'd):

$$
\begin{aligned}
F_Z(z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-y} p_X(x) \mathrm{d}x \right) p_Y(y) \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z} p_X(t-y) p_Y(y) \mathrm{d}t \mathrm{d}y \\
&= \int_{-\infty}^{z} \int_{-\infty}^{\infty} p_X(t-y) p_Y(y) \mathrm{d}y \mathrm{d}t
\end{aligned}
$$

Thus, the p.d.f. of Z is

$$
p_Z(z) = \int_{-\infty}^{\infty} p_X(z-y) p_Y(y) \mathrm{d}y.
$$

Similarly, $p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z-y) \mathrm{d}x$.

Sum of variables

Remark

- Binomial Distribution:
  $b(n_1, p) * b(n_2, p) * \cdots * b(n_k, p) = b(n_1 + n_2 + \cdots + n_k, p)$;
- Normal Distribution: If $X_i \sim N(\mu_i, \sigma_i^2)$, then

  $$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n \sim N(\mu_0, \sigma_0^2)$$

  where $\mu_0 = \sum_{i=1}^n a_i \mu_i$ and $\sigma_0^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$;
- Gamma Distribution:
  $Ga(\alpha_1, \lambda) * Ga(\alpha_2, \lambda) * \cdots * Ga(\alpha_n, \lambda) = Ga(\sum_{i=1}^n \alpha_i, \lambda)$
  - Exponential Distribution:
    $\underbrace{Exp(\lambda) * Exp(\lambda) * \cdots * Exp(\lambda)}_{m} = Ga(m, \lambda)$;
  - Chi-square Distribution:
    $\chi^2(n_1) * \chi^2(n_2) * \cdots * \chi^2(n_m) = \chi^2(\sum_{i=1}^m n_i)$

Maximum random variable

Theorem

Suppose that $X_1, X_2, \cdots, X_n$ are n mutually independent variates. Let $Y = \max\{X_1, X_2, \cdots, X_n\}$.

- If $X_i \sim F_i(x)$, then the c.d.f. of Y is $\prod_{i=1}^n F_i(y)$;
- If X_i's are identically distributed, i.e. $X_i \sim F(x)$, then the c.d.f. of Y is $(F(y))^n$;
- If X_i are identically distributed and continuous r.v.s with the p.d.f. $f(x)$, then the p.d.f. of Y is

$$p_Y(y) = F'_Y(y) = n[F(y)]^{n-1} f(y).$$

Minimum random variable

Theorem

Suppose that $X_1, X_2, \cdots, X_n$ are n mutually independent variates. Let $Y = \min\{X_1, X_2, \cdots, X_n\}$.

- If $X_i \sim F_i(x)$, then the c.d.f. of Y is $1 - \prod_{i=1}^{n}(1 - F_i(y))$;
- If X_i's are identically distributed, i.e. $X_i \sim F(x)$, then the c.d.f. of Y is $1 - (1 - F(y))^n$;
- If X_i are identically distributed and continuous r.v.s with the p.d.f. $f(x)$, then the p.d.f. of Y is

$$p_Y(y) = F_Y'(y) = n[1 - F(y)]^{n-1} f(y).$$

Transformation of a Bivariate

Suppose that the joint p.d.f. of two continuous bivariate vector (X, Y) is $f(x, y)$. If the function

$$\begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases}$$

has continuous partial derivatives and the inverse function

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \end{cases}$$

exists and it is unique with a Jacobian determinant

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \left(\frac{\partial(u, v)}{\partial(x, y)}\right)^{-1} = \left(\begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}\right)^{-1} \neq 0$$

Transformation of a Bivariate

Let

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases}$$

Then the joint p.d.f. of (U, V) is

$$f_{(U,V)}(u, v) = f_{(X,Y)}(x(u, v), y(u, v))|J|.$$

Transformation of a Bivariate

Example

Suppose that X and Y are independently and identically distributed with a normal distribution $N(\mu, \sigma^2)$. Let

$$\begin{cases} U = X + Y \\ V = X - Y \end{cases}$$

Find the joint p.d.f. of (U, V). Is U independent from V?

Solution: Since

$$\begin{cases} u = x + y \\ v = x - y \end{cases},$$

the inverse function is

$$\begin{cases} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{cases}.$$

Transformation of a Bivariate

Solution (Con'd): Then the Jacobian determinant is

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

Thus, the joint p.d.f. of (U, V) is

$$\begin{aligned} f(u,v) &= f(x(u,v), y(u,v))|J| = f_X\left(\frac{u+v}{2}\right) f_Y\left(\frac{u-v}{2}\right) \cdot \left|-\frac{1}{2}\right| \\ &= \frac{1}{4\pi\sigma^2} \exp\left\{-\frac{[(u+v)/2-\mu]^2}{2\sigma^2}\right\} \exp\left\{-\frac{[(u-v)/2-\mu]^2}{2\sigma^2}\right\} \\ &= \frac{1}{4\pi\sigma^2} \exp\left\{-\frac{(u-2\mu)^2+v^2}{4\sigma^2}\right\} \end{aligned}$$

Transformation of a Bivariate

Remark

- Joint: $(U, V) \sim N(2\mu, 0, 2\sigma^2, 2\sigma^2, 0)$;
- Marginal: $U \sim N(2\mu, 2\sigma^2)$, $V \sim N(0, 2\sigma^2)$;
- U and V are independent.

Transformation of a Bivariate

Special Case I:

Suppose that X and Y are independent and the p.d.f. of X and Y are respectively $f_X(x)$ and $f_Y(y)$. Then the p.d.f. of $U = XY$ is

$$f_U(u) = \int_{-\infty}^{\infty} f_X\left(\frac{u}{v}\right) f_Y(v) \frac{1}{|v|} \mathrm{d}v$$

Solution: Let $V = Y$. Then $\begin{cases} u = xy \\ v = y \end{cases}$ and the inverse function is $\begin{cases} x = \frac{u}{v} \\ y = v \end{cases}$. The Jacobian determinant is $J = \begin{vmatrix} 1/v & -u/v^2 \\ 0 & 1 \end{vmatrix} = \frac{1}{v}$. Then the joint p.d.f. of (U, V) is

$$f(u, v) = f_X\left(\frac{u}{v}\right) f_Y(v)|J| = f_X\left(\frac{u}{v}\right) f_Y(v)\frac{1}{|v|}.$$

Transformation of a Bivariate

Special Case II:

Suppose that X and Y are independent and the p.d.f. of X and Y are respectively $f_X(x)$ and $f_Y(y)$. Then the p.d.f. of $U = X/Y$ is

$$f_U(u) = \int_{-\infty}^{\infty} f_X(uv)\, f_Y(v)|v|\mathrm{d}v$$

Solution: Let $V = Y$. Then $\begin{cases} u = x/y \\ v = y \end{cases}$ and the inverse function is $\begin{cases} x = uv \\ y = v \end{cases}$. The Jacobian determinant is $J = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$. Then the joint p.d.f. of (U, V) is

$$f(u, v) = f_X(uv)\, f_Y(v)|J| = f_X(uv)\, f_Y(v)|v|.$$

Expectation & Variance

Theorem

Suppose the joint p.m.f. or p.d.f. of a bivariate vector (X, Y) is $f(x, y)$. The expectation of $Z = g(X, Y)$ is

$$E(Z) = \begin{cases} \sum_x \sum_y g(x, y) f(x, y), & X \text{ and } Y \text{ are discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \mathrm{d}x \mathrm{d}y, & X \text{ and } Y \text{ are continuous.} \end{cases}$$

Expectation & Variance

Remark (Con'd)

- If $g(X,Y) = X$, then

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$$

- If $g(X,Y) = (X - E(X))^2$, then

$$\begin{aligned} Var(X) &= E(X - E(X))^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - E(X))^2 f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x)\mathrm{d}x \end{aligned}$$

Expectation & Variance

Property

Suppose that $X_1, X_2, \cdots, X_n$ are n random variables.

- $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X)$;
- If X's are independent,
    - $E(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} E(X_i)$;
    - $Var(X_1 \pm X_2 \pm \cdots \pm X_n) = \sum_{i=1}^{n} Var(X_i)$;
- Particularly, $X_1, X_2, \cdots, X_n$ are independently and identically distributed with the variance σ^2. Then,

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$$

Covariance

Definition

Suppose (X, Y) is a bivariate vector. The **covariance** of X and Y is defined as

$$E\left((X - E(X))(Y - E(Y))\right).$$

Remark

- $Cor(X, X) = Var(X)$;
- If $Cov(X, Y) > 0$, X and Y are **positively** correlated;
- If $Cov(X, Y) < 0$, X and Y are **negatively** correlated;
- If $Cov(X, Y) = 0$, X and Y are **not** correlated;

Covariance

Property

Suppose X and Y are two variables.

- $Cov(X, Y) = E(XY) - E(X)E(Y)$;
- If X and Y are independent, $Cov(X, Y) = 0$; but not vice versa.
- X and Y are not correlated $\Leftrightarrow E(XY) = E(X)E(Y)$;
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$;
- $Cov(X, Y) = Cov(Y, X)$;
- If a is a constant, $Cov(X, a) = 0$;
- For any constants a and b, $Cov(aX, bY) = abCov(X, Y)$;
- Suppose Z is another variable.
  $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$;

Correlation

Definition

Suppose (X, Y) is a bivariate vector. The **correlation** of X and Y is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Remark

Suppose that μ_X and μ_Y are respectively the expectations of X and Y and σ_X and σ_Y are respectively the standard deviations. Let

$$X^* = \frac{X - \mu_X}{\sigma_X} \text{ and } Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

Then, $Cov(X^*, Y^*) = Corr(X, Y)$.

Correlation

Lemma (Schwarz Inequality)

Suppose that X and Y are two random variables. If the variances of X and Y exist, then

$$(Cov(X,Y))^2 \leq Var(X)Var(Y)$$

Property

Suppose that X and Y are two variables.

- $-1 \leq Corr(X,Y) \leq 1$, i.e. $|Corr(X,Y)| \leq 1$;
- There exist $a(\neq 0)$ and b such that $P(Y = aX + b) = 1$ $\Leftrightarrow Corr(X,Y) = \pm 1$;

Example

Suppose that (X, Y) is distributed as a bivariate normal distribution $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Then the correlation of X and Y is ρ.

Solution: The covariance of (X, Y) is

$$
\begin{aligned}
Cov(X,Y) &= E(X - E(X))(Y - E(Y)) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_1)(y-\mu_2) \\
&\quad \cdot \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right.\right. \\
&\qquad \left.\left. + \frac{(y-\mu_2)^2}{\sigma_2^2} \right) \right\} \mathrm{d}x\mathrm{d}y
\end{aligned}
$$

Solution (Con'd): We know

$$\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)$$
$$= \left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2 + \left(\sqrt{1-\rho^2}\frac{y-\mu_2}{\sigma_2}\right)^2.$$

Let

$$\begin{cases} u = \dfrac{1}{\sqrt{1-\rho^2}}\left(\dfrac{x-\mu_1}{\sigma_1} - \rho\dfrac{y-\mu_2}{\sigma_2}\right) \\ v = \dfrac{y-\mu_2}{\sigma_2} \end{cases}$$

Correlation

Solution (Con'd): Then

$$\begin{cases} x - \mu_1 = \sigma(u\sqrt{1-\rho^2} + \rho v) \\ y - \mu_2 = \sigma_2 v \end{cases}$$

and

$$\mathrm{d}x\mathrm{d}y = |J|\mathrm{d}u\mathrm{d}v = \sigma_1\sigma_2\sqrt{1-\rho^2}\mathrm{d}u\mathrm{d}v$$

Thus,

$$Cov(X,Y) = \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(uv\sqrt{1-\rho^2} + \rho v^2)\exp\left\{-\frac{1}{2}(u^2+v^2)\right\}\mathrm{d}u\mathrm{d}v.$$

As we know,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv\exp\left\{-\frac{1}{2}(u^2+v^2)\right\}\mathrm{d}u\mathrm{d}v = 0$$

Correlation

Solution (Con'd):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v^2 \exp\left\{-\frac{1}{2}(u^2+v^2)\right\} \mathrm{d}u\mathrm{d}v = 2\pi$$

Therefore,

$$\begin{aligned} Cov(X,Y) &= \frac{\sigma_1\sigma_2}{2\pi}\cdot\rho\cdot 2\pi = \rho\sigma_1\sigma_2 \\ Corr(X,Y) &= \frac{Cov(X,Y)}{\sigma_1\sigma_2} = \rho \end{aligned}$$

Remark

If (X,Y) is a bivariate normal distribution, then X and Y are independent if and only if $\rho = 0$.

Matrix Form

Definition

Suppose $\mathbf{X} = (X_1, X_2, \cdots, X_n)'$ is a $n-$dimensional random vector.

- The **expectation vector** of $\mathbf{X}$ is defined as

$$E(\mathbf{X}) = (E(X_1), E(X_2), \cdots, E(X_n))'$$

- The **variance-covariance matrix** of $\mathbf{X}$ is defined as

$$\begin{aligned} Cov(\mathbf{X}) &= E(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))' \\ &= \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & Var(X_n) \end{pmatrix} \end{aligned}$$

Matrix Form

Theorem

The variance-covariance matrix $Cov(\mathbf{X}) = (Cov(X_i, X_j))_{n\times n}$ is a symmetric and non-negative definite matrix.

Proof: It is obvious that $Cov(\mathbf{X})$ is symmetric since $Cov(X_i, X_j) = Cov(X_j, X_i)$. Then, we want to prove that this matrix is non-negative definite. For any real-valued vector $\mathbf{c} = (c_1, c_2, \cdots, c_n)'$,

$$\begin{aligned}
\mathbf{c}'Cov(\mathbf{X})\mathbf{c} &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j Cov(X_i, X_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left(c_i(X_i - E(X_i))\right)(c_j(X_j - E(X_j))) \\
&= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}(c_i(X_i - E(X_i)))(c_j(X_j - E(X_j)))\right)
\end{aligned}$$

Matrix Form

Proof (Con'd):

$$\begin{aligned}
\mathbf{c}' Cov(\mathbf{X})\mathbf{c} &= E\left(\sum_{i=1}^{n}(c_i(X_i - E(X_i)))\right)\left(\sum_{j=1}^{n}(c_j(X_j - E(X_j)))\right) \\
&= E\left(\sum_{i=1}^{n}(c_i(X_i - E(X_i)))\right)^2 \geq 0
\end{aligned}$$

Thus, $Cov(\mathbf{X})$ is non-negative definite.

Matrix Form

Example

Suppose that $\mathbf{X} = (X_1, X_2, \cdots, X_n)'$ is a $n-$dimensional random variable vector. The expectation vector is $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)'$ and the covariance matrix is $\mathbf{\Sigma} = Cov(\mathbf{X})$. If the joint p.d.f. of $\mathbf{X}$ is

$$f(\mathbf{x}) = \frac{\mathbf{1}}{(\mathbf{2}\pi)^{\mathbf{n/2}}|\mathbf{\Sigma}|^{\mathbf{1/2}}} \exp\left\{-\frac{\mathbf{1}}{\mathbf{2}}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-\mathbf{1}}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

where $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{-1}$ is the inverse function of Σ and $(\mathbf{x} - \boldsymbol{\mu})'$ is the transpose of $(x - \boldsymbol{\mu})$, then $\mathbf{X}$ is said to be $n-$dimensional variate normal distribution.

Conditional Distribution

Definition

Let (X, Y) be a discrete bivariate random vector with joint p.m.f. $f(x, y)$ and marginal p.m.f.s $f_X(x)$ and $f_Y(y)$.

- For any x such that $P(X = x) = f_X(x) > 0$, the **conditional p.m.f.** of Y given that $X = x$ is the function of y denoted as $f(y|x)$ and 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 p.m.f.** of X given that $Y = y$ is the function of x denoted as $f(x|y)$ and defined by

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

Conditional Distribution

Definition (Con'd)

- Given $X = x$, the **conditional c.d.f.** of Y is

$$F(y|x) = \sum_{t \leq y} P(Y = t | X = x)$$

- Given $Y = y$, the **conditional c.d.f.** of X is

$$F(x|y) = \sum_{t \leq x} P(X = t | Y = y)$$

Conditional Distribution

Example

Suppose that X and Y are independent and $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$. Given $X + Y = n$, what is the conditional distribution of X?

Solution: As we know, $X + Y \sim P(\lambda_1 + \lambda_2)$. Then,

$$P(X = k|X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)}$$

$$= \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} = \frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1} \cdot \frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}}$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}$$

Then, given $X + Y = n$, $X \sim b(n, p)$, where $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$.

Conditional Distribution

Definition

Let (X, Y) be a continuous bivariate random vector with joint p.d.f. $f(x, y)$ and marginal p.d.f.s $f_X(x)$ and $f_Y(y)$.

- For any x such that $P(X = x) = f_X(x) > 0$, the **conditional p.d.f.** of Y given that $X = x$ is the function of y denoted as $f(y|x)$ and 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 p.d.f.** of X given that $Y = y$ is the function of x denoted as $f(x|y)$ and defined by

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

Conditional Distribution

Definition (Con'd)

- Given $X = x$, the **conditional c.d.f.** of Y is

$$F(y|x) = \int_{-\infty}^{y} \frac{f(x,t)}{f_X(x)} \mathrm{d}t$$

- Given $Y = y$, the **conditional c.d.f.** of X is

$$F(x|y) = \int_{-\infty}^{x} \frac{f(t,y)}{f_Y(y)} \mathrm{d}t$$

Conditional Distribution

Example

Suppose $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. What is the condition p.d.f. of X given $Y = y$?

Solution: As we know, the marginal distribution of Y is $N(\mu_2, \sigma_2^2)$. Then,

$$
\begin{aligned}
f(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right)^2\right\}.
\end{aligned}
$$

Thus, given $Y = y$, the conditional distribution of X is a normal distribution with the expectation $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$ and the variance $\sigma_1^2(1-\rho^2)$.

Conditional Distribution

From the definition of the conditional p.d.f. of a continuous variable,

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

- **Law of Total Probability:**

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x)f(x,y)\mathrm{d}x \text{ and } f_X(x) = \int_{-\infty}^{\infty} f_Y(y)f(x,y)\mathrm{d}y$$

- **Bayes' Formula:**

$$f(x|y) = \frac{f_X(x)f(y|x)}{\int_{-\infty}^{\infty} f_X(x)f(y|x)\mathrm{d}x} \text{ and } f(y|x) = \frac{f_Y(y)f(x|y)}{\int_{-\infty}^{\infty} f_Y(y)f(x|Y)\mathrm{d}y}$$

Conditional Expectation

Definition

Suppose $f(x|y)$ is the conditional p.m.f or p.d.f. of X given $Y = y$. Given $Y = y$, the **conditional expectation** of X is denoted by $E(X|Y = y)$ and is defined by

$$E(X|Y = y) = \begin{cases} \sum_x x f(x|y) & X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} x f(x|y) \mathrm{d}x & X \text{ and } Y \text{ are continuous} \end{cases}$$

Remark

- Properties of the expectations;
- $E(X|Y = y)$ is a function of y;
- $E(X|Y)$ is also a random variable;

Conditional Expectation

Theorem

- $E(Y) = E(E(Y|X))$;
- $Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$;

Example

Let $T = \sum_{i=1}^{N} X_i$, where N is a random variable with a finite expectation and the X_i are random variable that are independent of N and have the common expectation $E(X)$. Then

$$
\begin{aligned}
E(T) &= E\left(\sum_{i=1}^{N} T_i\right) = E\left(E\left(\sum_{i=1}^{N} T_i|N\right)\right) \\
&= \sum_{i=1}^{\infty} E\left(\sum_{i=1}^{N} T_i|N=n\right) P(N=n) = E(X)E(N)
\end{aligned}
$$

Example (Con'd)

We further assume that X_i are independent random variables with the same expectation $E(X)$, and the same variance $Var(X)$, and that $Var(N) < \infty$.

$$\begin{aligned} Var(T) &= E(Var(T|N)) + Var(E(T|N)) \\ &= E(N)Var(X) + (E(X))^2 Var(N). \end{aligned}$$