9/10: Sample Spaces, Events and Conditioning

Probabilistic Reasoning

- Continous probability spaces and measure theory are important, but only confuse basic probabilistic intuitions
- Main probabilistic thinking and results can be based on finite probability spaces
  - Can be made precise using some nonstandard analysis: Nelson, "Radically Elementary Probability," PUP 1988.
  - "I am sure that many other probabilists teaching a beginning graduate course have also had the feeling that these measure-theoretic foundations serve more to salve our mathematical consciences than to provide an incisive tool for the scientist who wishes to apply probability theory."

Probabilistic Reasoning

- Should only require ability to add, multiply and reason (but some more advanced computational tools are helpful...)
- However even probabilistic reasoning with simple finite event spaces can be subtle and confusing – beware!

Interpretations of Probability

- Subjective vs. Frequentist

For now, this distinction will not be important. We'll proceed with the standard axiomatic formulation that gives probability a solid foundation.

Experiments and Events

Probability can usually be thought of in terms of an *experiment*.

- In an experiment where a coin is tossed 100 times, what is the probability that at least 70 heads are obtained?
- In an experiment where the a Web log is monitored during a half-hour interval for 60 consecutive days, what is the probability that the average number of pages visited during the half-hour will be greater than 1,000?

- If collection of 1000 documents is hashed with a particular string hash function, what is the probability that there will be a collision of the hash values?
- In an experiment where a user is asked to select a query and a Web page that is relevant to it, what is the probability the document will have a URL containing a domain name `.edu`?
- (DeGroot) From information relating to the life of Thomas Jefferson, a person might want to determine the probability that Jefferson was born in the year 1741. (!)

Sometimes the interpretation of this probability may be subtle, or it may not even be clear how to set up an "experiment" interpretation. But one must be there...

Axioms of Probability

Sample space — possible outcomes of an experiment; $S = \{\omega_1, \omega_2, \omega_3, \ldots\}$ (finite or infinite)

Event — subset of possible outcomes

$\text{Pr}(A)$ — assignment of a number to each event

Axiom 1: $\text{Pr}(A) \geq 0$

Axiom 2: $\text{Pr}(S) = 1$

Axiom 3: For every sequence of disjoint events

$$\text{Pr}\left(\cup_i A_i\right) = \sum_i \text{Pr}(A_i)$$

Most Important Case

Sample space $S = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n\}$ (finite)

$$\begin{aligned} p_i &= \text{Pr}(\omega_i) \\ \sum_{i=1}^{n} p_i &= 1 \\ \text{Pr}(A) &= \sum_{\omega_i \in A} p_i \end{aligned}$$

Example

Experiment: flip a coin three times, and record outcomes

Sample space: $S = \{HHH, HHT, HTH, HTT, \ldots, TTT\}$

Example event: $A =$ event that at least two heads are obtained

Example

Suppose we flip a coin and roll a die. The sample space is

$$S = \{(H,1),(H,2),\ldots,(H,6),(T,1),(T,2),\ldots,(T,6)\}$$

and we could consider the event that a head is flipped on the coin and an even number rolled on the die:

$$A = \{(H,2),(H,4),(H,6)\}$$

Subsets have smaller probability

For two events A and B we use the shorthand $AB \equiv A \cap B$ and $B^c = \{x \in S : x \notin B\}$.

If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.

Proof:

$$\begin{aligned}\Pr(B) &= \Pr(A \cup BA^c) \\ &= \Pr(A) + \Pr(BA^c) \quad \text{(Axiom 3)} \\ &\geq \Pr(A) \quad \text{(Axiom 1)}\end{aligned}$$

Subtracting out the Intersection

$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$

Proof:

$$\begin{aligned}&\Pr(A \cup B) = \\ &\quad = \Pr(AB^c) + \Pr(AB) + \Pr(A^cB) \quad \text{(A3)} \\ &\quad = \Pr(A) - \Pr(AB) + \Pr(AB) + \Pr(B) - \Pr(AB) \quad \text{(A3)} \\ &\quad = \Pr(A) + \Pr(B) - \Pr(AB)\end{aligned}$$

Independence

Two events A and B are *independent* in case

$$\text{Pr}(AB) = \text{Pr}(A)\text{Pr}(B)$$

A set of events $\{A_i\}$ is independent in case

$$\text{Pr}(\cap_i A_i) = \prod_i \text{Pr}(A_i)$$

Example

Suppose that the event space is $S = \{1, 2, 3, 4, 5, 6\}$ with uniform probabilities, and $A = \{2, 4, 6\}$ and $B = \{1, 2, 3, 4\}$.

Are these independent events?

Example

Flip a coin 10 times, and *assume* independent. Let $A =$ "At least one head". Then

$$\begin{aligned} \text{Pr}(A) &= 1 - \text{Pr}(A^c) \\ &= 1 - \text{Pr}(\text{"All tails"}) \\ &= 1 - \left(\frac{1}{2}\right)^{10} \\ &\approx 0.999 \end{aligned}$$

Union Bound

For any events $A_1, A_2, \ldots, A_n$,

$$\text{Pr}\left(\cup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \text{Pr}(A_i)$$

Proof:

$$\begin{aligned} \text{Pr}(A \cup B) &= \text{Pr}(A \cup BA^c) \\ &= \text{Pr}(A) + \text{Pr}(BA^c) \\ &\leq \text{Pr}(A) + \text{Pr}(B) \end{aligned}$$

Now use induction on n.

Exercise: Show the same for $n = \infty$

Conditioning

If A and B are events with Pr$(A) > 0$, the *conditional probability of B given A* is

$$\text{Pr}(B \mid A) = \frac{\text{Pr}(AB)}{\text{Pr}(A)}$$

Conditioning

Conditioning obeys the probability axioms:

$\text{Pr}(B \mid A) \geq 0$

$\text{Pr}(S \mid A) = 1$

$\text{Pr}(\cup_i B_i \mid A) = \sum_i \text{Pr}(B_i \mid A)$ if B_i are disjoint

Conditioning

If A and B are independent

$$\begin{aligned} \text{Pr}(B \mid A) &= \frac{\text{Pr}(BA)}{A} \\ &= \text{Pr}(B) \end{aligned}$$

Conversely, since

$$\text{Pr}(AB) = \text{Pr}(A)\text{Pr}(B \mid A)$$

if $\text{Pr}(B \mid A) = \text{Pr}(B)$ then A and B are independent

Conditioning

Conditioning is easy to think of in terms of frequencies of events.

In a frequentist interpretation of the probabilities, if $N(AB)$ is the number of times events A and B both occur, and $N(A)$ is the number of times A occurs, in N trials, then

$$\frac{N(AB)}{N(A)} = \frac{N(AB)/N}{N(A)/N}$$

should be the conditional probability $\text{Pr}(B \mid A)$

Multinomial Coefficients

Suppose there are n items, organized into k types. For example n words in a documents, from a k word vocabulary.

The number of different ways the n items can be divided into k groups is the *multinomial coefficient*

$$\begin{aligned}\binom{n}{n_1\, n_2\, \cdots\, n_k} &= \binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n_{k-1}+n_k}{n_{k-1}} \\ &= \frac{n!}{n_1!\, n_2! \cdots\, n_k!}\end{aligned}$$

Multinomial Theorem

$$(x_1+x_2+\cdots+x_k)^n = \sum_{n_1,n_2,\ldots,n_k} \binom{n}{n_1\, n_2 \cdots n_k} x^{n_1} x^{n_2} \cdots x_k^{n_k}$$

Example

In 7 bridge hands, what is the probability that in at least one hand each player is dealt an ace?

Example

In 7 bridge hands, what is the probability that in at least one hand each player is dealt an ace?

Number of hands is

$$\frac{52!}{13!^4}$$

Number of hands with each player holding an ace is

$$4!\frac{48!}{12!^4}$$

Example

Probability of a 1-ace-each hand is

$$4!\frac{48!}{52!}13^4 \approx \frac{1}{10}$$

Assuming independence of the hands,

$$\begin{aligned}\Pr(A) &= 1 - \Pr(A^c) \\ &= 1 - \left(\frac{9}{10}\right)^7 \\ &\approx \frac{1}{2}\end{aligned}$$

To Do

- Read Chapter 1 in DeGroot
- Go over a Matlab tutorial
- Assignment 1 due next Tuesday
- No class on Thursday, 9/12