Econ 2110, fall 2016, Part Ia
Review of Probability Theory

Maximilian Kasy

Department of Economics, Harvard University

Textbooks

Main reference for part I of class:

- **Casella, G. and Berger, R. (2001).** *Statistical inference*. Duxbury Press, chapters 1-4.

Alternative references

- Advanced undergrad text; many exercises:
  J. Blitzstein and Hwang J (2014). *Introduction to Probability*. Chapman & Hall
- More advanced / mathematical than this class:
  P. Billingsley (2012). *Probability and Measure*. Wiley

Roadmap

- Ia
  - Basic definitions
  - Conditional probability and independence
- Ib
  - Random Variables
  - Expectations
  - Transformation of variables
- Ic
  - Selected probability distributions
  - Inequalities

Part Ia

Practice problem

What is a “probability?”

Alternative approaches

1. A population share
2. A subjective assessment
3. An abstraction for coherent decision making under uncertainty
4. A mathematical object, mapping subsets of some set into $[0, 1]$

1. A population share:
    - "**frequentist**" perspective
    - actual population (students in this class) or more often hypothetical population (infinitely repeated throws of a coin)
    - useful for intuition – probabilities behave like population shares
    - no probabilities for one-time events ("is there life on Mars?")
    - "weaker" notion than the following two
2. A subjective assessment
    - Subjective "**Bayesian**" perspective
    - "psychological" entity
    - one-time events have probabilities

3. An abstraction for coherent decision making under uncertainty
    - **decision theoretic** perspective – part III of this class!
    - one-time events (states of the world) are assigned “probabilities” for the purpose of decision making
    - formally equivalent to subjective perspective, different interpretation and purpose
4. A mathematical object, mapping subsets of some set into $[0, 1]$
    - **purely formal** perspective
    - axioms satisfied by the mapping justifiable by corresponding properties of population shares
    - perspective we will take in part I of class

Key definitions

Sample Space Ω

- Set of all possible outcomes, not necessarily numerical.
- Specific outcomes denoted ω.
- Examples:
  - survey 10 people on their employment status; outcome: number of unemployed among the surveyed
    $\Omega = \{0, 1, 2, \ldots, 10\}$
  - ask a random person about her income
    $\Omega = \mathbb{R}^+$

Events

- Subsets of Ω, typically denoted with capital letters, such as A
- Examples:
  - survey: more than 30% of interviewees are unemployed
    $A = \{4, 5, 6 \ldots, 10\}$
  - income: person earns between 30.000$ and 40.000$ per year
    $A = [30.000, 40.000]$

σ-Algebra (or σ-field)

- Let $\mathscr{F}$ be a set of subsets of Ω (i.e. $\mathscr{F}$ is a set of events). $\mathscr{F}$ is a σ-algebra if and only if
    1. $\Omega \in \mathscr{F}$
    2. if $A \in \mathscr{F}$, then $A^c \in \mathscr{F}$ (A^c is the complement of A, i.e. $A^c = \Omega \backslash A$)
    3. if $A_1, A_2, \ldots \in \mathscr{F}$, then $\left(\bigcup_{j=1}^{\infty} A_j\right) \in \mathscr{F}$
- Property (2) is called 'closed under complements'
- property (3) is called 'closed under countable unions'.

Example

- σ-Algebras allow to model "information"
- health insurance example:
  consider individuals who are young or old and healthy or sick
- $\Omega = \{YH, YS, OH, OS\}$
- only age is public information
- insurer decisions can only condition on public information, that is on the σ-Algebra

$$\mathscr{F} = \{\emptyset, \{YH, YS\}, \{OH, OS\}, \Omega\}$$

- individual decisions can condition on the full σ-Algebra $\mathscr{F}'$ of all subsets of Ω

- Note that we can set $\varnothing = A_{k+1} = A_{k+2} = \ldots$, in which case $\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{k} A_j$.
- Recall de Morgan's Law $(A \cup B)^c = (A^c \cap B^c)$.
- $\Rightarrow (\bigcup_{j=1}^{\infty} A_j^c)^c = \bigcap_{j=1}^{\infty} A_j$,
  and σ-algebras are also closed under countable intersections.

Probability measure P

- A function that maps elements of the σ-algebra $\mathscr{F}$ (i.e. certain subsets of Ω) into real numbers: $P : \mathscr{F} \mapsto \mathbb{R}$ with the following properties
  1. $P(A) \geq 0$
  2. $P(\Omega) = 1$
  3. If $A_1, A_2, \ldots \in \mathscr{F}$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then $P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$
- Example: the probability of at most 1 person surveyed being unemployed

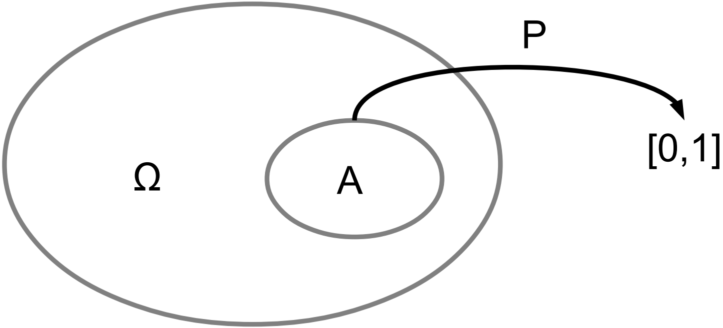
$$P(\{0,1\}) = (1-p)^{10} + 10 \cdot p \cdot (1-p)^9$$

where p is the unemployment rate

Probability space

- The triple $(\Omega, \mathscr{F}, P)$ is called a probability space.

Figure: probability space



Remarks

- The same random experiment can be described by different σ-algebras.
- all possible subsets of Ω are a σ-Algebra
- Why restrict the domain of P to a σ-algebra? Why not define P to map all possible subsets of Ω to $[0, 1]$?
  - Fine for experiments with finite or countably many outcomes
  - Fairly complicated problems arise for sample spaces with uncountably many outcomes. We will basically ignore them.
  - Pretty much all things of interest to us are "measurable," that is in suitably defined σ-algebras.

Some useful properties

1. $P(A) = 1 - P(A^c)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. $P(A \cup B) \geq P(A)$

Practice problem

Show that these properties hold, based on our definition of a probability space.

Conditional probability

- Let A, B be events in $(\Omega, \mathscr{F}, P)$, with $P(B) > 0$.
- The conditional probability of A, given B, is defined as

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

- Conditional probabilities can be understood as generating a new probability measure P', where $P'(A) = \frac{P(A \cap B)}{P(B)}$.
- Insurance example: probability of being healthy conditional on being old

$$P(H|O) = \frac{P(OH)}{P(\{OH, OS\})}$$

Practice problem

Show that P' is a probability measure.

Solution:

1. $P'(A) = \frac{P(A \cap B)}{P(B)} \geq 0$
2. $P'(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$
3.

$$\begin{aligned} P'(\bigcup_{j=1}^{\infty} A_j) &= P(B)^{-1} P((\bigcup_{j=1}^{\infty} A_j) \cap B) \\ &= P(B)^{-1} P(\bigcup_{j=1}^{\infty} (A_j \cap B)) \\ &= P(B)^{-1} \sum_{j=1}^{\infty} P(A_j \cap B) = \sum_{j=1}^{\infty} P'(A_j) \end{aligned}$$

- all properties of probability measures carry over to conditional probabilities
  e.g. $P(A \cup B|C) \geq P(A|C)$
  and $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
- frequentist intuition:
  probability is a population share among everyone in Ω
  conditional probability is a population share among everyone in B
- multiplication rule:

$$P(A \cap B) = P(A|B)P(B)$$

Bayes' Rule

- Suppose we know $P(B)$, $P(A|B)$ and $P(A|B^c)$, but we are interested in $P(B|A)$.
- Claim:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Practice problem

Show this is true.

Solution:

- Apply the definition of conditional probability repeatedly
  1. $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$
  2. numerator:
  $$P(A \cap B) = P(A|B)P(B)$$
  3. denominator:
  $$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B^c)) \\ &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \end{aligned}$$

Example

- Suppose 1 in 10,000 people have a certain virus infection
- A medical test has the following properties
  - If somebody is actually infected, the test yields a "positive" result with a probability of 99%
  - If somebody is not infected, the test yields a "positive" result with a probability of 5%

Practice problem

If someone is tested positive, what is the probability that she is actually infected?

Solution:

- Denote T the event of a positive test result, D the event of being infected with the disease.
- $$\begin{aligned} P(D|T) &= \frac{P(D,T)}{P(T)} \\ &= \frac{P(T|D)P(D)}{P(T,D)+P(T,D^c)} \\ &= \frac{P(T|D)P(D)}{P(T|D)P(D)+P(T|D^c)P(D^c)} \\ &= \frac{.99 \cdot .0001}{.99 \cdot .0001 + .05 \cdot .9999} \approx .002. \end{aligned}$$
- the test seems very good (correct result at least 95% of the time)
- but the probability of actually having the disease once you test positive is still very small (.002)

Example

Practice problem

Survey 2 random people
What is the probability of both being female given that at least one is female?

Solution:

- $E_1 = \{FF, FM, MF\}$, with probability $3/4$, $E_2 = \{FF\}$
- so $E_1 \cap E_2 = \{FF\}$ with probability $1/4$,
- therefore
$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{1/4}{3/4} = \frac{1}{3},$$
- (not $1/2$ as many people think at first.)

Independence

- The events A and B are independent if

$$P(A \cap B) = P(A)P(B).$$

- Claim:
  - If $P(A) = 0$ or $P(B) = 0$, then A and B are independent.
  - If $P(B) > 0$, then independence of A and B implies that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A).$$

  - If A and B are independent, then so are A^c and B, A^c and B^c, and A and B^c

Practice problem

Verify these claims.

Joint independence

- Three events E_1, E_2 and E_3 are jointly independent if :
  1. 1.1 E_1 and E_2 are independent,
     1.2 E_1 and E_3 are independent,
     1.3 E_2 and E_3 are independent.
  2. $$P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2) \cdot P(E_3).$$
- Joint independence of four events:
  1. all combinations of three events are jointly independent
  2. the probability of the intersection is equal to the product of the probabilities.
- etc.

Practice problem

Construct an example of three events which are pairwise independent but not jointly independent.

Example - unbreakable cryptography

- Suppose you want to transmit a binary message ($X = 0$ or $X = 1$)
- Take a random number $Y \in \{0, 1\}$ ("fair coin toss") which you shared with your recipient beforehand
- transmit the encrypted message $Z = 1$ if $X = Y$ and $Z = 0$ if $X \neq Y$

Practice problem

Verify that

- the events $\{X = 1\}$, $\{Y = 1\}$, and $\{Z = 1\}$ are pairwise independent but not mutually independent
- in particular $P(X = 1|Z = 1) = P(X = 1)$ ("the NSA won't learn anything about X if they intercept your Z")
- but your recipient can easily decode the message.

Conditional Independence

- events A and B are conditionally independent given $\{C, C^c\}$ if

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$
$$P(A \cap B|C^c) = P(A|C^c) \cdot P(B|C^c)$$

- important in part II of class (causality), regression with controls,...
- conditional independence does not imply independence
- independence does not imply conditional independence

Example

- conditional probabilities given $\{C, C^c\}$:

	$A \cap B$	$A \cap B^c$	$A^c \cap B$	$A^c \cap B^c$
$P(.\|C)$	4/9	2/9	2/9	1/9
$P(.\|C^c)$	1/9	2/9	2/9	4/9

- $P(C) = 1/2$
- here A and B are conditionally independent but not independent
- verify!
- intuition: C makes both A and B more likely, but otherwise there is no connection between A and B

Example

- conditional probabilities given $\{C, C^c\}$:

	$A \cap B$	$A \cap B^c$	$A^c \cap B$	$A^c \cap B^c$
$P(.\|C)$	1/3	1/3	1/3	0
$P(.\|C^c)$	0	0	0	1

- $P(C) = 3/4$
- here A and B are independent but not conditionally independent
- verify!
- in this example: C holds if A or B holds
  for instance: getting into some school (C) requires that you fulfill at least criterion A or B