Workshop Overview
Applied Probabilistic Reasoning: A Vade Mecum to Accompany a First Course in Statistics

Lawrence Hubert
The University of Illinois
email: lhubert@illinois.edu

Vade Mecum: a handbook or guide that is kept constantly at hand for consultation

I have never heard any of your lectures, but from what I can learn I should say that for people who like the kind of lectures you deliver, they are just the kind of lectures such people like – Artemus Ward (from a newspaper advertisement, 1863)

The Modules of the Vade Mecum

The Vade Mecum consists of twelve separate modules placed at

http://cda.psych.uiuc.edu/applied_probabilistic_reasoning

Module 1: A Brief Primer on Applied Probabilistic Reasoning

This Module 1 will be the source for the material we do today

When you registered for this workshop, you may have also received by return email a manuscript form of Module 1:

probabilistic_reasoning_chapter_one_apa.pdf

You may have also received a book that Howard Wainer and I wrote, entitled "A Statistical Guide for the Ethically Perplexed": sgep_published_version.pdf

If we have time today or tomorrow, I will review the materials for a course on ethics and statistics based on this book:

http://cda.psych.uiuc.edu/sgep_course_material

Module 2: The (Un)reliability of Clinical and Actuarial Predictions of Dangerous Behavior

Module 3: The Analysis of $2 \times 2 \times 2$ (Multiway) Contingency Tables: Explaining Simpson's Paradox and Demonstrating Racial Bias in the Imposition of the Death Penalty

Module 4: Probabilistic Reasoning and Diagnostic Testing

Modules 2, 3, and 4 will be the source for most of the material we do tomorrow

Module 5: Probabilistic Reasoning in the Service of Gambling

Module 6: Probabilistic Reasoning Through the Basic Sampling Model

Module 7: Probabilistic (Mis)Reasoning and Related Confusions

Module 8: Probabilistic Reasoning, Forensic Evidence, and the Relevance of Base Rates

Module 9: Probability and Litigation

Module 10: Sleuthing with Probability and Statistics

Module 11: Cross-validation and the Control of Error Rates

Module 12: An Olio of Topics in Applied Probabilistic Reasoning

Part I Topic Areas: The Basics of Probabilistic Reasoning

The topic of probabilistic reasoning will be approached through the language of "events" either occurring or not occurring;

And how the occurrence of one event might change the likelihood (or probability) of another event occurring.

This idea will be phrased as one event being *facilitative* or *inhibitive* of another event.

To make all of this concrete, the specific terminology will be introduced through three examples.

1) Example 1: Colorectal Cancer Screening

2) Example 2: A Risque Legend of Cinderella

3) Example 3: The O. J. Simpson Murder Trial

Part I Topic Areas Continued

4) The Sally Clark Case: Improper Use of Statistical Independence

5) Statistical Syllogisms

6) The Charles Peirce Notion of Abductive Reasoning

7) "If p, then q" Statements

8) Genetic Probabilistic Reasoning

9) Summary: Arguing Probabilistic Causation

Example 1: Colorectal Cancer Screening

The data are from Gerd Gigerenzer, *Calculated Risks* (in the form of what is called a 2×2 contingency table):

	$+$CC: A	$-$CC: $\bar{A}$	Row Sums
$+$FOBT: B	15	299	314
$-$FOBT: $\bar{B}$	15	9671	9686
Column Sums	30	9970	10,000

These data are on putative group of 10,000 individuals cross-classified as to whether a Fecal Occult Blood Test (FOBT) is positive [B: $+$FOBT] or negative [$\bar{B}$: $-$FOBT], and the presence of Colorectal Cancer [A: $+$CC] or its absence [$\bar{A}$: $-$CC].

The Language of Events

Suppose we pick at random one person from the pool of 10,000, and note whether this person is +CC or −CC, and also whether this person is +FOBT or −FOBT.

Or, to use alternative "event" terminology as labels for the rows and columns, whether for this person the event A or $\bar{A}$ occurs, and whether for this person the event B or $\bar{B}$ occurs.

Note that the symbol $\bar{A}$ is read either as "not A", or as "A bar" (similarly for $\bar{B}$)

Example 1: Moving to Probabilities

In carrying out the random picking process it might be useful to rephrase the outcomes observed in terms of "probabilities":

	+CC: A	−CC: $\bar{A}$	Row Sums
+FOBT: B	$P(A \cap B) \equiv \frac{15}{10{,}000}$	$P(\bar{A} \cap B) \equiv \frac{299}{10{,}000}$	$P(B) \equiv \frac{314}{10{,}000}$
−FOBT: $\bar{B}$	$P(A \cap \bar{B}) \equiv \frac{15}{10{,}000}$	$P(\bar{A} \cap \bar{B}) \equiv \frac{9671}{10{,}000}$	$P(\bar{B}) \equiv \frac{9686}{10{,}000}$
Column Sums	$P(A) \equiv \frac{30}{10{,}000}$	$P(\bar{A}) \equiv \frac{9970}{10{,}000}$	

The symbol "$\cap$" refers to the word "and"; $A \cap B$ is a "joint" event (as are $\bar{A} \cap B$, $A \cap \bar{B}$, and $\bar{A} \cap \bar{B}$); "$\equiv$" stands for "defined as"

Also, note that probabilities are numbers between zero and one; and are attached to events (or to joint events).

Example 1: Conditional Probability

Typically, the substantive questions of interest are stated in a "conditional" way:

for example, suppose the FOBT is positive (event B has occurred, i.e., one of the 314 people in row +FOBT has been picked)

what is the probability that the person is also +CC (that the event A has also occurred, i.e., the person picked is one of the 15 in the cell (+FOBT, +CC)?

Restate as a conditional probability:

$P(A|B) = 15/314$

where "|" is read as "given" – restating this, if we know the result is in the first row (event B has occurred), what is the probability we are also in the first column (that event A also occurs)?

Conditional Probability Formula

A formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{15/10,000}{314/10,000} = \frac{15}{314}$$

Also, for purposes of the next slide, note that

$$P(A|B) = \frac{15}{314} = .048 > P(A) = \frac{30}{10,000} = .003$$

Example 1: Event Facilitation and Inhibition

$P(A)$ ($= P(+CC)$) $= 30/10{,}000$ is the "prior" (or "marginal") probability of the person having colorectal cancer irrespective of the FOBT results.

Note that $P(A|B) = .048 > P(A) = .003$, so knowing that the test is positive (that B has occurred) increases the probability of A occurring over its marginal value of $P(A)$

When this happens, we say that event B is "facilitative" of A

If $P(A|B) < P(A)$, event B is said to be "inhibitive" of A

Example 1: False Positives

Because $P(+CC|+FOBT) > P(+CC)$, +FOBT is facilitative of +CC.

The size of the difference, $P(+CC|+FOBT) - P(+CC) = +.045$, may not be large in any absolute sense, but the change does represent a fifteenfold increase over the marginal probability of .003 for $P(+CC)$.

But note that even if you have a positive FOBT, over 95% ($= \frac{299}{314}$) of the time you don't have cancer; that is, there are 95% false positives.

Unfortunately, dismal results such as these appear regularly. Even though an event may be facilitative or inhibitive of another, this can be a very weak condition by itself. The degree of facilitation or inhibition may be so weak in absolute terms that reliance on it is mistaken both practically and ethically.

Example 1: Event (Statistical) Independence

Suppose event B is neither facilitative or inhibitive of A and (therefore)

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

When this occurs, we say that the events A and B are statistically independent

Equivalently, if $P(A|B) = P(A)$, then $P(A \cap B) = P(A)P(B)$

The same result will be true for any combination of A and $\bar{A}$ with B and $\bar{B}$

Also, *all* joint cell probabilities in the 2×2 contingency table can be constructed by multiplying the two appropriate marginal probabilities together; for example, $P(A \cap \bar{B}) = P(A)P(\bar{B})$

Example 2: A Risque Legend of Cinderella

On her hurried way out of the castle just before midnight, Cinderella drops the one glass slipper (but, say, she holds on to the other one).

Also, suppose Cinderella loses all of her fitted clothes and jewelry including tiara, bra, and so on.

When the Prince sets off to find Cinderella, the following events are of interest:

slipper fits: event A occurs
slipper doesn't fit: event $\bar{A}$ occurs

person is Cinderella: event B occurs
person is not Cinderella: event $\bar{B}$

Example 2: Sequential Weight-of-the-Evidence Argument

- Clearly, the occurrence of the event A (that the slipper fits) increases the probability that event B occurs (that the person is Cinderella); or that A is facilitative of B
- Note that the Prince has an array of fitted clothes and jewelery that also could be tried on sequentially, with each item that fits Cinderella being facilitative of the event B of being Cinderella
- We may never get to a "sure thing" (a probability of 1.0 for the event B of being Cinderella), and thus, have her identified "beyond a shadow of a doubt"
- However, the sequential process may get up eventually to another level of a "burden of proof" – "beyond a reasonable doubt"
- But then again, we may have the proverbial "smoking gun" if Cinderella pulls out the identical matching slipper she held onto that night

Example 3: The O. J. Simpson Murder Trial

One of the most publicized criminal trials in American history was the 1995 O. J. Simpson murder case in Los Angeles.

You might remember Simpson not being able to fit easily into the blood-splattered leather glove that was found at the crime scene.

This allowed defense counsel Johnny Cochran to issue the famous line:

“if it doesn’t fit, you must acquit”

We will deal with this statement probabilistically using the language of events occurring or not occurring.

Example 3: Rephrasing Cochran's Assertion Probabilistically

glove fits: event A occurs
glove doesn't fit: event $\bar{A}$ occurs

jury convicts: event B occurs
jury acquits: event $\bar{B}$ occurs

"if it doesn't fit, you must acquit" $\Leftrightarrow$

"if $\bar{A}$ occurs, then $\bar{B}$ occurs" $\Leftrightarrow$

$P(\bar{B}|\bar{A}) = 1.0$

To have such a "sure thing" of acquittal seems too strong

But how about: $P(\bar{B}|\bar{A}) > P(\bar{B})$?

Example 3: Continued

Or, in words:

the occurrence of $\bar{A}$ (the glove not fitting) should increase the probability of acquittal to somewhere above the prior (or marginal) value of $P(\bar{B})$;

there is no specification of low large this increase should be other than being short of the value of 1.0 (for a "sure thing" of acquittal)

As we said before in our colorectal cancer example, and in the risque Cinderella legend, when $P(\bar{B}|\bar{A}) > P(\bar{B})$, $\bar{A}$ is said to be facilitative of $\bar{B}$: the glove not fitting is facilitative of acquittal.

Example 3: Relevance of Evidence

First of all, note that if $\bar{A}$ is facilitative of $\bar{B}$ (that is, $P(\bar{B}|\bar{A}) > P(\bar{B})$), a whole collection of such statements hold:

$\bar{B}$ is facilitative of $\bar{A}$ and inhibitive of A;
$\bar{A}$ is facilitative of $\bar{B}$ and inhibitive of B;
B is facilitative of A and inhibitive of $\bar{A}$;
A is facilitative of B and inhibitive of $\bar{B}$.

Alternative legal phrases for the words "facilitative" and "inhibitive" are "positively relevant" and "negatively relevant"

Rule 401 in the Federal Rules of Evidence (FRE) defines evidence relevance as follows: Evidence is relevant if it has any tendency to make a fact more or less probable than it would be without the evidence

However, just because evidence may be relevant doesn't automatically then make it admissible under FRE Rule 403; examples would the race, sex, or age of a defendant.

The Sally Clark Case: Improper Use of Statistical Independence

Statistical independence of two events, A and B, was defined earlier by

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

Sally Clark was convicted in England of killing her two children (who both died in their cribs – "cot" death or sudden infant death syndrome (SIDS)).

The conviction was based partially on an inappropriate assumption of statistical independence.

The purveyor of statistical misinformation in this case was Sir Roy Meadow, famous for Meadow's Law:

Meadow's Law

"One sudden infant death is a tragedy, two is suspicious, and three is murder unless proved otherwise"

What Meadow did in the Sally Clark case was to first obtain an estimate of one SIDS death in a family (1 in 8500)

To estimate the probability of two deaths in the same family, Meadow assumed the two deaths to be independent events, and merely squared the estimated probability value for one death to obtain the probability of two deaths – a value of 1 in 73 million.

Meadow then went on to assert that this value of 1 in 73 million was the probability that Sally Clark was innocent.

This, by the way, is another error called the "prosecutor's fallacy" that we will return to later.

Royal Statistical Society Press Release (October 23, 2001)

The Royal Statistical Society today issued a statement, prompted by issues raised by the Sally Clark case, expressing its concern at the misuse of statistics in the courts.
In the recent highly-publicised case of *R v. Sally Clark*, a medical expert witness drew on published studies to obtain a figure for the frequency of sudden infant death syndrome (SIDS, or 'cot death') in families having some of the characteristics of the defendant's family. He went on to square this figure to obtain a value of 1 in 73 million for the frequency of two cases of SIDS in such a family.

Press Release Continued

This approach is, in general, statistically invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong a priori reasons for supposing that the assumption will be false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely. **The well-publicised figure of 1 in 73 million thus has no statistical basis. Its use cannot reasonably be justified as a 'ballpark' figure because the error involved is likely to be very large, and in one particular direction. The true frequency of families with two cases of SIDS may be very much less incriminating than the figure presented to the jury at trial.**

Statistical Syllogisms

A statistical syllogism argues from a generalization that is true for the most part to a particular case.

1) (Major Premise) Almost all people are taller than 26 inches
2) Donald Trump is a person
3) Therefore, Donald Trump is almost certainly taller than 26 inches

It is posible for the major premise to be true and the conclusion false, but that is not very likely.

Confidence Interval Interpretation

A confidence interval is a range of numbers that would encompass a fixed but unknown (parameter) value (for example, the proportion of the electorate who will vote for Donald Trump).

To interpret a given confidence interval, we can use a statistical syllogism:

(1) (Major Premise) "if this particular confidence interval construction method were repeated for multiple samples, the collection of all such random intervals would encompass the true population parameter, say, 95% of the time";
(2) "this is one such constructed interval";
(3) "it is very likely that this interval contains the true population value."

Federal Rules of Evidence: Rule 403

In a legal context one has to be careful about the use of statistical syllogism that generally involve a judgement as to which groups commit which crimes.

For example, stating that the event of being black is facilitative of the occurrence of a future event (or act) of violence.

These types of inferences are forbidden under Rule 403 of the Federal Rules of Evidence:

Rule 403. Exclusion of Relevant Evidence on Grounds of Prejudice, Confusion, or Waste of Time: Although relevant, evidence may be excluded if its probative [i.e., its legal proof] value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.

An Example From Texas

A Texas fact:

In Texas capital murder cases, a prediction of future dangerous behavior is needed to have a death penalty imposed.

A prominent Texas psychologist Walter Quijano has regularly testified that because a defendant is black or Hispanic, there is an increased probability of a future event (or act) of violence occurring.

We give a redaction in an appendix to the first module of the Supreme Court case of *Duane Edward Buck v. Rick Taylor* (2011).

The defendant, Duane Buck, was attempting to avoid the imposition of the death penalty sentence.

The Argument in Buck's Case

Buck's lawyers argued that the death penalty should be lifted because Quijano stated at Buck's trial that because he was black, there was an increased probability he would engage in future acts of violence.

The Supreme Court refused to hear the case (that is, to grant what is called *certiorari*), not because Buck didn't have a case of prejudicial racial evidence being introduced (in violation of Rule 403), but because, incredibly, Quijano was a witness for the defense (that is, for Buck).

More on Quijano

From *The Texas Tribune*, October 31, 2011: Texas Ends Deal With Psychologist Over Race Testimony (Brandi Grissom):

The Texas Youth Commission terminated its contract Friday with a psychologist who testified repeatedly in death penalty cases that Hispanic and black men were more likely to be dangerous in the future. The termination followed a Texas Tribune inquiry into the agency's six-year agreement with the doctor. ...

In 2000, then-Texas Attorney General John Cornyn admitted the state had erred in the 1996 trial of Victor Hugo Saldana, along with six other death penalty cases, in which Quijano testified regarding race and future dangerousness.

"Because the use of race in Saldano's sentencing seriously undermined the fairness, integrity, or public reputation of the judicial process, Texas confesses error," Cornyn wrote in a court filing. ...

Quijano said he was confused about the new concerns regarding his testimony. The doctor said his past testimony has been taken out of context and misconstrued. He said he has testified in about 150 death penalty cases. He explained that in the six controversial cases, he said people of African American and Hispanic backgrounds were more likely to be less educated, to have fewer work opportunities and to have grown up around negative influences in disadvantaged neighborhoods.

"It doesn't mean the color makes the person violent," Quijano said.

The Charles Peirce Notion of Abductive Reasoning

Another way of interpreting whether one event is facilitative or inhibitive of another is through the idea of abductive reasoning introduced by Charles Peirce in the late 19th century.

Simply put, abductive reasoning is "guessing" (or "abducing") some explanation from some observed event or circumstance.

An example: suppose we get up in the morning and observe that the lawn is wet (event A has occurred)

By abduction, we guess that it rained last night (event B has occurred)

The event A of the lawn being wet is facilitative of the event B of it raining last night: $P(B|A) > P(B)$

Deduction, Induction, and Abduction

Charles Peirce has a nice beanbag analogy to illustrate the three reasoning modes of deduction, induction, and abduction:

Deduction

(Step 1) Rule: All the beans from this bag are white
(Step 2) Case: These beans are from this bag
Therefore,
(Step 3) Result: These beans are white

Induction

(Step 1) Case: These beans are from this bag
(Step 2) Result: These beans are white
Therefore,
(Step 3) Rule: All the beans from this bag are white

Beanbag Abduction Example

Abduction

(Step 1) Rule: All the beans from this bag are white
(Step 2) Result: These beans are white
Therefore,
(Step 3) Case: These beans are from this bag

In this last abduction example, we abduce ("guess") that "these beans are from this bag" from the observation that "these beans are white"

"If p, then q" Statements

In beginning statistics we often encounter "if p, then q" statements, where p and q are two propositions.

An example: let p be "the animal is a Yellow Labrador Retriever," and q be "the animal is in the order *Carnivora*"

We know that "if p (the antecedent proposition), then q (the consequent proposition)" is true.

We also know that the contrapositive is true: "if not q, then not p" – if "the animal is not in the order *Carnivora*," then "the animal is not a Yellow Labrador Retriever"

Logical Fallacies

There are two fallacies awaiting the unsuspecting:

denying the antecedent: if not p, then not q (if "the animal is not a Yellow Labrador Retriever," then "the animal is not in the order *Carnivora*");

affirming the consequent: if q, then p (if "the animal is in the order *Carnivora*," then "the animal is a Yellow Labrador Retriever").

In a probabilistic context, we reinterpret the phrase "if p, then q" as B being facilitative of A; that is, $P(A|B) > P(A)$, where p is identified with B and q with A.

With such a probabilistic reinterpretation, we no longer have the fallacies of denying the antecedent (that is, $P(\bar{A}|\bar{B}) > P(\bar{A})$), or of affirming the consequent (that is, $P(B|A) > P(B)$); all of these are now necessary consequences of the first statement that B is facilitative of A.

Genetic Probabilistic Reasoning

In reasoning about various medical contexts, it would be very rare to rely on the simple logic of "if p, then q"

More likely, we are given problems characterized by fallible data, and subject to all sorts of probabilistic processes.

As an example, suppose someone has a genetic marker for some disease (for example, the BRCA1 mutation for breast cancer); those with the marker have a higher probability of getting the disease.

But an "if p then q" situation does not hold – it is not true that if you have the marker, then you must get the disease

Continued: Genetic Probabilistic Reasoning

Genetic probabilistic reasoning might best be done through simple 2×2 contingency tables, like the one used earlier in our colorectal cancer screening example.

Here, A and $\bar{A}$ denote the presence/absence of the marker, and B and $\bar{B}$ denote the presence/absence of the disease

If is assumed that A is facilitative of B: $P(B|A) > P(B)$, then possible measures of the strength of facilitation might be

$P(B|A) - P(B)$

or

$P(B|A)/P(B)$

Penetrance

In the context of genetics, for example, the conditional probability, $P(B|A)$, is typically reported by itself as a measure of the strength of facilitation;

this is called "penetrance" – the probability of disease occurrence given the presence of the marker.

A fairly recent and high profile instance of the BRCA1 mutation being assessed as strongly facilitative of breast cancer (that is, having high "penetrance") was for the actress Angelina Jolie, who opted for a prophylactic double mastectomy to reduce her chances of contracting breast cancer.

A few excerpts follow from her Op-Ed article, "My Medical Choice," that appeared in the *New York Times* (May 14, 2013):

Angelina Jolie Excerpt

… My doctors estimated that I had an 87 percent risk of breast cancer and a 50 percent risk of ovarian cancer, although the risk is different in the case of each woman.
Only a fraction of breast cancers result from an inherited gene mutation. Those with a defect in BRCA1 have a 65 percent risk of getting it, on average.
Once I knew that this was my reality, I decided to be proactive and to minimize the risk as much I could. I made a decision to have a preventive double mastectomy. I started with the breasts, as my risk of breast cancer is higher than my risk of ovarian cancer, and the surgery is more complex. …

In Summary: Arguing Probabilistic Causation

The idea of arguing probabilistic causation is, in effect, the notion of one event being facilitative or inhibitive of another.

If a collection of "q" conditions is observed that would be the consequence of a single "p," one may be more prone to conjecture the presence of "p," much like we could do in the Cinderella example.

Although this process may seem like merely affirming the consequent, in a probabilistic context this is not a fallacy, and could be referred to as "inference to the best explanation," or as we have noted above, an interpretation of the Charles Peirce notion of abductive reasoning.

Most uses of information in contexts that are legal (forensic) or medical (screening) need to be assessed probabilistically