Several random variables

Multivariate distributions

Francesco Corona

Department of Computer Science
Federal University of Ceará, Fortaleza

Several RVs
Distributions
Variance-covariance
Transformations
Linear combinations

Several random variables

Distributions

The notions developed for two random variable can be extended to n RVs

⇝ We start by defining the space of n random variables

Distributions (cont.)

Definition

Consider a random experiment with sample space $\mathcal{C}$

Let the random variable X_i assign to each element $c \in \mathcal{C}$ one and only one real number $X_i(c) = x_i,\ i = 1, 2, \dots, n$

*We say that $(X_1, \dots, X_n)$ is a n-dimensional **random vector***

*The **space**/**range** of the random vector is the set of ordered n-tuples*

$$\mathcal{D} = \big\{(x_1, x_2, \dots, x_n) : x_1 = X_1(c), \dots, x_n = X_n(c), c \in \mathcal{C}\big\}$$

Distributions (cont.)

Definition

Let A be a subset of the space $\mathcal{D}$

Then, $P\big[(X_1, \ldots, X_n) \in A\big] = P(C)$

$$C = \big\{c : c \in \mathcal{C} \text{ and } (X_1(c), X_2(c), \ldots, X_n(c) \in A)\big\}$$

Distributions (cont.)

We denote $(X_1, \ldots, X_n)'$ by the n-dimensional (column) vector $\mathbf{X}$

- The observed values $(x_1, \ldots, x_n)'$ of the RV are $\mathbf{x}$

Distributions

Several random variables

Distributions (cont.)

The joint CDF is defined as always

$$F_{\mathbf{X}}(\mathbf{x}) = P\big[X_1 \leq x_1, \dots, X_n \leq x_n\big] \tag{1}$$

The n RVs $X_1, X_2, \dots, X_n$ can of the discrete type or of the continuous type

- The (cumulative) distribution

$$\rightsquigarrow F_{\mathbf{X}}(\mathbf{x}) = \sum_{w_1 \leq x_1} \sum_{w_2 \leq x_2} \cdots \sum_{w_n \leq x_n} p(w_1, w_2, \dots, w_n)$$

$$\rightsquigarrow F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(w_1, w_2, \dots, w_n) \mathrm{d}w_n \cdots \mathrm{d}w_2 \mathrm{d}w_1$$

Distributions (cont.)

For the continuous case

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x}) \tag{2}$$

(Except, possibly, at points with zero probability)

Distributions (cont.)

A continuous function f must satisfy the conditions of a PDF

- f is defined and is non-negative for all real values of its argument(s)
- Its integral over all real values of its argument(s) is 1

A point function p must satisfy the conditions of a PMF

- p is defined and is non-negative for all real values of its argument(s)
- Its sum over all real values of its argument(s) is 1

Distributions (cont.)

As always, it is convenient to define the support of a random vector

Discrete case

- All points in $\mathcal{D}$ that have positive probability mass

Continuous case

- All points in $\mathcal{D}$ that can be placed in an open set of positive probability

We use $\mathcal{S}$ to denote support sets

Distributions (cont.)

Example

Let X, Y and Z be random variables with the joint PDF

$$f(x,y,z) = \begin{cases} e^{-(x+y+z)}, & 0 < x, y, z < \infty \\ 0, & \text{elsewhere} \end{cases}$$

The distribution function of X, Y and Z

$$\begin{aligned} F(x,y,z) = P(X \leq x, Y \leq y, Z \leq z) &= \int_0^z \int_0^y \int_0^x e^{-u-v-w} \mathrm{d}u \mathrm{d}v \mathrm{d}w \\ &= \Big(1 - \frac{1}{e^x}\Big)\Big(1 - \frac{1}{e^y}\Big)\Big(1 - \frac{1}{e^z}\Big), \quad 0 \leq x, y, z < \infty \end{aligned} \tag{3}$$

and zero elsewhere

$$\rightsquigarrow \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x})$$

Distributions (cont.)

Let $(X_1, X_2, \dots, X_n)$ be a random vector, let $Y = u(X_1, X_2, \dots, X_n)$

- (For some function u)

As with the bivariate case, the expected value of the RV may exist

Continuous case

$\rightsquigarrow$ The n-fold integral must exist

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \big|u(x_1, x_2, \dots, x_n)\big| f(x_1, x_2, \dots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

Discrete case

$\rightsquigarrow$ The n-fold sum must exist

$$\sum_{x_n} \cdots \sum_{x_1} \big|u(x_1, x_2, \dots, x_n)\big| p(x_1, x_2, \dots, x_n)$$

Distributions (cont.)

If the expected value of Y exists, then we can determine its expectation

- Continuous case

$$E(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \dots, x_n) f(x_1, \dots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \qquad (4)$$

- Discrete case

$$E(Y) = \sum_{x_n} \cdots \sum_{x_1} u(x_1, \dots, x_n) p(x_1, \dots, x_n) \qquad (5)$$

Distributions (cont.)

The usual properties of expectation hold for the n-dimensional case

- $\rightsquigarrow$ E is a linear operator

Let $Y_j = u_j(X_1, \dots, X_n)$, for $j = 1, \dots, m$, suppose that each $E(Y_i)$ exist

$$E\Big[\sum_{j=1}^{m} k_j Y_j\Big] = \sum_{j=1}^{m} k_j E(Y_j), \tag{6}$$

for some constants $k_1, \dots, k_m$

Distributions (cont.)

Notions of marginal and conditional probability density functions for n RVs

- Previous definitions can be generalised to n-variables

Distributions (cont.)

Let the RVs $X_1, X_2, \ldots, X_n$ be of the continuous type with joint PDF

$$f(x_1, x_2, \ldots, x_n)$$

As with the bivariate case, we have for every b

$$F_{X_1}(b) = P(X_1 \leq b) = \int_{-\infty}^{b} f_1(x_1) \mathrm{d}x_1$$

- $f_1(x_1)$ is defined by the $(n-1)$-fold integral

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

$f_1(x_1)$ is the PDF of the RV X_1, the marginal PDF of X_1

Distributions (cont.)

Remark

The marginal probability density functions $f_2(x_2), \dots, f_n(x_n)$ of $X_2, \dots, X_n$

- They are similar $(n-1)$-fold integrals

Each marginal PDF is a PDF of one random variable

Distributions (cont.)

It is possible and convenient to extend the terminology to joint PDFs

Let $f(x_1, \dots, x_n)$ be the joint PDF of n random variables $X_1, \dots, X_n$

Let us take any group of $k < n$ of these RVs

We wish to find their joint PDF

- $\rightsquigarrow$ This joint PDF is the marginal PDF of the k-group of RVs

Distributions (cont.)

Example

Let $n = 6$, $X_1, X_2, X_3, X_4, X_5, X_6$ are some RVs with some joint PDF

$$f(x_1, x_2, x_3, x_4, x_5, x_6)$$

Let $k = 3$ and let us select the group X_2, X_4, X_5

The marginal PDF of X_2, X_4, X_5 is the joint PDF of the group

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4, x_5, x_6)\mathrm{d}x_1\mathrm{d}x_3\mathrm{d}x_6$$

(if the random variables are of the continuous type)

Distributions (cont.)

As for the conditional PDF, suppose $f_1(x_1) > 0$, then define the symbol

$$f_{2,\dots,n|1}(x_2,\dots,x_n|x_1) = \frac{f(x_1,x_2,\dots,x_n)}{f_1(x_1)}$$

This is the **joint conditional PDF** of $X_2,\dots,X_n$, given $X_1 = x_1$

Joint conditional PDF of $(n-1)$ RVs, $X_1,\dots,X_{i-1},X_{i+1},\dots,X_n$, $X_i = x_i$

- The joint PDF of $X_1,\dots,X_n$ divided by the marginal PDF $f_i(x_i)$
- (Provided that $f_i(x_i) > 0$)

Distributions (cont.)

Remark

Or, more generally

The joint conditional PDF of any $(n-k)$ of the RVs for given values of the remaining k RVs is defined as the joint PDF of the n RVs divided by the marginal PDF of the group of k RVs (provided the latter PDF is positive)

Distributions (cont.)

A joint conditional PDF is a PDF of a certain set of random variables

- $\rightsquigarrow$ The expectation of a function of these RVs is defined

We must emphasise that a particular conditional PDF is considered

- Such expectations are called conditional expectations

Distributions (cont.)

The conditional expectation of $u(X_2, \dots, X_n)$, given $X_1 = x_1$

For the case of RVs of the continuous type

$$E\big[u(X_2, \dots, X_n)\big|x_1\big] = \int_{-\infty}^{\infty} \cdots \int_{\infty}^{\infty} u(x_2, \dots, x_n) f_{2,\dots,n|1}(x_2, \dots, x_n|x_1) \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

- (Provided that $f_1(x_1) > 0$ and that the integral converges absolutely)

$h(X_1) = E\big[u(X_2, \dots, X_n)\big|X_1\big]$ is a (useful) random variable

Distributions (cont.)

The concept of marginal/conditional distributions generalise to discrete RVs

- Use PMFs and summations instead of PDFs and integrals

Distributions (cont.)

Let the RVs $(X_1, X_2, \ldots, X_n)$ have the joint PDF $f(x_1, x_2, \ldots, x_n)$

Let $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$ be the marginal PDFs

We generalise the bivariate definition of independence of RVs X_1 and X_2

- Mutual independence of $X_1, X_2, \ldots, X_n$

RVs $X_1, X_2, \ldots, X_n$ are said to be **mutually independent** if and only if

⇝ Continuous case

$$f(x_1, \ldots, x_n) \equiv f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$

⇝ Discrete case

$$p(x_1, \ldots, x_n) \equiv p_1(x_1) p_2(x_2) \cdots p_n(x_n)$$

Distributions (cont.)

Suppose $X_1, X_2 \dots, X_n$ are mutually independent

Then,

$$P\big[(a_1 < X_1 < b_1), \dots, (a_n < X_n < b_n)\big]$$
$$= P(a_1 < X_1 < b_1) \cdots P(a_n < X_n < b_n) = \prod_{i=1}^{n} P(a_i < X_i < b_i)$$

We define the symbol $\prod_{i=1}^{n} \varphi(i)$ as always

$$\prod_{i=1}^{n} \varphi(i) = \varphi(1)\varphi(2) \cdots \varphi(n)$$

Distributions (cont.)

For two independent random variables X_1 and X_2

$$E\Big[u(X_1)v(X_2)\Big] = E\Big[u(X_1)\Big]E\Big[v(X_2)\Big]$$

For n mutually independent random variables $X_1, X_2, \dots, X_n$

$$E\Big[u_1(X_1)u_2(X_2)\cdots u_n(X_n)\Big] = E\Big[u_1(X_1)\Big]E\Big[u_2(X_2)\Big]\cdots E\Big[u_n(X_n)\Big]$$

or, compactly

$$E\Big[\prod_{i=1}^{n} u_i(X_i)\Big] = \prod_{i=1}^{n} E\Big[u_i(X_i)\Big]$$

Distributions (cont.)

We can define the MGF of the joint distribution of n RVs $X_1, X_2, \dots, X_n$

Suppose that the expectation $M(t_1, t_2, \dots, t_n)$ exist

$$E\left[\exp\left(t_1 X_1 + t_2 X_2 + \cdots + t_n X_n\right)\right]$$

For $-h_i < t_i < h_i$, $i = 1, 2, \dots, n$, with each h_i positive

This expectation is the MGF of the joint distribution of $X_1, X_2, \dots, X_n$

- It is unique

It uniquely determines the joint distribution of the n variables

$\rightsquigarrow$ (hence, also all marginal distributions)

Distributions (cont.)

$\rightsquigarrow$ The MGF of the marginal distributions of X_i

$$\rightsquigarrow M(0, \ldots, 0, t_i, 0, \ldots, 0)$$

$\rightsquigarrow$ The MGF of the marginal distributions of X_i and X_j

$$\rightsquigarrow M(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0)$$

Distributions (cont.)

We can also generalise the bivariate theorem

Theorem

Suppose the joint MGF $M(t_1, t_2)$ exists for random variables X_1 and X_2

Then, X_1 and X_2 are independent if and only if

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2)$$

The joint MGF is identically equal to the product of the marginals MGFs

Distributions (cont.)

Consider the mutual independence of $X_1, X_2, \dots, X_n$

- $\rightsquigarrow$ The factorisation is a necessary and sufficient condition

$$M(t_1, t_2, \dots, t_n) = \prod_{i=1}^{n} M(0, \dots, 0, t_i, 0, \dots, 0) \tag{7}$$

The joint MGF in vector notation reads

$$M(\mathbf{t}) = E\big[\exp(\mathbf{t}'\mathbf{X})\big], \text{ for } \mathbf{t} \in B \subset \mathcal{R}^n$$

$B = \big\{\mathbf{t} : -h_i < t_i < +h_i, i = 1, \dots, n\big\}$

Distributions (cont.)

Theorem 1.1

Suppose $X_1, X_2, \ldots, X_n$ are n mutually independent random variables

Suppose X_i has MGF $M_i(t)$ for $-h_i < t < h_i$, $h_i > 0$ $(i = 1, 2, \ldots, n)$

Let $T = \sum_{i=1}^n k_i X_i$, where $k_1, k_2, \ldots, k_n$ are constants

Then, T has the MGF

$$M_T(t) = \prod_{i=1}^n M_i(k_i t), \quad -\min_i \{h_i\} < t < +\min_i \{h_i\}, \tag{8}$$

Distributions (cont.)

Proof

Assume t is in the interval $(-\min_i \{h_i\}, +\min_i \{h_i\})$

Then, by independence

$$M_T(t) = E\Big[e^{\sum_{i=1}^n tk_i X_i}\Big] = E\Big[\prod_{i=1}^n e^{(tk_i)X_i}\Big]$$
$$= \prod_{i=1}^n E\big[e^{tk_i X_i}\big] = \prod_{i=1}^n M_i(k_i t)$$

which concludes our proof

$\blacksquare$

Distributions (cont.)

Example

Let X_1,X_2 and X_3 be three mutually independent random variables

Let each have the PDF

$$f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases} \tag{9}$$

The joint PDF of X_1, X_2, X_3

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = 8x_1x_2x_3, \quad 0 < x_i < 1, i = 1, 2, 3$$

- and zero elsewhere

Distributions (cont.)

The expected value of $5X_1X_2^3 + 3X_2X_3^4$

$$\int_0^1\int_0^1\int_0^1 (5x_1x_2^3 + 3x_2x_3^4)(8x_1x_2x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = 2$$

Let Y be the maximum of X_1, X_2 and X_3

Then, we have

$$\begin{aligned} P(Y \leq 1/2) &= P(X_1 \leq 1/2, X_2 \leq 1/2, X_3 \leq 1/2) \\ &= \int_0^{1/2}\int_0^{1/2}\int_0^{1/2} 8x_1x_2x_3\mathrm{d}x_3\mathrm{d}x_2\mathrm{d}x_1 \qquad (10) \\ &= \big(1/2\big)^6 = 1/64 \end{aligned}$$

Distributions (cont.)

In a similar manner, we find the CDF of Y

$$G(y) = P(Y \leq y) = \begin{cases} 0, & y < 0 \\ y^6 & 0 \leq y < 1 \\ 1, & 1 \leq y \end{cases}$$

Accordingly, the PDF of Y

$$g(y) = \begin{cases} 6y^5, & 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

■

Distributions (cont.)

Remark

If X_1, X_2 and X_3 are mutually independent, they **pairwise independent**

- X_i and X_j, $i \neq j$, $i, j = 1, 2, 3$ are independent

Pairwise independence does not necessarily mean mutual independence

Let X_1, X_2 and X_3 have the joint PMF

$$p(x_1, x_2, x_3) = \begin{cases} 1/4, & (x_1, x_2, x_3) \in \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\} \\ 0, & \text{elsewhere} \end{cases}$$

The joint PMF of X_i and X_j, $i \neq j$

$$p_{ij}(x_i, x_j) = \begin{cases} 1/4, & (x_i, x_j) \in \{(0,0), (1,0), (0,1), (1,1)\} \\ 0, & \text{elsewhere} \end{cases}$$

Distributions (cont.)

The marginal PMF of X_i

$$p_i(x_i) = \begin{cases} 1/2, & x_1 = 0, 1 \\ 0, & \text{elsewhere} \end{cases}$$

Obviously, if $i \neq j$, then we have

$$p_{ij}(x_i, x_j) \equiv p_i(x_i) p_j(x_j)$$

Thus X_i and X_j are independent

However,

$$p(x_1, x_2, x_3) \not\equiv p_1(x_1) p_2(x_2) p_3(x_3)$$

Thus, X_1, X_2 and X_3 are not mutually independent

∎

Distributions (cont.)

If several variables are mutually independent and have the same distribution, we say that they are **independent and identically distributed**, or **iid**

Distributions (cont.)

We state a useful corollary to Theorem 1.1 for iid RVs

Corollary

Let $X_1, X_2, \dots, X_n$ be iid RVs each with MGF $M(t)$, $t \in (-h, +h)$, $h > 0$

Let $T = \sum_{i=1}^{n} X_i$

Then, T has MGF given by

$$M_T(t) = \big[M(t)\big]^n, \quad -h < t < +h \qquad (11)$$

■

Variance-covariance matrices

Several random variables

Multivariate variance-covariance matrix

Extend the discussion on the covariance between two RVs to n-variate case

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a n-dimensional random vector

- We have defined the expectation of a random vector
- It is the vector of the expectations of its components

$\rightsquigarrow$ $E(\mathbf{X}) = \big[E(X_1), \ldots, E(X_n)\big]'$

Suppose that $\mathbf{W}$ is a $m \times n$ matrix of random variables

- $\mathbf{W} = \big[W\big]_{ij}$ for the random variables W_{ij}
- $1 \leq i \leq m$ and $1 \leq j \leq n$

We can always roll out the matrix into a $mn \times 1$ vector

We define the expectation of a random matrix

$$E\big[\mathbf{W}\big] = \big[E(W_{ij})\big] \tag{12}$$

Multivariate variance-covariance matrix (cont.)

Theorem 1.2

Let $\mathbf{W}_1$ *and* $\mathbf{W}_2$ *be* $m \times n$ *matrices of random variables*

Let $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *be* $k \times m$ *matrices of constants*

Let $\mathbf{B}$ *be a* $n \times l$ *matrix of constants*

Then,

$$E\big[\mathbf{A}_1\mathbf{W}_1 + \mathbf{A}_2\mathbf{W}_2\big] = \mathbf{A}_1 E\big[\mathbf{W}_1\big] + \mathbf{A}_2 E\big[\mathbf{W}_2\big] \tag{13}$$

$$E\big[\mathbf{A}_1\mathbf{W}_1\mathbf{B}\big] = \mathbf{A}_1 E\big[\mathbf{W}_1\big]\mathbf{B} \tag{14}$$

Multivariate variance-covariance matrix (cont.)

Proof

Consider the linearity of the operator E on RVs

$\rightsquigarrow$ for the (i, j)-th components of expression (13)

$$E\Big[\sum_{s=1}^{m} a_{1is}\, W_{1sj} + \sum_{s=1}^{m} a_{2is}\, W_{2sj}\Big] = \sum_{s=1}^{m} a_{1is}\, E[\, W_{1sj}] + \sum_{s=1}^{m} a_{2is}\, E[\, W_{2sj}]$$

Hence, by Equation (12), expression (13) holds true

The derivation of expression (14) is analogous

■

Multivariate variance-covariance matrix (cont.)

Let $\mathbf{X} = (X_1, \dots, X_n)'$ be a random n-vector

$\rightsquigarrow$ Suppose that $\sigma^2 = \text{Var}(X_i) < \infty$

The **mean** of $\mathbf{X}$ is $\boldsymbol{\mu} = E(\mathbf{X})$

We define its **variance-covariance matrix**

$$\text{Cov}(\mathbf{X}) = E\big[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\big] = [\sigma_{ij}] \tag{15}$$

The i-th diagonal entry

$$\rightsquigarrow \sigma_{ii} = \sigma_i^2 = \text{Var}(X_i)$$

The (i, j)-th off-diagonal entry

$$\rightsquigarrow \text{Cov}(X_i, X_j)$$

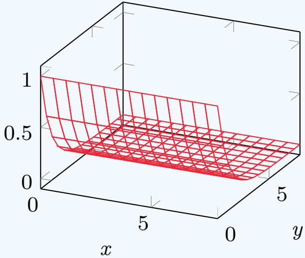
$(\star)$

Multivariate variance-covariance matrix (cont.)

Example

Let the continuous-type random variables X and Y have the joint PDF

$f_{X,Y}(x,y)$ with $(x,y) \in \mathcal{R}^2$



$$f(x,y) = \begin{cases} \exp(-y), & 0 < x < y < \infty \\ 0, & \text{elsewhere} \end{cases}.$$

We have determined the joint MGF

$$M(t_1,t_2) = \frac{1}{(1-t_1-t_2)(1-t_2)}, \quad t_1+t_2 < 1, t_2 < 1$$

Multivariate variance-covariance matrix (cont.)

The first two moments

$$\begin{aligned} \mu_1 &= 1 \\ \mu_2 &= 2 \\ \sigma_1^2 &= 1 \\ \sigma_2^2 &= 2 \\ E\big[(X-\mu_1)(Y-\mu_2)\big] &= 1 \end{aligned} \tag{16}$$

Let $\mathbf{Z} = (X, Y)'$, then we have

$$\begin{aligned} E[\mathbf{Z}] &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ \mathrm{Cov}(\mathbf{Z}) &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \end{aligned} \tag{17}$$

■

Multivariate variance-covariance matrix (cont.)

We now present two useful properties of $\text{Cov}(X_i, X_j)$

Multivariate variance-covariance matrix (cont.)

Theorem

Let $\mathbf{X} = (X_1, \dots, X_n)'$ *be a random* n*-vector*

- *Suppose that* $\sigma_i^2 = Var(X_i) < \infty$

Then,

$$Cov(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}' \tag{18}$$

Proof

We can use Theorem 1.2 to derive Equation (18)

$$\begin{aligned}
\mathrm{Cov}(\mathbf{X}) &= E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\
&= E[\mathbf{X}\mathbf{X}' - \boldsymbol{\mu}\mathbf{X}' - \mathbf{X}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}'] \\
&= E[\mathbf{X}\mathbf{X}'] - \boldsymbol{\mu}E[\mathbf{X}'] - E[\mathbf{X}]\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\mu}'
\end{aligned}$$

■

Multivariate variance-covariance matrix (cont.)

Theorem

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a random n-vector

- *Suppose that $\sigma_i^2 = Var(X_i) < \infty$*

Let $\mathbf{A}$ be a $m \times n$ matrix of constants

Then,

$$Cov(\mathbf{A}\mathbf{X}) = \mathbf{A}\,Cov(\mathbf{X})\mathbf{A}' \tag{19}$$

■

Multivariate variance-covariance matrix (cont.)

All variance-covariance matrices are positive semi-definite

- $\mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a} \geq 0$, for all vectors $\mathbf{a} \in \mathcal{R}^n$

Let $\mathbf{X}$ be a random vector and let $\mathbf{a}$ be any $n \times 1$ vector of constants

Then, $Y = \mathbf{a}'\mathbf{X}$ is a random variable

$\rightsquigarrow$ Hence, its variance is non-negative

$$0 \leq \text{Var}(Y) = \text{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a} \tag{20}$$

$\rightsquigarrow$ Hence, $\text{Cov}(\mathbf{X})$ is positive semi-definite

Several RVs
Distributions
Variance-covariance
Transformations
Linear combinations

Transformations

Several random variables

Transformations

Technically, the determination of the joint PDF of two functions of two random variables of the continuous type is a corollary to a theorem in analysis

$\rightsquigarrow$ Change of variables in a two-fold integral

The theorem naturally extends to n-fold integrals

Transformations (cont.)

Consider an integral of the form

$$\int \cdots \int_A f(x_1, x_2, \dots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

A is a subset of a n-dimensional space $\mathcal{S}$

Let

$$\begin{aligned} y_1 &= u_1(x_1, x_2, \dots, x_n) \\ y_2 &= u_2(x_1, x_2, \dots, x_n) \\ &\cdots \\ y_n &= u_n(x_1, x_2, \dots, x_n) \end{aligned}$$

and their inverse functions

$$\begin{aligned} x_1 &= w_1(y_1, y_2, \dots, y_n) \\ x_2 &= w_2(y_1, y_2, \dots, y_n) \\ &\cdots \\ x_n &= w_n(y_1, y_2, \dots, y_n) \end{aligned}$$

They define a 1-to-1 transformation that maps $\mathcal{S}$ onto $\mathcal{T}$

Transformations (cont.)

Vector function $\mathbf{u}(\mathbf{x})$ and inverse vector function $\mathbf{w}(\mathbf{y})$ define a 1-to-1 map

- $\mathcal{S}$ in $x_1, x_2, \dots, x_n$ space is mapped onto $\mathcal{T}$ in $y_1, y_2, \dots, y_n$ space

The transformation maps subsets A of $\mathcal{S}$ onto subsets B of $\mathcal{T}$

Let the first partial derivative of the inverse functions be continuous

Let J be the $n \times n$ Jacobian determinant of the inverse transformation

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \cdots & \dfrac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \dfrac{\partial x_n}{\partial y_2} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix}$$

We assume that J not be identically zero in $\mathcal{T}$

Transformations (cont.)

Then,

$$\int \cdots \int_A f(x_1, \ldots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$
$$= \int \cdots \int_B f\big[\underbrace{w_1(y_1, \ldots, y_n)}_{x_1}, \ldots, \underbrace{w_n(y_1, \ldots, y_n)}_{x_n}\big] \big|J\big| \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_n$$

Thus, we are able to determine the joint PDF of n functions of n RVs

- Whenever the conditions of the theorem are satisfied

Transformations (cont.)

Proper changes in notation to denote n-spaces v 2-spaces are needed

The joint PDF of RVs $Y_1 = u_1(X_1, \dots, X_n), \dots, Y_n = u_n(X_1, \dots, X_n)$

$$
\begin{aligned}
& g(y_1, y_2, \dots, y_n) \\
& \qquad = f\big[\underbrace{w_1(y_1, \dots, y_n)}_{x_1}, \dots, \underbrace{w_n(y_1, \dots, y_n)}_{x_n}\big]\big|J\big|, \text{ for } (y_1, \dots, y_n) \in \mathcal{T}
\end{aligned}
$$

and zero elsewhere

Transformations (cont.)

Example

Let X_1, X_2 and X_3 have the joint PDF

$$f(x_1, x_2, x_3) = \begin{cases} 48 x_1 x_2 x_3, & 0 < x_1 < x_2 < x_3 < 1 \\ 0, & \text{elsewhere} \end{cases} \tag{21}$$

Let $Y_1 = X_1/X_2$, $Y_2 = X_2/X_3$ and $Y_3 = X_3$

The associated inverse transformations

$$x_1 = y_1 y_2 y_3$$
$$x_2 = y_2 y_3$$
$$x_3 = y_3$$

The determinant of the Jacobian of the inverse transformation

$$J = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ 0 & y_3 & y_2 \\ 0 & 0 & 1 \end{vmatrix} = y_2 y_3^2$$

Transformations (cont.)

The inequalities define the support

$$\begin{aligned} 0 &< y_1 y_2 y_3 \\ y_1 y_2 y_3 &< y_2 y_3 \\ y_2 y_3 &< y_3 \\ y_3 &< 1 \end{aligned}$$

This gives the unit-cube as support $\mathcal{T}$ of (Y_1, Y_2, Y_3)

$$\mathcal{T} = \big\{(y_1, y_2, y_3) : 0 < y_i < 1, i = 1, 2, 3\big\}$$

Transformations (cont.)

Hence, the joint PDF of Y_1, Y_2, Y_3

$$g(y_1, y_2, y_3) = 48(\underbrace{y_1 y_2 y_3}_{x_1})(\underbrace{y_2 y_3}_{x_2})(\underbrace{\ y_3\ }_{x_3})|y_2 y_3^2|$$

$$= \begin{cases} 48 y_1 y_2^3 y_3^5, & 0 < y_i < 1, i = 1, 2, 3 \\ 0, & \text{elsewhere} \end{cases}$$

The marginal PDFs

$$g_1(y_1) = 2y_1, \quad 0 < y_1 < 1, \text{ zero elsewhere}$$
$$g_2(y_2) = 4y_2^3, \quad 0 < y_2 < 1, \text{ zero elsewhere}$$
$$g_3(y_3) = 6y_3^5, \quad 0 < y_3 < 1, \text{ zero elsewhere}$$

$$\rightsquigarrow g(y_1, y_2, y_3) = g(y_1)g(y_2)g(y_3)$$

The random variables Y_1, Y_2, Y_3 are mutually independent

Transformations (cont.)

Example

Let X_1, X_2, X_3 be three IID random variables with common PDF

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \\ 0, & \text{elsewhere} \end{cases}$$

The joint PDF of X_1, X_2, X_3

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \begin{cases} e^{-\sum_{i=1}^{3} x_i}, & 0 < x_i < \infty, i = 1, 2, 3 \\ 0, & \text{elsewhere} \end{cases}$$

Consider the random variables Y_1, Y_2 and Y_3

$$Y_1 = \frac{X_1}{X_1 + X_2 + X_3}$$
$$Y_2 = \frac{X_2}{X_1 + X_2 + X_3}$$
$$Y_3 = X_1 + X_2 + X_3$$

Transformations (cont.)

Hence, the inverse transformation

$$
\begin{aligned}
x_1 &= y_1 y_3 \\
x_2 &= y_2 y_3 \\
x_3 &= y_3 - y_1 y_3 - y_2 y_3
\end{aligned}
$$

The determinant of the Jacobian

$$
J = \begin{vmatrix} y_3 & 0 & y_1 \\ 0 & y_3 & y_2 \\ -y_3 & -y_3 & 1 - y_1 - y_2 \end{vmatrix} = y_3^2
$$

Transformations (cont.)

The support of $\mathcal{S}$ of X_1, X_2, X_3 maps onto $\mathcal{T}$ of Y_1, Y_2, Y_3

$$0 < y_1 y_3 < \infty$$
$$0 < y_2 y_3 < \infty$$
$$0 < y_3(1 - y_1 - y_2) < \infty$$

The support $\mathcal{T}$

$$\mathcal{T} = \big\{(y_1, y_2, y_3) : 0 < y_1, 0 < y_2, 0 < y_1 - y_2, 0 < y_3 < \infty\big\}$$

Transformations (cont.)

Hence, the joint PDF of Y_1, Y_2, Y_3

$$g(y_1, y_2, y_3) = y_3^2 e^{-y_3}, \quad (y_1, y_2, y_3) \in \mathcal{T}$$

- The marginal PDF of Y_1

$$g_1(y_1) = \int_0^{1-y_1} \int_0^{\infty} y_3^2 e^{-y_3} \mathrm{d}y_3 \mathrm{d}y_2 = 2(1 - y_1)$$

for $0 < y_1 < 1$ (zero elsewhere)

- The marginal PDF of Y_2

$$g_2(y_2) = 2(1 - y_2), \quad 0 < y_2 < 1 \text{ (zero elsewhere)}$$

- The marginal PDF of Y_3

$$g_3(y_3) = \int_0^1 \int_0^{1-y_1} y_3^2 e^{-y_3} \mathrm{d}y_2 \mathrm{d}y_1 = 1/2 y_3^2 e^{-y_3}$$

for $0 < y_3 < \infty$ (zero elsewhere)

Transformations (cont.)

$$\rightsquigarrow g(y_1, y_2, y_3) \neq g_1(y_1) g_2(y_2) g_3(y_3)$$

The random variables Y_1, Y_2, Y_3 are dependent

Transformations (cont.)

The joint PDF of Y_1 and Y_3

$$g_{13}(y_1, y_3) = \int_0^{1-y_1} y_3^2 e^{-y_3} \mathrm{d}y_2 = (1 - y_1) y_3^2 e^{-y_3}$$

$$\text{for } 0 < y_1 < 1, 0 < y_3 < \infty \text{ (zero elsewhere)}$$

$\rightsquigarrow$ Thus, Y_1 and Y_2 are independent

The joint PDF of Y_2 and Y_3

$$g_{12}(y_1, y_2) = \int_0^{\infty} y_3^2 e^{-y_3} \mathrm{d}y_3 = 2$$

$$0 < y_1, 0 < y_2, y_1 + y_2 < 1 \text{ (zero elsewhere)}$$

$\rightsquigarrow$ Thus, Y_2 and Y_3 are independent

Transformations (cont.)

As in the bivariate case, we could use the MGF technique

Consider the case in which $Y = g(X_1, X_2, \dots, X_n)$ is a function of the RVs

Then, in the continuous case, the MGF of Y

$$E(e^{tX}) = \int_{-\infty}^{\infty} \int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} e^{tg(x_1, x_2, \dots, x_n)} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$$

$f(x_1, x_2, \dots, x_n)$ is the joint PDF

In the discrete case, summation replaces integration

Transformations (cont.)

Example

Let X_1, X_2, X_3 and X_4 be independent random variables with common PDF

$$p(x_1, x_2, x_3) = \begin{cases} \dfrac{\mu_1^{x_1} \mu_2^{x_2} \mu_3^{x_3} e^{-\mu_1} e^{-\mu_2} e^{-\mu_3}}{x_1! x_2! x_3!}, & x_i = 0, 1, 2, \dots, i = 1, 2, 3 \\ 0, & \text{elsewhere} \end{cases}$$

Let $Y = X_1 + X_2 + X_3$ be a random variable with the MGF

$$\begin{aligned} E\big(e^{tY}\big) &= E\big[e^{t(X_1+X_2+X_3)}\big] \\ &= E\big[e^{tX_1} e^{tX_2} e^{tX_3}\big] \\ &= E\big(e^{tX_1}\big) E\big(e^{tX_2}\big) E\big(e^{tX_3}\big) \end{aligned}$$

Because of the independence of X_1, X_2 and X_3

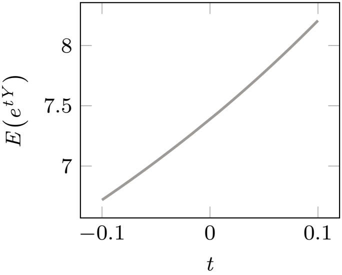
Transformations (cont.)

Earlier, we found

$$E\left(e^{tX_i}\right) = e^{\left[\mu_i(e^t-1)\right]}, \quad i = 1, 2, 3$$

Hence,

$$E\left(e^{tY}\right) = e^{\left[(\mu_1+\mu_2+\mu_3)(e^t-1)\right]}$$



$$E\left(e^{tY}\right) = e^{\left[(\mu_1+\mu_2+\mu_3)(e^t-1)\right]}$$

(In the plot, $\mu_1 = 1$, $\mu_2 = 1$ and $\mu_3 = 1$)

Transformations (cont.)

This is the MGF of a random variable Y with the PMF

$$p_Y(y) = \begin{cases} \dfrac{(\mu_1 + \mu_2 + \mu_3)^y \, e^{-\big(\mu_1+\mu_2+\mu_3\big)}}{y!}, & y = 0, 1, 2, \dots \\ 0, & \text{elsewhere} \end{cases}$$

Thus, this is the distribution of $Y = X_1 + X_2 + X_3$

$\blacksquare$

Transformations (cont.)

Example

Let X_1, X_2, X_3 and X_4 be independent random variables with common PDF

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Let $Y = X_1 + X_2 + X_3 + X_4$

Because of the independence of X_1, X_2, X_3 and X_4

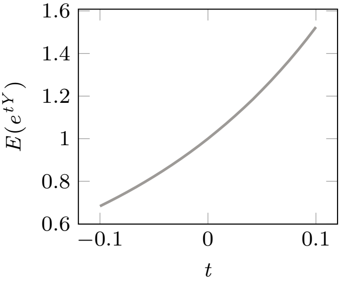
$$E(e^{tY}) = E(e^{tX_1})E(e^{tX_2})E(e^{tX_3})E(e^{tX_4})$$

We have

$$E(e^{tX_i}) = (1-t)^{-1}, \quad \text{for } t < 1, \quad i = 1, 2, 3, 4$$

Transformations (cont.)

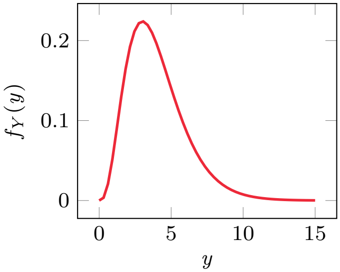
Hence,



$$E(e^{tY}) = (1 - t)^{-4}$$

Transformations (cont.)

This is the MGF of a distribution with PDF



$$f_Y(y) = \begin{cases} \dfrac{1}{3!} y^3 e^{-y}, & 0 < y < \infty \\ 0, & \text{elsewhere} \end{cases}$$

Accordingly, this is the distribution of Y

Linear combinations

Several random variables

Linear combinations

Let $(X_1, \dots, X_n)'$ indicate a random vector

We are often interested in some function of $T = T(X_1, \dots, X_n)$

Let us consider a linear combination of the variables

$$T = \sum_{i=1}^{n} a_i X_i$$

$\mathbf{a} = (a_1, \dots, a_n)'$ is some specified vector

Linear combinations (cont.)

The mean of T follows from linearity of the expectation operator

Theorem 3.1

Let $(X_1, \ldots, X_n)'$ *indicate a random vector*

Let $T = \sum_{i=1}^{n} a_i X_i$

Then,

$$E(T) = \sum_{i=1}^{n} a_i E(X_i)$$

Provided $E\big[|X_i|\big] < \infty$, *for* $i = 1, \ldots, n$

Linear combinations (cont.)

For the variance of T, we first state a general result about covariances

Theorem 3.2

Let $(X_1, \ldots, X_n)'$ and $(Y_1, \ldots, Y_m)'$ indicate two random vectors

Let $T = \sum_{i=1}^{n} a_i X_i$

Let $W = \sum_{j=1}^{m} b_j Y_j$

Suppose that $E\big[X_i^2\big] < \infty$, for $i = 1, \ldots, n$

Suppose that $E\big[Y_j^2\big] < \infty$, for $j = 1, \ldots, m$

Then,

$$Cov(T, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$$

Linear combinations (cont.)

Proof

Using the definition of covariance and Theorem 3.1, we have the first equality

$$
\begin{aligned}
\text{Cov}(T, W) &= E\Big\{ \sum_{i=1}^{n} \sum_{j=1}^{m} \Big[a_i X_i - a_i E(X_i)\Big] \Big[b_j Y_j - b_j E(Y_j)\Big] \Big\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j E\Big\{ \Big[X_i - E(X_i)\Big] \Big[Y_j - E(Y_j)\Big] \Big\}
\end{aligned}
$$

The second equality follows from the linearity of E

$\blacksquare$

Linear combinations (cont.)

For the variance of T, we replace W by T

Corollary

Let $(X_1, \dots, X_n)'$ and let $T = \sum_{i=1}^{n} a_i X_i$

Suppose that $E\big[X_i^2\big] < \infty$, for $i = 1, \dots, n$

$$\text{Var}(T) = \text{Cov}(T, T) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j) \tag{22}$$

If $X_1, \dots, X_n$ are independent RVs, then the covariance $\text{Cov}(X_i, X_j) = 0$

$\rightsquigarrow$ Equation (22) gets simplified

Linear combinations (cont.)

Corollary 3.1

Let $X_1, \dots, X_n$ be independent random variables with finite variances

Then,

$$\text{Var}(T) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) \tag{23}$$

To obtain this result, only X_i and X_j need be uncorrelated for all $i \neq j$

- $\rightsquigarrow$ $\text{Cov}(X_i, X_j) = 0$, $i \neq j$, true when $X_1, \dots, X_n$ are independent

∎

Linear combinations (cont.)

Let the RVs $X_1, \ldots, X_n$ are independent and identically distributed

The RVs make a **random sample** of size n from the common distribution

Two commonly used statistics of the random sample

- $\rightsquigarrow$ **Sample mean**
- $\rightsquigarrow$ **Sample variance**

Linear combinations (cont.)

Example

Sample mean

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables

- Let μ and σ^2 be the common mean and variance

The **sample mean**

$$\rightsquigarrow \overline{X} = n^{-1} \sum_{i=1}^{n} X_i \tag{24}$$

This is a linear combination of the sample observations, with $a_i \equiv n^{-1}$

Linear combinations (cont.)

By Theorem 3.1 and Corollary 3.1

$$\begin{aligned} E(\overline{X}) &= \mu \\ \text{Var}(\overline{X}) &= \sigma^2/n \end{aligned} \tag{25}$$

We say that $\overline{X}$ is an unbiased estimator of μ

■

Linear combinations (cont.)

Example

Sample variance

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables

- Let μ and σ^2 be the common mean and variance

The **sample variance**

$$S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = (n-1)^{-1} \Big(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \Big) \qquad (26)$$

The second equality follows after some algebra ($\star$)

Linear combinations (cont.)

By Theorem 3.1 and Corollary 3.1, by the results from the previous example

$$\begin{aligned} E(S^2) &= (n-1)^{-1}\Big[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\Big] \\ &= (n-1)^{-1}\Big\{n\sigma^2 + n\mu^2 - n\Big[(\sigma^2/n) + \mu^2\Big]\Big\} \\ &= \sigma^2 \end{aligned} \tag{27}$$

We used the fact that $E(X^2) = \sigma^2 + \mu^2$

Hence, S^2 is an unbiased estimator of σ^2

■