STA257

Neil Montgomery | HTML is official | PDF versions good for in-class use only
Last edited: 2016-09-12 14:49

admin

contact, websites, notes

date format	ISO8601
instructor	Neil Montgomery
email	neilm@mie.utoronto.ca
office	BA8137
office hours	any time
website	portal (announcements, grades, etc.)
github	https://github.com/sta257-fall-2016 (lecture material, code, etc.)

Official lecture notes are HTML.

PDFs will be uploaded before classes for people who like to annotate them during lecture on a PED. **But they will never be updated after class.**

evaluation, book, tutorials

what	when	how much
midterm 1	2016-10-03 (OOPS!) during class time	25%
midterm 2	2016-11-14 during class time	25%
exam	TBA	50%

The book is *Mathematical Statistics and Data Analysis, 3rd Edition* by Rice (also used in STA261).

I will suggest exercises from this book each week.

Your TA will work through some of them in tutorial each week.

other resources

Other (free, downloadable PDF through library) similar books:

- *Introduction to Probability with Statistical Applications* by Schay (similar level)
- *An Intermediate Course in Probability* by Gut (more advanced)
- *Introduction to probability models* by S. Ross (a bit more engineer-y)

Avoid books with titles like *Introduction to Probability and Statistics For <Some Certain Specific Type of Student or Career>*

A very nice (free PDF through library) book for those really interested in the mathematics of all this is: *Elementary Analysis* by K. Ross. (Subtitled "The Theory of Calculus".)

Wikipedia, youtube, stackexchange, google. Lots of good stuff. Lots of bad stuff.

meanings of probability

no need to write this down

- Physical, long term relative frequency, "repeated experiments", "frequentist", "propensity"
- Evidential, subjective, "Bayesian", inductive

Visit Department of Philosophy for more information.

Whatever the interpretation, the mathematical rules are the same, based on axioms that define how probabilities can be assiged to events.

axiomatic approach to probability

elements and sets

We can think of an element ω belonging to a set A. We can think of sets A and B along with a universal set S. We have the following notions, and more:

- Membership $\omega \in A$
- Union "or" $A \cup B$; Intersection "and" $A \cap B$; works for infinitely many
- Complement $A^c = \{w \in S : w \notin A\}$
- Empty set has no elements: $\emptyset$
- Disjointness: $A \cap B = \emptyset$ (notice: not a probability concept)
- Subset: $A \subseteq B$ (and "proper subset")
- Set difference: $A \backslash B = A \cap B^c$

sample space

Probability starts with a sample space S, a collection with all possible outcomes of the random process. Often cumbersome and arbitrary; mainly used this week. Examples:

- Coin toss: $\{H, T\}$
- Picking a card: $\{A\spadesuit, A\heartsuit, A\diamondsuit, A\spadesuit\}$
- Toss two coins: $\{HH, HT, TH, TT\}$. Or possibly: $\{\{H, H\}, \{H, T\}, \{T, T\}\}$
- A race among 8 horses: ?
- Toss a coin until a head appears: $\{T, TH, THH, THHH, \ldots\}$
- Pick a real number between 0 and 1 "uniformly": $(0, 1)$ (A "continuous" sample space.)

event

An event is a collection of outcomes; equivalently a subset of the sample space S.

Naming conventions: Roman letter near the beginning $A, B, C, \ldots$ or $A_1, \ldots, A_n$ or $A_1, A_2, A_3, \ldots$ as required.

Many examples possible from the example sample spaces.

it's really more complicated than that

This is an elementary course, so we will not concern ourselves with the fact that not *all* subsets of a sample space are allowed to be called "events".

Really an event has to be a "suitable" collection of outcomes.

For finite and countable (i.e. "list-able") sample spaces, in fact *all* events are "suitable".

But not for uncountable sample spaces, such as $(0, 1)$.

The "space" of suitable events can be called $\mathcal{A}$.

the probability axioms

A *probability measure* is a real-valued function (usually called) P. Its domain is $\mathcal{A}$, a space of suitable events in S. In addition, it has the following properties:

1. $P(S) = 1$
2. $P(A) \geq 0$ for all *events* $A \in \mathcal{A}$.
3. If $A_1, A_2, A_3, \ldots$ is a *disjoint* collection of events, then:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

The last property is called "$\sigma-$additivity".

There is a notion of "probability triple" $(S, \mathcal{A}, P)$ when needed to be absolutly clear (which isn't often in this course.)

the axiomatic approach

In the fussiest possible treatment, the first question is: are these axioms *consistent*, which is the same as asking "Are there any probability measures at all?"

Theorem 0: the axioms are consistent.

Proof: $\ldots$

Advanced note…when the sample space is something like $S = (0, 1)$ and if we were to allow $\mathcal{A}$ to be the collection of *all* subsets of S, then the axioms are *inconsistent*.

everyday properties of P

Continuing with total and absolute fussiness:

Theorem 1: $P(\emptyset) = 0$

Proof: $\ldots$

Theorem 2: If A_1 and A_2 are disjoint then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Proof: $\ldots$

Theorem 2a: If $A_1, A_2, \ldots, A_n$ are disjoint then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ (called "finite additivity")

Proof: "Induction" (Note: the book lists this Theorem as an "axiom", which is not technically wrong but…)

more everyday properties of P, with proofs

Theorem 3: $P(A^c) = 1 - P(A)$

Theorem 4: If $A \subseteq B$ then $P(A) \le P(B)$. "P is monotone."

Theorem 4a: $P(B \backslash A) = P(B) - P(B \cap A^c)$

Theorem 4b: If $A \subseteq B$ then $P(B \backslash A) = P(B) - P(A)$.

Theorem 5: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (and generalizations)

This is admittedly getting dull.

a non-everyday application of this axiomatic approach to P

towards showing the "continuity" of P

The culmination of our axiomatic approach will be to define the notion of "continuity" for P and prove that the defined property holds.

Recall from the prerequiste the notion of a *continuous* function. There are several equivalent definitions, one of which uses left- and right-continuity.

A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is left-continuous at x if for any non-decreasing sequence $x_1 \leq x_2 \leq x_3 \leq \ldots$ that converges to x, then $f(x_i) \rightarrow f(x)$.

(Right continuity is the same but with a non-increasing sequence.)

f is continuous at x if it is left- and right-continuous at x.

f is continuous if it is continuous at every point in its domain.

increasing sequences of events that converge

The domain of P is a collection of events $\mathcal{A}$. We need a notion of the following for events:

$$A_1, A_2, \ldots \text{ increases to } A$$

Definition: $A_n \nearrow A$ means $A_i \subseteq A_{i+1}$ and $\bigcup_{i=1}^{\infty} A_i = A$

Example: Consider $S = (0, 1)$. Let $A_n = \left(0, \frac{1}{2} - \frac{1}{2^{n+1}}\right)$ for $n \geq 1$ and $A = \left(0, \frac{1}{2}\right)$

What about the probabilities of these events under the uniform model?

the continuity theorem

Theorem 6 (The Continuity Theorem): If A_n and A are events in $\mathcal{A}$ and $A_n \nearrow A$, then $P(A_n) \longrightarrow P(A)$.

Proof: $\ldots$

This is analogous to left-continuity. There is also a right-continuity:

Corallary: Suppose A_n and A are events in $\mathcal{A}$ with $A_i \supseteq A_{i+1}$ and $\bigcap_{i=1}^{\infty} A_i = A$. Then $P(A_n) \longrightarrow P(A)$.

Proof: The Continuity Theorem, a de Morgan's Law, and "Theorem 3".

Something to try if you like: finite additivity together with The Continuity Theorem implies $\sigma-$additivity.

application to the continuous sample space example

Reconsider the uniform pick on $S = (0, 1)$, where the probability of choosing a number in any $0 < a < b < 1$ is $b - a$.

What is the probability of choosing exactly $\frac{1}{2}$?

Let A be the event that the number chosen is *rational*. What is P(A)?

some computations for finite and countable sample spaces

finite and countable S in general

This is hard to prove with elementary methods, but suppose:

$$S = \{\omega_1, \ldots, \omega_n\} \qquad \text{(finite), or,}$$
$$S = \{\omega_1, \omega_2, \omega_3, \ldots\} \qquad \text{(countable)}$$

then a valid probability can always be based on $P(\{\omega_i\}) = p_i$ with $0 \leq p_i \leq 1$ and $\sum p_i = 1$.

An important special case for finite S is the uniform case: $p_i = \frac{1}{n}$.

In this case many problems can be solved by counting the number of outcomes in S and counting the number of outcomes in an event.

Some people enjoy these problems. Others don't. Fortunately for you, I do not!

permutations and combinations

At the very least we'll need to recall (or learn!) these.

Number of ways to choose k items out of n where order matters:

$$ {}_nP_k = \begin{cases} 0 & \text{if k > n,} \\ \frac{n!}{(n-k)!} & \text{otherwise.} \end{cases} $$

and when order doesn't matter:

$$ {}_nC_r = \binom{n}{k} = \frac{n!}{k!(n-k)!} $$

Two classic examples: "The Birthday Problem" and "Lotto"

conditional probability

partial information

I'll role a six-sided die. $S = \{1, 2, 3, 4, 5, 6\}$. Consider these events:

$$\begin{aligned} A &= \{2, 5\}, \\ B &= \{2, 4, 6\}, \\ C &= \{1, 2\}. \end{aligned}$$

So $P(A) = \frac{2}{6} = \frac{1}{3}$.

What if I peek and tell you "Actually, B occurred". What is the probabality of A given this partial information? It is $\frac{1}{3}$.

I roll the die again, peek, and tell you "Actually, C occurred". Now the probability of A is $\frac{1}{2}$.

Intuitively we used a "sample space restriction" approach.

elementary definition of conditional probability

Given B with $P(B) > 0$,

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

"The conditional probability of A given B"

The answers for the previous example coincide with the intuitive approach.

Theorem 7: For a fixed B with $P(B) > 0$, the function $P_B(A) = P(A|B)$ is a probability measure.

Proof: exercise.

useful expressions for calculation - I

$P(A \cap B) = P(A|B)P(B)$ often comes in handy.

Consider the testing for, and prevalence of, a viral infection such as HIV.

Denote by A the event "tests positive for HIV", and by B the event "is HIV positive."

For the ELISA screening test, $P(A|B)$ is about 0.995. The prevalence of HIV in Canada is about $P(B) = 0.00212$.

useful expressions for calculation - II

Take some event B. The sample space can be divided in two into B and B^C.

This is an example of a *partition* of S, which is generally a collection $B_1, B_2, \dots$ of disjoint events such that $\bigcup_{i=1}^{\infty} B_i = S$.

Theorem 8: If $B_1, B_2, \dots$ is a partition of S with all $P(B_i) > 0$, then

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$$

Proof: …

Continuing with the HIV example, suppose we also know $P(A|B^c) = 0.005$ ("false positive").

We can now calculate $P(A)$.

useful expressions for calculation - III

Much to my amusement, Theorem 8 gets a grandiose title: ***"THE! LAW! OF! TOTAL! PROBABILITY!!!"***

Now, in the HIV example, we also might be interested in $P(B|A)$, the chance of an HIV+ person testing positive.

A little algebra:

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

In our example this is $\frac{0.0021094}{0.0070988} = 0.2971$.

Bayes' rule

Theorem 9: If $B_1, B_2, \ldots$ is a partition of S with all $P(B_i) > 0$, then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}$$