Introduction to Bayesian Inference

• What are we trying to do?
  • Make inferences about hypotheses based on all information at our disposal
  • See how new data affects our inferences
  • Need to identify all hypotheses (states of Nature) that may be true.
  • Need to know what observations each hypothesis predicts
  • Need to know how to compute the consequences
  • Our approach is Bayesian

Introduction to Bayesian Inference

• The mathematics we will be using is actually very simple. There is nothing hard or deep or difficult, no measure theory, nothing complex.
• The most difficult part of Bayesian inference is learning to think in a Bayesian manner. Once you “get it,” things become much easier
• People who have already learned some classical statistics experience may some difficulty in shifting to a Bayesian mode of thinking. Some things may have to be “unlearned”. However, once you “get it,” you’ll find it much easier to understand standard statistical ideas as well as the Bayesian ones. So it is an advantage for a classical statistician to understand Bayesian statistics.

Introduction to Bayesian Inference

• Finally, Bayesian inference is extraordinarily powerful. Since a very simple approach can be applied in many diverse and complicated situations, it is often the method of choice even for those whose background in statistics is more classical.

Introduction to Bayesian Inference

• Why consider Bayesian inference?
  • Conceptually simple
  • Logically consistent
  • Uniform approach
  • Powerful
  • Elegant
What do hypotheses predict about potential data? How does data support (or undermine) hypotheses?

• How new data support hypotheses

H1
H2
H3
D1
D2
D3

• What do we infer if we observe D1? D2? D3?

Introduction to Bayesian Inference

Deduction: Deduce outcomes from hypotheses

A → B
A
Therefore B

Induction (“statistical syllogism”): Infer hypotheses from outcomes

If A then we are likely to observe B and C
B and C are observed
Therefore A is supported more than E (as indicated by the width of the lines)

Observing D1 refutes H1, supports H2 a little and H3 strongly
Observing D2 supports H1 and H3 a little and H2 moderately.
Introduction to Bayesian Inference

- Statistical inference is not magic
  - Cannot get information that isn’t present in the data
  - Statistical inference is easily misused
  - Watch out for “Garbage in, Garbage out”
  - Often the right solution is a better experiment or better observations, not slick statistical procedures!

Bayesian Inference 1/13/09 9

Herman Rubin’s Five Commandments

- For the client:
  1. Thou shalt know that thou must make assumptions.
  2. Thou shalt not believe thy assumptions.

- For the consultant:
  3. Thou shalt not make thy client’s assumptions for him.
  4. Thou shalt inform thy client of the consequences of his assumptions.

- For the person who is both (e.g., a biostatistician or psychometrician):
  5. Thou shalt keep thy roles distinct, lest thou violate some of the other commandments.

Bayesian Inference 1/13/09 11

Probability

- What is the probability that a fair coin comes up heads?
  - Before tossing the coin

Bayesian Inference 1/13/09 12
What is the probability that a fair coin comes up heads?
- Before tossing the coin
- After tossing but before looking
- After the professor looks but before the professor says what he saw
- After the professor says what he saw
- After a student looks and reports what she saw
What is the probability that a fair coin comes up heads?
• Before tossing the coin
• After tossing but before looking
• After the professor looks but before the professor says what he saw
• After the professor says what he saw
• After a student looks and reports what she saw
• After you personally look

For a Bayesian, probability is conditioned on what each individual knows. It can vary from individual to individual.
• Probability is not “out there”. It is in your head.
  • “Probability does not exist” — Bruno deFinetti
• Probability is about epistemology, not ontology.

In the Bayesian view, probability describes Your degree of belief, given what you know.
• Betting odds allow us to elicit probabilities
• Can apply to hypotheses, not just events
• Probability is sometimes viewed as the frequency of occurrence after a finite number of repeated trials
  • We will call this frequency, not probability
• Probability is sometimes viewed as the frequency after a hypothetical infinite set of trials
  • This is the basis of “classical” or “frequentist” statistics
  • Bayesian statistics does not use this notion (but we will discuss it where appropriate)

Consider the following propositions:
• If I toss this coin, will it come up heads?
• What team will win the Superbowl next year?
• Is the Suez canal longer than the Panama canal?
• Will it rain tomorrow?
• Is the millionth digit of π the digit ‘3’?
• What will Microsoft stock sell for one year from now?
Probability

- Consider the following propositions:
  - If I toss this coin, will it come up heads?
  - What team will win the Superbowl next year?
  - Is the Suez canal longer than the Panama canal?
  - Will it rain tomorrow?
  - Is the millionth digit of π the digit ‘3’?
  - What will Microsoft stock sell for one year from now?
- Note that some of these propositions have perfectly definite answers: They involve Bayesian probabilities because their answers are unknown to us, not because they involve some “random” event.

The Conditional Nature of Probability

- Background information is information that is assumed but not always stated. Such information could include things like: How to do mathematics, basic knowledge that You may have about physics, genetics, etc., stuff You learned while growing up, data that You have gathered and already included in Your calculations, etc., etc. There is always background information, which we write I.
- For simplicity, we usually omit writing I explicitly; thus, for P(AlI) we will usually write just P(A).
- But, if You calculate probabilities in a way that does not take this background information I into account properly, this improper conditioning may generate apparent paradoxes in probability theory.

Probability

- Probability is always conditional, that is, the probability that You assign to something is always dependent on information that You have.
- So, we write P(AlC) to mean “the probability that A is true, given that C is true.”
- Someone else, who will usually have different information C from You, will usually, and legitimately, assign a different probability to a proposition than You will; thus, P(AlC) is generally not equal to P(AlC′).

Probability Axioms

- P(AlI) is a real number. P(AlI) satisfies the following axioms of probability:
  - 0 ≤ P(AlI) ≤ 1, and P(AlI)=1 means I → A
  - P(AlA, I) = 1
  - P(AlI) + P(A′l I) = 1 where A′ is the negation of A
  - P(AlB l I) = P(AlB, I)P(Bl I) product law, and definition of conditional probability
    » P(AlB, I) = P(AlB l I)P(Bl I) if P(Bl I)=0
  - If A and B are mutually exclusive propositions, then P(A ∨ Bl I) = P(Al I) + P(Bl I)
    [Derivable from the product law, hence not an independent axiom]
### Probability Axioms

- If we omit the background information $I$, we write for example $P(A)$, and have:
  - $0 \leq P(A) \leq 1$
  - $P(A|A) = 1$
  - $P(A) + P(A^c) = 1$
  - etc.

### Compound Probabilities

- If $P(A,B)=0$ then the proposition “$A$ and $B$” is false and we say that $A$ and $B$ are mutually exclusive.
- This axiom is therefore not independent of the others.

### Compound Probabilities

- We have:
  $$P(A \cup B) = P(\overline{(A^c \land B^c)}) = P((\overline{A^c}, \overline{B^c}))$$
  $$= 1 - P(A^c, B^c)$$
  $$= 1 - P(A^c)P(B^c | A^c)$$
  $$= 1 - P(A^c)[1 - P(B | A^c)]$$
  $$= P(A) + P(A^c, B)$$
  $$= ... = P(A) + P(B) - P(A,B)$$

- You should be able to fill in the missing steps! Hint: Start by using the product law on the term $P(A^c, B)$

### Compound Probabilities

- If $A_1, A_2, \ldots$ are mutually exclusive and exhaustive (i.e., they include all the possibilities) then
  $$\sum P(A_i) = P(A_1 \lor A_2 \lor \ldots \lor A_n)$$
  $$= P(\text{something happens}) = 1$$

- Also in this case, for any $B$
  $$P(B) = P(B \land (A_1 \lor A_2 \lor \ldots \lor A_n))$$
  $$= \sum P(B \land A_i) = \sum P(B | A_i)P(A_i)$$

- This last is a very important result because it will enable us to ignore irrelevant possibilities $A_1, A_2, \ldots$ by summing over them (a process known as marginalization)