# ECON2125/4021/8013

## Lecture 11

John Stachurski

Semester 1, 2015

# Announcements

- Midterm exam date finalized
  - Date: 23rd April
  - Place: COP G30
  - Time: 6pm (writing time 6:30–8:30pm)

# Quadratic Forms

Up till now we have studied linear functions extensively

Next level of complexity is quadratic maps

Let $\mathbf{A}$ be $N \times N$ and symmetric, and let $\mathbf{x}$ be $N \times 1$

The **quadratic function** on $\mathbb{R}^N$ associated with $\mathbf{A}$ is the map

$$Q \colon \mathbb{R}^N \to \mathbb{R}, \qquad Q(\mathbf{x}) := \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{j=1}^N \sum_{i=1}^N a_{ij} x_i x_j$$

The properties of $Q$ depend on $\mathbf{A}$

An $N \times N$ symmetric matrix $\mathbf{A}$ is called

1. **nonnegative definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N$
2. **positive definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$
3. **nonpositive definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^N$
4. **negative definite** if $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$

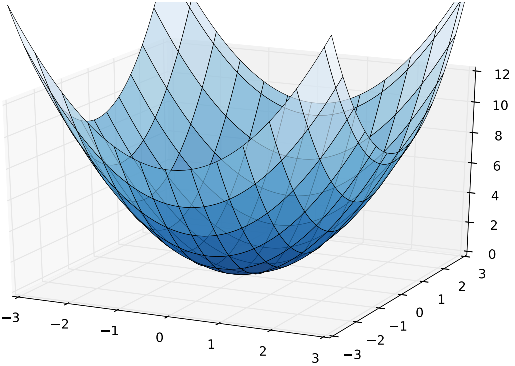



Figure : A positive definite case: $Q(\mathbf{x}) = \mathbf{x}'\mathbf{I}\mathbf{x}$

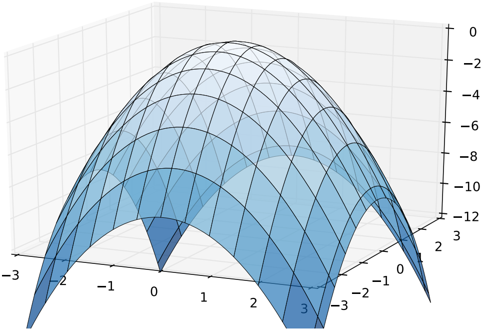



Figure : A negative definite case: $Q(\mathbf{x}) = \mathbf{x}'(-\mathbf{I})\mathbf{x}$

Note that some matrices have none of these properties

- $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$ for some $\mathbf{x}$
- $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for other $\mathbf{x}$

In this case $\mathbf{A}$ is called **indefinite**

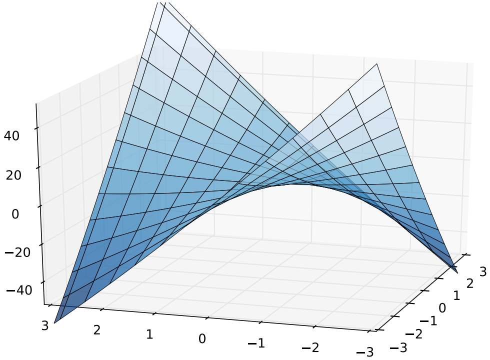



Figure : Indefinite quadratic function $Q(\mathbf{x}) = x_1^2/2 + 8x_1x_2 + x_2^2/2$

**Fact.** A symmetric matrix $\mathbf{A}$ is

1. positive definite $\iff$ all eigenvalues are strictly positive
2. negative definite $\iff$ all eigenvalues are strictly negative
3. nonpositive definite $\iff$ all eigenvalues are nonpositive
4. nonnegative definite $\iff$ all eigenvalues are nonnegative

It follows that

- $\mathbf{A}$ is positive definite $\implies \det(\mathbf{A}) > 0$

In particular, $\mathbf{A}$ is nonsingular

# New Topic

# PROBABILITY

# Topics

- Probability models
- Random variables
- Expectations
- Distributions
- Independence and dependence
- Asymptotics
- Multivariate models

# Motivation

The real world is messy relative to models

- especially econ / finance

In physics / chemistry / engineering, many theories are quite precise

- Hooke's law
- $E = mc^2$
- Ideal gas law
- etc.

The same is not true of models in econ / finance

Data is "noisy" relative to models

- Not everything can be explained by a given model
- Some events are "unpredictable"

Implication: We should model noise explicitly in order to

- Better match models to data
- Prepare for statistical analysis
- Add information we have about the noise

Good news: noise / randomness itself contains patterns

- Bursts of volatility in financial markets
- Bell shaped curve in abilities, test outcomes, etc.
- "Power law" in size of cities, firms
- Return on equities higher than bonds "on average"

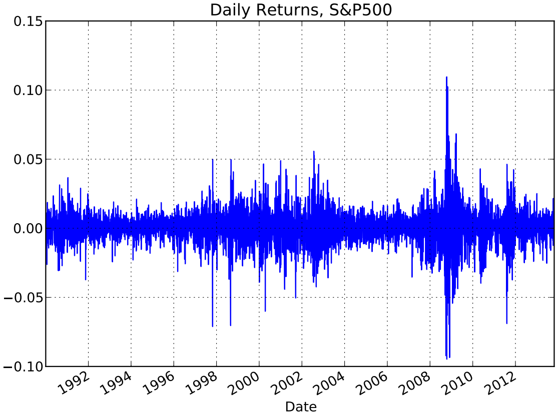



Figure : Volatility of daily returns

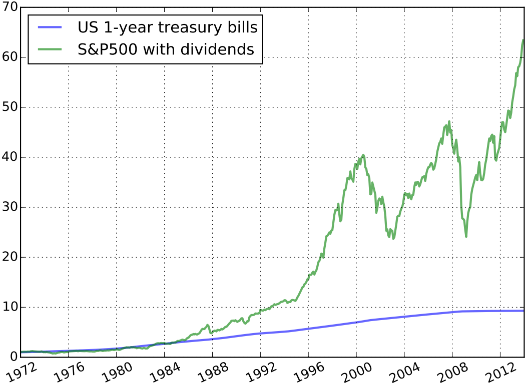



Figure : Cumulative return, 1$ invested in equities or bonds

The role of probability theory:

- Model phenomena that are “not fully predictable”
- Provide concepts for analyzing such phenomena
- Facilitate deductive reasoning in this setting

Example. Oil futures are “riskier” than US treasuries

Example. If event $A$ occurs whenever event $B$ occurs, then the probability of $A$ should be at least as high

Example. A monkey typing randomly at a keyboard will eventually reproduce the entire works of Shakespeare word for word

# Sample Spaces

First step of modeling: list all the things that can happen

In probability theory this is called the **sample space**

$= $ set of all possible outcomes in a random experiment

- can be any nonempty set
- typically denoted $\Omega$
- typical element of $\Omega$ denoted $\omega$

A subset of $\Omega$ is also called an **event**

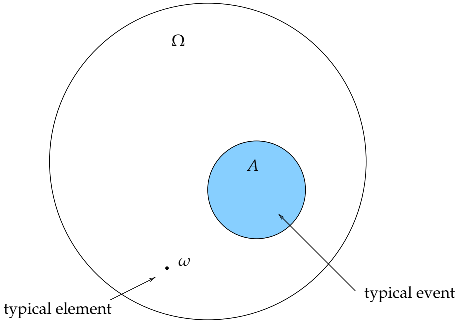



Figure : Sample space

Let $\mathcal{F}$ denote set of all events

For example, $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$

## Example

Consider an experiment where we roll a dice

We let $\Omega := \{1, \ldots, 6\}$ represent the set of possible outcomes

A typical outcome is

$$\omega = 4$$

A typical element of $\mathcal{F}$ is

$$A := \{2, 4, 6\} = \{ \text{ an even face } \}$$

The idea "event $A$ occurs" means that

$$\text{when } \omega \in \Omega \text{ is selected by "nature," } \omega \in A$$

Example

Consider again the experiment where we roll a dice

As before let $\Omega := \{1, \ldots, 6\}$

Let $A$ be the event

$$\{2, 4, 6\} = \{ \text{ an even face } \}$$

"$A$ occurs" means $\omega$ is one of $2, 4, 6$

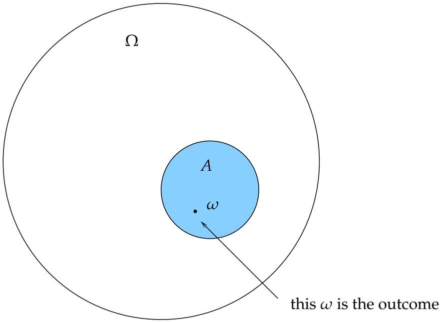



Figure : Event $A$ occurs

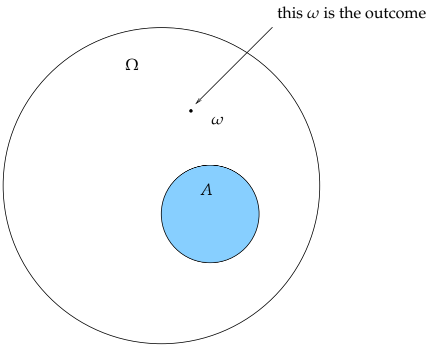



Figure : Event $A$ does not occur (but $A^c$ does)

# Probabilities

In probability theory, we first assign probability to events

Not individual outcomes—that can be problematic!

- See course notes for details

To each event $A \in \mathcal{F}$, we assign a probability $\mathbb{P}(A)$

$\mathbb{P}(A)$ represents the "probability that event $A$ occurs"

## Example

Consider again rolling a dice

The sample space is $\Omega := \{1, \ldots, 6\}$

We want to assign a probability to any event — any $A \in \mathcal{F}$

To this end we set

$$\mathbb{P}(A) := \frac{\# A}{6} \quad \text{for each} \quad A \in \mathcal{F}$$

- $\# A :=$ number of elements in $A$

For example,

$$\mathbb{P}\{2, 4, 6\} = \frac{3}{6} = \frac{1}{2}$$

We want $\mathbb{P}$ to satisfy some axioms...

A **probability** on $(\Omega, \mathcal{F})$ is a function $\mathbb{P}\colon \mathcal{F} \to [0,1]$ that satisfies

1. $\mathbb{P}(\Omega) = 1$, and
2. If $A$ and $B$ are disjoint events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Second property is called **additivity**

Note: Some technical details omitted — see course notes

## Example

As before let $\Omega := \{1, \ldots, 6\}$ and $\mathbb{P}(A) := \#A/6$

Are the axioms satisfied?

1. $\mathbb{P}(\Omega) = \mathbb{P}\{1, \ldots, 6\} = 6/6 = 1$
2. Additivity also holds:

First observe that $A \cap B = \emptyset \implies \#(A \cup B) = \#A + \#B$

$$\therefore \quad \mathbb{P}(A \cup B) = \frac{\#(A \cup B)}{6} = \frac{\#A}{6} + \frac{\#B}{6} = \mathbb{P}(A) + \mathbb{P}(B)$$

## Example

Memory chip is made up of billions of tiny switches/bits

- Switches can be off or on (0 or 1)

Random number generator accesses $N$ bits, switching each one on or off

We take

- $\Omega := \{(b_1, \ldots, b_N) : \text{where } b_n \text{ is 0 or 1 for each } n\}$
- $\mathbb{P}(A) := 2^{-N}(\#A)$

**Ex.** Show that $\mathbb{P}$ is a probability

**Fact.** If $\mathbb{P}$ is a probability and $A_1, \ldots, A_J$ are disjoint, then

$$\mathbb{P}\left(\cup_{j=1}^{J} A_j\right) = \sum_{j=1}^{J} \mathbb{P}(A_j)$$

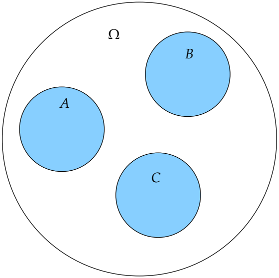



Figure : $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$

Proof for $J = 3$: Fixing disjoint $A$, $B$, $C$ and observing that

$$A \cup B \cup C = (A \cup B) \cup C$$

we have

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}((A \cup B) \cup C)$$

Clearly $A$, $B$, $C$ disjoint $\implies$ $A \cup B$ and $C$ disjoint

Hence

$$\begin{aligned} \mathbb{P}((A \cup B) \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \end{aligned}$$

## Example

Let $\Omega := \{1, \ldots, 6\}$ and $\mathbb{P}(A) := \# A/6$ for $A \in \mathcal{F}$

Prob of even is sum of probs of distinct ways we can get an even

$$\begin{aligned}
\mathbb{P}\{2,4,6\} &= \mathbb{P}[\{2\} \cup \{4\} \cup \{6\}] \\
&= \mathbb{P}\{2\} + \mathbb{P}\{4\} + \mathbb{P}\{6\} \\
&= 1/6 + 1/6 + 1/6 \\
&= 1/2
\end{aligned}$$

**Fact.** If $\mathbb{P}$ is a probability on $\mathcal{F}$ and $A, B \in \mathcal{F}$ with $A \subset B$, then

1. $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$
2. $\mathbb{P}(A) \leq \mathbb{P}(B)$
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
4. $\mathbb{P}(\emptyset) = 0$

Proof: When $A \subset B$, we have $B = (B \setminus A) \cup A$ and hence

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A)$$

All results follow (why!?)

Remark: Item 2 is called **monotonicity**

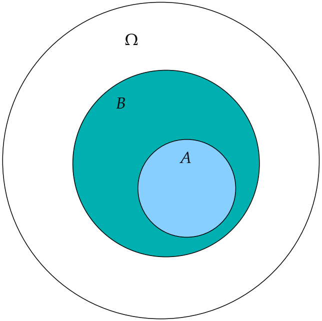



Figure : Monotonicity: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

**Fact.** If $A$ and $B$ are any events, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

**Ex.** Check the fact using $A = [(A \cup B) \setminus B] \cup (A \cap B)$

Implication: For any $A, B \in \mathcal{F}$, we have

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

- This is called **sub-additivity**
- What is the connection with additivity?

# Conditional Probability

Let $A$ and $B$ be two events and let $\mathbb{P}$ be a probability

The **conditional probability of $A$ given $B$** is defined as

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Defined only when $\mathbb{P}(B) > 0$

Intuitively,

- We don't know the actual outcome $\omega$
- But we do know that $\omega \in B$
- So what's the probability that $\omega \in A$?

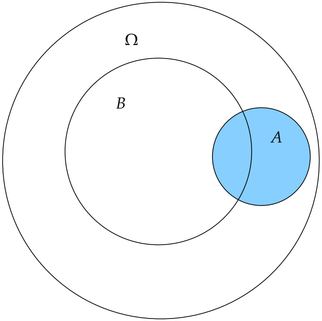



Figure : $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$

# Independent Events

Events $A$ and $B$ are called **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Intuitively, conditioning on independent events provides no additional information

In particular, when $\mathbb{P}(B) > 0$,

$$A,\ B \text{ independent} \iff \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)}$$

$$\iff \mathbb{P}(A \mid B) = \mathbb{P}(A)$$

## Example

Experiment: roll a dice twice.

$$\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\} \quad \text{and} \quad \mathbb{P}(A) := \#A/36$$

Now consider the events

$$A := \{(i,j) \in \Omega : i \text{ is even}\} \quad \text{and} \quad B := \{(i,j) \in \Omega : j \text{ is even}\}$$

In this case we have

$$A \cap B = \{(i,j) \in \Omega : i \text{ and } j \text{ are even}\}$$

We now show that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

This proves that $A$ and $B$ are independent under the probability $\mathbb{P}$

Recall that

$$\# \text{ of possible } (i,j) = \# \text{ of possible } i \ \times \ \# \text{ of possible } j$$

Applying this rule gives

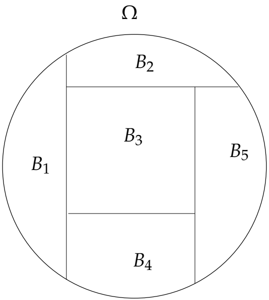
- $\#A = 3 \times 6 = 18$
- $\#B = 6 \times 3 = 18$
- $\#(A \cap B) = 3 \times 3 = 9$

$$\therefore \quad \mathbb{P}(A \cap B) = \frac{9}{36} = \frac{1}{4} = \frac{18}{36} \times \frac{18}{36} = \mathbb{P}(A)\mathbb{P}(B)$$

# Law of Total Probability

A collection of events $\{B_1, \ldots, B_M\}$ is called a **partition** of $\Omega$ if

$$i \neq j \implies B_i \cap B_j = \emptyset \quad \text{and} \quad \cup_{m=1}^{M} B_m = \Omega$$



**Fact.** If $A \in \mathcal{F}$ and $B_1, \ldots, B_M$ is a partition of $\Omega$ with $\mathbb{P}(B_m) > 0$ for all $m$, then

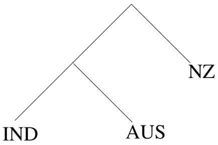
$$\mathbb{P}(A) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m)$$

Proof: Given any such $A$ and partition $B_1, \ldots, B_M$, we have

$$\mathbb{P}(A) = \mathbb{P}[A \cap (\cup_{m=1}^{M} B_m)] = \mathbb{P}[\cup_{m=1}^{M} (A \cap B_m)]$$

$$= \sum_{m=1}^{M} \mathbb{P}(A \cap B_m) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m)$$

Example. Suppose NZ in final of WC and IND, AUS in semi




I figure that $\mathbb{P}(\text{IND beats AUS}) = 0.35$ and

$$\mathbb{P}(\text{NZ beats AUS}) = 0.4, \qquad \mathbb{P}(\text{NZ beats IND}) = 0.5$$

Hence

$$\begin{aligned}\mathbb{P}(\text{NZ wins}) &= \mathbb{P}(\text{NZ wins} \mid \text{plays AUS})\mathbb{P}(\text{NZ plays AUS}) \\ &\quad + \mathbb{P}(\text{NZ wins} \mid \text{plays IND})\mathbb{P}(\text{NZ plays IND}) \\ &= 0.4 \times 0.65 + 0.5 \times 0.35 = 0.435\end{aligned}$$

# Bayes' Theorem

The Bayesian approach to statistics rapidly growing in popularity

Example. The Signal and the Noise by Nate Silver

- Successful in forecasting complex events like elections
- Advocates a Bayesian approach to statistics / forecasting

To understand the Bayesian approach consider the saying

"When you hear hooves think horses not zebras"

Meaning: Assess new information through lens of prior knowledge

**Fact.** If $A, B$ are events with nonzero probability, then

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{1}$$

Proof: From

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{and} \quad \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(B|A)\,\mathbb{P}(A)$$

Rearranging yields (1)

Example. Banks use automated systems to try to detect fraudulent or illegal transactions

- A field of statistics called novelty detection

Consider a test that responds to each transaction with $P$ or $N$

- $P$ means "positive" — transaction flagged as fraudulent
- $N$ means "negative" — transaction flagged as normal

Letting $F$ mean fraudulent, we suppose that

- $\mathbb{P}(P \mid F) = 0.99$ – flags 99% of fraudulent transactions
- $\mathbb{P}(P \mid F^c) = 0.01$ – false positives
- $\mathbb{P}(F) = 0.001$ – prevalence of fraud

What is the probability of fraud given a positive test?

We use Bayes rule

$$\mathbb{P}(F \mid P) = \frac{\mathbb{P}(P \mid F)\mathbb{P}(F)}{\mathbb{P}(P)}$$

and the law of total probability

$$\mathbb{P}(P) = \mathbb{P}(P \mid F)\mathbb{P}(F) + \mathbb{P}(P \mid F^c)\mathbb{P}(F^c)$$

to get

$$\mathbb{P}(F \mid P) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = \frac{11}{122} \approx \frac{1}{11}$$

Less than one in ten positives are actually fraudulent