

---

# Advanced Mathematical Methods

WS 2022/23

## 4 Mathematical Statistics

Dr. Julie Schnaitmann

*Department of Statistics, Econometrics and Empirical  
Economics*

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



WIRTSCHAFTS- UND  
SOZIALWISSENSCHAFTLICHE  
FAKULTÄT

---

# Outline: Mathematical Statistics

- 4.6 Joint distributions
- 4.7 Marginal Distributions
- 4.8 Covariance and correlation
- 4.9 Conditional Distributions
- 4.10 Conditional Moments
- 4.11 The bivariate normal distribution
- 4.12 Multivariate Distributions

---

# Readings

- A. Papoulis and A. U. Pillai. *Probability, Random Variables and Stochastic Processes*.  
Mc Graw Hill, fourth edition, 2002, Chapter 6

---

## Online References

MIT Course on Probabilistic Systems Analysis and Applied Probability (by John Tsitsiklis)

- Discrete RVs II: Functions of RV, conditional probabilities, specific distribution, total expectation theorem, joint probabilities

<https://www.youtube.com/watch?v=-qCEoqpwjf4>

- Discrete RVs III: Conditional distributions and joint distributions continued

<https://www.youtube.com/watch?v=EObHWIEKGjA>

- Multiple Continuous RVs: conditional pdf and cdf, joint pdf and cdf

<https://www.youtube.com/watch?v=CadZXGNauY0>

## 4.6 Joint distributions

### Definition: Joint density function

The joint density for two discrete random variables  $X_1$  and  $X_2$  is given as

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} P(X_1 = x_{1i} \cap X_2 = x_{2j}) & \forall i, j \\ 0 & \text{else} \end{cases}$$

Properties:

- $f_{\mathbf{X}}(x_1, x_2) \geq 0 \quad \forall (x_1, x_2) \in \mathbb{R}^2$
- $\sum_{x_i} \sum_{x_j} f_{\mathbf{X}}(x_{1i}, x_{2j}) = 1$

## 4.6 Joint distributions

### Definition: Joint cumulative distribution function

The cdf for two discrete random variables  $X_1$  and  $X_2$  is given as

$$F_{\mathbf{X}}(x_1, x_2) = P(X_1 \leq x_1 \cap X_2 \leq x_2) = \sum_{x_{1i} \leq x_1} \sum_{x_{2i} \leq x_2} f_{\mathbf{X}}(x_{1i}, x_{2i})$$

it follows that

$$P(a \leq X_1 \leq b \cap c \leq X_2 \leq d) = \sum_{a \leq x_1 \leq b} \sum_{c \leq x_2 \leq d} f_{\mathbf{X}}(x_{1i}, x_{2i})$$

---

## 4.6 Joint distributions

If  $X_1$  and  $X_2$  are two continuous random variables, the following holds:

$$\text{pdf} \quad f_{\mathbf{X}}(x_1, x_2) = \frac{\partial^2 F_{\mathbf{X}}(x_1, x_2)}{\partial x_1 \partial x_2}$$

$$\text{cdf} \quad F_{\mathbf{X}}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{\mathbf{X}}(u_1, u_2) du_2 du_1$$

---

## 4.7 Marginal Distributions

Derive the distribution of the individual variable from the joint distribution function:

→ sum or integrate out the other variable

$$f_{X_1}(x_1) = \begin{cases} \sum_{x_{2j}} f_{\mathbf{X}}(x_{1i}, x_{2j}) & \text{if } \mathbf{X} \text{ is discrete} \\ \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2) dx_2 & \text{if } \mathbf{X} \text{ is continuous} \end{cases}$$



---

## 4.7 Marginal Distributions

Two random variables are statistically independent if their joint density is the product of the marginal densities:

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \Leftrightarrow X_1 \text{ and } X_2 \text{ are independent.}$$

Under independence the cdf factors as well:

$$F_{\mathbf{X}}(x_1, x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2).$$

Expectations in a joint distribution are computed with respect to the marginals.

---

## 4.7 Marginal Distributions

Two random variables are statistically independent if their joint density is the product of the marginal densities:

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \Leftrightarrow X_1 \text{ and } X_2 \text{ are independent.}$$

Under independence the cdf factors as well:

$$F_{\mathbf{X}}(x_1, x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2).$$

Expectations in a joint distribution are computed with respect to the marginals.

---

## 4.8 Covariance and correlation

$$\text{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

Properties:

- symmetry:  $\text{Cov}[X_1, X_2] = \text{Cov}[X_2, X_1]$
- linear transformation:

$$\begin{aligned} Y_1 &= b_0 + b_1 X_1 & Y_2 &= c_0 + c_1 X_2 \\ \Rightarrow \text{Cov}[Y_1, Y_2] &= b_1 c_1 \text{Cov}[X_1, X_2] \end{aligned}$$

- calculation:

$$\text{Cov}[X_1, X_2] = \begin{cases} \sum_{x_{1i}} \sum_{x_{2j}} x_{1i} x_{2j} f_{\mathbf{X}}(x_{1i}, x_{2j}) - E[X_1]E[X_2] \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{\mathbf{X}}(x_1, x_2) dx_2 dx_1 - E[X_1]E[X_2] \end{cases}$$

---

## 4.8 Covariance and correlation

### Pearson's correlation coefficient

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \cdot \text{Var}(X_2)}} = \frac{\sigma_{X_1, X_2}}{\sigma_{X_1} \sigma_{X_2}}$$

- If  $X_1$  and  $X_2$  are independent, they are also uncorrelated.
- Uncorrelated does not imply independence!
- Exception: normal distribution, characterized by 1st and 2nd moment.

---

## 4.9 Conditional Distributions

- Distribution of the variable  $X_1$  given that  $X_2$  takes on a certain value  $x_2$ .
- Closely related to conditional probabilities:

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1 \cap X_2 = x_2)}{P(X_2 = x_2)}$$

conditional pdf of  $X_1$  given  $X_2 = x_2$ :

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

---

## 4.9 Conditional Distributions

conditional cdf of  $X_1$  given  $X_2 = x_2$ :

$$P(X_1 = x_1 | X_2 = x_2) = \sum_{x_{1i} \leq x_1} f_{X_1|X_2}(x_{1i}|x_2) = F_{X_1|X_2}(x_1|x_2).$$

If  $X_1$  and  $X_2$  are independent, the conditional probability and the marginal probability coincide:

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1)$$

because

$$f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

---

## 4.9 Conditional Distributions

The joint pdf can be derived from conditional and marginal densities in 2 ways:

$$f_{X_1 X_2} = f_{X_1|X_2}(x_1|x_2) \cdot f_{X_2}(x_2) = f_{X_2|X_1}(x_2|x_1) \cdot f_{X_1}(x_1)$$

---

## 4.10 Conditional Moments

$$\begin{aligned}E[Y^k|X = x] &= \sum_j y_j^k \cdot \frac{P(X = x \cap Y = y_j)}{P(X = x)} \\&= \sum_{y_j} y_j^k \cdot P(Y = y_j|X = x) \\&= \sum_{y_j} y_j^k \cdot f_{Y|X}(y_j|x) \\&= \sum_{y_j} y_j^k \cdot \frac{f_{XY}(x, y_j)}{f_X(x)} \quad \text{if } Y \text{ is discrete} \\E[Y^k|X = x] &= \int_{-\infty}^{\infty} y^k \cdot \frac{f_{XY}(x, y)}{f_X(x)} dy \quad \text{if } Y \text{ is continuous}\end{aligned}$$



---

## 4.10 Conditional Moments

$$\begin{aligned} \text{Var}[Y|X = x] &= E_{Y|X}[(Y - E[Y|X = x])^2] \\ &= \sum_{y_j} (y_j - E[Y|X = x])^2 \cdot f_{Y|X}(y_j|x) \end{aligned}$$

if  $Y$  is discrete

$$\begin{aligned} \text{Var}[Y|X = x] &= E_{Y|X}[(Y - E[Y|X = x])^2] \\ &= \int_{-\infty}^{\infty} (y - E[Y|X = x])^2 \cdot f_{Y|X}(y|x) dy \end{aligned}$$

if  $Y$  is continuous

---

## 4.10 Conditional Moments

Law of Total Expectations/ Law of Iterated Expectations

$$E[Y] = E_X [E[Y|X]]$$

$$E_X [E_{Y|X}[Y|X]] = E[Y] = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} y \cdot \frac{f_{XY}(x, y)}{f_X(x)} dy \right] f_X(x) dx$$

$E_{Y|X}$  is a random value as  $X$  is a random variable.

## 4.11 The bivariate normal distribution

### Definition: Bivariate normal distribution

Two random variables  $X_1$  and  $X_2$  are jointly normally distributed if they are described by the joint pdf

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{1}{2}q(x_1, x_2)\right]$$

where

$$q(x_1, x_2) = \frac{1}{1-\rho^2} \left[ \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1}\right) \left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 \right].$$

---

## 4.11 The bivariate normal distribution

If  $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ , then

- $f_{X_1}(x_1) \sim N(\mu_1, \sigma_1^2)$ ,  
 $f_{X_2}(x_2) \sim N(\mu_2, \sigma_2^2)$ ,
- $f_{X_1|X_2} \sim N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2))$ ,  
 $f_{X_2|X_1} \sim N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1 - \rho^2))$ .

---

## 4.12 Multivariate Distributions

$\mathbf{x}$  a random vector with joint density  $f_{\mathbf{X}}(\mathbf{x})$

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(\mathbf{t}) dt_1 dt_2 \dots dt_{n-1} dt_n$$

Expected Value:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

---

## 4.12 Multivariate Distributions

### Covariance Matrix

$$\begin{aligned} & E [(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] \\ &= \begin{pmatrix} (x_1 - \mu_1)(x_1 - \mu_1) & (x_1 - \mu_1)(x_2 - \mu_2) & \dots & (x_1 - \mu_1)(x_n - \mu_n) \\ (x_2 - \mu_2)(x_1 - \mu_1) & (x_2 - \mu_2)(x_2 - \mu_2) & \dots & (x_2 - \mu_2)(x_n - \mu_n) \\ \vdots & & & \\ (x_n - \mu_n)(x_1 - \mu_1) & (x_n - \mu_n)(x_2 - \mu_2) & \dots & (x_n - \mu_n)(x_n - \mu_n) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{n1} & \dots & \dots & \sigma_n^2 \end{pmatrix} = E [\mathbf{x}\mathbf{x}'] - \boldsymbol{\mu}\boldsymbol{\mu}' = \boldsymbol{\Sigma} \end{aligned}$$

---

## 4.12 Multivariate Distributions

Linear Transformation: sum of  $n$  random variables  $\sum_{i=1}^n a_i x_i$

$$\begin{aligned} E[a_1 x_1 + a_2 x_2 + \dots + a_n x_n] &= E[\mathbf{a}'\mathbf{x}] \\ &= \mathbf{a}'E[\mathbf{x}] = \mathbf{a}'\boldsymbol{\mu} \end{aligned}$$

$$\begin{aligned} \text{Var}[\mathbf{a}'\mathbf{x}] &= E[(\mathbf{a}'\mathbf{x} - E[\mathbf{a}'\mathbf{x}])^2] \\ &= E[(\mathbf{a}'(\mathbf{x} - E[\mathbf{x}]))^2] \\ &= E[(\mathbf{a}'(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{a})] \\ &= \mathbf{a}'E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']\mathbf{a} \\ &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij} \end{aligned}$$

---

## 4.12 Multivariate Distributions

Linear transformation:  $\mathbf{y} = \mathbf{A}\mathbf{x}$

$i$ -th element in  $\mathbf{y} = \mathbf{A}\mathbf{x}$  is  $y_i = \mathbf{a}_i\mathbf{x}$  with  $\mathbf{a}_i$   $i$ -th row in  $\mathbf{A}$

$\Rightarrow E[y_i] = E[\mathbf{a}_i\mathbf{x}] = \mathbf{a}_i\boldsymbol{\mu}$  as before

$$\begin{aligned}E[\mathbf{y}] &= E[\mathbf{A}\mathbf{x}] = \mathbf{A}E[\mathbf{x}] = \mathbf{A}\boldsymbol{\mu} \\ \text{Var}[\mathbf{y}] &= E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])'] \\ &= E[(\mathbf{A}\mathbf{x} - \mathbf{A}\boldsymbol{\mu})(\mathbf{A}\mathbf{x} - \mathbf{A}\boldsymbol{\mu})'] \\ &= E[(\mathbf{A}(\mathbf{x} - \boldsymbol{\mu}))(\mathbf{A}(\mathbf{x} - \boldsymbol{\mu}))'] \\ &= E[\mathbf{A}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{A}'] \\ &= \mathbf{A}E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']\mathbf{A}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\end{aligned}$$