Principles in Economics and Mathematics: the mathematical part

Bram De Rock

Practicalities about me

- Bram De Rock
- Office: R.42.6.218
- E-mail: bderock@ulb.ac.be
- Phone: 02 650 4214
- Mathematician who is doing research in Economics
- Homepage: http://www.revealedpreferences.org/bram.php

Practicalities about the course

- 12 hours on the mathematical part
- Micael Castanheira: 12 hours on the economics part
- Slides are available at MySBS and on http://mathecosolvay.com/spma/
- Course evaluation
  - Written exam to verify if you can apply the concepts discussed in class
  - Compulsory for students in *Financial Markets*
  - On a voluntary basis for students in *Quantitative Finance*

Course objectives and content

- Refresh some useful concepts needed in your other coursework
  - No thorough or coherent study
  - Interested student: see references for relevant material
- Content:
  1. Calculus (functions, derivatives, optimization, concavity)
  2. Financial mathematics (sequences, series)
  3. Linear algebra (solving system of linear equations, matrices, linear (in)dependence)
  4. Fundamentals on probability (probability and cumulative distributions, expectations of a random variable, correlation)

References

- Chiang, A.C. and K. Wainwright, "*Fundamental Methods of Mathematical Economics*", Economic series, McGraw-Hill.
- Green, W.H., "*Econometric Analysis, Seventh Edition*", Pearson Education limited.
- Luderer, B., V. Nollau and K. Vetters, "*Mathematical Formulas for Economists*", Springer, New York. ULB-link
- Simon, C.P. and L. Blume "*Mathematics for Economists*", Norton & Company, New York.
- Sydsaeter, K., A. Strom and P. Berck, "*Economists' Mathematical Manual*", Springer, New York. ULB-link

Outline

Role of functions

- Calculus = "the study of functions"
- Functions allow to exploit mathematical tools in Economics
- E.g. make consumption decisions
  - max $U(x_1, x_2)$ s.t. $p_1x_1 + p_2x_2 = Y$
  - Characterization: $x_1 = f(p_1, p_2, Y)$
  - Econometrics: estimate f
  - Allows to model/predict consumption behavior
- Warning about identification
  - Causality: what is driving what?
  - Functional structure: what is driving the result?
  - Does the model allow to identify

Some important functions of one variable

- The straight line: $y = A + Bx$
  - A is the intercept or intersection with the $y-$ axis
  - B is the slope
  - The impact of changes in x is constant
  - E.g. the effect on demand of a price change
- Polynomial functions: $y = A_nx^n + \cdots + A_0$
  - Quadratic and cubic functions are special cases
  - Non-linear functions to capture more advance patterns due to changes in x
  - E.g. profit as a function of sold quantities
- Hyperbolic functions: $y = \frac{A}{x}$
  - The impact of changes in x goes to infinity around zero

Some important functions of one variable

- Exponential functions: a^x and e^x
  - Used as growth ($a > 1$) or decay curves ($0 < a < 1$)
  - Always positive
  - The relative growth/decay remains constant
  - E.g. the growth of capital at constant interest rate
  - Remember: $a^x a^y = a^{x+y}$, $(a^x)^y = a^{xy}$ and $a^0 = 1$
- Logarithmic functions: $\log_a(x)$ or $\ln(x)$
  - The inverse of the exponential function: $y = \log_a(x)$ if and only if $a^y = x$
  - Can only be applied to positive numbers
  - Remember: $\log_a(xy) = \log_a(x) + \log_a(y)$, $\log_a(\frac{x}{y}) = \log_a(x) - \log_a(y)$, $\log_a(x^k) = k\log_a(x)$ and $\log_a(1) = 0$

Derivatives

- Marginal changes are important in Economics
  - The impact of a infinitesimally small change of one of the variables
  - Comparative statistics: what is the impact of a price change?
  - Optimization: what is the optimal consumption bundle?
- Marginal changes are mostly studied by taking derivatives
- Characterizing the impact depends on the function
  - $f : D \subseteq \mathbb{R}^n \to \mathbb{R}^k : (x_1, \ldots, x_n) \mapsto (y_1, \ldots y_k) = f(x_1, \ldots, x_n)$
  - We will always take $k = 1$
  - First look at $n = 1$ and then generalize
  - Note: $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$

Outline

Functions of one variable: $f : D \subseteq \mathbb{R} \to \mathbb{R}$

$$\frac{df}{dx} = f' = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

- Limit of quotient of differences
- If it exists, then it is called the derivative
- f' is again a function
- E.g. $f(x) = 3x^2 - 4$
- E.g. discontinuous functions, border of domain, $f(x) = |x|$

Some important derivatives and rules

Let us abstract from specifying the domain D and assume that $c, n \in \mathbb{R}_0$

- If $f(x) = c$, then $f'(x) = 0$
- If $f(x) = cx^n$, then $f'(x) = ncx^{n-1}$
- If $f(x) = ce^x$, then $f'(x) = ce^x$
- If $f(x) = c\ln(x)$, then $f(x) = c\frac{1}{x}$
- $(f(x) \pm g(x))' = f'(x) \pm g'(x)$
- $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x) \neq f'(x)g'(x)$
- $(\frac{f(x)}{g(x)})' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$

Application: the link with marginal changes

- By definition it is the limit of changes
- Slope of the tangent line
  - Increasing or decreasing function (and thus impact)
  - Does the inverse function exist?
- First order approximation in some point c
  - Based on expression for the tangent line in c
  - $f(c + \Delta x) \approx f(c) + f'(c)(\Delta x)$
  - More general approximation: Taylor expansion

Application: elasticities

- The elasticity of f in x: $\frac{f'(x)x}{f(x)}$
- The limit of the quotient of changes in terms of percentage
  - Percentage change of the function: $\frac{f(x+\Delta x) - f(x)}{f(x)}$
  - Percentage change of the variable: $\frac{\Delta x}{x}$
  - Quotient: $\frac{f(x+\Delta x) - f(x)}{\Delta x} \frac{x}{f(x)}$
- Is a unit independent informative number
  - E.g. the (price) elasticity of demand

Application: comparative statics for a simple market model

- Demand: $P = \frac{10}{4} - \frac{Q}{4}$ or $Q = 10 - 4P$
- Supply: $P = \frac{Q}{\alpha} - \frac{2}{\alpha}$ or $Q = 2 + \alpha P$
- $P^* = \frac{8}{4+\alpha}$ and $Q^* = \frac{8+10\alpha}{4+\alpha}$
- $\frac{dP^*}{d\alpha} = \frac{-8}{(4+\alpha)^2}$ and $\frac{dQ^*}{d\alpha} = \frac{32}{(4+\alpha)^2}$
- The (price) elasticity of demand is $-\frac{-4P}{10-4P}$

Some exercises

- Compute the derivative of the following functions (defined on $\mathbb{R}^+$)
  - $f(x) = 17x^2 + 5x + 7$
  - $f(x) = -\sqrt{x} + 3$
  - $f(x) = \frac{1}{x^2}$
  - $f(x) = 17x^2 e^x$
  - $f(x) = \frac{x ln(x)}{x^2 - 4}$
- Let $f(x) : \mathbb{R} \to \mathbb{R} : x \mapsto x^2 + 5x$.
  - Determine on which region f is increasing
  - Is f invertible?
  - Approximate f in 1 and derive an expression for the approximation error
  - Compute the elasticity in 3 and 5

The chain rule

- Often we have to combine functions
  - If $z = f(y)$ and $y = g(x)$, then $z = h(x) = f(g(x))$
- We have to be careful with the derivative
- A small change in x causes a chain reaction
  - It changes y and this in turn changes z
- That is why $\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = f'(y)g'(x)$
  - Don't be confused: these are not fractions
  - Can easily be generalized to compositions of more than two functions
  - $\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{du} \cdots \frac{dv}{dx}$
- E.g. if $h(x) = e^{x^2}$, then $h'(x) = e^{x^2}2x$
  - I.e. $z = e^y$ and $y = x^2$

Higher order derivatives

- The derivative is again a function of which we can take derivatives
- Higher order derivatives describe the changes of the changes
- Notation
  - $f''(x)$ or more generally $f^{(n)}(x)$
  - $\frac{d}{dx}(\frac{df}{dx})$ or more generally $\frac{d^n}{dx^n} f(x)$
- E.g. if $f(x) = 5x^3 + 2x$, then $f'''(x) = f^{(3)}(x) = 30$

Application: concave and convex functions

$$f : D \subset \mathbb{R}^n \to \mathbb{R}$$

- f is concave
  - $\forall x, y \in D, \forall \lambda \in [0,1] : f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$
  - If $n = 1, \forall x \in D : f''(x) \leq 0$
- f is convex
  - $\forall x, y \in D, \forall \lambda \in [0,1] : f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
  - If $n = 1, \forall x \in D : f''(x) \geq 0$

Application: concave and convex functions

- Very popular and convenient assumptions in Economics
  - E.g. optimization
- Sometimes intuitive interpretation
  - E.g. risk-neutral, -loving, -averse
- Don't be confused with a convex set
  - S is a set $\Leftrightarrow \forall x, y \in S, \forall \lambda \in [0,1] : \lambda x + (1-\lambda)y \in S$

Some exercises

- Compute the first and second order derivative of the following functions (defined on $\mathbb{R}^+$)
  - $f(x) = -\pi$
  - $f(x) = -\sqrt{5x} + 3$
  - $f(x) = e^{-3x}$
  - $f(x) = \ln(5x)$
  - $f(x) = x^3 - 6x^2 + 17$
- Determine which of these functions are concave or convex

Outline

Functions of more than one variable: $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$

- Same applications in mind but now several variables
  - E.g. what is the marginal impact of changing x_1, while controlling for other variables?
- Look at the partial impact: partial derivatives
  - $\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n) = f_{x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1,\ldots,x_i+\Delta x_i,\ldots,x_n) - f(x_1,\ldots,x_i,\ldots,x_n)}{\Delta x_i}$
  - Same interpretation as before, but now fixing remaining variables
- E.g. $f(x_1, x_2, x_3) = 2x_1^2 x_2 - 5x_3$
  - $\frac{\partial}{\partial x_1} f(x_1, x_2, x_3) = 4x_1 x_2$
  - $\frac{\partial}{\partial x_2} f(x_1, x_2, x_3) = 2x_1^2$
  - $\frac{\partial}{\partial x_3} f(x_1, x_2, x_3) = -5$

Partial derivative

- Geometric interpretation: slope of tangent line in the x_i direction
- Same rules hold
- Higher order derivatives
  - $\frac{\partial^2}{\partial x_i^2} f(x_1, \ldots, x_n)$
  - $\frac{\partial^2}{\partial x_i x_j} f(x_1, \ldots, x_n) = \frac{\partial^2}{\partial x_j x_i} f(x_1, \ldots, x_n)$
  - E.g. $\frac{\partial^2}{\partial x_1^2} f(x_1, x_2, x_3) = 4x_2$ and $\frac{\partial^2}{\partial x_1 \partial x_3} f(x_1, x_2, x_3) = 0$

Some remarks

- Gradient: $\nabla f(x_1, \ldots, x_n) = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$
- Chain rule: special case
  - $x_1 = g_1(t), \ldots, x_n = g_n(t)$ and $f(x_1, \ldots, x_n)$
  - $h(t) = f(x_1, \ldots, x_n) = f(g_1(t), \ldots, g_n(t))$
  - $\frac{dh(t)}{dt} = h'(t) = \frac{\partial f(x_1, \ldots, x_n)}{\partial x_1} \frac{dx_1}{dt} + \cdots + \frac{\partial f(x_1, \ldots, x_n)}{\partial x_n} \frac{dx_n}{dt}$
  - E.g. $f(x_1, x_2) = x_1 x_2$, $g_1(t) = e^t$ and $g_2(t) = t^2$
  - $h'(t) = e^t t^2 + e^t 2t$

Some remarks

- Slope of indifference curve of $f(x_1, x_2)$
  - Indifference curve: all (x_1, x_2) for which $f(x_1, x_2) = C$ (with C some give number)
  - Implicit function theorem: $f(x_1, g(x_1)) = C$
  - $\frac{\partial}{\partial x_1} f(x_1, x_2) + \frac{\partial}{\partial x_2} f(x_1, x_2) \frac{dg}{dx_1} = 0$
  - Slope $= -\frac{\frac{\partial}{\partial x_1} f(x_1, x_2)}{\frac{\partial}{\partial x_2} f(x_1, x_2)}$

Some exercises

- Compute the gradient and all second order partial derivatives for the following functions (defined on $\mathbb{R}^+$)
  - $f(x_1, x_2) = x_1^2 - 2x_1 x_2 + 3x_2^2$
  - $f(x_1, x_2) = \ln(x_1 x_2)$
  - $f(x_1, x_2, x_3) = e^{x_1 + 2x_2} - 3x_1 x_3$
- Compute the marginal rate of substitution for the utility function $U(x_1, x_2) = x_1^\alpha x_2^\beta$

Outline

Optimization: important use of derivatives

- Many models in economics entail optimizing behavior
  - Maximize/Minimize objective subject to constraints
- Characterize the points that solve these models
- Note on Mathematics vs Economics
  - Profit = Revenue - Cost
  - Marginal revenue = marginal cost
  - Marginal profit = zero

Optimization: formal problem

$$
\begin{aligned}
&\max / \min f(x_1, \ldots, x_n) \\
&s.t. \\
&g_1(x_1, \ldots, x_n) = c_1 \\
&\ldots \\
&g_m(x_1, \ldots, x_n) = c_m \\
&x_1, \ \ldots x_n \geq 0
\end{aligned}
$$

- Inequality constraints are also possible
- Kuhn-Tucker conditions

Necessity and sufficiency

- Necessary conditions based on first order derivatives
  - Local candidate for an optimum
- Sufficient conditions based on second order derivatives
- Necessary condition is sufficient if
  - The constraints are convex functions
  - E.g. no constraints, linear constraints, ...
  - The objective function is concave: global maximum is obtained
  - The objective function is convex: global minimum is obtained
  - Often the “real” motivation in Economics

Necessary conditions

1. Free optimization
   - No constraints
   - $f'(x^*) = 0$ if $n = 1$
   - $\frac{\partial}{\partial x_i} f(x_1^*, \ldots, x_n^*) = 0$ for $i = 1, \ldots, n$
   - Intuitive given our geometric interpretation

Necessary conditions

2. Optimization with positivity constraints
   - No g_i constraints
   - On the boundary extra optima are possible
   - Often ignored: interior solutions
   - $x_i^* \geq 0$ for $i = 1, \ldots, n$
   - $x_i^* \frac{\partial}{\partial x_i} f(x_1^*, \ldots, x_n^*) = 0$ for $i = 1, \ldots, n$
   - $\frac{\partial}{\partial x_i} f(x_1^*, \ldots, x_n^*) \leq 0$ for all $i = 1, \ldots, n$ simultaneously OR $\frac{\partial}{\partial x_i} f(x_1^*, \ldots, x_n^*) \geq 0$ for all $i = 1, \ldots, n$ simultaneously

Necessary conditions

3. Constrained optimization without positivity constraints
   - Define Lagrangian: $L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = f(x_1, \ldots, x_n) - \lambda_1(g_1(x_1, \ldots, x_n)) - \cdots - \lambda_m(g_m(x_1, \ldots, x_n))$
   - $\frac{\partial}{\partial x_i} L(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*) = 0$ for all $i = 1, \ldots, n$
   - $\frac{\partial}{\partial \lambda_j} L(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_m^*) = 0$ for all $j = 1, \ldots, m$
   - Alternatively: $\nabla f(x_1^*, \ldots, x_n^*) = \lambda_1^* \nabla g_1(x_1^*, \ldots, x_n^*) + \cdots + \lambda_m^* \nabla g_m(x_1^*, \ldots, x_n^*)$
   - Some intuition: geometric interpretation
   - Lagrange multiplier = shadow price

Application: utility maximization

$$\max U(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha} \quad s.t. \quad p_1 x_1 + p_2 x_2 = Y$$

- $L(x_1, x_2, \lambda_1) = x_1^{\alpha} x_2^{1-\alpha} - \lambda_1(p_1 x_1 + p_2 x_2 - Y)$
- $\frac{\partial L}{\partial x_1} = \alpha x_1^{\alpha-1} x_2^{1-\alpha} - \lambda_1 p_1 = 0$
- $\frac{\partial L}{\partial x_2} = (1-\alpha) x_1^{\alpha} x_2^{-\alpha} - \lambda_1 p_2 = 0$
- $\frac{\partial L}{\partial \lambda_1} = p_1 x_1 + p_2 x_2 - Y = 0$
- $x_1^* = \frac{\alpha Y}{p_1}, x_2^* = \frac{(1-\alpha)Y}{p_2}$ and $\lambda_1^* = (\frac{\alpha}{p_1})^{\alpha} (\frac{(1-\alpha)}{p_2})^{(1-\alpha)}$

Some exercises

- Find the optima for the following problems
  - $\max / \min x^3 - 12x^2 + 36x + 8$
  - $\max / \min x_1^3 - x_2^3 + 9x_1x_2$
  - $\min 2x_1^2 + x_1x_2 + 4x_2^2 + x_1x_3 + x_3^2 - 15x_1$
  - $\max x_1x_2$ s.t. $x_1 + 4x_2 = 16$
  - $\max x_2x_3 + x_1x_3$ s.t. $x_2^2 + x_3^2 = 1$ and $x_1x_3 = 3$
- Add positivity constraints to the above unconstrained problems and do the same

Outline

Motivation

- Sequences and series are frequently used in Finance
- E.g. a stream of dividends is a sequence of numbers
- E.g. the price of a stock is the sum of all future dividends

Outline

Sequences

- A sequence is simply an infinite list of numbers
  - $a_1, a_2, a_3, \dots$
  - E.g. $1, 3, -\sqrt{2}, \dots$
- Often there is a systematic pattern
  - There is formula describing the sequence
  - E.g. $1, \frac{1}{2}, \frac{1}{3}, \dots$ or $a_n = \frac{1}{n}$

Two useful types of sequences

- Arithmetic sequence: $a_n = a + (n-1)d$
  - $a, a+d, a+2d, \dots$
  - There is a constant difference between the terms
  - E.g. -3, -1, 1, 3, ...
  - E.g. weekly evolution of the stock if the firm does not sell and produces d units every week
- Geometric sequence $a_n = ar^n - 1$
  - $a, ar, ar^2, \dots$
  - The ratio between the terms is constant
  - E.g. 7, 14, 28, ...
  - E.g. yearly evolution of capital at constant interest rate

Series

- A series is the sum of all the terms of a sequence
- This can be a finite number or an infinite number
  - E.g. $1 + 2 + 3 + \dots = +\infty$
  - E.g. $1 + \frac{1}{2} + \frac{1}{3} + \dots = +\infty$
  - E.g. $1 + \frac{1}{2} + \frac{1}{4} + \dots = 2$
- Partial sum S_N is the sum of the first N elements of the sequence
  - Finite version of the series
  - Evolves to the series if N gets bigger

Two useful types of series

- Arithmetic series
  - Sum of arithmetic sequence: $a_n = a + (n-1)d$
  - Partial sum: $S_N = Na + (1 + 2 + \dots + N - 1)d = Na + \frac{N(N-1)}{2}d$
  - Series is useless: 0 or $\pm\infty$, depending on d and a
- Geometric series
  - Sum of geometric sequence: $a_n = ar^{n-1}$
  - Partial sum: $S_N = a(1 + r + \dots + r^{N-1}) = a\frac{1-r^N}{1-r}$
  - Series: $\frac{a}{1-r}$ if $|r| < 1$, else $\pm\infty$

Exercises

- Consider the sequence $26, 22, 18, \ldots$
  - Give the sum of the first 8 elements
  - Give a formula for the partial sums
- Consider the sequence $\frac{1}{3}, \frac{1}{9}, \frac{1}{27}, \ldots$
  - Give the sum of the first 8 elements
  - Give a formula for the partial sums
- Let a_n be an arithmetic sequence for which the sum of the first 12 terms is 222 and the sum of the first 5 terms is 40. What is the general formula of this sequence?
- Let a_n be a geometric sequence for which the fourth term is 56 and the sixth term is $\frac{7}{8}$. What is the series of this sequence?

Outline

Compounded interest

- Compound interest on yearly basis
  - Assume capital K and yearly interest rate of r%
  - You receive interests only at the end of the year
  - Capital after N years: $K(1+r)^N$
  - I.e. interest on interests also matter
- Interest is compounded several times per year
  - m times per year you receive interests
  - Of course the interest rate is adapted: $\frac{r}{m}$
  - Capital after one year: $K(1+\frac{r}{m})^m$
  - More capital since more interests on interests

Compounded interest

- Interest is compounded continuously
  - Often used in Macro
  - m goes to infinity
  - Capital after one year: Ke^r
- Note $Ke^r > K(1+\frac{r}{m})^m > K(1+r)$
  - Nominal interest rate is r
  - Annual percentage rate: $(1+\frac{r}{m})^m - 1$ or $e^r - 1$
- Depreciation calculations are very similar
  - Depreciation rate r
  - Use $1-r$ instead of $1+r$

Net present value

- 1000 Euro in 2014 does not have the same value as 1000 Euro in 2024
- We need to discount future amounts to make them comparable
  - We use compounded interest to do this
- Example
  - Assume interest rate at saving accounts is 2% and you receive interests on a yearly basis
  - Capital K after 10 years: $K(1.02)^{10}$
  - $820,35(1.02)^{10} = 1000$ or $820.35 = \frac{1000}{(1.02)^{10}}$
  - The discounted value of the 1000 Euro of 2024 is 820.35

Evaluating investments

- Assume that an investment of K Euro will give a yearly return of 1000 Euros for the next 5 years
- For which K is this an interesting investment if the interest rate is 2%
- Answer
  - We need to discount the 1000 Euro of every year
  - $\frac{1000}{1.02} + \frac{1000}{(1.02)^2} + \cdots + \frac{1000}{(1.02)^5} = \frac{1000}{1.02}(1 + \frac{1}{1.02} + \cdots + \frac{1}{(1.02)^4})$
  - Geometric sequence/series: $\frac{1000}{1.02}\frac{1-(\frac{1}{1.02})^5}{1-\frac{1}{1.02}} = 1000\frac{1-(\frac{1}{1.02})^5}{0.02} = 4713.46$
  - So K should be less than 4713.46 Euro to make this investment profitable

Exercises

Consider a stock or bond that gives you a yearly dividend of 10 Euro

- Assume that you receive dividends for 10 years, what is the price you want to pay for this stock/bond if the interest rate is 2% (compounded yearly)?
- Assume now that you receive dividends forever, what is then the price you want to pay?
- Due to uncertainty, you want to add a risk premium of 2%, meaning that you now discount with 4% instead of 2%. What is the impact on both prices?

Outline

Motivation

- Matrices allow to formalize notation
- Useful in solving system of linear equations
- Useful in deriving estimators in econometrics
- Allows us to make the link with vector spaces

Outline

Matrices

$$A = (a_{ij})_{i=1,\dots,n;j=1,\dots,m} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

- $a_{ij} \in \mathbb{R}$ and $A \in \mathbb{R}^{n\times m}$
  - n rows and m columns
- Square matrix if $n = m$
- Notable square matrices
  - Symmetric matrix: $a_{ij} = a_{ji}$ for all $i, j = 1, \dots, n$
  - Diagonal matrix: $a_{ij} = 0$ for all $i, j = 1, \dots, n$ and $i \neq j$
  - Triangular matrix: only non-zero elements above (or below) the diagonal

Matrix manipulations

Let $A, B \in \mathbb{R}^{n\times m}$ and $k \in \mathbb{R}$

- Equality: $A = B \Leftrightarrow a_{ij} = b_{ij}$ for all $i, j = 1, \dots, n$
- Scalar multiplication: $kA = (ka_{ij})_{i=1,\dots,n;j=1,\dots,m}$
- Addition: $A \pm B = (a_{ij} \pm b_{ij})_{i=1,\dots,n;j=1,\dots,m}$
  - Dimensions must be equal
- Transposition: $A' = A^t = (a_{ji})_{j=1,\dots,m;i=1,\dots,n}$
  - $A \in \mathbb{R}^{n\times m}$ and $A^t \in \mathbb{R}^{m\times n}$
  - $(A \pm B)^t = A^t \pm B^t$
  - $(kA)^t = kA^t$
  - $(A^t)^t = A$

Matrix multiplication

Let $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times k}$

- $AB = (\sum_{h=1}^{m} a_{ih} b_{hj})_{i=1,\dots,n; j=1,\dots,k}$
- Multiply the row vector of A with the column vector of B
  - Aside: scalar/inner product and norm of vectors
  - Orthogonal vectors
- Number of columns of A must be equal to number of rows of B
- $AB \neq BA$, even if both are square matrices
- $(AB)^t = B^t A^t$

Example

Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$

- $A^t = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, $B^t \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$ and $C^t = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}$
- $AB = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$ and $BA = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}$
- $AC = \begin{pmatrix} 5 & 6 & 7 \\ 4 & 4 & 4 \end{pmatrix}$
- $(AB)^t = B^t A^t = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$

Exercises

- Let $A = \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $D = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 4 & 0 & -3 & 1 \end{pmatrix}$
  - Compute $-3C$, $A + B$, $A - D$ and D^t
  - Compute AB, BA, AC, CA, AD and DA
- Let A be a symmetric matrix, show then that $A^t = A$
- A square matrix A is called idempotent if $A^2 = A$
  - Verify which of the above matrices are idempotent
  - Find the value of α that makes the following matrix idempotent: $\begin{pmatrix} -1 & 2 \\ \alpha & 2 \end{pmatrix}$

Two numbers associated to square matrices: trace

Let $A, B, C \in \mathbb{R}^{n \times n}$

- Trace(A) $= tr(A) = \sum_{i=1}^{n} a_{ii}$
- Used in econometrics
- Properties
  - $tr(A^t) = tr(A)$
  - $tr(A \pm B) = tr(A) \pm tr(B)$
  - $tr(cA) = ctr(A)$ for any $c \in \mathbb{R}$
  - $tr(AB) = tr(BA)$
  - $tr(ABC) = tr(BCA) = tr(CAB)$ $\neq tr(ACB) (= tr(BAC) = tr(CBA))$

Two numbers associated to square matrices: trace

Example

- Let $A = \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$
- Then $AB = \begin{pmatrix} 21 & 0 \\ -7 & 0 \end{pmatrix}$, $BA = \begin{pmatrix} 1 & 2 \\ 10 & 20 \end{pmatrix}$ and $A + B = \begin{pmatrix} 4 & 8 \\ 2 & -3 \end{pmatrix}$
- $tr(A) = 1$, $tr(B) = 0$ and $tr(A + B) = 1$
- $tr(AB) = 21 = tr(BA)$

Two numbers associated to square matrices: determinant

Let $A \in \mathbb{R}^{n\times n}$

- If $n = 1$, then $\det(A) = a_{11}$
- If $n = 2$, then $\det(A) = a_{11}a_{22} - a_{12}a_{21} = a_{11}\det(a_{22}) - a_{12}\det(a_{21})$
- If $n = 3$, then $\det(A) = a_{11}\det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12}\det\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13}\det\begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$
- Can be generalized to any n
- Works with columns too

Two numbers associated to square matrices: determinant

Let $A, B \in \mathbb{R}^{n\times n}$

- $\det(A^t) = \det(A)$
- $\det(A \pm B) \neq \det(A) \pm \det(B)$
- $\det(cA) = c^n \det(A)$ for any $c \in \mathbb{R}$
- $\det(AB) = \det(BA)$
- A is non-singular (or regular) if A^{-1} exists
  - I.e. $AA^{-1} = A^{-1}A = I_n$
  - I_n is a diagonal matrix with 1 on the diagonal
  - Does not always exist
  - $\det(A) \neq 0$

Two numbers associated to square matrices: determinant

Example

- Let $A = \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$
- Then $AB = \begin{pmatrix} 21 & 0 \\ -7 & 0 \end{pmatrix}$, $BA = \begin{pmatrix} 1 & 2 \\ 10 & 20 \end{pmatrix}$ and $A + B = \begin{pmatrix} 4 & 8 \\ 2 & -3 \end{pmatrix}$
- $\det(A) = 0$, $\det(B) = -7$ and $\det(A + B) = -28$
- $\det(AB) = 0 = \det(BA)$

Exercises

- Let $A = \begin{pmatrix} 2 & 4 & 5 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}$
  - Compute $tr(A)$ and $\det(-2A)$
- Show that for any triangular matrix A, we have that $\det(A)$ is equal to the product of the elements on the diagonal
- Let $A, B \in \mathbb{R}^{n \times n}$ and assume that B is non-singular
  - Show that $tr(B^{-1}AB) = tr(A)$
  - Show that $tr(B(B^tB)^{-1}B^t) = n$
- Let $A, B \in \mathbb{R}^{n \times n}$ be two non-singular matrices
  - Show that AB is then also invertible
  - Give an expression for $(AB)^{-1}$

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

Outline

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

A notion of vector spaces

- A set of vectors V is a vector space if
  - Addition of vectors is well-defined
  - $\forall a, b \in V : a + b \in V$
  - Scalar multiplication is well-defined
  - $\forall k \in \mathbb{R}, \forall a \in V : ka \in V$
- We can take linear combinations
  - $\forall k_1, k_2 \in \mathbb{R}, \forall a, b \in V : k_1 a + k_2 b \in V$
- E.g. $\mathbb{R}^2$ or more generally $\mathbb{R}^n$
- Counterexample $\mathbb{R}^2_+$

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

Linear (in)dependence

Let V be a vector space

- A set of vectors $v_1, \ldots, v_n \in V$ is *linear dependent* if one of the vectors can be written as a linear combination of the others
  - $\exists k_1, \ldots k_{n-1} \in \mathbb{R} : v_n = k_1 v_1 + \cdots + k_{n-1} v_{n-1}$
- A set of vectors are *linear independent* if they are not linear dependent
  - $\forall k_1, \ldots k_n \in \mathbb{R} : k_1 v_1 + \cdots + k_n v_n = 0 \Rightarrow k_1 = \cdots = k_n = 0$
- In a vector space of dimension n, the number of linear independent vectors cannot be higher than n

Linear (in)dependence

Example

- $\mathbb{R}^2$ is a vector space of dimension 2
- $v_1 = (1,0)$, $v_2 = (1,2)$, $v_3 = (-1,4)$ and $v_4 = (2,4)$
- $v_3 = -3v_1 + 2v_2$, so v_1, v_2, v_3 are linear dependent
- $v_4 = 2v_2$, so v_2, v_4 are linear dependent
- v_1, v_2 are linear independent
- v_3 is linear independent

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

Link with matrices: rank

Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times k}$

- The row rank of A is the maximal number of linear independent rows of A
- The column rank of A is the maximal number of linear independent columns of A
- Rank of A = column rank of A = row rank of A
- Properties
  - $rank(A) \leq \min(n, m)$
  - $rank(AB) \leq \min(rank(A), rank(B))$
  - $rank(A) = rank(A^tA) = rank(AA^t)$
  - If $n = m$, then A has maximal rank if and only if $\det(A) \neq 0$

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

Link with matrices: rank

Example

- Let $A = \begin{pmatrix} 3 & 6 \\ -1 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$
- Then $AB = \begin{pmatrix} 21 & 0 \\ -7 & 0 \end{pmatrix}$, $BA = \begin{pmatrix} 1 & 2 \\ 10 & 20 \end{pmatrix}$
- $rank(A) = 1$ and $rank(B) = 2$
- $rank(AB) = 1$

Introduction
Calculus
Financial mathematics
Linear algebra
Fundamentals of probability theory
Motivation
Matrix algebra
The link with vector spaces
Application: solving a system of linear equations

Exercises

- Let $A = \begin{pmatrix} 2 & 4 & 5 \\ 0 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}$
  - Show in two ways that A has maximal rank
- Let $A, B \in \mathbb{R}^{n\times m}$
  - Show that there need not be any relation between $rank(A+B)$, $rank(A)$ and $rank(B)$
- Show that if A is invertible, then it needs to have a maximal rank

Outline

System of linear equations

- Linear equations in the unknowns $x_1, \ldots, x_m$
  - Not $x_1 x_2$, x_m^2,...
- Constraints hold with equality
  - Not $2x_1 + 5x_2 \leq 3$
- E.g. $\begin{cases} 2x_1 + 3x_2 - x_3 = 5 \\ -x_1 + 4x_2 + x_3 = 0 \end{cases}$

Using matrix notation

- m unknowns: $x_1, \ldots, x_m$
- n linear constraints: $a_{i1}x_1 + \cdots + a_{im}x_m = b_i$ with $a_{i1}, \ldots a_{im}, b_i \in \mathbb{R}$ and $i = 1, \ldots, n$
- $Ax = b$
  - $A = (a_{ij})_{i=1,\ldots,n, j=1,\ldots,m}$
  - $x = (x_i)_{i=1,\ldots,m}$
  - $b = (b_i)_{i=1,\ldots,n}$
- Homogeneous if $b_i = 0$ for all $i = 1, \ldots, n$

Solving a system of linear equations

- Solve this by logical reasoning
  - Eliminate or substitute variables
  - Can also be used for non-linear systems of equations
  - Can be cumbersome for larger systems
- Use matrix notation
  - Gaussian elimination of the augmented matrix $(A|b)$
  - Can be programmed
  - Only for systems of linear equations
  - Theoretical statements are possible

Example

$$\begin{cases} -x_1 + 4x_2 = 0 \\ 2x_1 + 3x_2 = 5 \end{cases}$$

- $x_1 = 4x_2 \Rightarrow 11x_2 = 5 \Rightarrow x_2 = \frac{5}{11}$ and $x_1 = \frac{20}{11}$
- $\begin{pmatrix} -1 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}$
- Take linear combinations of rows of the augmented matrix
  - Is the same as taking linear combinations of the equations
  - $\begin{pmatrix} -1 & 4 & |0 \\ 2 & 3 & |5 \end{pmatrix} \Rightarrow \begin{pmatrix} -1 & 4 & | & 0 \\ 0 & 11 & | & 5 \end{pmatrix} \Rightarrow \begin{pmatrix} -1 & 0 & | & \frac{-20}{11} \\ 0 & 11 & | & 5 \end{pmatrix}$

Theoretical results

Consider the linear system $Ax = b$ with $A \in \mathbb{R}^{n \times m}$

- This system has a solution if and only if $rank(A) = rank(A|b)$
  - $rank(A) \leq rank(A|b)$ by definition
  - Is the (column) vector b a linear combination of the column vectors of A?
  - If $rank(A) < rank(A|B)$, the answer is no
  - If $rank(A) = rank(A|B)$, the answer is yes
- The solution is unique if $rank(A) = rank(A|B) = m$
  - $n \geq m$
- There are ∞ many solutions if $rank(A) = rank(A|B) < m$
  - $n < m$ or too many constraints are ‘redundant’

Exercises

- Solve the following systems of linear equations
  - $\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ 3x_1 + 2x_2 + x_3 = 1 \end{cases}$ and $\begin{cases} x_1 - x_2 + x_3 = 1 \\ 3x_1 + x_2 + x_3 = 0 \\ 4x_1 + 2x_3 = -1 \end{cases}$
- For which values of k does the following system of linear equations have a unique solution?
  - $\begin{cases} x_1 + x_2 = 1 \\ x_1 - kx_2 = 1 \end{cases}$
- Consider the linear system $Ax = b$ with $A \in \mathbb{R}^{n \times n}$
  - Show that this system has a unique solution if and only if A is invertible
  - Give a formula for this unique solution
- Show that homogeneous systems of linear equations always have a (possibly non-unique) solution

Outline

Motivation

In econometrics/statistics we want

- To draw conclusions about a random variable X
  - Data Generating Process
  - Determines the random outcome for X
  - Possibly an infinite population
  - E.g. X is the income of a person
- And we can only use a limited set of observations
  - Due to randomness there is always uncertainty
  - Same holds because of the finite set of observations
  - E.g. we observe the income of 1000 persons

Motivation

We will recall the basic notations and tools

- Allows to quantify the uncertainty
- Refreshes some of the important concepts
- E.g. with 95% certainty we can conclude that the average income for the population lies between 1900 and 2100 Euros

Outline

Random variable

Let X be a random variable

- The outcome of a random data generating process
- Univariate versus multivariate
- Discrete versus continuous
  - Indivisible or countably infinite
- Probabilities are associated to the possible outcomes
  - $Prob(X = x)$ or $Prob(a \leq X \leq b)$ with $a, b \in \mathbb{R}$
  - Probability distributions $f(x)$
- Examples
  - The outcome of the throw of a dice
  - The temperature on September 24

Probability density function (pdf)

Let X be a discrete random variable

- The probability density function is a function satisfying
  - $f(x) = Prob(X = x)$
  - $0 \leq Prob(X = x) \leq 1$
  - $\sum_x f(x) = 1$
- Formalizes our intuitive notion of probability
  - Has a direct interpretation
  - Probabilities are positive
  - Total probability cannot exceed 1
  - E.g. pick a random number out of $\{1, 2, 3\}$

Probability density function (pdf)

Let X be a continuous random variable

- The probability density function $f(x)$ is a function satisfying
  - $Prob(a \leq x \leq b) = \int_a^b f(x)dx$
  - $f(x) \geq 0$
  - $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Extends our machinery to the continuous case
  - Because of indivisibility we have that $Prob(X = x) = 0$
  - Probabilities are surfaces (and positive by construction)
  - E.g. pick a random number in $[0, 1]$

Cumulative distribution function (cdf)

- X is a discrete random variable
  - $F(x) = \sum_{X \leq x} f(X) = \sum_{X \leq x} Prob(X = x) = Prob(X \leq x)$
- X is a continuous random variable
  - $F(x) = \int_{-\infty}^{x} f(t)dt$
  - $f(x) = \frac{dF(x)}{dx}$
- Note
  - $0 \leq F(x) \leq 1$
  - F is an increasing function
  - $Prob(a \leq X \leq b) = F(b) - F(a)$

Quantile function

- Quantile function Q is the “inverse function” of the cdf
- This function defines the relative position in the distribution
  - E.g. first quartile, second decile, median, ...
- X is a continuous random variable
  - $Q(p) = x$ if $F(x) = p$
  - Or if $F(x) = p$, then $Prob(X \leq Q(p)) = p$
- X is a discrete random variable
  - The cdf is a step function in the discrete case
  - If $F(x) = p$, then $Q(p)$ is the smallest value for which $Prob(X \leq Q(p)) \geq p$

Measure of central tendacy

- The expected value or mean
  - $E(X) = \sum_x xf(x)$ if X is discrete
  - $E(X) = \int_x xf(x)dx$ if X is continuous
  - It the value that we expect on average
- The median is $Med(X) = Q(0.5)$
  - For symmetric distributions: mean $\approx$ median
  - For right skewed distributions : mean $>$ median
  - For left skewed distributions: mean $<$ median
  - Less sensitive for outliers
- The mode is $\arg\max f(x)$
  - The value of X that has the highest probability of occurring

Measure of central tendacy

- Let g be an increasing function
  - $E(g(X)) \neq g(E(X))$
  - $Med(g(X)) = g(Med(X))$
  - The mode becomes $g(mode)$
- Only exception: $g(x) = a + bx$
  - Then $E(g(X)) = g(E(X)) = a + bE(X)$
- Example
  - Pick a random number from $\{1, 2, 3\}$
  - $g(x) = x^2$
  - $E(X) = 2$, $Med(X) = 2$ and $mode = \{1, 2, 3\}$
  - $E(g(X)) = \frac{14}{3}$, $Med(X) = 2$ and $mode = \{1, 4, 9\}$

Measure of dispersion

- The variance
  - $Var(X) = E((X - E(X))^2) = \sum_x (x - E(X))^2 f(x)$ if X is discrete
  - $Var(X) = \int_x (x - E(X))^2 f(x)dx$ if X is continuous
  - How far is x from the average
  - Squared deviations since too small or too big
- Standard deviation = $(Var(X))^{\frac{1}{2}}$
- The inter quartile range: $Q(0.75) - Q(0.25)$
  - Less sensitive for outliers

Measure of dispersion

- Let g be an increasing function
  - $Var(g(X)) \neq g(Var(X))$
- However if $g(x) = a + bx$, then $Var(g(X)) = b^2 Var(X)$
  - Squared deviations
  - Adding a constant does not change dispersion
- Example
  - Pick a random number from $\{1, 2, 3\}$
  - $g(x) = x^2$
  - $Var(X) = \frac{2}{3}$
  - $Var(g(X)) = \frac{98}{3}$

Central moments

Let X be a random variable

- $Var(X)$ is an example of a central moment of X
- $\mu_r = E((x - E(X))^r)$
- Related to skewness if $r = 3$
  - Is zero for symmetric distributions
  - Puts less weight on outcomes that are closer to $E(X)$
- Kurtosis if $r = 4$
  - Measure for the thickness of the tails
  - Puts more weight on the extreme observations

Exercises

- Compute the first four central moments for the following random variables
  - $X \in \{0, 1\}$ and $f(X = 0) = \frac{1}{4}$
  - $X \in \{a\}$ and $f(X = a) = 1$, with $a \in \mathbb{R}$
- Show that $Prob(a < X < b) = Prob(a < X \leq b)$ $= Prob(a \leq X \leq b)$ if X is a continuous variable
- Argue that the same does not need to hold if X is discrete

Some remarks

- See the document overview_distributions.pdf for some important distributions
- Two more important functions for a continuous random variable X
  - The survival function: $S(x) = 1 - F(x)$
  - E.g. x stands for time until transition
  - The hazard function: $h(x) = \frac{f(x)}{S(x)}$
  - E.g. x stands for duration of an event

Outline

Importance of multivariate setting

- In principle same machinery
  - Probability density function, expected value, ...
  - But relative position does not exist in general
  - Slightly more technical
- Allows to formally study new concepts
  - The independence of random variables
  - More general, the correlation between random variables
  - But also marginal and conditional pdf's
- We will focus on the bivariate case
  - Everything can of course be generalized

Two discrete random variables

Let X and Y be two discrete random variables

- E.g. $X =$ male/female (i.e. $X \in \{0, 1\}$) and $Y =$ score at the exam (i.e. $Y \in \{1, 2, \ldots, 20\}$)
- The joint pdf $f(x, y)$
  - $f(x, y) = Prob(X = x, Y = y)$
  - $f(x, y) \geq 0$
  - $\sum_x \sum_y f(x, y) = 1$
- The joint cdf $F(x, y)$
  - $F(x, y) = Prob(X \leq x, Y \leq y) = \sum_{X \leq x} \sum_{Y \leq y} f(x, y)$
- Expected generalizations
  - Quantile function is not well-defined
  - Relative position? Inverse function?

Two continuous random variables

Let X and Y be two continuous random variables

- E.g. $X =$ temperature on September 24 and $Y =$ liters of rain per square meter on September 24
- The joint pdf $f(x, y)$
  - $f(x, y) = Prob(a \leq x \leq b, c \leq y \leq d)$
  - $f(x, y) \geq 0$
  - $\int_x \int_y f(x, y) dydx = 1$
- The joint cdf $F(x, y)$
  - $F(x, y) = Prob(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dtds$

The marginal pdf

Let X and Y be two random variables

- The pdf for one variable, irrespective of the value of the other variable
- E.g. the pdf for the exam score, irrespective of the sex of the student
  - The probability of having 12 is the sum of the probability of a female having 12 and a male having 12
- Formally
  - E.g. $f_X(x) = \sum_y Prob(x = X, y = Y)$ if X and Y are discrete
  - E.g. $f_Y(y) = \int_x f(x, y) dx$ if X and Y are continuous

Independent variables

Let X and Y be two random variables

- The marginal distributions allow us to define independence
- X and Y are independent if and only if $f(x,y) = f_X(x)f_Y(y)$ for all values of x and y
- Remark that for dependent variables a similar relation between the joint pdf and the marginal pdf's does not exist

Independent variables

Example

- E.g. X = male/female and Y = score at the exam
- $Prob(X = Male) = 0.4\ (= f_X(Male))$
- $Prob(X = Female) = 0.6\ (= f_X(Female))$
- $Prob(Y = 12) = 0.3\ (= f_Y(12))$
- $Prob(X = Male, Y = 12) = 0.10 \neq 0.4 \times 0.3$
- $Prob(X = Female, Y = 12) = 0.20 \neq 0.6 \times 0.3$
- There is dependence
  - E.g. females have a higher probability of obtaining 12
  - Since $0.20 > 0.18$, not since $0.20 > 0.10$!

The conditional pdf

Let X and Y be two random variables

- The pdf of one variable for a given value of the other variable
- E.g. what is the probability of having 12, conditional on being female
- Formally: $f(y|x) = \frac{f(x,y)}{f_X(x)}$

The conditional pdf

Let X and Y be two random variables

- Let X and Y be independent
  - Then $f(y|x) = f_Y(y)$ and $f(x|y) = f_X(x)$
  - Conditioning on x or y does not give extra information
- Reformulating the above:
  - $f(x,y) = f(y|x)f_X(x) = f(x|y)f_Y(y)$
  - This is the factorization of the joint distribution that takes dependence into account

The conditional pdf

Example

- E.g. $X =$ male/female and $Y =$ score at the exam
- $Prob(X = Male) = 0.4$ and $Prob(X = Female) = 0.6$
- $Prob(Y = 12) = 0.3$
- $Prob(X = Male, Y = 12) = 0.10$ and $Prob(X = Female, Y = 12) = 0.20$
- $Prob(Y = 12|X = Male) = \frac{0.10}{0.4} = 0.25$
- $Prob(Y = 12|X = Female) = \frac{0.20}{0.6} = 0.33$
- $Prob(X = Female|Y = 12) = \frac{0.20}{0.3} = 0.66$
- This is formally confirming our previous intuitive conclusion

Expected value

- The marginal and conditional pdf allow to compute the same numbers as before
- Expected value or mean
  - The expected value for X, irrespective of the value of Y
  - $E(X) = \sum_x xf_X(x) = \sum_x \sum_y xf(x, y)$ if X and Y are discrete
  - $E(Y) = \int_y yf_Y(y)dy = \int_x \int_y yf(x, y)dydx$ if X and Y are continuous
- Conditional expected value or mean
  - The expected value for X, conditional on the value of Y
  - $E(X|Y) = \sum_x xf(x|y)$ if X and Y are discrete
  - $E(Y|X) = \int_y yf(y|x)dy$ if X and Y are continuous
  - Regression: $y = E(Y|X) + (y - E(Y|X)) = E(Y|X) + \epsilon$

Variance

- Variance
  - The dispersion of X, irrespective of the value of Y
  - $Var(X) = \sum_x (x - E(X))^2 f_X(x) = \sum_x \sum_y (x - E(X))^2 f(x, y)$ if X and Y are discrete
  - $E(Y) = \int_y (y - E(Y))^2 f_Y(y)dy = \int_x \int_y (y - E(Y))^2 f(x, y)dydx$ if X and Y are continuous
- Conditional variance
  - The dispersion of X, conditional on the value of Y
  - $Var(X|Y) = \sum_x (x - E(X))^2 f(x|y)$ if X and Y are discrete
  - $Var(Y|X) = \int_y (y - E(Y))^2 f(y|x)dy$ if X and Y are continuous
  - Homoscedasticity: the conditional variance does not vary

Covariance and correlation

Let X and Y be two random variables

- Summarize the dependence between X and Y in a single number
- The covariance of X and Y
  - $Cov(X, Y) = E((X - E(X))(Y - E(Y)))$
  - Compare to $Var(X) = E((X - E(X))^2)$
  - A positive/negative number indicates a positive/negative dependence
  - $Cov(X, Y) = 0$ if X and Y are independent

Covariance and correlation

Let X and Y be two random variables

- Only sign of $Cov(X, Y)$ has a meaning
  - Rescaling of X and Y changes $Cov(X, Y)$ but of course not their dependence
- Correlation
  - $r(X, Y) = \rho(X, Y) = \frac{Cov(X,Y)}{(Var(X))^{\frac{1}{2}}(Var(X))^{\frac{1}{2}}}$
  - $-1 \leq r(X, Y) \leq 1$
  - Both size and sign have a meaning
  - This is not about causality!

Some remarks

Let X and Y be two random variables and $a, b, c, d \in \mathbb{R}$

- $E(aX + bY + c) = aE(X) + bE(Y) + c$
  - Similar as before and not influenced by (in)dependence
- $Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$
  - Extra term capturing the dependence of X and Y
- $Cov(aX + bY, cX + dY) = acVar(X) + bdVar(Y) + (ad + bc)Cov(X, Y)$
- Let X and Y be independent and g_1 and g_2 two functions
  - $E(g_1(X)g_2(Y)) = E(g_1(X))E(g_2(Y))$
  - Independence is crucial
  - In the above properties the linearity is crucial

A final example: the bivariate normal distribution

- The joint distribution of two variables that are normally distributed
- The joint pdf:

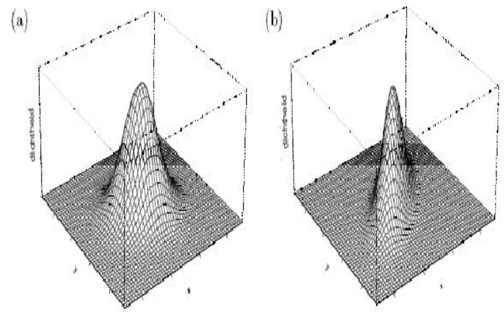
$$f(x, y) = \frac{1}{2\pi\sqrt{det(\Sigma)}} e^{-\frac{1}{2}(x-\mu_X, y-\mu_Y)\Sigma^{-1}(x-\mu_X, y-\mu_Y)^t}$$

  - μ_X and μ_Y are the expected values of X and Y
  - $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$ is the covariance matrix
  - σ_X and σ_Y are the standard deviations of X and Y
  - ρ is the correlation

A final example: the bivariate normal distribution

Expected generalization

- Same structure as in the univariate setting
- X and Y can be dependent



A final example: the bivariate normal distribution

Some results that **only** hold for the bivariate normal setting

- X and Y are independent if and only if $\rho = 0$
- The marginal pdf is again a normal distribution
  - $f_X : X \sim N(\mu_X, \sigma_X^2)$
- The conditional distribution is also a normal distribution
  - $f_{X|Y} : X|Y \sim N(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2))$