Remarks on Improper “Ignorance” Priors

- **As a limit of proper priors**

- **Two caveats relating to computations with *improper* priors, based on their relationship with finitely-additive, but not countably-additive probability.**
  1. **Failure of the law of conditional expectations and “integrating-out.”**
  2. **Equivalent (transformed) random variables and “integrating-out.”**

You have seen (HW #1) that for $X \sim \boldsymbol{N}(\theta, \sigma^2)$, with the variance specified, if θ has a (Normal) conjugate prior $\theta \sim \boldsymbol{N}(\mu, \tau^2)$, then the posterior for θ satisfies

$$\mathbf{P}(\theta \mid x) \sim \boldsymbol{N}(\mu', \tau'^2)$$

where $\mu' = (\sigma^2\mu + \tau^2 x) / (\sigma^2+\tau^2)$ and $\tau'^2 = \sigma^2\tau^2 / (\tau^2+\sigma^2)$

If we let $\tau^2 \to \infty$ in the posterior probability, $\mathbf{P}(\theta \mid x) \to \boldsymbol{N}(x, \sigma^2)$.

This is a case where the Confidence Interval theory, and the Bayesian posterior probability for these intervals agree.

The corresponding "improper" prior is the uniform (Lebesgue) density for θ, $d\theta$.

However, though Lebesgue measure is a σ-finite *measure*, it corresponds to a (class of) finitely, but not countably additive *probabilities*.

That is, with Lebesgue measure used to depict a probability distribution for θ, the resulting *probability* satisfies shift-invariance: for all i and j

$$\mathbf{P}(i \leq \theta \leq i+1) = \mathbf{P}(j \leq \theta \leq j+1).$$

Then, even by finitely additive reasoning,

$$\mathbf{P}(\mathrm{i} \leq \theta \leq \mathrm{i}+1) = 0$$

As the line is a countable union of unit-intervals, **P** is not countably additive!

What are the consequences of using a finitely additive, but not countably additive prior probability for θ, disguised as an *improper* prior?

Anomalous Example 1:

Let X and Y be independent *Poisson* random variables where

X has mean θ and Y has mean $\lambda\theta$

Let the prior for the parameters be the product of

the uniform *improper* Lebesgue density for $\lambda > 0$

and a *proper* Gamma $\Gamma(2,1)$ density for θ.

Observe $Y = y$ first.

For computing a formal Bayes' posterior,

product of prior and likelihood densities satisfies:

$$\boldsymbol{f}(y, \theta, \lambda) = \exp[-\theta(\lambda+1)](\theta\lambda)^{y}\theta/y!$$

- Not surprisingly, one observation of Y proves uninformative about λ, as the integral of $\boldsymbol{f}(y, \theta, \lambda)$ with respect to θ is 1, *improper*.

- To find the marginal posterior for θ given Y, evaluate the integral of $\boldsymbol{f}(y, \theta, \lambda)$ with respect to λ, leading to

$$\boldsymbol{g}(y,\theta) = \exp[-\theta], \text{ a } \textit{proper } \Gamma(1,1) \text{ distribution.}$$

BUT, this formal application of Bayes' theorem with the (partially) *improper* prior yields the **same** *proper* Gamma $\Gamma(1,1)$ marginal posterior for θ given y, *regardless the value of Y observed*!

The (*proper*) marginal posterior distribution for θ, Γ(1,1), is a different distribution than the (*proper*) marginal prior distribution for θ, Γ(2,1).

Evidently, $\mathrm{E}[\theta] \neq \mathrm{E}[\ \mathrm{E}[\theta|Y]\ ]$

But the anomaly is not restricted to inference about the parameters:
Consider the distribution of *X*. Prior to observing Y = y, we have

$$\mathbf{P}(X = 2) \ = \ \int_0^\infty \left(\frac{\theta^2}{2}\exp[-\theta]\right)(\theta\exp[-\theta])\,d\theta \ = 3/16.$$

However, conditioned by Y = y, as *X* and *Y* are independent given (λ,θ), we get

$$\mathbf{P}(X = 2 \mid Y = \mathrm{y}) \ = \ \int_0^\infty \left(\frac{\theta^2}{2}\exp[-\theta]\right)\exp[-\theta]\,d\theta \ = 1/8$$

independent of the observed value y.

This, too, is an evident violation of the law of conditional expectations as:

$$\mathrm{E}[X] \neq \mathrm{E}[\ \mathrm{E}[X|Y]\ ]$$

Even though these posterior distributions are proper, that they are the consequence of a finitely, but not countably additive (*improper*) prior probability affects them as follows.

Theorem: Every finitely additive but not countably additive probability **P** admits an infinite partition $\pi = \{h_1, h_2, \ldots, h_n, \ldots\}$, event F, and $\delta > 0$ where:

$$\mathbf{P}(F) \geq k \text{ and yet } \mathbf{P}(F \mid h_i) < k - \delta \quad (i = 1, 2, \ldots)$$

Evidently, then, $\mathrm{E}[F] \neq \mathrm{E}[\ \mathrm{E}[F|h]\ ]$ and one *cannot* write

$$\mathbf{P}(F) = \int_h \mathbf{P}(F \mid h_i)\, d\mathbf{P}(h_i)$$

as always is possible for the countably additive case.

In the example, above, this anomaly occurs in the partition of the random variable *Y*, and affects the conditional probability for the parameter θ and the conditional probability for the other observable *X*, given *Y*.

Here is a second version of the same problem using Jeffreys' improper prior for Normally distributed data.

Anomalous Example 2 (Buehler-Feddersen, 1963):

Let (X_1,X_2) be conditionally *iid* $\boldsymbol{N}(\mu,\sigma^2)$.

Trivially, as μ is the median of the distribution, for each pair (μ,σ^2),

$$\mathbf{P}(X_{\min} \le \mu \le X_{\max} \mid \mu,\sigma^2) = .50.$$

It is a straightforward calculation that, using H.Jeffreys' *improper* prior

$$\boldsymbol{p}(\mu,\sigma^2) \;\propto\; d\mu\; d\sigma/\sigma,$$

which also is a limit of conjugate priors, then, for each pair (x_1, x_2)

$$\mathbf{P}(x_{\min} \le \mu \le x_{\max} \mid x_1, x_2) \;=\; .50.$$

Define a statistic $t = (x_1+x_2)/(x_1-x_2)$. So, for pairs (x_1, x_2) satisfying $|t| \le 1.5$,

$$\mathbf{P}(x_{\min} \le \mu \le x_{\max} \mid x_1, x_2, |t| \le 1.5) \;=\; .50. \qquad (*)$$

Buehler and Feddersen (1963) show that, for each pair (μ,σ^2), however,

$$\mathbf{P}(X_{\min} \le \mu \le X_{\max} \mid \mu,\sigma^2, |t| \le 1.5) \;>\; .518. \qquad (**)$$

- Integrating-out in the (X_1, X_2)-partition, using (*), we get:

$$\mathbf{P}(X_{\text{min}} \leq \mu \leq X_{\text{max}} \mid |t| \leq 1.5) = .50.$$

- Integrating-out in the (μ,σ^2)-partition, using (**), we get:

$$\mathbf{P}(X_{\text{min}} \leq \mu \leq X_{\text{max}} \mid |t| \leq 1.5) > .518.$$

Evidently, with Jeffreys' improper prior, at least one of these two partitions must fail to allow "integrating-out."

But which one is the culprit, or is "integrating-out" illicit in them both?

(2) Equivalent random variables and "integrating-out" a nuisance parameter under an *improper* prior.

Here is an illustration of the general problem.

Let (X_1, X_2) be positive, real-valued random variables.

Consider the (uniform) *improper* prior density

$$\mathbf{p}(x_1, x_2) \propto dx_1 dx_2$$

By usual rules for factoring: $\mathbf{p}(x_1, x_2) = \mathbf{p}(x_1 \mid x_2)\mathbf{p}(x_2) = \mathbf{p}(x_2 \mid x_1)\mathbf{p}(x_1)$

the joint density is the independent product of two *improper* uniform marginal

distributions $\mathbf{p}(x_1, x_2) \propto \mathbf{p}(\mathrm{x}_1)\ \mathbf{p}(\mathrm{x}_2).$

Transform to the equivalent pair (X_1, Y), where $Y = X_2/X_1$.

The *improper* joint density for (X_1, Y) is

$$\mathbf{p}(x_1, y) \propto x_1 dx_1 dy.$$

By usual rules for factoring this is the independent product of the two *improper* marginal priors

$$\mathbf{p}(x_1, y) = \mathbf{p}^*(x_1)\, \mathrm{p}(x_2),$$

where

$$\mathbf{p}^*(x_1) \propto x_1 dx_1 \neq \mathbf{p}(x_1) \propto dx_1$$

and $\mathbf{p}^*(x_1)$ is *off* by the Jacobian of the transformation from (X_1, X_2) to (X_1, Y).

Note: This is a magnification of a familiar (measure 0) problem with σ-additive probability – the so-called "Borel" paradox. Where, for a set of measure 0, it is possible that, though (X, Y) is an equivalent pair to (X, Z), and though $Y = y_0$ is equivalent to $Z = z_0$, nonetheless

$$\mathbf{P}(X \mid Y = y_0) \neq \mathbf{P}(X \mid Z = z_0)$$

Example (Dawid and Stone, 1972):

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ (n > 1) be conditionally *iid* $\boldsymbol{N}(\mu,\sigma^2)$, and use Jeffreys' improper (joint) prior density $\quad \mathbf{p}(\mu,\sigma) \;\propto\; d\mu d\sigma/\sigma$,

which can be arrived at as a limit of conjugate priors.

Recall that $(\bar{x}, s^2)$ are jointly sufficient statistics for the parameters.

The joint posterior density $\quad \mathbf{p}(\mu,\sigma \mid \bar{x}, s^2)$ (where $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$)

is proportional to $\quad \sigma^{-(n+1)}\, exp[-n(\mu-\bar{x})^2/2\sigma^2 - (n-1)s^2/2\sigma^2]\; d\mu d\sigma$

which can be described as follows in terms of the equivalent pair (μ,σ^2)

$$\mathbf{p}(\mu \mid \sigma^2, \bar{x}, s^2) \text{ is Normal } \boldsymbol{N}(\bar{x}, \sigma^2/n)$$

and

$$\mathbf{p}(\sigma^2 \mid \bar{x}, s^2) \text{ is Inverse Gamma } \Gamma([\tfrac{n-1}{2}], [\tfrac{n-1}{2}]s^2)$$

Recall that, by integrating out s from the joint posterior, $\mathbf{p}(\mu \mid \bar{x}, s^2)$ is such that, $\sqrt{n}(\mu-\bar{x})/s$ has a Student's *t*-distribution with *n-1* d.f.

Let $\theta = \mu/\sigma$ and transform the joint posterior $\mathbf{p}(\mu,\sigma \mid \bar{x}, s^2)$ to the joint posterior for the equivalent pair (θ, σ):

$$\mathbf{p}(\theta, \sigma \mid \bar{x}, s^2) \propto \sigma^{-n}\, exp[-n\theta^2/2 + n\theta\bar{x}/\sigma - R^2/2\sigma^2]\; d\theta d\sigma$$

where $R^2 = \Sigma x_i^2$.

Integrate-out σ (again!) to give the marginal posterior

$$\mathbf{p}(\theta \mid \boldsymbol{x}) \propto exp[-n\theta^2/2] \int_0^\infty \omega^{n-2} \exp[-\tfrac{1}{2}\omega^2 + r\theta\omega]d\omega = \boldsymbol{g}(\theta, r) \qquad (*)$$

where $r = (n/n-1)\bar{x}/s^2$. Thus, $\mathbf{p}(\theta \mid \boldsymbol{x})$ depends solely on the data through (r,n).

The distribution of r, likewise depends solely on θ, given the ancillary n. (See Stone & Dawid, 1972 for the details)

$$\mathbf{p}(r \mid \theta,\sigma) \propto exp[-n\theta^2/2](1- \mathrm{r}^2/\mathrm{n})^{(\mathrm{n}-3)/2} \int_0^\infty \omega^{n-1} \exp[-\tfrac{1}{2}\omega^2 + r\theta\omega]d\omega \qquad (**)$$

However, we *cannot* write $\boldsymbol{g}(\theta, r) = \mathbf{p}(r \mid \theta,\sigma)\, \boldsymbol{h}(\theta)$ for any prior $\boldsymbol{h}$ on θ.

That is, we cannot interpret (*) as a Bayesian *posterior* for θ, given the data r, its likelihood (**), and some prior $\boldsymbol{h}$ on θ.

However, if we compute $\mathbf{p}(\theta \mid \boldsymbol{x})$ using the improper prior $d\mu d\sigma/\sigma^2$, no such anomaly arises. Of course, this is a "change" in the prior matching the Jacobian of the transformation from (μ, σ) to (θ, σ).

Summary

(1) In general, it is not valid to "integrate-out" random variables from *proper* posterior distributions that are based on *improper* prior distributions. Just when this is allowed depends upon the details of the (merely) finitely additive probability that the improper prior represents. The theory of finitely additive probability does not yet offer a transparent answer when integration over improper priors is permitted.

(2) *This is a controversial point*: In general, familiar mathematical rules for manipulating transformations of a proper joint posterior density may not apply when that posterior is the result of an improper prior. That is, even when integrating out a (nuisance) parameter is warranted in a particular problem, how to do this when equivalent parameters are used is not settled!

References

Kadane, J.B., Schervish,M.J., and Seidenfeld, T. (1996) "Reasoning to a Foregone Conclusion," *J. Amer. Stat. Assoc.* **91**, 1228-1236.

Seidenfeld, T. (1982) "Paradoxes of Conglomerability and Fiducial Inference," in *Logic, Methodology and Philosophy of Science VI: Proceedings of the 6th International Congress of Logic, Methodology and Philosophy of Science, Hannover 1979*. North Holland: pp. 395-412.

Stone, M. and Dawid, P. (1972) "Un-Bayesian implications of improper Bayes inference in routine statistical problems," *Biometrika* **59**, 369-375.