BAYESIAN OBJECTIONS TO OBJECTIVE BAYESIANISM

Philip Dawid
University College London

Parametric Inference

- Parameter: Θ
- Experiment $\mathcal{E}$
- Observable: X
- Statistical model: $X \mid \Theta \sim p_{\mathcal{E}}(x \mid \theta)$ (conditional density)

+ *perhaps further structure…*

Bayesianism

- All uncertainty can be expressed by means of *PROBABILITY*
- So introduce *(prior) density* $p(\theta)$ for Θ
- ➢ *Joint* density $p(x, \theta) = p(x \mid \theta)\, p(\theta)$
  - ➢ *Posterior* $p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta)$
  - ➢ *Predictive* $p(x) = \int p(x \mid \theta)\, p(\theta)\, d\theta$

SUBJECTIVE BAYESIANISM

- Prior distribution $p(\theta)$ represents individual's beliefs about Θ
  - – depends on *meaning* of Θ
  - – independent of *experiment* $\mathcal{E}$ and *form* of model $p_{\mathcal{E}}(x \mid \theta)$
- Joint distribution $p(x, \theta)$ obeys all the laws of probability
  - – *proper*: all densities must integrate to 1
  - – different ways of calculating the same thing must give the same answer (!)

Likelihood Principle

$\mathcal{E}_1$: Toss a penny **7** times (binomial experiment)

➢ x_1: get **4** heads

$\mathcal{E}_2$: Toss same penny until **4**th head (negative binomial experiment)

➢ x_2: takes **7** tosses

In either case $p_{\mathcal{E}_i}(x_i \mid \theta) \propto \theta^4 (1-\theta)^3$ (likelihood)

- – so get same posterior
  - – since same prior (same penny, same Θ)

Sampling consistency

- Model $X \mid \Theta = \theta \sim N(\theta, 1)$
- Have $E(X - \Theta \mid \theta) = 0$ all θ
- Problem if *e.g.* $E(X - \Theta \mid x) > 0$ all x
- This can *not* happen
  - – *though posterior distribution of* $X - \Theta$ *will differ from its* $N(0, 1)$ *sampling distribution*

OBJECTIVE (?) BAYESIANISM

1) Eschew subjectivity, celebrate ignorance
2) Retain probability (perhaps "improper")
3) Construct prior (or posterior) in a way that only depends on the form of the model $p(x \mid \theta)$, not on the meanings of X and Θ
4) Attempt to mimic sampling properties

   e.g. so that posterior of $X - \Theta$ is $N(0, 1)$

My attitude

- Great idea!
- It would be nice if it could be done…
- But it *can't* be done, while continuing to observe all the laws of probability
  - as soon as we assume it can, we lay ourselves open to paradox and inconsistency

FORM WITHOUT SUBSTANCE?

- "Invariant" prior distributions
  - e.g. Jeffreys, group-structural,…
- For binomial

$$p_J(\theta) \propto \theta^{-1/2} (1-\theta)^{-1/2}$$

- For negative binomial

$$p_J(\theta) \propto \theta^{-1} (1-\theta)^{-1/2}$$

- So likelihood principle is violated
- What prior for an *unending* sequence of coin tosses??

Sampling consistency?

$$X \sim N(\Theta, 1)$$

Take (improper) uniform prior $p(\theta) \propto 1$

Posterior mimics "sampling property":

$$X - \Theta \sim N(0, 1)$$

BUT THEN $E(X^2 - \Theta^2) =$

+1 in sampling distribution

−1 in posterior distribution

Bivariate Normal

$$\mathbf{X}_i \sim N(\mathrm{M}, \Sigma), \quad (i = 1, \ldots, n) \qquad \text{where} \quad \mathbf{X}_i = (X_{i1}, X_{i2})'$$

Standard estimators: $\qquad \mathrm{M} = (M_1, M_2)'$

$$\overline{\mathbf{X}} = \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \end{pmatrix}, \ \mathbf{S} \text{ of } \mathrm{M} = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \ \Sigma \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Sampling distributions:

$$\frac{n(\overline{X}_1 - M_1)^2}{S_{11}} \sim F_{1,n-1} \qquad (1)$$

$$\frac{n(n-2)}{2(n-1)} (\overline{\mathbf{X}} - \mathrm{M})' \mathbf{S}^{-1} (\overline{\mathbf{X}} - \mathrm{M}) \sim F_{2,n-2} \qquad (2)$$

Can mimic *either* (1) *or* (2) in the posterior

– *but not both at once*

wrong d.f ⇒ "strong inconsistency"

A way out?

- Allow prior distribution to vary with the "parameter of interest"
  - e.g. "reference" priors
- But then we get different answers to the same question, e.g. $E(\Theta^2 \mid x)$

–incoherent!

Marginalization Paradox

Parameter **Θ,** observable **X**

Statistical model p(**x** | **θ**), prior p(**θ**)

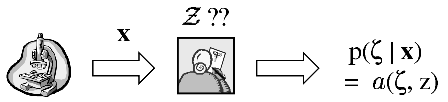
→ posterior p(**θ** | **x**)

Interested in $\mathcal{Z}$, a function of **Θ**

Calculate marginal posterior density p(ζ | **x**) of $\mathcal{Z}$

Find this depends on **x** only through the value z of some statistic Z:

$$p(\zeta \mid \mathbf{x}) = a(\zeta, z)$$



Semantics or Syntax?

You program a computer to read x and calculate $a(\zeta, z)$

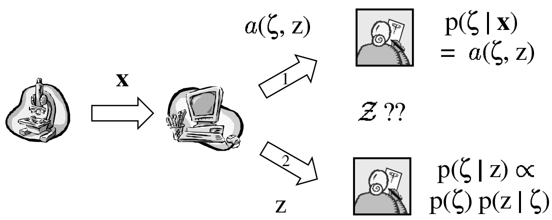
What should you do with its output?

1. Use it directly as your own posterior for $\mathcal{Z}$

   or

2. Treat *experiment* + *computer* as a compound experiment, outputting reduced information (equivalent to) Z
   - Apply probability calculations to this reduced model to obtain posterior for $\mathcal{Z}$

Semantics or Syntax?



Semantics or Syntax?

- If the prior is proper, you will get the same answer by either method
- But this can FAIL for improper priors, including many recommended for "objective" use
  - "MARGINALIZATION PARADOX" (MP)

Example: Scaled Means

$X_{ij} \sim N(M_i, \Sigma^2) \quad (i = 1, 2; j = 1, \ldots, n)$

Standard estimators $\overline{X}_1, \overline{X}_2, S$ of M_1, M_2, Σ

Interested in $\mathcal{Z}_i = M_i / \Sigma \quad (i = 1, 2)$

"Relatively invariant" priors have form $p(\mu_1, \mu_2, \sigma) \propto \sigma^{\lambda}$ $\begin{cases} \lambda = -3 & \text{Jeffreys} \\ \lambda = -1 & \text{"recommended"} \end{cases}$

Posterior density of $\mathcal{Z}_1$ only depends on $Z_1 = \overline{X}_1 / S$

- MP unless $\lambda = -2$

Posterior density of $(\mathcal{Z}_1, \mathcal{Z}_2)$ only depends on (Z_1, Z_2)

- MP unless $\lambda = -3$

Model selection

- Proper Bayes ⇒ consistency
  - though sensitive to prior specification
- Improper Bayes not well-defined
- Various "objective" get-outs suggested
  - *but these sacrifice consistency!*

What is left?

- Need "default" priors for when we are in a hurry
- Important to investigate their good and bad properties
- They *will* have some bad properties
- It is time to abandon the search for a fully self-consistent theory of objective Bayesian inference

Further Reading

- Dawid, A. P. (1983). Invariant Prior Distributions. *Encyclopedia of Statistical Sciences* vol. 4, edited by S. Kotz, N. L. Johnson and C. B. Read. Wiley-Interscience, 228–236.
- Dawid, A. P. and Stone, M. (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. *Biometrika* **59**, 369–375.
- Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference (with Discussion). *J. Roy. Statist. Soc.* B **35**, 189–233.
- Dawid, A. P., Stone, M. and Zidek, J. V. (1980). Comments on Jaynes' paper 'Marginalization and prior probabilities'. *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, edited by A. Zellner. North-Holland Publishing Company, 79–82.
- Dawid, A. P., Stone, M. and Zidek, J. V. (1996). Critique of E.T.Jaynes's "Paradoxes of Probability Theory". Research Report 172, Department of Statistical Science, University College London. http://www.ucl.ac.uk/Stats/research/Resrprts/abs96.html#172
- Stone, M. (1976). Strong inconsistency from uniform priors (with Discussion). *J. Amer. Statist. Ass.* **71**, 114–125.