

MIT OpenCourseWare
<http://ocw.mit.edu>

14.30 Introduction to Statistical Methods in Economics
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

14.30 Introduction to Statistical Methods in Economics

Lecture Notes 1

Konrad Menzel

February 3, 2009

1 Introduction and Overview

This class will give you an introduction to Probability Theory and the main tools of Statistics. Probability is a mathematical formalism to describe and analyze situations in which we do not have perfect knowledge of all relevant factors. In modern life, we are all routine consumers of statistical studies in fields ranging from medicine to sociology, and probabilistic reasoning is crucial to follow most of the recent debates in economics and finance.

In the first half of this class, we'll talk about probabilities as a way of describing genuine risk - or our subjective lack of information - over events.

Example 1 *In subprime lending, banks offered mortgages to borrowers who were much less likely to repay than their usual clientele. In order to manage the risks involved in lending to prospective home-owners who do not own much that could serve as collateral, thousands of these loans were bundled and resold as "mortgage backed securities," i.e. the bank which made the original loans promised to pay the holder of that paper whatever repayment it received on the loans. Eventually, there were more complicated financing schemes under which the pool was divided into several "tranches", where a first tranche was served first, i.e. if the tranche had a nominal value of, say, 10 million dollars, anyone holding a corresponding claim got repaid whenever the total of repayments in the pool surpassed 10 million dollars. The lower tranches were paid out according to whatever money was left after serving the high-priority claims. How could it be that the first "tranche" from a pool with many very risky loans was considered to be "safe" when each of the underlying mortgages was not? The low-priority tranches were considered riskier - why? And why did in the end even the "safe" securities turn out to be much riskier in retrospect than what everyone in the market anticipated? We'll get back to this when we talk about the Law of Large Numbers, and under which conditions it works, and when it doesn't.*

Usually in order to answer this type of question, you'll have to know a lot about the distribution (i.e. the relative likelihood) of outcomes, but in some cases you'll actually get by with much less: in some cases you are only concerned with "typical" values of the outcome, like expectations or other moments of a distribution. In other cases you may only be interested in an average over many repetitions of a random experiment, and in this situation the law of large numbers and the central limit theorem can sometimes give you good approximations without having to know much about the likelihood of different outcomes in each individual experiment.

The second half of the class deals with the question how we can learn about populations and probability distributions from data. In any empirical science you'll come across the problem of *inductive reasoning*, which means drawing general conclusions from a few (or even very many) observed special cases. In political polls ("who would you vote for in the next presidential election?"), a survey institute typically

polls at most a few thousand out of over a hundred potential voters. In medical trials, we try to draw conclusions about the effectiveness of a drug for an entire population from only a few dozen study participants.

If we only observe a subset of individuals (e.g. a random sample of voters in an exit poll) from a population of interest (e.g. all voters who turned out for a primary election), there will be some uncertainty over whether this sample was actually *representative* for the whole population with respect to the question we're after (e.g. the vote share for a particular candidate). Formalizing this for practical use makes heavy use of probability theory.

Example 2 *The controversial first Lancet study on civilian casualties in the Iraq war surveyed a random sample of 1849 households (which had in total 12,801 members) across the country three and a half years after the invasion in March 2003, and extrapolated the number of deaths reported by the surveyed household to the total population of the whole country, which is about 29 million. The authors of the study arrived at an estimate of about 112,000 "excess" deaths for the first 18 months after the invasion and stated that "with 95% confidence", the actual number was between 69,000 and 155,000. We will see later on in the course what exactly this statement means. The width of the confidence interval around the estimate gives a measure of the uncertainty inherent in extrapolating from a small sub-population to the entire country. Since this is a politically and emotionally charged subject, the study triggered an intense debate in scientific publications and the blogosphere - reading up on the debate will teach you a lot about how statistics is actually "done" in practice.*

2 Set Theory and Events

2.1 Random Experiments

Definition 1 *A random experiment is any procedure which can - at least theoretically - (a) be repeated arbitrarily often and under identical conditions, and (b) has a well-defined set of possible outcomes.*

A standard example for this would be a coin flip which can produce two different outcomes, heads H or tails T (we'll neglect the possibility that the coin ends up standing upright on its edge). Another important type of experiments in the realm of statistics would be an exit poll after an election: say we ask 2,000 randomly selected voters as they exit the poll stations to report which candidate they voted for. We could in principle redraw arbitrarily many further samples of 2,000 individuals from the total population that turned out to vote on election day, so condition (a) is satisfied. The set of outcomes for this experiment is the respective number of interviewed persons which reports to have voted for each candidate on the ballot.

Definition 2 *The sample space S is the collection of all possible outcomes of an experiment.*

For many purposes, we are not primarily interested in single outcomes, but instead group collections of outcomes together into *events*. Therefore we will in the following describe the experiment in terms of *sets*.

Definition 3 *An event A can be any collection of outcomes (this includes individual outcomes, the empty set, or the entire sample space).*

If the realized outcome is a member of the event A , then A is said to occur.

Let's look at last year's presidential race as an example. At the most elementary level, we could describe the sample space S as the collection of individuals who may - as a logical possibility - be elected

as president in November, i.e. as a first try we could look at the main candidates at the early stages of the primaries:

$$S = \{\text{Clinton, Huckabee, McCain, Obama, Paul, Romney, Schwarzenegger}\}$$

Is this a good description? - probably not: even though these are the most likely outcomes, we can't logically rule out that some other candidate (from either party or independent) enters the race later on, so a more fool-proof definition of the random experiment would augment the sample space

$$\tilde{S} = \{\text{Clinton, Huckabee, McCain, Obama, Paul, Romney, Schwarzenegger, other Democrat, other Republican, other Independent}\}$$

but to keep things simple, let's just ignore this possibility for now. Some events of interest could be e.g.

$$\text{"The 44th president of the US will be Republican"} = \{\text{Huckabee, McCain, Paul, Romney, Schwarzenegger, other Republican}\}$$

$$\text{"The 44th president is married to the 42nd president"} = \{\text{Clinton}\}$$

2.2 More about Sets and Events

2.2.1 Set Inclusion " \subset "

The event B is *contained* in A if every outcome in B also belongs to A , or in symbols

$$B \subset A \text{ if } (s \in B \implies s \in A)$$

Clearly, any event C is contained in the sample space S , i.e.

$$C \subset S \text{ for any event } C$$

and every event includes the impossible event

$$\emptyset \subset C \text{ for any event } C$$

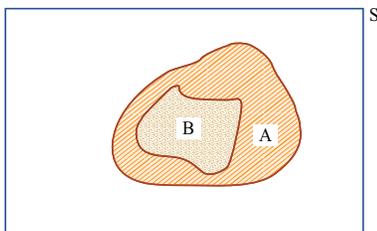


Image by MIT OpenCourseWare.

Figure 1: $B \subset A$ - " B implies A "

If A and B contain each other, they are equal,

$$A \subset B \text{ and } B \subset A \implies A = B$$

and set inclusion is transitive, i.e.

$$A \subset B \text{ and } B \subset C \implies A \subset C$$

In our presidential elections example, since (if we can trust Wikipedia, McCain was born on an American air base in the Panama Canal zone)

$$\begin{aligned} \text{"The 44th President was born in Panama"} &= \{\text{McCain}\} \\ &\subset \text{"The 44th President was born abroad"} = \{\text{McCain, Schwarzenegger}\} \end{aligned}$$

and

$$\{\text{McCain, Schwarzenegger}\} \subset \text{"The 44th President is a Republican"}$$

we can conclude that

$$\text{"The 44th President was born in Panama"} \subset \text{"The 44th President is a Republican"}$$

2.2.2 Unions of Sets, "∪"

The *union* of A and B is the collection of all outcomes that are members of A or B (or both, this is the logical, inclusive "or" corresponding to the symbol \vee). In symbols

$$A \cup B = \{s \in S \mid s \in A \vee s \in B\}$$

The set union is symmetric:

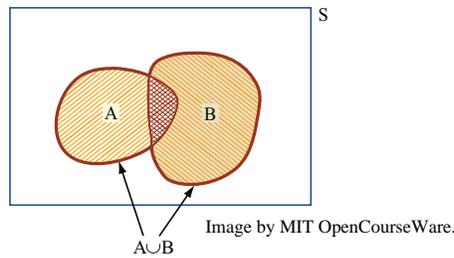


Figure 2: Union of A and B - " A or B "

$$A \cup B = B \cup A$$

Furthermore,

$$B \subset A \implies A \cup B = A \text{ for any events } A, B \subset S$$

In particular,

$$\begin{aligned} A \cup \emptyset &= A \\ A \cup A &= A \\ A \cup S &= S \end{aligned}$$

It also doesn't matter in which order we take union of sets/events (associative property):

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

2.2.3 Intersections of Sets, "∩"

The *intersection* of A and B is the (possibly empty) collection of outcomes that are members of both A and B , written as

$$A \cap B = \{s \in S \mid s \in A \wedge s \in B\}$$

where " \wedge " denotes the logical "and". Some texts use the alternative notation $A \cap B = AB$. As the set

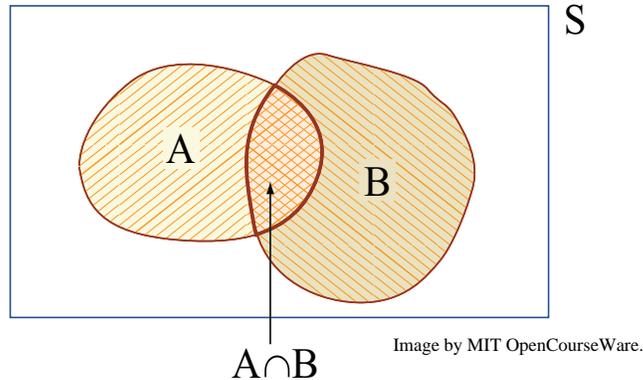


Figure 3: Intersection of A and B - " A and B "

union, the intersection is symmetric,

$$A \cap B = B \cap A$$

Also

$$B \subset A \implies A \cap B = B \text{ for any events } A, B \subset S$$

From this it follows that

$$\begin{aligned} A \cap \emptyset &= \emptyset \\ A \cap A &= A \\ A \cap S &= A \end{aligned}$$

Again, like the set union, the intersection of sets has the associative property

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

In addition, set intersection and union have the *distributive* properties

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

and

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

As an example, for the events

- A = "President 44 is a Woman" = {Clinton}
- B = "President 44 was born in Midwest" = {Clinton, Romney}
- C = "President 44 is a Republican" = {Huckabee, McCain, Paul, Romney, Schwarzenegger}

we can see that

$$\begin{aligned}A \cap (B \cup C) &= \{\text{Clinton}\} \cap \{\text{Clinton}, \text{Huckabee}, \text{McCain}, \text{Paul}, \text{Romney}, \text{Schwarzenegger}\} \\ &= \{\text{Clinton}\} \\ (A \cap B) \cup (A \cap C) &= \{\text{Clinton}\} \cup \emptyset = \{\text{Clinton}\}\end{aligned}$$

as it should be according to the first distributive property.

2.2.4 Set Complement, A^c

The *complement* A^c of A is the set of outcomes in S which do not belong to A , i.e.

$$A^c = \{s \in S \mid s \notin A\} = S \setminus A$$

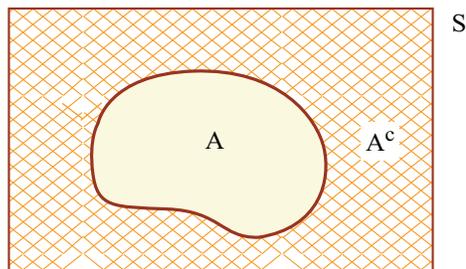


Image by MIT OpenCourseWare.

Figure 4: Complement to A - "not A "

From the definition, you can easily check that complements have the following properties

$$\begin{aligned}(A^c)^c &= A \\ A \cup A^c &= S \\ A \cap A^c &= \emptyset\end{aligned}$$

From the last statement, it follows that

$$S^c = \emptyset$$

and together with the first property, this implies

$$\emptyset^c = S$$

One useful set of relationships between intersections and unions is the following

$$\begin{aligned}(A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c\end{aligned}$$

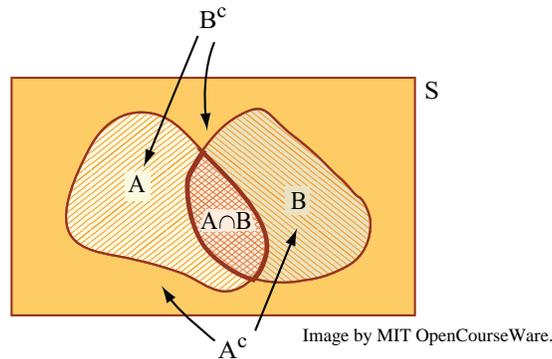


Figure 5: Illustration of $(A \cap B)^C = A^C \cup B^C$ - you can see the rule $(A \cup B)^C = A^C \cap B^C$ from the same graph.

2.2.5 Partitions of Events

A and B are *disjoint* (or *mutually exclusive*) if they have no outcomes in common, i.e.

$$A \cap B = \emptyset$$

A collection of events A_1, A_2, \dots is said to be *exhaustive* if their union equals S , i.e.

$$\bigcup_{i \geq 1} A_i = A_1 \cup A_2 \cup \dots = S$$

A collection of events A_1, A_2, \dots is called a *partition* of the sample space if (1) any two *distinct* events A_i, A_j (with $i \neq j$), A_i and A_j are disjoint $A_i \cap A_j = \emptyset$, and (2) the collection A_1, A_2, \dots is exhaustive. In a similar fashion we can define *partitions of an event B* as collections of mutually exclusive subevents whose union equals B .

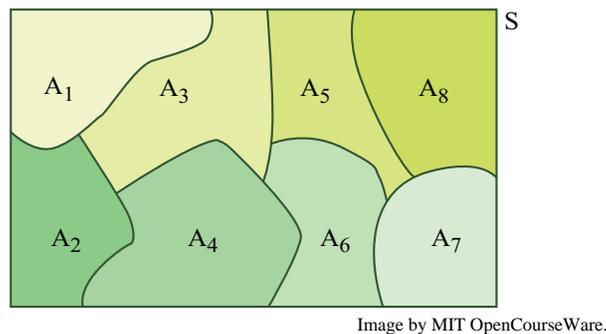


Figure 6: Partition of S into A_1, A_2, \dots, A_8