Probability Theory: Application Areas

Psychology (Statistics) 484

Statistics, Ethics, and the Social and Behavioral Sciences

June 14, 2013

Week 3: Probability Theory—Application Areas

— how subjective probabilities might be related to the four levels of a "legal burden of proof": "preponderance of the evidence"; "clear and convincing evidence"; "clear, unequivocal, and convincing evidence"; and "proof beyond a reasonable doubt"

— the distinction between "general causation" and "specific causation"; the common legal standard for arguing specific causation as an "attributable proportion of risk" of 50% or more

— issues of probability, risk, and gambling; spread betting and point shaving; parimutuel betting; the importance of context and framing in risky choice and decision-making

Required Reading:
SGEP (87–118) —
Some Probability Considerations in Discrimination and Classification
Probability and Litigation
Probability of causation
Probability scales and rulers
The cases of Vincent Gigante and Agent Orange
Betting, Gaming, and Risk
Spread betting
Parimutuel betting
Some psychological considerations in gambling

Popular Articles —

Better Decisions Through Science, John A. Swets, Robyn M. Dawes, and John Monahan (*Scientific American*, October 2000)

Do Fingerprints Lie? Michael Specter (*New Yorker*, May 27, 2002)

Under Suspicion, Atul Gawande (*New Yorker*, January 8, 2001)

Suggested Reading:

Suggested Reading on Agent Orange and Judge Weinstein

Appendix: The Redacted Text of Judge Weinstein's Opinion in the Fatico Case

Appendix: Guidelines for Determining the Probability of Causation and Methods for Radiation Dose Reconstruction Under the Employees Occupational Illness Compensation Program Act of 2000

Appendix: District of Columbia Court of Appeals, In Re As. H (Decided: June 10, 2004)

Suggested Reading on Issues of Risk

Suggested Reading on Issues of Betting and Gaming

Film: *The Central Park Five* (2 hours)

Discrimination and Classification

The term *discrimination* can refer to the task of separating groups through linear combinations of variables maximizing a criterion, such as an F-ratio.

The linear combinations themselves are commonly called Fisher's linear discriminant functions.

The related term *classification* refers to the task of allocating observations to existing groups, typically to minimize the cost and/or probability of misclassification.

These two topics are intertwined, but here we briefly comment only on the topic of classification.

In the simplest situation, we have two populations, π_1 and π_2;

π_1 is assumed to be characterized by a normal distribution with mean μ_1 and variance σ_X^2 (the density is denoted by $f_1(x)$);

π_2 is characterized by a normal distribution with mean μ_2 and (common) variance σ_X^2 (the density is denoted by $f_2(x)$).

Given an observation, say x_0, we wish to decide whether it should be assigned to π_1 or to π_2.

Assuming that $\mu_1 \leq \mu_2$, a criterion point c is chosen; the rule then becomes: allocate to π_1 if $x_0 \leq c$, and to π_2 if $> c$.

The probabilities of misclassification are given in the following chart:

		True State π_1	π_2
Decision	π_1	$1-\alpha$	β
	π_2	α	$1-\beta$

In the terminology of our previous usage of Bayes' rule to obtain the positive predictive value of a test, and assuming that π_1 refers to a person having "it," and π_2 to not having "it," the sensitivity of the test is $1-\alpha$ (true positive);

specificity is $1-\beta$, and thus, β refers to a false positive.

To choose c so that $\alpha + \beta$ is smallest, select the point at which the densities are equal.

A more complicated way of stating this decision rule is to allocate to π_1 if $f_1(x_0)/f_2(x_0) \geq 1$; if < 1, then allocate to π_2.

Suppose now that the prior probabilities of being drawn from π_1 and π_2 are p_1 and p_2, respectively, where $p_1 + p_2 = 1$. If c is chosen so the Total Probability of Misclassification (TPM) is minimized (that is, $p_1\alpha + p_2\beta$), the rule would be to allocate to π_1 if $f_1(x_0)/f_2(x_0) \geq p_2/p_1$; if $< p_2/p_1$, then allocate to π_2.

Finally, to include costs of misclassification, $c(1|2)$ (for assigning to π_1 when actually coming from π_2), and $c(2|1)$ (for assigning to π_2 when actually coming from π_1),

choose c to minimize the Expected Cost of Misclassification (ECM), $c(2|1)p_1\alpha + c(1|2)p_1\beta$, by the rule of allocating to π_1 if $f_1(x_0)/f_2(x_0) \geq (c(1|2)/c(2|1))(p_2/p_1)$;

if $< (c(1|2)/c(2|1))(p_2/p_1)$, then allocate to π_2.

ROC Curves

In the terminology of signal detection theory and the general problem of yes/no diagnostic decisions, a plot of sensitivity (true positive probability) on the y-axis against $1-$ specificity on the x-axis as c varies, is an ROC curve (for Receiver Operating Characteristic).

This ROC terminology originated in World War II in detecting enemy planes by radar (group π_1) from the noise generated by random interference (group π_2).

The ROC curve is bowed from the origin of (0, 0) at the lower-left corner to (1.0, 1.0) at the upper right; it indicates the trade-off between increasing the probability of true positives and the increase of false positives.

Generally, the adequacy of a particular diagnostic decision strategy is measured by the area under the ROC curve.

Probability and Litigation

Jack Weinstein is a sitting federal judge in the Eastern District of New York (Brooklyn).

He also may be the only federal judge ever to publish an article in a major statistics journal (*Statistical Science*, 1988, *3*, 286–297, "Litigation and Statistics").

This last work developed out of Weinstein's association in the middle 1980s with the National Academy of Science's Panel on Statistical Assessment as Evidence in the Courts.

This panel produced the comprehensive Springer-Verlag volume. *The Evolving Role of Statistical Assessments as Evidence in the Courts* (1988; Stephen E. Fienberg, Editor).

290th Commandment

The importance that Weinstein gives to the role of probability and statistics in the judicial process is best expressed by Weinstein himself (we quote from his *Statistical Science* article):

The use of probability and statistics in the legal process is not unique to our times. Two thousand years ago, Jewish law, as stated in the Talmud, cautioned about the use of probabilistic inference. The medieval Jewish commentator Maimonides summarized this traditional view in favor of certainty when he noted:

"The 290th Commandment is a prohibition to carry out punishment on a high probability, even close to certainty . . . No punishment [should] be carried out except where . . . the matter is established in certainty beyond any doubt . . . "

That view, requiring certainty, is not acceptable to the courts. We deal not with the truth, but with probabilities, in criminal as well as civil cases. Probabilities, express and implied, support every factual decision and inference we make in court.

Maimonides' description of the 290th Negative Commandment is given in its entirety in an appendix.

According to this commandment, an absolute certainty of guilt is guaranteed by having two witnesses to exactly the same crime.

Such a probability of guilt being identically one is what is meant by the contemporary phrase "without any shadow of a doubt."

Two points need to be emphasized about this Mitzvah (Jewish commandment).

One is the explicit unequalness of costs attached to the false positive and negative errors:

"it is preferable that a thousand guilty people be set free than to execute one innocent person."

The second is in dealing with what would now be characterized as the (un)reliability of eyewitness testimony.

Two eyewitnesses are required, neither is allowed to make just an inference about what happened but must have observed it directly, and exactly the same crime must be observed by both eyewitnesses.

Fatico Case

Judge Weinstein's interest in how probabilities could be part of a judicial process goes back some years before the National Research Council Panel.

In one relevant opinion from 1978, *United States v. Fatico*, he wrestled with how subjective probabilities might be related to the four levels of a "legal burden of proof"; what level was required in this particular case; and, finally, was it then met.

The four (ordered) levels are: preponderance of the evidence; clear and convincing evidence; clear, unequivocal, and convincing evidence; and proof beyond a reasonable doubt.

The case in point involved proving that Daniel Fatico was a "made" member of the Gambino organized crime family, and thus could be given a "Special Offender" status.

Other Standards

Other common standards used for police searches or arrests might also be related to an explicit probability scale.

The lowest standard (perhaps a probability of 20%) would be "reasonable suspicion" to determine whether a brief investigative stop or search by any governmental agent is warranted (in the 2010 "Papers, Please" law in Arizona, a "reasonable suspicion" standard is set for requesting documentation).

A higher standard would be "probable cause" to assess whether a search or arrest is warranted, or whether a grand jury should issue an indictment.

A value of, say, 40% might indicate a "probable cause" level that would put it somewhat below a "preponderance of the evidence" criterion.

Probability of Causation

Judge Weinstein is best known for the mass (toxic) tort cases he has presided over for the last four decades (for example, asbestos, breast implants, Agent Orange).

In all of these kinds of torts, there is a need to establish, in a legally acceptable fashion, some notion of causation.

There is first a concept of *general causation* concerned with whether an agent can increase the incidence of disease in a group;

because of individual variation, a toxic agent will not generally cause disease in every exposed individual.

Specific causation deals with an individual's disease being attributable to exposure from an agent.

Cohort Studies

The establishment of general causation (and a necessary requirement for establishing specific causation) typically relies on a *cohort study*.

	Disease	No Disease	Row Sums
Exposed	N_{11}	N_{12}	N_{1+}
Not Exposed	N_{21}	N_{22}	N_{2+}

Here, N_{11}, N_{12}, N_{21}, and N_{22} are the cell frequencies; N_{1+} and N_{2+} are the row frequencies.

Conceptually, these data are considered generated from two (statistically independent) binomial distributions for the "Exposed" and "Not Exposed" conditions.

If we let p_E and p_{NE} denote the two underlying probabilities of getting the disease for particular cases within the conditions, respectively, the ratio $\frac{p_E}{p_{NE}}$ is referred to as the relative risk (RR), and may be estimated with the data as follows:

estimated relative risk $= \widehat{\mathrm{RR}} = \frac{\hat{p}_E}{\hat{p}_{NE}} = \frac{N_{11}/N_{1+}}{N_{21}/N_{2+}}$.

A measure commonly referred to in tort litigations is attributable risk (AR), defined as

AR $= \frac{p_E \; - \; p_{NE}}{p_E}$, and estimated by

$\widehat{\mathrm{AR}} = \frac{\hat{p}_E \; - \; \hat{p}_{NE}}{\hat{p}_E} = 1 - \frac{1}{\widehat{\mathrm{RR}}}$.

Attributable Risk

Attributable risk, also known as the "attributable proportion of risk" or the "etiologic fraction," represents the amount of disease among exposed individuals assignable to the exposure.

It measures the maximum proportion of the disease attributable to exposure from an agent, and consequently, the maximum proportion of disease that could be potentially prevented by blocking the exposure's effect or eliminating the exposure itself.

If the association is causal, AR is the proportion of disease in an exposed population that might be caused by the agent, and therefore, that might be prevented by eliminating exposure to the agent.

The common legal standard used to argue for both specific and general causation is an RR of 2.0, or an AR of 50%.

At this level, it is "as likely as not" that exposure "caused" the disease (or "as likely to be true as not," or "the balance of the probabilities").

Obviously, one can never be absolutely certain that a particular agent was "the" cause of a disease in any particular individual, but to allow an idea of "probabilistic causation" or "attributable risk" to enter into legal arguments provides a justifiable basis for compensation.

It has now become routine to do this in the courts.

Probability Scales and Rulers

The topic of relating a legal understanding of burdens of proof to numerical probability values has been around for a very long time.

Fienberg (1988) provides a short discussion of Jeremy Bentham's (1827) suggestion of a "persuasion thermometer," and some contemporary reaction to this idea from Thomas Starkie (1833):

Jeremy Bentham appears to have been the first jurist to seriously propose that witnesses and judges numerically estimate their degrees of persuasion. Bentham envisioned a kind of moral thermometer:

The scale being understood to be composed of ten degrees—in the language applied by the French philosophers to thermometers, a decigrade scale—a man says, My persuasion is at 10 or 9, etc. affirmative, or at least 10, etc. negative . . .

Several particularly knotty problems and (mis)interpretations when it comes to assigning numbers to the possibility of guilt arise most markedly in eyewitness identification.

Because cases involving eyewitness testimony are typically criminal cases, they demand burdens of proof "beyond a reasonable doubt";

thus, the (un)reliability of eyewitness identification becomes problematic when it is the primary (or only) evidence presented to meet this standard.

As discussed extensively in the judgment and decision-making literature, there is a distinction between making a subjective estimate of some quantity, and one's confidence in that estimate once made.

For example, suppose someone picks a suspect out of a lineup, and is then asked the (Bentham) question,

"on a scale of from one to ten, characterize your level of 'certainty'."

Does an answer of "seven or eight" translate into a probability of innocence of two or three out of ten?

Exactly such confusing situations, however, arise.

We give a fairly extensive redaction in an appendix of an opinion from the District of Columbia Court of Appeals in a case named "In re As.H" (2004).

It combines extremely well both the issues of eyewitness (un)reliability and the attempt to quantify that which may be better left in words;

the dissenting Associate Judge Farrel noted pointedly:

"I believe that the entire effort to quantify the standard of proof beyond a reasonable doubt is a search for fool's gold."

Betting, Gaming, and Risk

Antoine Gombaud, better known as the Chevalier de Méré, was a French writer and amateur mathematician from the early 17th century.

He is important to the development of probability theory because of one specific thing; he asked a mathematician, Blaise Pascal, about a gambling problem dating from the Middle Ages, named "the problem of points."

The question was one of fairly dividing the stakes among individuals who had agreed to play a certain number of games, but for whatever reason had to stop before they were finished.

Pascal in a series of letters with Pierre de Fermat, solved this equitable division task, and in the process laid out the foundations for a modern theory of probability.

Pascal and Fermat also provided the Chevalier with a solution to a vexing problem he was having in his own personal gambling.

Apparently, the Chevalier had been very successful in making even money bets that a six would be rolled at least once in four throws of a single die.

But when he tried a similar bet based on tossing two dice 24 times and looking for a double-six to occur, he was singularly unsuccessful in making any money.

The reason for this difference between the Chevalier's two wagers was clarified by the formalization developed by Pascal and Fermat for such games of chance.

Some Useful Concepts

A *simple experiment* is some process that we engage in that leads to one single outcome from a set of possible outcomes that could occur.

For example, a simple experiment could consist of rolling a single die once, where the set of possible outcomes is $\{1, 2, 3, 4, 5, 6\}$ (note that curly braces will be used consistently to denote a set).

Or, two dice could be tossed and the number of spots occurring on each die noted; here, the possible outcomes are integer number pairs: $\{(a, b) \mid 1 \leq a \leq 6; 1 \leq b \leq 6\}$.

Flipping a single coin would give the set of outcomes, $\{H, T\}$, with "H" for "heads" and "T" for "tails";

picking a card from a normal deck could give a set of outcomes containing 52 objects, or if we were only interested in the particular suit for a card chosen, the possible outcomes could be $\{H, D, C, S\}$, corresponding to heart, diamond, club, and spade, respectively.

The set of possible outcomes for a simple experiment is the *sample space* (which we denote by the script letter $\mathcal{S}$).

An object in a sample space is a *sample point*.

An *event* is defined as a subset of the sample space, and an event containing just a single sample point is an *elementary event*.

A particular event is said to occur when the outcome of the simple experiment is a sample point belonging to the defining subset for that event.

As a simple example, consider the toss of a single die, where $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

The event of obtaining an even number is the subset $\{2, 4, 6\}$;

the event of obtaining an odd number is $\{1, 3, 5\}$;

the (elementary) event of tossing a 5 is a subset with a single sample point, $\{5\}$, and so on.

For a sample space containing K sample points, there are 2^K possible events (that is, there are 2^K possible subsets of the sample space).

This includes the "impossible event" (usually denoted by $\emptyset$), characterized as that subset of $\mathcal{S}$ containing no sample points and which therefore can never occur;

and the "sure event," defined as that subset of $\mathcal{S}$ containing all sample points (that is, $\mathcal{S}$ itself), which therefore must always occur.

In our single die example, there are $2^6 = 64$ possible events, including $\emptyset$ and $\mathcal{S}$.

The motivation for introducing the idea of a simple experiment and sundry concepts is to use this structure as an intuitively reasonable mechanism for assigning probabilities to the occurrence of events.

These probabilities are usually assigned through an assumption that sample points are equally likely to occur, assuming we have characterized appropriately what is to be in $\mathcal{S}$.

Generally, only the probabilities are needed for the K elementary events containing single sample points.

The probability for any other event is merely the sum of the probabilities for all those elementary events defined by the sample points making up that particular event.

This last fact is due to the disjoint set property of probability introduced at the beginning of the last chapter.

In the specific instance in which the sample points are equally likely to occur, the probability assigned to any event is merely the number of sample points defining the event divided by K.

As special cases, we obtain a probability of 0 for the impossible event, and 1 for the sure event.

The Chevalier Games

One particularly helpful use of the sample space/event concepts is when a simple experiment is carried out multiple times (for, say, N replications), and the outcomes defining the sample space are the ordered N-tuples formed from the results obtained for the individual simple experiments.

The Chevalier who rolls a single die four times, generates the sample space

$$\{(D_1, D_2, D_3, D_4) \mid 1 \leq D_i \leq 6, 1 \leq i \leq 4\} ,$$

that is, all 4-tuples containing the integers from 1 to 6.

Generally, in a replicated simple experiment with K possible outcomes on each trial, the number of different N-tuples is K^N (using a well-known arithmetic multiplication rule).

Thus, for the Chevalier example, there are $6^4 = 1296$ possible 4-tuples, and each such 4-tuple should be equally likely to occur (given the "fairness" of the die being used).

To define the event of "no sixes rolled in four replications," we would use the subset (event)

$$\{(D_1, D_2, D_3, D_4) \mid 1 \leq D_i \leq 5, 1 \leq i \leq 4\} \ ,$$

containing $5^4 = 625$ sample points.

Thus, the probability of "no sixes rolled in four replications" is $625/1296 = .4822$.

As we will see formally below, the fact that this latter probability is strictly less than 1/2 gives the Chevalier a distinct advantage in playing an even money game defined by his being able to roll at least one six in four tosses of a die.

The other game that was not as successful for the Chevalier, was tossing two dice 24 times and betting on obtaining a double-six somewhere in the sequence.

The sample space here is
$\{(P_1, P_2, \ldots, P_{24})\}$, where
$P_i = \{(a_i, b_i) \mid 1 \leq a_i \leq 6; 1 \leq b_i \leq 6\}$,
and has 36^{24} possible sample points.

The event of "not obtaining a double-six somewhere in the sequence" would look like the sample space just defined except that the $(6, 6)$ pair would be excluded from each P_i.

Thus, there are 35^{24} members in this event.

The probability of "not obtaining a double-six somewhere in the sequence" is

$$\frac{35^{24}}{36^{24}} = (\frac{35}{36})^{24} = .5086 \ .$$

Because this latter value is greater than $1/2$ (in contrast to the previous gamble), the Chevalier would now be at a disadvantage making an even money bet.

Random Variables to Evaluate Bets

The best way to evaluate the perils or benefits present in a wager is through the device of a discrete random variable.

Suppose X denotes the outcome of some bet; and let $a_1, \ldots, a_T$ represent the T possible payoffs from one wager, where positive values reflect gain and negative values reflect loss.

In addition, we know the probability distribution for X; that is, $P(X = a_t)$ for $1 \leq t \leq T$.

What one expects to realize from one observation on X (or from one play of the game) is its expected value,

$$E(X) = \sum_{t=1}^{T} a_t P(X = a_t).$$

If $E(X)$ is negative, we would expect to lose this much on each bet; if positive, this is the expected gain on each bet.

When $E(X)$ is 0, the term "fair game" is applied to the gamble, implying that one neither expects to win or lose anything on each trial; one expects to "break even."

When $E(X) \neq 0$, the game is "unfair" but it could be unfair in your favor ($E(X) > 0$), or unfair against you ($E(X) < 0$).

To evaluate the Chevalier's two games, suppose X takes on the values of $+1$ and -1 (the winning or losing of one dollar, say).

For the single die rolled four times, $E(X) = (+1)(.5178) + (-1)(.4822) = .0356 \approx .04$.

Thus, the game is unfair in the Chevalier's favor because he expects to win a little less than four cents on each wager.

For the 24 tosses of two dice, $E(X) = (+1)(.4914) + (-1)(.5086) = -.0172 \approx -.02$.

Here, the Chevalier is at a disadvantage.

The game is unfair against him, and he expects to lose about two cents on each play of the game.

Spread Betting

The type of wagering that occurs in roulette or craps is often referred to as fixed-odds betting; you know your chances of winning when you place your bet.

A different type of wager is spread betting, invented by a mathematics teacher from Connecticut, Charles McNeil, who became a Chicago bookmaker in the 1940s.

Here, a payoff is based on the wager's accuracy; it is no longer a simple "win or lose" situation.

Generally, a spread is a range of outcomes, and the bet itself is on whether the outcome will be above or below the spread.

In common sports betting (for example, NCAA college basketball), a "point spread" for some contest is typically advertised by a bookmaker.

If the gambler chooses to bet on the "underdog," he is said to "take the points" and will win if the underdog's score plus the point spread is greater than that of the favored team;

conversely, if the gambler bets on the favorite, he "gives the points" and wins only if the favorite's score minus the point spread is greater than the underdog's score.

In general, the announcement of a point spread is an attempt to even out the market for the bookmaker, and to generate an equal amount of money bet on each side.

The commission that a bookmaker charges will ensure a livelihood, and thus, the bookmaker can be unconcerned about the actual outcome.

Parimutuel Betting

The term *parimutuel betting* (based on the French for "mutual betting") characterizes the type of wagering system used in horse racing, dog tracks, jai alai, and similar contests where the participants end up in a rank order.

It was devised in 1867 by Joseph Oller, a Catalan impresario (he was also a bookmaker and founder of the Paris Moulin Rouge in 1889).

Very simply, all bets of a particular type are first pooled together;

the house then takes its commission and the taxes it has to pay from this aggregate;

finally, the payoff odds are calculated by sharing the residual pool among the winning bets.

To explain using some notation, suppose there are T contestants and bets are made of $W_1, W_2, \ldots, W_T$ on an outright "win."

The total pool is $T_{pool} = \sum_{t=1}^{T} W_t$.

If the commission and tax rate is a proportion, R, the residual pool, R_{pool}, to be allocated among the winning bettors is $R_{pool} = T_{pool}(1 - R)$.

If the winner is denoted by $t*$, and the money bet on the winner is W_{t*}, the payoff per dollar for a successful bet is R_{pool}/W_{t*}.

We refer to the odds on outcome $t*$ as

$$\left(\frac{R_{pool}}{W_{t*}} - 1\right) \text{ to } 1 .$$

For example, if $\frac{R_{pool}}{W_{t*}}$ had a value of 9.0, the odds would be 8 to 1: you get 8 dollars back for every dollar bet plus the original dollar.

In comparison with casino gambling, parimutuel betting pits one gambler against other gamblers, and not against the house.

Also, the odds are not fixed but calculated only after the betting pools have closed (thus, odds cannot be turned into real probabilities legitimately; they are empirically generated based on the amounts of money bet).

A skilled horse player (or "handicapper") can make a steady income, particularly in the newer Internet "rebate" shops that return to the bettor some percentage of every bet made.

Because of lower overhead, these latter Internet gaming concerns can reduce their "take" considerably (from, say, 15% to 2%), making a good handicapper an even better living than before.

Psychological Considerations in Gambling

As shown in the work of Tversky and Kahneman, the psychology of choice is dictated to a great extent by the framing of a decision problem;

that is, the context into which a particular decision problem is placed.

The power of framing in how decision situations are assessed, can be illustrated well though an example and the associated discussion provided by Tversky and Kahneman (1981, p. 453) and given in the text.

The Value of Information

The most relevant aspect of any decision-making proposition involving risky alternatives is the information one has, both on the probabilities that might be associated with the gambles and what the payoffs might be.

In the 1987 movie, *Wall Street*, the character playing Gordon Gekko states:

“The most valuable commodity I know of is information.”

Dirty Harry

I know what you're thinkin'. "Did he fire six shots or only five?" Well, to tell you the truth, in all this excitement I kind of lost track myself.
— Harry Callahan (*Dirty Harry*)

The movie quotation just given from *Dirty Harry* illustrates the crucial importance of who has information and who doesn't.

At the end of Callahan's statement to the bank robber as to whether he felt lucky, the bank robber says:

"I gots to know!"

Harry puts the .44 Magnum to the robber's head and pulls the trigger; Harry knew that he had fired six shots and not five.