# Subjective Beliefs

Larry Blume

Cornell University & IHS Wien



2021-03-17 01:42:59-04:00

# What is Probability For?

- Evidential support

  "Having evaluated the data with our current meteorological models, I conclude there is a 75% chance of rain tomorrow."

- Degree of belief

  "I am 50% sure that the 100th digit of $\pi$ is 9."

- A fact about the world.

  "The probability that a radium atom decays in one year is roughly 0.0004."

# How To Measure Uncertainty

- qualitative probability
- Dutch book argument
- probability from preferences

- $S$ A state space (decision theory), sample space (statistics), universe of discourse (philosophy)
- Events are sets of states; subsets of $S$. The collection $\mathscr{S}$ of events has the following properties. If $A$ and $B$ are in $\mathscr{S}$, then

  If $A$ and $B$ are in $\mathscr{S}$, then so is $A \cap B$

  $A^c \in \mathscr{S}$.

  $S \in \mathscr{S}$.

  This allows us to derive $A \cup B = (A^c \cap B^c)^c \in \mathscr{S}$ and $\emptyset \in \mathscr{S}$.
- A probability measure is a function $p : \mathscr{S} \to [0, 1]$ with the following properties
  - $P(S) = 1$.
  - If $A$ and $B$ are disjoint, then $p(A \cup B) = p(A) + p(B)$.

  This allows us to derive $p(\emptyset) = 0$, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$, etc.

# Probability?

Three sources.

- Frequentist
- Subjective
- Logical

# Frequentist Probability

*The theory of chances consists in reducing all events of the same kind to a certain number of equally possible cases, that is to say, to cases whose existence we are equally uncertain of, and in determining the number of cases favourable to the event whose probability is sought. The ratio of this number to that of all possible cases is the measure of this probability, which is thus only a fraction whose numerator is the number of favourable cases, and whose denominator is the number of all possible cases.*

*Laplace (1814)*

# Frequentist Probability

*There is no place in our system for speculations concerning the probability that the sun will rise tomorrow. Before speaking of it we should have to agree on an (idealized) model which would presumably run along the lines "out of infinitely many worlds one is selected at random..." Little imagination is required to construct such a model, but it appears both uninteresting and meaningless.*

*William Feller (1957)*

# Frequentist Probability

The modern version of the frequentist approach is the law of large numbers. The LLN describes the outcome of repeatedly performing the same experiment. The average of the outcomes should be close to the expected value of the outcome, and closer as the number of repetitions increase. One version says that if the repeats are independently run, each with expected value $\mu$, and if the variances of the experimental outcomes are uniformly bounded, then

$$\lim_{n \to \infty} \mathrm{P}\left( \left| \frac{1}{n} \sum_{k=1}^{n} X_n - \mu \right| > \epsilon \right) = 0.$$

# Logical Probability

Were the dinosaurs killed by an asteroid? Probability is one way of describing how confident we are in the truth of this claim. This is what is meant by epistemic probability. In this scheme, $P(A|B)$ is the degree to which $B$ supports the claim $A$.

There are different views as to where these probabilities come from.

- Keynes, Carnap — probabilities are objective relations between sentences, and therefore are distinct from degree of belief.
- Ramsey sees probability as "the logic of partial belief."

It has proven very hard to get the logical view off the ground. For the last half-century or so the subjectivists have held sway.

# Subjective Probability

> *The so-called magnitudes or degrees of knowledge or probability, in virtue of which one is greater and another less, really arise out of an order in which it is possible to place them.*
>
> *J. M. Keynes*

Following Keynes we can imagine a binary order $\succ$ on $\mathscr{S}$ whose interpretation is “more likely than.” Keynes imagines this is a partial order and discusses the possibility of partial numerical representations.

# Qualitative Probability

**Definition.** A qualitative probability structure is a triple $(S, \mathscr{S}, \succ)$ such that

- A.1. $\succ$ is a preference relation;
- A.2. For all $A \in \mathscr{S}$, $\emptyset \not\succ A$;
- A.3. $S \succ \emptyset$.
- A.4. if $A$ is disjoint from both $B$ and $C$, then $C \not\succ B$ iff $A \cup C \not\succ A \cup B$.

**Definition.** A probability representation for $\succ$ is a representation for $\succ$ that respects disjoint unions: If $A$ and $B$ are disjoint, then

$$p(A \cup B) = p(A) + p(B).$$

# Properties of Representations

**Theorem.** If $p$ represents $\succ$ satisfying A.1–4, then

a) $p(\varnothing) = 0$,

b) $p(S) > 0$.

c) if $A \subset B$ then $p(A) \leq p(B)$,

**Proof.**

a) $\varnothing \cap \varnothing = \varnothing$ so A.4 implies $p(\varnothing) = p(\varnothing \cup \varnothing) = 2p(\varnothing)$.

b) A.3.

c) If $A \subset B$, then $B = A \cup B/A$. A.2 has $B/A \succsim \varnothing$ so A.4 implies $B \equiv B/A \cup A \succsim A$.

The representation is preserved upon multiplication by a positive constant, so w.l.o.g. take $p(S) = 1$.

# Qualitative Probability

If $S$ is finite, a numerical representation always exists. Does A.4 guarantee the existence of a probability representation?

**NO!**

There is a five-state counterexample. See Krantz, Luce, Suppes and Tversky (1971) sec. 5.2.2.

# Existence Of A Probability Representation

A.5. For any two lists of sets $\mathscr{A} = \{A_1, \ldots, A_Q\}$ and $\mathscr{B} = \{B_1, \ldots, B_Q\}$ in $\mathscr{S}$ such that

a) for all $q < Q$ $A_q \succcurlyeq B_q$,

b) for some $q < Q$, $A_q \succ B_q$,

c) for all $s$, $\#\{A_q : s \in A_q\} = \#\{B_q : s \in B_q\}$,

then $a_Q \prec b_Q$.

Claim. A.5 implies A.4. Take $\mathscr{A} = \{A, B \cup C\}$, $\mathscr{B} = \{B, A \cup C\}$, and suppose first $A \succ B$, then $A \cup C \succ B \cup C$.

Theorem. A qualitative probability order has a probability representation iff it satisfies axioms A.1–3 and A.5.

# A Theorem Of The Alternative

This is a variation on Farkas' Lemma.

Lemma. Let $v_1 \ldots, v_m$ and $w_1, \ldots, w_n$ be vectors in a $k$-dimensional Euclidean space. One and only one of the following two inequality systems has a solution:

$$(1) \quad \begin{aligned} v_i \cdot x > 0 &\text{ for all } i \le m, \\ w_j \cdot x = 0 &\text{ for all } j \le n, \\ x \in \mathbb{R}^k& \end{aligned} \qquad (2) \quad \begin{aligned} \sum_1^m y_i v_i + \sum_1^n z_j w_j = 0, \\ \sum_i y_i > 0 \\ y \in \mathbb{R}^m_+, z \in \mathbb{R}^n \end{aligned}$$

This is an immediate consequence of Farkas' Lemma, noting that for a homogeneous system, $Ax > 0$ has a solution iff $Ax \ge e$ has a solution. ($e$ is the vector of all 1s.)

# Existence Of A Probability Representation

Proof of the Theorem. Label the states 1 through $S$. Each state is $A$ is represented by its indicator vector $r^A$ such that $r_s^A = 1$ if $s \in A$ and 0 otherwise. Let $V$ be the matrix whose rows $r^A - r^B$ where $A \succ B$. Let $W$ be the matrix whose rows are $r^A - r^B$ where $A \sim B$. Have each equally likely pair appear only once, and if $\{s\} \sim \emptyset$, the corresponding $w_j$ is $r^s$. Consider a solution $x$ to the inequality system.

$$Vx > 0, \quad Wx = 0.$$

Since each $\{s\} \not\succ \emptyset$, if $r^{\{s\}}$ is not a row in $V$, then it can appear in $W$, and so $x \geq 0$. Thus $x$ can be scaled to be a probability distribution.

A solution $x$ does not exist iff there is a $y \geq 0$ and a $z$ such that $\sum_1^k y_i v_i + \sum_j z_j w_j = 0$, $y \geq 0$ and $\sum_i y_i = 1$. W.l.o.g., $z_j$ can be made non-negative by choice of the sign for each $w_j$. The coefficients of $y$ and $z$ in this inequality system are either 1, 0, or $-1$. By Gaussian elimination one can see that if this inequality system has a solution, it has a rational, and hence an integer solution.

Each $v_i$ and $w_j$ is difference $a_i - b_i$ and $a_j - b_j$, where each vector in the differences is an indicator vector for some set. Now build the list $A$ by taking $y_1$ copies of the set whose indicator vector is $a_1$, followed by $y_2$ copies of the set whose indicator vector is $a_2$, and so forth through all the $v$'s and then through all the $w$'s. Construct list $B$ the same way. Then each element of $A$ is at least as good as the corresponding element of $B$. Since $\sum_i y_i = 1$, at least one $y_i > 0$ and so there is some element in $A$ better than it's corresponding element of $B$, and the remaining equality in (2) guarantees that the total number of times each $s$ appears in the $A$ list equals the total in $B$. So this list violates A.5. ■

# Why Probability?

There is nothing natural about A.5, so why should we use probability to represent belief?

# Making Book

What does it mean to say, “I will give you 2 to 1 odds on classes being in-person in the fall?"

# Making Book

What does it mean to say, “I will give you 2 to 1 odds on classes being in-person in the fall?"

I would accept a deal that pays $1 if classes open in the fall and costs me $2 otherwise.

I put $2 in an envelope and you put in $1. Whoever wins the bet gets the envelope. The envelope’s contents is the stake. Here the stake is $3.

# Making Book

I give you \$2. You give me \$3 if classes open, and nothing otherwise.

Define the price of the bet to be

$$p = \frac{\text{potential loss}}{\text{stake}} \qquad = \frac{2}{3}$$

\$0.66 is the most I would pay for a claim on \$1 if in-person instruction returns in the fall.

# How to Make Money Making Book


Alexei and Blanche want to bet on Silver Blaze in a coming race. Alexei believes with probability 3/4 that Silver Blaze will win. Blanche believes Silver Blaze will win with probability only 1/4. The bookmaker can sell Alexei a ticket that pays off \$1 if Silver Blaze wins for a price of $1/2+\epsilon$. He can sell Blanche a ticket that pays off \$1 if Silver Blaze loses for a price of 1/2. The book looks like this:

| | SB Wins | SB Loses |
|---|---|---|
| A | $1/2+\epsilon-1$ | $1/2+\epsilon$ |
| B | $1/2$ | $1/2-1$ |
| Total | $\epsilon$ | $\epsilon$ |

This is a Dutch book.

# Dutch Books And Rational Beliefs

The book maker makes a sure profit because Alexei's and Blanche's beliefs are inconsistent. If a single individual has inconsistent beliefs, a book maker can make a profit by offering him bets on different events.

A betting system is a set of prices $p(E)$ for a bet on each event $E \in \mathcal{S}$. It is fair if it does not offer a sure profit. It is strictly fair if it does not offer a loss, and sometimes offers a profit.

Theorem. A betting system is (strictly) fair iff there is a (strictly positive) vector $q$ of probabilities on states $s \in \mathcal{S}$ such that for all events $E \in \mathcal{S}$, $p(E) = \sum_{s \in E} q(s)$.

# Proof

The Proof relies on the following fact:

Lemma: One and only one of the following two systems has a solution:

$$Ax \gg 0 \qquad\qquad \begin{aligned} yA &= 0 \\ y &> 0 \end{aligned}$$

and if $\gg$is weakened to $>$ in the first system, then then $<$ is strengthened to $\ll$ in the second system.

The lemma is a consequence of Farkas' lemma.

Notation: $\ll$ strictly less than; $<$ at least as small as and not equal to, and similarly for $>$ and $\gg$.

# Proof

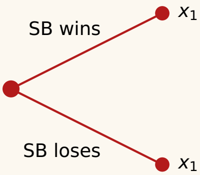
For each $E \in \mathscr{S}$ let $E(s)$ denote the indicator function which is 1 for $s \in E$ and 0 otherwise. Let $P_E$ denote the price of the bet on $E$. Let $A$ be a $S \times |\mathscr{S}|$ matrix with columns indexed by events and rows by states, and entries $a_{Es} = p_E - E(s)$. The bookmaker can buy or sell to the bettor a portfolio $z$ of bets on events, where $z_E$ is the stake in the event $E$ and $z_E$ is positive if he sells and negative if he buys. Then $Az$ is the vector whose $s$'th element is his earnings in state $s$. The betting system is fair if there is no $Z$ such that $Az \gg 0$.

If the inequality has no solution, then there is a $q > 0$ such that $qA = 0$ and that for each $E \in \mathscr{S}$, $\sum_s q_s p_E = \sum_s q_s E(s)$. Since the system is homogeneous we can take $\sum_s q_s = 1$. Then $p_E = \sum_{s \in E} q_s$, and so $p_E$ is the probability of $E$ given by the probability distribution $q$.

For the case of strict fairness, the second set of alternatives applies, and gives a strictly positive $q$. ■

# Semi-Subjective Expected Utility

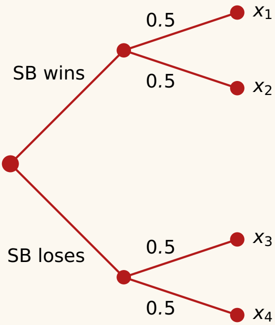
Suppose we have gambles of the form




How would probabilities appear?

How could risk be quantified?

Anscombe and Aumann's idea is to mix horse race gambles together with roulette wheel gambles which can be used to measure the risk.



# The Choice Problem

There is a finite set $S$ of states, and a finite set $X$ of prizes. Denote by $\mathscr{R}$ the set of roulette wheels, probability distributions on $X$. A horse race is a function $f : S \to \mathscr{R}$. The set of horse races is $\mathscr{H}$. The decision maker has a preference order $\succ$ on $\mathscr{H}$.

Convex combinations of horse races are horse races $(f\alpha g)(s) = \alpha f(s) + (1-\alpha)g(s)$. So $\mathscr{H}$ with convex combinations is a mixture space.

A.1. $\succ$ on $H$ is a preference relation.

A.2. if $h \succ k$ then for any $g$ and $0 < \alpha \leq 1$, $h\alpha g \succ k\alpha g$.

A.3. the Archimedean axiom.

# The Representation Theorem

Theorem. There exist functions $u_s : X \to \mathbb{R}$ such that $h \succ k$ iff

$$\sum_{s \in S} \sum_{x} u_s(x) h(s)(x) > \sum_{s \in S} \sum_{x} u_s(x) k(s)(x)$$

If the functions $v_s$ give another representation, then $v_s = a u_s + b_s$ where $a > 0$.

Proof. From the mixture space theorem we know there is a mixture-preserving function $V : \mathscr{H} \to \mathbb{R}$. We do some calculations to see that it has the desired form. See Kreps Ch 8. ■

Notice! We don't yet have probabilities.

# From State-Dependent EU to State-Independent EU

A.4. Preferences are non-trivial; there are $h, k$ such that $h \succ k$.

Definition. A state $s$ is null if for all $h$, $k$ such that $h(t) = k(t)$ for $t \neq s$, $h \sim k$.

If $s$ is null, then $u_s$ is a constant function. A.4 implies some state is not null.

A.5. (Non-null state independence). Suppose that $h$ and $k$ are two acts such that $h(t) = k(t)$ for all $t \neq s$, $h(s) = p$, and $k(s) = q$, and suppose that $h \succ k$. Then for any non-null $s'$, for any $h'$ and $k'$ that agree except on $s'$, $h'(s') = p$ and $k'(s') = q$, $h' \succ k'$.

# From State-Dependent EU to State-Independent EU

Theorem. If $\succ$ satisfies A.1–5, then there is a payoff function $u : X \to \mathbb{R}$ and a probability distribution $p$ on $S$ such that

$$h \succ k \text{ iff } \sum_s p(s) \sum_x u(x) h(s)(x) > \sum_s p(s) \sum_x u(x) k(s)(x).$$

Furthermore, if $(q, v)$ also represent $\succ$, then $q = p$ and $v = au + b$ with $a > 0$.

# Savage

A & A assign probabilities by comparing bets on states to gambles with objective probabilities. But suppose there are none such?

Savage has two ideas:

- Suppose we could divide the state space into $2^n$ equally likely events. Then each event is like the outcome of a flip of $n$ coins.

- If preferences over outcomes are independent of the state of the world in which they occur, then the likelihood of events can be determined by how the DM ranks bets on them. If she likes skiing, and if she prefers a ski trip in event $A$ to a ski trip in event $B$, then we infer that she believes $A$ is more likely than $B$.

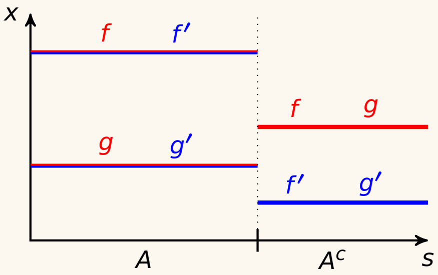
See Kreps Ch 8 pp 120–25.

# Savage

A.1. $\succ$ is a preference relation.

A.2. There are two constant acts $x$ and $y$ such that $x \succ y$.

A.3. In the picture, $f \succ g$ iff $f' \succ g'$.



# Savage

A.3 allows us to define conditional preferences $f \succ_A g$ by changing $f$ and $g$ to a common consequence on $A^c$. Each $\succ_A$ will be a preference relation.

Definition. A state is null if $f \sim_A g$ for all $f, g$.

A.4. (Non-null state-independence.) If $A$ is not null, then $x \succ_A y$ iff $x \succ y$.

# Savage

A.5. Suppose $x \succ y$ and $x' \succ y'$. Suppose $f = x$ on $A$ and $y$ on $A^c$, and $f' = x$ on $A$ and $y'$ on $A^c$. Suppose $g = x$ on $B$ and $y$ on $B^c$, and $g' = x$ on $B$ and $y'$ on $B^c$. Then $f \succ g$ iff $f' \succ g'$.

This gives us a well-defined “more likely than” relation on subsets of $S$.

A.6 (Sure thing principle.) For all $A \subset S$: If for all $s \in A$ $f \succ g(s)$ then $f \succcurlyeq_A g$. If for all $s \in A$ $g(s) \succ f$ then $g \succcurlyeq_A f$.

# Savage — The Denouement

All of the above, and a few more axioms to make the subdivision work, give us the following conclusion:

Theorem.

a) There is a qualitative probability $\gg$ relation on the subsets of $S$ and a unique probability distribution $p$ such that $A \gg B$ iff $p(A) > p(B)$.

b) For all sets $A$ and $0 \leq r \leq 1$ there is a $B \subset A$ such that $p(B) = rp(A)$.

c) There is a bounded utility function $u : X \to \mathbb{R}$ such that $f \succ g$ iff

$$\mathrm{E}\left\{u(f(s))\right\} > \mathrm{E}\left\{u(g(s))\right\}.$$

Furthermore, $u$ is unique up to positive affine transformations.

# Comparisons

**Dutch books** Does not inform situations where DM's don't bet. "How likely is it that Grandmother is going to die tomorrow?" "Gee, I don't know, let's call up Blume's cousin Earl and make book on it."

**A&A** (1) We are really only concerned with probabilities over horse races that pay off in sure things. Non-trivial roulette wheel acts are fictional. (2) A preexisting probability space bootstraps the probabilities over horse races. Probability not derived — it is a primitive.

**Savage** Roulette wheels are gone, but in its place is a gigantic state space for even the most simple problems. Where does that come from? Savage says, add arbitrary length sequences of coin flips. Probability on basic states is still bootstrapped.

# Conditional Probability

Kolmogorov defines conditional probabilities as follows:

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

Example: A disease infects 1% of the population. A test for the disease gives a correct yes 99% of the time and a correct no 98% of the time. What is the conditional probability of being ill having received a positive test result?

$$\begin{aligned}
\mathrm{P}(\text{ill}|\text{test}+) &= \frac{\mathrm{P}(\text{test}+|\,\text{ill})\,\mathrm{P}(\text{ill})}{\mathrm{P}(\text{test}+)} \\
&= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.01 \times 0.99} \\
&= 1/2
\end{aligned}$$

# Bayes Rule

$$\mathrm{P}(A)\,\mathrm{P}(B|A) = \mathrm{P}(B)\,\mathrm{P}(A|B)$$

Conditioning provides a way of revising probabilistic beliefs. What is wrong with this?

- It only accounts for partitional information.

  A friend says to me, “The likelihood that the 100th digit of $\pi$ is odd is 80%." How should I change my initial belief

# Bayes Rule

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

Conditioning provides a way of revising probabilistic beliefs. What is wrong with this?

- It cannot respond to learning something that was *ex ante* impossible.

# Alternative Foundations for Probability

$(S, \mathscr{S})$ are given as before. Also given is a collection of non-empty sets $\mathscr{B} \subset \mathscr{S}$. A conditional probability system is a function $p : S \times \mathscr{B} \to [0, 1]$ such that

- For each $B \in \mathscr{B}$, $p(\cdot|B) : \mathscr{S} \to [0, 1]$ satisfies the disjoint union property, and that $p(B|B) = 1$.
- If $A, B \in \mathscr{S}$, $C \in \mathscr{B}$ and $B \cap C \in \mathscr{B}$, then

$$p(A|B \cap C) \cdot p(B|C) = p(A \cap B|C).$$

This is consistent with Kolmogorov's axioms and definitions of probability and conditional probability, and extends it since it defines conditional probability when $B$ has Kolmogorov 0-probability.

# Dutch Books and Conditional Probability

One can introduce the notion of a bet conditional on F, that pays off 1 on $E \cap F$, 0 on $F/E$, and returns the price $p_{E|F}$ on $F^c$. The same proof applied to these bets shows that no Dutch book can be made against a decision maker if and only if there is a probability distribution $q$ on $\mathscr{S}$ such that $p_{E|F} = q\{E|F\}$.

# Expected Utility and Conditional Probability

Suppose we have acts $f$ and $g$, discover that $s \in B$, and want to compute the conditional preference. Savage says modify $f$ and $g$ to a common consequence $h$ on $B^c$. The expected utility comparison of $f$ to $g$ given $B$ is the sign of

$$\left\{\sum_{s\in B} u(f(s))p(s) + \sum_{s\in B^c} u(h(s))p(s)\right\} - \left\{\sum_{s\in B} u(g(s))p(s) + \sum_{s\in B^c} u(h(s))p(s)\right\}$$

$$= \sum_{s\in B} u(f(s))p(s) - \sum_{s\in B} u(g(s))p(s)$$

$$= \sum_{s\in B} u(f(s))\frac{p(s)}{\sum_{t\in B} p(t)} - \sum_{s\in B} u(g(s))\frac{p(s)}{\sum_{t\in B} p(t)}$$

$$= \mathrm{E}\left\{u(f(s))|B\right\} - \mathrm{E}\left\{u(g(s))|B\right\}$$