Phil/LPS 31 Introduction to Inductive Logic
Lecture 17

David Mwakima
dmwakima@uci.edu
Department of Logic and Philosophy of Science
University of California, Irvine

June 5th 2023

Topics

- Subjective Bayesian Interpretation of Probability
- Evaluating Evidence within the Two Interpretations

Motivation for the subjective view

- The subjective interpretation of probability is partly motivated by the problem of the single case, which is about the applicability of the frequentist interpretation of probability.

Motivation for the subjective view

- The subjective interpretation of probability is partly motivated by the problem of the single case, which is about the applicability of the frequentist interpretation of probability.
- Recall that this was the problem raised by the fact that we talk about the probability of single and non-repeatable events.

Motivation for the subjective view

- The subjective interpretation of probability is partly motivated by the problem of the single case, which is about the applicability of the frequentist interpretation of probability.
- Recall that this was the problem raised by the fact that we talk about the probability of single and non-repeatable events.
- Moreover, different people can attach different probability judgments on the likelihood of events. For example, I thought that the probability that Manchester City would win this past weekend's derby was 50% given how well Manchester United were currently playing. But you could have thought that this probability was higher than 50%, say 80%, because you think Pep Guardiola is a better coach than Erik ten Hag. Is there an objective fact about whose belief is right?

Probabilities don't exist (objectively)

- The subjective interpretation of probability is jointly attributed to the work of Bruno de Finetti, Frank Ramsey, and Leonard Savage.

Probabilities don't exist (objectively)

- The subjective interpretation of probability is jointly attributed to the work of Bruno de Finetti, Frank Ramsey, and Leonard Savage.
- de Finetti is famous for saying:

Probabilities don't exist (objectively)

- The subjective interpretation of probability is jointly attributed to the work of Bruno de Finetti, Frank Ramsey, and Leonard Savage.
- de Finetti is famous for saying:

  My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this: PROBABILITY DOES NOT EXIST . . . [I]n the conception we follow and sustain here, only subjective probabilities exist – i.e., the degree of belief in the occurrence of an event attributed by a given person at a given instant and with a given set of information.

Coherent fair betting quotients are subjective probabilities

- All three authors proposed essentially the same approach to a subjective probability, which has roots in Rev. Thomas Bayes's essay from 1763.

Coherent fair betting quotients are subjective probabilities

- All three authors proposed essentially the same approach to a subjective probability, which has roots in Rev. Thomas Bayes's essay from 1763.
- For an event E, subjective epistemic probability of E, $P(E)$, is the rate at which an individual is willing to bet on the occurrence E. So fair betting rates or quotients are measurements that reveal your subjective probabilities or someone else's probabilities.

Coherent fair betting quotients are subjective probabilities

- All three authors proposed essentially the same approach to a subjective probability, which has roots in Rev. Thomas Bayes's essay from 1763.
- For an event E, subjective epistemic probability of E, $P(E)$, is the rate at which an individual is willing to bet on the occurrence E. So fair betting rates or quotients are measurements that reveal your subjective probabilities or someone else's probabilities.
- Starting with this idea, some of the subjective theorists I have mentioned proved what are known as representation theorems

Coherent fair betting quotients are subjective probabilities

- All three authors proposed essentially the same approach to a subjective probability, which has roots in Rev. Thomas Bayes's essay from 1763.
- For an event E, subjective epistemic probability of E, $P(E)$, is the rate at which an individual is willing to bet on the occurrence E. So fair betting rates or quotients are measurements that reveal your subjective probabilities or someone else's probabilities.
- Starting with this idea, some of the subjective theorists I have mentioned proved what are known as representation theorems
- Essentially what representation theorems show is that an agent's preference ordering on bets she is willing to make represent that agent's subjective graded beliefs about the events the bets are on.

Coherent fair betting quotients are subjective probabilities

- Although these graded beliefs, degrees of belief or epistemic probabilities are purely subjective; you have seen in Homework 7 that these beliefs have to be constrained by coherence.

Coherent fair betting quotients are subjective probabilities

- Although these graded beliefs, degrees of belief or epistemic probabilities are purely subjective; you have seen in Homework 7 that these beliefs have to be constrained by coherence.
- The constraint of coherence says:

Coherent fair betting quotients are subjective probabilities

- Although these graded beliefs, degrees of belief or epistemic probabilities are purely subjective; you have seen in Homework 7 that these beliefs have to be constrained by coherence.
- The constraint of coherence says:

  A person's graded beliefs are coherent if and only if they are consistent with the Kolmogorov Axioms (or the rules of probability).

Coherent fair betting quotients are subjective probabilities

- Although these graded beliefs, degrees of belief or epistemic probabilities are purely subjective; you have seen in Homework 7 that these beliefs have to be constrained by coherence.
- The constraint of coherence says:

  A person's graded beliefs are coherent if and only if they are consistent with the Kolmogorov Axioms (or the rules of probability).

- You have also seen in your homework that one argument why the constraint of coherence is desirable is the Dutch-Book Argument. This argument shows that if an agent has incoherent graded beliefs, then it is possible for them to enter into a sure-loss contract (i.e., a Dutch Book).

Subjective Bayesian Interpretation of Probability

- The subjective interpretation of probability becomes a subjective Bayesian interpretation of probability when one adds Bayes' theorem as an inductive rule.

$$
\begin{aligned}
P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\
&= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}
\end{aligned}
$$

Subjective Bayesian Interpretation of Probability

- The subjective interpretation of probability becomes a subjective Bayesian interpretation of probability when one adds Bayes' theorem as an inductive rule.

$$
\begin{aligned}
P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\
&= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}
\end{aligned}
$$

- The subjective Bayesian interpretation of probability also goes by the name of Bayesianism or Bayesian confirmation theory when it is used as a philosophical approach to questions in the theory of knowledge and confirmation theory.

Bayesian Confirmation Theory

- In the context of Bayesianism, the terms that appear in Bayes' Theorem receive special names, which you may see in the literature.

$$
\begin{aligned}
P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\
&= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}
\end{aligned}
$$

Bayesian Confirmation Theory

- In the context of Bayesianism, the terms that appear in Bayes' Theorem receive special names, which you may see in the literature.

$$\begin{aligned} P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\ &= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)} \end{aligned}$$

- H is called a hypothesis; E is called evidence.

Bayesian Confirmation Theory

- In the context of Bayesianism, the terms that appear in Bayes' Theorem receive special names, which you may see in the literature.

$$\begin{aligned} P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\ &= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)} \end{aligned}$$

- H is called a hypothesis; E is called evidence.
- $P(H)$ is called the prior probability of H. It is the epistemic probability one has about H before one sees the evidence.

Bayesian Confirmation Theory

- In the context of Bayesianism, the terms that appear in Bayes' Theorem receive special names, which you may see in the literature.

$$
\begin{aligned}
P(H \mid E) &= \frac{P(E|H)P(H)}{P(E)} \\
&= \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}
\end{aligned}
$$

- H is called a hypothesis; E is called evidence.
- $P(H)$ is called the prior probability of H. It is the epistemic probability one has about H before one sees the evidence.
- $P(H|E)$ is called the posterior probability of H. It is the epistemic probability one has about H after one sees the evidence.

Bayesian Confirmation Theory as an Inductive Logic

- The inductive rule of going from $P(H)$ to $P(H|E)$ through Bayes' theorem is known as conditionalization or updating (on the evidence) and it is the other distinctive feature of Bayesianism apart from its subjective interpretation of probability.

Bayesian Confirmation Theory as an Inductive Logic

- The inductive rule of going from $P(H)$ to $P(H|E)$ through Bayes' theorem is known as conditionalization or updating (on the evidence) and it is the other distinctive feature of Bayesianism apart from its subjective interpretation of probability.
- There are different kinds of conditionalization or updating rules. We shall not cover all the different kinds in this introductory class. The thing they have in common is that they are all inductive rules of learning from what has been observed.

Bayesian Confirmation Theory as an Inductive Logic

- The inductive rule of going from $P(H)$ to $P(H|E)$ through Bayes' theorem is known as conditionalization or updating (on the evidence) and it is the other distinctive feature of Bayesianism apart from its subjective interpretation of probability.
- There are different kinds of conditionalization or updating rules. We shall not cover all the different kinds in this introductory class. The thing they have in common is that they are all inductive rules of learning from what has been observed.
- Bayesian Confirmation Theory is an inductive logic in our sense! Although it is not perfect, it is progress.

Advantages of the Subjective Bayesian Interpretation

- Provided a person's epistemic probabilities or graded beliefs are coherent, the subjective Bayesian interpretation of probability is a faithful interpretation of the Kolmogorov Axioms.

Advantages of the Subjective Bayesian Interpretation

- Provided a person's epistemic probabilities or graded beliefs are coherent, the subjective Bayesian interpretation of probability is a faithful interpretation of the Kolmogorov Axioms.
- It avoids the problem of the single case, which was an obstacle in the application of the frequentist interpretation.

Advantages of the Subjective Bayesian Interpretation

- Provided a person's epistemic probabilities or graded beliefs are coherent, the subjective Bayesian interpretation of probability is a faithful interpretation of the Kolmogorov Axioms.
- It avoids the problem of the single case, which was an obstacle in the application of the frequentist interpretation.
- There are lots of advantages of using Bayesian methods in statistical decision theory, but that subject is beyond the scope of this introductory class.

Disadvantages of the Subjective Bayesian Interpretation

- The main disadvantage of the subjective Bayesian interpretation is that it requires us to have prior epistemic probabilities. Often two or more individuals can differ radically on their prior epistemic probabilities about a hypothesis H. In this case, even for the same evidence, they could still differ on their posterior epistemic probabilities. (Is this a feature or a bug?)

Disadvantages of the Subjective Bayesian Interpretation

- The main disadvantage of the subjective Bayesian interpretation is that it requires us to have prior epistemic probabilities. Often two or more individuals can differ radically on their prior epistemic probabilities about a hypothesis H. In this case, even for the same evidence, they could still differ on their posterior epistemic probabilities. (Is this a feature or a bug?)
- Although there are well-known responses to this disadvantage (these are the merging of opinion theorems), a more serious problem for Bayesianism is the problem of the Catch-All Hypothesis

The problem of the Catch-All Hypothesis

- Bayes' Theorem looks simple when we're just looking at two competing hypotheses H and H^c.

$$P(H \mid E) = \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}$$

The problem of the Catch-All Hypothesis

- Bayes' Theorem looks simple when we're just looking at two competing hypotheses H and H^c.

$$P(H \mid E) = \frac{P(E|H)P(H)}{P(E \mid H)P(H) + P(E \mid H^c)P(H^c)}$$

- Typically, however, we don't know whether the set of hypotheses we are considering is a genuine partition of the space of hypotheses. So in order to apply the general form of Bayes' theorem to update our graded belief about a hypothesis, say H_1:

$$P(H_1 \mid E) = \frac{P(E|H_1)P(H_1)}{\sum_{i=1}^{n} P(H|E_i)P(H_i)}$$

we need not only $P(H_1)$ and $P(H_2)$ but also the probability of every other hypothesis besides H_1 and H_2. This probability of "everything else besides" is the probability of the catch-all.

Frequentist Measure of Evidence

- We can now compare the two interpretations of probability with respect to how they quantify or evaluate statistical evidence in the context of statistical hypothesis testing or model comparison.

Frequentist Measure of Evidence

- We can now compare the two interpretations of probability with respect to how they quantify or evaluate statistical evidence in the context of statistical hypothesis testing or model comparison.
- Let's begin by fixing our terms and notation. H_0 denotes the null hypothesis, which is the hypothesis we wish to test with our data. H_1 denotes the alternative hypothesis, which we would consider if the null hypothesis is inconsistent with the data.

Frequentist Measure of Evidence

- We can now compare the two interpretations of probability with respect to how they quantify or evaluate statistical evidence in the context of statistical hypothesis testing or model comparison.
- Let's begin by fixing our terms and notation. H_0 denotes the null hypothesis, which is the hypothesis we wish to test with our data. H_1 denotes the alternative hypothesis, which we would consider if the null hypothesis is inconsistent with the data.
- In testing statistical hypotheses, H_0 may say that a parameter $\boldsymbol{\theta}$ about the population is in some set Ω_0 or Ω_1 where $\{\Omega_0, \Omega_1\}$ is a partition of the parameter space Ω. So we are testing:

$$H_0 : \boldsymbol{\theta} \in \Omega_0$$
$$H_1 : \boldsymbol{\theta} \in \Omega_1$$

Frequentist Measure of Evidence

- In evaluating statistical evidence within the frequentist interpretation of probability, we do so with the goal of minimizing two kinds of error probabilities. Type I and Type II error probabilities.

Frequentist Measure of Evidence

- In evaluating statistical evidence within the frequentist interpretation of probability, we do so with the goal of minimizing two kinds of error probabilities. Type I and Type II error probabilities.
- A Type I error probability is the probability of rejecting H_0 when $\boldsymbol{\theta} \in \Omega_0$. A Type II error probability is the probability of not rejecting H_0 when $\boldsymbol{\theta} \in \Omega_1$. Think of the error probabilities as the risk with a 0-1 loss function.

	$\boldsymbol{\theta} \in \Omega_0$	$\boldsymbol{\theta} \in \Omega_1$
Reject H_0	1	0
Do not reject H_0	0	1

Frequentist Measure of Evidence

- The weights for calculating the risk depend on the probability of seeing data X either when $\boldsymbol{\theta} \in \Omega_0$ or when $\boldsymbol{\theta} \in \Omega_1$. This probability will come from the distribution of your observed sample. Known as the sampling model, $f(X)$.

Frequentist Measure of Evidence

- The weights for calculating the risk depend on the probability of seeing data X either when $\boldsymbol{\theta} \in \Omega_0$ or when $\boldsymbol{\theta} \in \Omega_1$. This probability will come from the distribution of your observed sample. Known as the sampling model, $f(X)$.
- Let α be the maximum risk of rejecting H_0 when $\boldsymbol{\theta} \in \Omega_0$, i.e., the Type I error rate. α is a measure of how often we expect to make a Type I error if we could repeat the experiment many many times under identical conditions. It is usually chosen in accordance to some optimality criterion to be small, usually 0.05 or 0.001.

Frequentist Measure of Evidence

- The weights for calculating the risk depend on the probability of seeing data X either when $\boldsymbol{\theta} \in \Omega_0$ or when $\boldsymbol{\theta} \in \Omega_1$. This probability will come from the distribution of your observed sample. Known as the sampling model, $f(X)$.
- Let α be the maximum risk of rejecting H_0 when $\boldsymbol{\theta} \in \Omega_0$, i.e., the Type I error rate. α is a measure of how often we expect to make a Type I error if we could repeat the experiment many many times under identical conditions. It is usually chosen in accordance to some optimality criterion to be small, usually 0.05 or 0.001.
- The frequentist interpretation will find evidence against H_0 if the probability of seeing data X when $\boldsymbol{\theta} \in \Omega_0$ is less than α. This probability is known as the p-value. It is the main measure of statistical evidence from a frequentist perspective.

Bayesian Measure of Evidence

- In quantifying statistical evidence from a Bayesian perspective, we use Bayes' Theorem, we get the following relation.(How?)

$$\text{posterior odds} \propto \text{likelihood} \times \text{prior odds}$$

Bayesian Measure of Evidence

- In quantifying statistical evidence from a Bayesian perspective, we use Bayes' Theorem, we get the following relation.(How?)

$$\text{posterior odds} \propto \text{likelihood} \times \text{prior odds}$$

- Suppose $\frac{\pi(H_0)}{\pi(H_1)}$ measures our **prior odds** for two hypothesis H_0 and H_1. Then we expect the **posterior odds** $\frac{p(H_0|X)}{p(H_1|X)}$ to be given by:

$$\frac{p(H_0|X)}{p(H_1|X)} = U\frac{\pi(H_0)}{\pi(H_1)}$$

Bayesian Measure of Evidence

- In quantifying statistical evidence from a Bayesian perspective, we use Bayes' Theorem, we get the following relation.(How?)

$$\text{posterior odds} \propto \text{likelihood} \times \text{prior odds}$$

- Suppose $\frac{\pi(H_0)}{\pi(H_1)}$ measures our **prior odds** for two hypothesis H_0 and H_1. Then we expect the **posterior odds** $\frac{p(H_0|X)}{p(H_1|X)}$ to be given by:

$$\frac{p(H_0|X)}{p(H_1|X)} = U\frac{\pi(H_0)}{\pi(H_1)}$$

- Bayes Factor B_{01} is the updating factor U that quantifies the evidence in favor of M_0. (How?)

$$B_{01} = \frac{f(X|H_0)}{f(X|H_1)}$$

Comparing the two measures

- The frequentist measure of evidence involves minimizing risk in making a decision to reject or to fail to reject a null hypothesis. It is thus less straightfoward than the Bayesian measure of evidence where no such risk calculations are made as of this day.

Comparing the two measures

- The frequentist measure of evidence involves minimizing risk in making a decision to reject or to fail to reject a null hypothesis. It is thus less straightfoward than the Bayesian measure of evidence where no such risk calculations are made as of this day.
- Because the Bayesian measure of evidence has no associated risk involved with it, philosophers object that is possible to draw erroneous conclusions for claims that have been poorly tested. See Deborah Mayo *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*

Comparing the two measures

- The frequentist measure of evidence involves minimizing risk in making a decision to reject or to fail to reject a null hypothesis. It is thus less straightfoward than the Bayesian measure of evidence where no such risk calculations are made as of this day.
- Because the Bayesian measure of evidence has no associated risk involved with it, philosophers object that is possible to draw erroneous conclusions for claims that have been poorly tested. See Deborah Mayo *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*
- The frequentist measure of evidence is not a conditional probability. This is because on the frequentist interpretation of probability, hypotheses are not random variables (i.e., cannot be modeled as repeatable events).

Comparing the two measures

- The Bayesian measure of evidence bears directly on our beliefs regarding hypotheses through the updating feature provided by Bayes' theorem. The Frequentist measure of evidence doesn't bear on our beliefs at all! Instead, it is a decision criterion based on minimizing risk.