Bridges between Frequentist Statistics and Objective Bayesianism

Jon Williamson

Multiplicity and Unification in Statistics and Probability
25 June 2009

Contents

1 The Statistical View

Stats works like this:

1. Conceptualise the problem. Isolate a set $\mathbb{M}$ of models for consideration.
    - Models are probability functions.
2. Gather evidence $\mathcal{E}$.
3. Apply statistical methods to evaluate models in $\mathbb{M}$ in the light of $\mathcal{E}$.
    - Inferences and decisions will be made on the basis of a set $\mathbb{E} \subseteq \mathbb{M}$ of models that are appropriate given $\mathcal{E}$.

Frequentist stats works like this:

1. Conceptualise the problem. Isolate a set $\mathbb{M}$ of models for consideration.
   - $\mathbb{M}$ is a set of candidates for physical probability P^*.
2. Gather evidence $\mathcal{E}$.
3. Apply statistical methods to evaluate models in $\mathbb{M}$.
   - Inferences and decisions will be made on the basis of a set $\mathbb{E} \subseteq \mathbb{M}$ of models that render the evidence sufficiently likely, assuming the evidence is gathered in an appropriate way etc.

Bayesianism works like this:

1. Choose appropriate variables, $\mathbb{M}$, and a prior function P over $\mathbb{M}$.
2. Gather evidence, ensuring $\mathcal{E}$ is in the domain of P.
3. Take a new stance $P' = P(\cdot|\mathcal{E})$ over $\mathbb{M}$ (conditionalisation).
    - Typically via Bayes' theorem.
    - $\mathbb{E}$ is the set of models with sufficiently high P'.

Subjective: P is a matter of personal choice.

Objective: P is maximally equivocal.

It looks like the two approaches are just incompatible ways of doing stats.

But what are the questions that the two approaches are trying to answer?

Frequentist: How does evidence impact on the set of candidate physical probability functions?

Bayesian: How does evidence impact on rational degree of belief?

- Not so incompatible after all?

2 The Epistemological View

Key Question. How strongly should an agent with evidence $\mathcal{E}$ believe the various propositions expressible in her language $\mathcal{L}$?

- E.g., a finite propositional language $\mathcal{L} = \{A_1, \ldots, A_n\}$.
- $\mathcal{E}$ is *total evidence*, everything that is taken for granted: data, assumptions, theoretical knowledge ...
  - Not necessarily expressible as propositions of $\mathcal{L}$.

The Bayesian Answer. An agent's belief function $P_{\mathcal{E}}$ over $\mathcal{L}$ should satisfy certain norms:

Probability: It should be a probability function.

Calibration: It should be compatible with evidence,

- in particular, calibrated with known physical probabilities.

Equivocation: It should otherwise equivocate sufficiently between the basic possibilities expressible in $\mathcal{L}$.

Strict subjectivism: Probability (+ conditionalisation)
Empirically-based subjectivism: Probability + Calibration (+ conditionalisation)
Objectivism: Probability + Calibration + Equivocation

Explicating the Norms

Probability: $P_{\mathcal{E}}$ should be a probability function:

- **P1**: $P_{\mathcal{E}}(\omega) \geq 0$ for each $\omega \in \Omega = \{\pm A_1 \wedge \cdots \wedge \pm A_n\}$,
- **P2**: $P_{\mathcal{E}}(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$, and
- **P3**: $P_{\mathcal{E}}(\theta) = \sum_{\omega \models \theta} P_{\mathcal{E}}(\omega)$ for each $\theta \in S\mathcal{L}$.

Calibration: It should be compatible with evidence,

- **C1**: $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$
  - $\mathbb{P}^*$ set of candidate physical probability functions.
  - $\langle \cdot \rangle$ convex hull.
  - $\mathbb{S}$ structural constraints.

Equivocation: It should otherwise equivocate sufficiently between the basic possibilities expressible in $\mathcal{L}$.

- **E1**: $P_{\mathcal{E}}$ sufficiently close to equivocator $P_{=}(\omega) = 1/2^n$.
  - $d(P, P_{=}) = \sum_{\omega \in \Omega} P(\omega) \log P(\omega)/P_{=}(\omega)$.
  - ▸ proximity to the equivocator = entropy.

Explicating the Norms

Probability: $P_{\mathcal{E}}$ should be a probability function:

- **P1**: $P_{\mathcal{E}}(\omega) \geq 0$ for each $\omega \in \Omega = \{\pm A_1 \wedge \cdots \wedge \pm A_n\}$,
- **P2**: $P_{\mathcal{E}}(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$, and
- **P3**: $P_{\mathcal{E}}(\theta) = \sum_{\omega \models \theta} P_{\mathcal{E}}(\omega)$ for each $\theta \in S\mathcal{L}$.

Calibration: It should be compatible with evidence,

- **C1**: $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$
  - $\mathbb{P}^*$ set of candidate physical probability functions. ⟵ **Frequentist statistics!**
  - $\langle \cdot \rangle$ convex hull.
  - $\mathbb{S}$ structural constraints.

Equivocation: It should otherwise equivocate sufficiently between the basic possibilities expressible in $\mathcal{L}$.

- **E1**: $P_{\mathcal{E}}$ sufficiently close to equivocator $P_=(\omega) = 1/2^n$.
  - $d(P, P_=) = \sum_{\omega \in \Omega} P(\omega) \log P(\omega)/P_=(\omega)$.
  - ▸ proximity to the equivocator = entropy.

3 Frequentist Statistics for Calibration

A new picture of the relation between frequentist statistics and objective Bayesianism:

1. Be explicit about language $\mathcal{L}$ and evidence $\mathcal{E}$.
2. $\mathbb{M} = \mathbb{P}^*$ = candidate physical probability functions, given $\mathcal{E}$ (**frequentist statistics**).
3. $\mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$ = belief functions compatible with evidence.
4. Choose a belief function $P_{\mathcal{E}}$ in $\mathbb{E}$ sufficiently close to the equivocator (**OBE**).

- Use $P_{\mathcal{E}}$ as a basis for action.

Objection:

- The nice thing about maxent is that it is precise enough to implement in a decision support system.

- Frequentist statistics is a hotch-potch of vaguely formulated methods.

- ▸ They are combined at the expense of precision and automation.

Not necessarily *...*

- **Evidential Probability** (EP) is Henry Kyburg's implementation of frequentist statistics,

- it can more or less plug into the above scheme.

  - EP can be thought of as determining $\langle \mathbb{P}^* \rangle$ rather than $\mathbb{P}^*$.

4 Evidential Probability and OBE

Joint work with Greg Wheeler (Wheeler and Williamson, 2009).

- Evidence $\mathcal{E}$ can include statistical statements

$$\%x(\tau(x), \rho(x), [l, u])$$

  - the proportion of ρ-s that satisfy τ is between l and u: $freq_{\rho}(\tau) \in [l, u]$.
  - ρ is the *reference class*.

- The language $\mathcal{L}$ of EP is a 1st order language in which such statements can be expressed.

- Apply rules for manipulating reference classes.

- ▸ *Evidential probability* $Prob(\theta, \Gamma_{\delta}) = [l, u]$.
  - An interval rather than a sharp probability.
  - Interpret this as bounds on $P^*(\theta)$, the sharp probability of θ.
  - $\langle \mathbb{P}^* \rangle$ is the set of functions satisfying EP-consequences of $\mathcal{E}$.
  - (Need some consistency maintenance procedure.)

Urn Example. It is known just that

- the proportion of White balls in an Urn is in $[l, u]$, $\%x(W(x), U(x), [l, u])$.
- ball t is drawn from the Urn, $U(t)$.

▸ Then we can derive $P^*(W(t)) \in [l, u]$.

  - If $l = u = n/N$ then we can derive $P^*(W(t)) = n/N$.
  - If there is conflicting statistical evidence then EP applies conflict resolution rules.
  - If there is no statistical evidence we derive $P^*(W(t)) \in [0, 1]$.

Conflict Resolution

Reference Class Problem: How can one calculate the probability that an individual satisfies τ when it belongs to several reference classes ρ_i with known statistics?

Suppose we know that

- $[p, q]$ is the smallest interval covering reports of the proportion of Rs that are Us.
- $[l, u]$ is the smallest interval covering reports of the proportion of Ss that are Vs.
- $U(t_1) \leftrightarrow V(t_2), R(t_1), S(t_2)$.

Say $[p, q]$ and $[l, u]$ *conflict* if neither interval is strictly contained in the other.
One can ignore the S, V evidence if

Richness: The intervals conflict and R measures S together with other properties.

- $R = (S, T, \ldots)$

Specificity: The intervals conflict and R holds of fewer individuals than S.

Strength: $[l, u]$ contains all intervals that survive Richness and Specificity.

The remaining statistics are the *relevant statistics*.

- ▸ The smallest interval covering these gives the evidential probability.

Example

- Evidential certainties:
  - Bob smokes a packet of cigarettes a day and is a politician.

1. The proportion of smokers that live to the age of 80 is in $[.5, .8]$
2. The proportion of obese smokers living to 80 is in $[.3, .7]$ while the proportion of non-obese smokers living to 80 is in $[.6, .7]$.
3. The proportion of those who smoke a packet a day living to 80 is in $[.4, .75]$.
4. The proportion of politicians living to 80 is in $[.6, .7]$.

- 1 is eliminated in favour of 2 by Richness.
- 2 is eliminated in favour of 3 by Specificity.
- 3 is eliminated in favour of 4 by Strength.

▸ We conclude that the probability that Bob will live to 80 is in $[.6, .7]$.
  - A claim about physical probability, $\langle \mathbb{P}^* \rangle$.

▸ Then maxent will give probability .6 to Bob living to 80.
  - A claim about rational degree of belief.

Second-Order EP

Where do the intervals come from?

- Marginals: $freq_\rho(\tau) = u$ implies $freq_\rho(\sigma \wedge \tau) \in [0, u]$.
- Higher order probability: e.g., probability at least .95 that $freq_\rho(\tau) \in [l, u]$.
  - Kyburg: if $.95 > 1 - \delta$, a given threshold, then infer $freq_\rho(\tau) \in [l, u]$.
  - Wheeler and Williamson (2009): second-order EP, $freq_\rho(\tau) \in [l, u]^{[.95,1]}$.

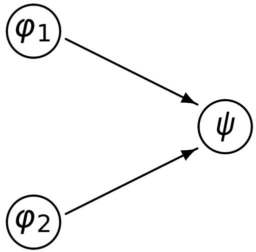
Won't calculations in 2oEP be impractical?

- Not necessarily: can use credal nets.
- Assumptions:
  - Items of evidence are by default independent.
  - 2o probability is by default uniformly distributed over the given interval.

Constructing a Credal Net

The structure of 1^o EP calculations determine the structure of the credal net:

- Evidence $\mathcal{E}$: $freq_R(F) \in [.2, .4]^{[.9,1]}, Rt^{\{1\}}$
- What is the probability that $P^*(Ft) \in [.2, .4]$?



The X_i and 1^o EP inferences determine the conditional probability constraints:

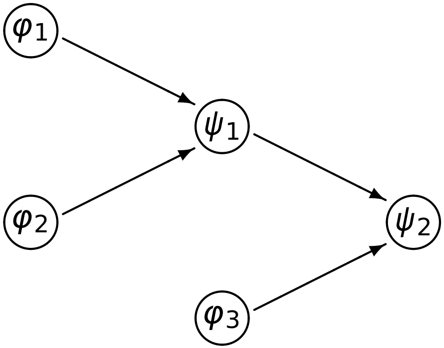
$$P(\varphi_1) \in [.9, 1], P(\varphi_2) = 1$$

$$P(\psi|\varphi_1 \wedge \varphi_2) = 1, P(\psi|\neg\varphi_1 \wedge \varphi_2) = .2 = P(\psi|\varphi_1 \wedge \neg\varphi_2) = P(\psi|\neg\varphi_1 \wedge \neg\varphi_2)$$

We can then chain inferences (won't work in 1^o EP):

$$freq_R(F) \in [.2,.4]^{[.9,1]}, Rt^{\{1\}} \approx\!\!\!| \; P(Ft) \in [.2,.4]^{Y_1}$$

$$freq_R(F) \in [.2,.4]^{[.9,1]}, Rt^{\{1\}}, freq_F(G) \in [.6,.7]^{[.9,1]} \approx\!\!\!| \; P(Gt) \in [0,.25]^{?}$$



$$P(\varphi_1) \in [.9,1], P(\varphi_2) = 1, P(\varphi_3) \in [0.9,1]$$

$$P(\psi_1|\varphi_1 \wedge \varphi_2) = 1, P(\psi_1|\neg\varphi_1 \wedge \varphi_2) = .2 = P(\psi_1|\varphi_1 \wedge \neg\varphi_2) = P(\psi_1|\neg\varphi_1 \wedge \neg\varphi_2)$$

$$P(\psi_2|\psi_1 \wedge \varphi_3) = \frac{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1] \cap [0,.25]|}{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1]|} = .31,$$

$$P(\psi_2|\neg\psi_1 \wedge \varphi_3) = .27, P(\psi_2|\psi_1 \wedge \neg\varphi_3) = P(\psi_2|\neg\psi_1 \wedge \neg\varphi_3) = .25$$

Conclusion

- From the epistemological point of view, frequentist statistics and objective Bayesianism are not just compatible, they are components of the same rational epistemology.

▸ We should move away from the standard statistical picture.

  – (We should anyway since the objective Bayesian need not conditionalise.)

- Introducing frequentist statistics need not be at the expense of precision and tractability:

  – Evidential probability and objective Bayesianism can be fruitfully combined;
  – Credal nets can be applied to both formalisms and their unification, making for tractable reasoning.

Related Work

- http://www.kent.ac.uk/secl/philosophy/jw

References

Wheeler, G. and Williamson, J. (2009). Evidential probability and objective Bayesian epistemology. In Bandyopadhyay, P. S. and Forster, M., editors, *Handbook of the Philosophy of Statistics*. Elsevier.