The Measurement of Statistical Evidence
Lecture 2 - part 1

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/sta4522/STA4522.html

2021

Probability

- recap

- we have the two problems **E** and/or **H** concerning a real-world object Ψ and our goal is to provide answers (inferences) to these problems based upon the evidence in the observed data x

- towards this end we have introduced the ingredients: the statistical model $\{f_\theta : \theta \in \Theta\}$ and the prior π on θ which leads to the joint probability distribution $(\theta, x) \sim \pi(\theta) f_\theta(x)$

- having observed x we replace the prior π by the conditional of θ given x, also called the posterior and denoted $\pi(\theta \,|\, x)$, for subsequent probability statements about the unknown value of θ

- for the object of interest we have $\psi = \Psi(\theta)$ (taking possibly different values for different θ values) which induces prior π_Ψ and posterior $\pi_\Psi(\theta \,|\, x)$ via marginalization

- probability is playing a big role

- what is probability?

- Kolmogorov's axioms for probability: a probability model $(\Omega, \mathcal{F}, P)$ is specified by a set Ω, a σ-algebra $\mathcal{F}$ on Ω and a function $P : \mathcal{F} \rightarrow [0, 1]$ satisfying $P(\Omega) = 1$ and for any sequence $A_1, A_2, \ldots$ of mutually disjoint elements of $\mathcal{F}$, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

- this successfully "rigorized" the concept of probability so it can be treated as a valid mathematical concept

- but this does not tell us what probability is or how to use it properly

- the following position is adopted here

> *No matter how probabilities are assigned (i.e. defining P), for $A \in \mathcal{F}$, the interpretation of $P(A)$ is that it represents our belief, within the context of $(\Omega, \mathcal{F}, P)$, of an unknown value $\omega \in \Omega$ being in A.*

Example *urn containing 75 White balls and 25 Black balls, after a thorough mixing a ball is drawn and its color observed*

- What is the probability that a W ball is drawn? most would say this probability is $3/4$ and, provided the mixing and drawing was done in a suitable way (more on this later), this seems reasonable

- Does this probability mean that if we repeated this many times that the proportion of balls that are W will converge to $3/4$? this limit argument may well be true but it doesn't have anything to do with the **single** draw we are concerned with

- $3/4$ measures our belief, on the $[0,1]$ scale with 0 and 1 representing categorical beliefs, that the outcome on this single occurrence will be W

- sometimes a distinction is made between probabilities as long-run relative frequencies and probabilities as degrees of belief but actually relative frequencies just gives us one way to **assign** probabilities as degrees of belief and there are other ways to make such assignments

- actually, the way of making the assignments is somewhat irrelevant, what matters in statistical problems is whether or not the assigned probabilities are contradicted by the observed data as nothing more can be said

Example *Bayes*

- recall that we have the prior predictive probability measure M for x before it is observed, namely, for $A \subset \mathcal{X}$

$$\begin{aligned} M(A) &= \int_A m(z)\,dz = \int_A \int_\Theta \pi(\theta) f_\theta(z)\,d\theta\,dz \\ &= \int_\Theta \pi(\theta)\left(\int_A f_\theta(z)\,dz\right)\,d\theta = \int_\Theta \pi(\theta) P_\theta(A)\,d\theta \end{aligned}$$

- so $M(A)$ represents our belief that the observed value of x will be in A before we observe it

- but suppose that we are in the archetypal context so the actual "long relative frequency" of this event is $P_{\theta_{true}}(A)$

- obviously, in general, $P_{\theta_{true}}(A) \neq M(A)$

- but θ_{true} is not known so we can't quote $P_{\theta_{true}}(A)$ and $M(A)$ reflects the uncertainty in the sampling process and the uncertainty due to θ_{true} being unknown and so represents the appropriate degrees of belief

- some will argue that, while x is "random" (or stochastic) θ is not and so $M(A)$ is not really a probability

- but we will subsequently argue that "randomness" has nothing to do with probability per se but rather is concerned with attaining objectivity

Conditional Probability

- undoubtedly the most important concept in statistics

> *Principle of Conditional Probability*
> *Suppose probability model* $(\Omega, \mathcal{F}, P)$ *holds and the fact that* $\omega \in C$, *where* $C \in \mathcal{F}$, *becomes known. Then for* $A \in \mathcal{F}$ *our initial degrees of belief* $P(A)$ *that* $\omega \in A$ *is replaced by* $P(A \,|\, C)$.

- $P(A \,|\, C) = P(A \cap C)/P(C)$ when $P(C) > 0$

- at times paradoxes are presented and these are generally resolved by being clear about how the information "$\omega \in C$" was generated

- this information must arise via an information generator $\Xi : \Omega \overset{onto}{\rightarrow} \Xi$ so that $C = \Xi^{-1}\{\xi_0\} = \{\omega : \Xi(\omega) = \xi_0\}$ for some $\xi_0 \in \Xi$ and the value ξ_0 is observed

Example *Bayes*

- $(\theta, x) \sim \pi(\theta) f_\theta(x)$ and x is observed so $\Omega = \Theta \times \mathcal{X}$ and $\Xi(\theta, x) = x$

Example *Monte Hall Problem (Let's Make A Deal game show)*

- a contestant is presented with 3 doors I, II and III, asked to pick one and picks I

- behind one door there is a desirable prize and behind the other two there are goats

- contestant chooses door I and then Monte goes and opens door II, revealing a goat and asks the contestant if they would like to switch their choice to door III

- should they switch?

- the set of possible outcomes is $\Omega = \{(I, II, III), (I, III, II), (II, III, I)\}$ where the first two coordinates of $\omega \in \Omega$ indicate the doors with goats

- so $A = \{(II, III, I)\}$ is the event the contestant wins by not switching

- the information provided "seems" to be $C = \{(I, II, III), (II, III, I)\}$

- using the *Principle of Insufficient Reason* (when completely ignorant about an unknown, assign equal probabilities to each possibility) then

$$\begin{aligned} P(A) &= 1/3,\ P(C) = 2/3,\ P(A \cap C) = P(A) = 1/3 \\ P(A \mid C) &= P(A \cap C)/P(C) = (1/3)/(2/3) = 1/2 \end{aligned}$$

and there is no reason to switch

- but what is the information generator Ξ (restrict definition to case contestant chooses I) used by Monte?

$$\Xi(I, II, III) = II,\ \Xi(I, III, II) = III,\ \Xi(II, III, I) = ?$$

- if $\Xi(II, III, I) = II$, then $C = \Xi^{-1}\{II\} = \{(I, II, III), (II, III, I)\}$ and $P(A \mid C) = 1/2$

- if $\Xi(II, III, I) = III$, then $C = \Xi^{-1}\{II\} = \{(I, II, III)\}$ and $P(A \mid C) = 0$

- so the conditional probability is ambiguous, although we can say that *if Monte is using a deterministic rule*, then we know $P(A^c \mid C) = 1/2$ or $P(A^c \mid C) = 1$ so switching seems sensible

- alternatively, suppose in the indeterminate case Monte chooses which door to open according to U where

$$\begin{aligned} P(U &= II \,|\, \text{contestant chose } I) = p \\ P(U &= III \,|\, \text{contestant chose } I) = 1-p \end{aligned}$$

and $p \in [0,1]$ is unknown to the contestant

- the information generator is now $\Xi^*(\omega, U)$ (given in the book) and the conditional probability that the contestant will win by not switching is $p/(1+p)$ and this can be any number in $[0, 1/2]$ so prob. of winning by switching is $> 1/2$ and switching is always correct (increases belief in winning the prize)

- often it is implicitly assumed (insufficient reason) that $p = 1/2$, so prob. of winning by switching is $2/3$

- if we apply insufficient reason and say $p \sim U(0,1)$ the probability of winning by switching is 0.693

- if the contestant is told how the information was generated, then the probability of winning by switching is clear, otherwise it is undefined, but it can still be concluded that switching is appropriate.

Probability via betting *(de Finetti)*

- you are a combination bettor/bookie who will buy/sell any gamble on unknown $\omega \in \Omega$ at some price

- a *gamble* X is a function $X : \Omega \to R^1$ where a purchaser of X receives $X(\omega)$ (in units of utility) when ω is revealed as the true value (negative values correspond to losses)

- $L(\Omega) =$ set of all bounded gambles (a linear space)

- $P : L(\Omega) \to R^1$ is a price function called a *prevision*

- you will buy or sell X for $P(X)$ with gain $X(\omega) - P(X)$ or $P(X) - X(\omega)$ respectively

- how is P determined?

> *Principle of Avoiding Sure Losses: A rational gambler will never price gambles on the value of an unknown $\omega \in \Omega$ so that there is a sure loss.*

- of course, a rational gambler will accept a sure gain

- a combination of gambles that guarantees a loss for a gambler is known as a *Dutch book*

Lemma The buying and selling price of $X \in L(\Omega)$ must be the same.

Proof: Suppose a gambler would buy X for p_1 and sell X for p_2 where $p_1 > p_2$. This combination results in the gain

$$(X - p_1) + (p_2 - X) = p_2 - p_1 < 0$$

and the gambler has a sure loss. So by the principle of avoiding sure loss we must have $p_1 \leq p_2$.

If $p_1 < p_2$, then for any $\varepsilon > 0$ the gambler will not pay $p_1 + \varepsilon$ for X and will not sell X for $p_2 - \varepsilon$. But this combination of gambles has gain

$$(X - p_1 - \varepsilon) + (p_2 - \varepsilon - X) = p_2 - p_1 - 2\varepsilon$$

and this is positive when ε is small enough ensuring a sure gain. Since the gambler is rational, we must have that $p_1 = p_2$.

Definition The prevision P is *coherent* if for every $m, n \in \mathbb{N}$ and every $X_1, \ldots, X_m, Y_1, , \ldots, Y_n \in L(\Omega)$,

$$\sup_{\omega \in \Omega} \left\{ \sum_{i=1}^{m} (X_i(\omega) - P(X_i)) - \sum_{i=1}^{n} (Y_i(\omega) - P(Y_i)) \right\} \geq 0.$$

- a gambler with a coherent prevision will never have a sure loss on any finite combination of gambles

- the proofs of the following are in the text

Lemma The prevision P is coherent if and only if

$$\sup_{\omega \in \Omega} \sum_{i=1}^{n} \lambda_i (X_i(\omega) - P(X_i)) \geq 0 \qquad (*)$$

for all $\lambda_1, \ldots, \lambda_n \in R^1$.

- P has the properties of an expectation operator and conversely.

Theorem The prevision P is coherent if and only if
(i) $P(X+Y) = P(X) + P(Y)$ for every $X, Y \in L(\Omega)$,
(ii) $P(\lambda X) = \lambda P(X)$ for every $X \in L(\Omega), \lambda \in R^1$,
(iii) $P(X) \geq 0$ whenever $X \in L(\Omega)$ satisfies $X \geq 0$,
(iv) $P(1) = 1$.

- so restricting P to indicator functions implies that P corresponds to a *finitely* additive probability measure.

Corollary A coherent prevision P gives a finitely additive probability measure on 2^{Ω}. If u is paid for the gamble with payoff $u+v$ whenever $\omega \in A$ and 0 otherwise, then $P(A) = u/(u+v)$.

- a *contingent gamble* for $B \subset \Omega$ arises when a gambler buys or sells X for $P(X \,|\, B)$, with the purchaser receiving $X(\omega)$ when $\omega \in B$ and, when $\omega \notin B$, the bet is *called-off* with the price $P(X \,|\, B)$ returned to the purchaser with payoff $I_B(X - P(X \,|\, B))$

- $P(\cdot \,|\, B) : L(\Omega) \to R^1$ is called a *conditional prevision* and it is required that $P(\cdot \,|\, B)$ satisfy the Thm so that it is coherent.

Theorem If $\mathcal{B} = \{B_1, \ldots, B_m\}$ is a finite partition of Ω, then $P(X) = \sum_{i=1}^{m} P(B_i) P(X \,|\, B_i)$.

- some argue, based on the betting formulation for probability, that Kolmogorov (iii) should be weakened to only require finite additivity but then we lose continuity of P

Example *Uniform Probability on* $\mathbb{N}$

- If $\Omega = \mathbb{N}$, then define $P(A) = \lim_{n\to\infty} \#(A \cap \{1, \ldots, n\})/n$ whenever this limit exists

- it can be shown that P can be extended to a finitely additive probability measure on 2^{Ω}

- clearly $P(A) = 0$ for any finite set A while $P(A) = 1/2$ if A is the subset of even natural numbers and so $\sum_{i=1}^{\infty} P(\{2i\}) = 0 \neq 1/2$
- so P is not countably additive

- also $P(\mathbb{N} \mid \{i\}) \equiv 1$ and $1 = P(\mathbb{N}) \neq \sum_{i=1}^{\infty} P(\mathbb{N} \mid \{i\}) P(\{i\}) = 0$ so no TTP

- some argue, such as de Finetti, that the betting formulation is the *correct* interpretation for probability and that if you don't reason in situations of uncertainty according to its precepts, then you are *incoherent*

- but this has similar consequences to insisting probability is only concerned with long-run relative frequencies and the bettor/bookie formulation is unrealistic

- everything works fine in the finite case, and so presents way of thinking of probability assignments

- it requires utilities

- for the infinite case it provides interesting math but the general results are not relevant for statistics

- no notion of "randomness", is that a problem?