Advanced Mathematical Methods

WS 2023/24

4 Mathematical Statistics

Dr. Julie Schnaitmann

Department of Statistics, Econometrics and Empirical Economics

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

WIRTSCHAFTS- UND SOZIALWISSENSCHAFTLICHE FAKULTÄT

Outline: Mathematical Statistics

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

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

Properties:

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

4.6 Joint distributions

Definition: Joint cumulative distribution function

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

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

it follows that

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

4.6 Joint distributions

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

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

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

4.7 Marginal Distributions

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

$\rightarrow$ sum or integrate out the other variable

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

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

$$Cov[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

Properties:

- symmetry: $Cov[X_1, X_2] = Cov[X_2, X_1]$
- linear transformation:

$$Y_1 = b_0 + b_1 X_1 \qquad Y_2 = c_0 + c_1 X_2$$
$$\Rightarrow Cov[Y_1, Y_2] = b_1 c_1 Cov[X_1, X_2]$$

- calculation:

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

4.8 Covariance and correlation

Pearson's correlation coefficient

$$\rho_{x_1,x_2} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1) \cdot Var(X_2)}} = \frac{\sigma_{x_1,x_2}}{\sigma_{x_1}\sigma_{x_2}}$$

- 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 varibale X_1 given that X_2 takes on a certain value x_1.
- 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

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

$$
E[Y^k|X=x] = \int_{-\infty}^{\infty} y^k \cdot \frac{f_{XY}(x,y)}{f_X(x)}\, dy \qquad \text{if } Y \text{ is continuous}
$$

4.10 Conditional Moments

$$Var[Y|X = x] = E_{Y|X}[(Y - E[Y|X = x])^2]$$

$$= \sum_{y_j} (y_j - E[Y|X = x])^2 \cdot f_{Y|X}(y_j|x)$$

if Y is discrete

$$Var[Y|X = x] = E_{Y|X}[(Y - E[Y|X = x])^2]$$

$$= \int_{-\infty}^{\infty} (y - E[Y|X = x])^2 \cdot f_{Y|X}(y|x) dy$$

if Y is continuous

4.10 Conditional Moments

Law of Total Expectations/ Law of Iterated Expectations

$$E[Y] = E_X\left[E[Y|X]\right]$$

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

$E_{Y|X}$ 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

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

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

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

$$E\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'\right]$$

$$= \begin{pmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) & \dots & (x_1-\mu_1)(x_n-\mu_n) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) & \dots & (x_2-\mu_2)(x_n-\mu_n) \\ \vdots & & & \\ (x_n-\mu_n)(x_1-\mu_1) & (x_n-\mu_2)(x_n-\mu_2) & \dots & (x_n-\mu_n)(x_n-\mu_n) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots \sigma_{2n} \\ \vdots & & \\ \sigma_{n1} & \dots & \dots \sigma_n^2 \end{pmatrix} = E\left[\boldsymbol{x}\boldsymbol{x}'\right] - \boldsymbol{\mu}\boldsymbol{\mu}' = \boldsymbol{\Sigma}$$

4.12 Multivariate Distributions

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

$$
\begin{aligned}
E[a_1x_1 + a_2x_2 + \ldots a_nx_n] &= E[\mathbf{a}'\mathbf{x}] \\
&= \mathbf{a}'E[\mathbf{x}] = \mathbf{a}'\boldsymbol{\mu} \\
Var[\mathbf{a}'\mathbf{x}] &= E[(\mathbf{a}'\mathbf{x} - E[\mathbf{a}'\mathbf{x}])^2] \\
&= E[(\mathbf{a}'(\mathbf{x} - E[\mathbf{x}])^2] \\
&= E[(\mathbf{a}'(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{a}] \\
&= \mathbf{a}'E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']\mathbf{a} \\
&= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij}
\end{aligned}
$$

4.12 Multivariate Distributions

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

i-th element in $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ is $y_i = \boldsymbol{a_i}\boldsymbol{x}$ with $\boldsymbol{a_i}$ i-th row in $\boldsymbol{A}$

$\Rightarrow E[y_i] = E[\boldsymbol{a_i}\boldsymbol{x}] = \boldsymbol{a_i}\boldsymbol{\mu}$ as before

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