

Notes on Jaynes's Treatment of the Marginalization Paradox

Kevin S. Van Horn
Leuther Analytics

November 23, 2003

In Chapter 15 of *Probability Theory: The Logic of Science* [2] (abbreviated PTLOS), Jaynes discusses the Marginalization Paradox (MP) of Dawid, Stone, and Zidek [1] (abbreviated DSZ), going into details for two specific examples. In the MP we have a model with four variables: ζ , η , y , and z . ζ is the parameter of interest, η is a nuisance parameter, and y and z are data. ζ and η are assigned independent priors $\pi_1(\zeta)$ and $\pi_2(\eta)$, and we are given $p(y, z | \zeta, \eta)$. Suppose now that $\pi_2(\eta)$ belongs to a certain class of improper priors; the MP is that, deriving $p(\zeta | z)$ in two different but (apparently) equally legitimate ways, we get different answers. This result has been used to argue that one should not use improper priors at all in Bayesian analyses. Jaynes argues that the problem is not the use of an improper prior, but that the two derivations implicitly condition on different states of information, and hence we should expect to get different answers. In this note I show that Jaynes is wrong: the problem does in fact stem from the use of an improper prior. To Jaynes's credit, however, the paradox is resolved by carefully following the procedures he advocates in PTLOS.

1 The change-point problem

In PTLOS section 15.8.1 Jaynes looks specifically at a particular parameterization of the change-point problem, wherein y is a scalar, z is a vector of length $n - 1$, and we have

$$p(y, z | \zeta, \eta, I_1) = c^{n-\zeta} \eta^n y^{n-1} \exp(-\eta y Q(\zeta, z)) \quad (1)$$

for a constant c and positive function Q . Here I_1 is the state of information of a Bayesian B_1 who knows about all four variables and the above sampling distribution. Integrating out y , we find

$$p(z | \zeta, \eta, I_1) = \frac{c^{n-\zeta} (n-1)!}{Q(\zeta, z)^n}. \quad (2)$$

Since the right-hand side of (2) has no dependence on η , Jaynes follows DSZ in concluding (PTLOS 15.67) that

$$p(z \mid \zeta, I_1) = \frac{c^{n-\zeta}(n-1)!}{Q(\zeta, z)^n}. \quad (3)$$

Now, this is the first point where we need to exercise some caution. How do we justify the last step above? Let's carefully apply the rules of probability theory. To get $p(z \mid \zeta, I_1)$ from $p(z \mid \zeta, \eta, I_1)$ we must multiply by $p(\eta \mid \zeta, I_1)$ and integrate out η :

$$\begin{aligned} p(z \mid \zeta, I_1) &= \int_0^\infty d\eta p(z, \eta \mid \zeta, I_1) \\ &= \int_0^\infty d\eta p(z \mid \eta, \zeta, I_1) p(\eta \mid \zeta, I_1) \\ &= p(z \mid \eta_0, \zeta, I_1) \int_0^\infty d\eta p(\eta \mid \zeta, I_1) \\ &= p(z \mid \eta_0, \zeta, I_1) \\ &= \frac{c^{n-\zeta}(n-1)!}{Q(\zeta, z)^n}, \end{aligned}$$

for any $0 < \eta_0 < \infty$. The third equality follows because $p(z \mid \eta, \zeta, I_1)$ has no dependence on η .

What if $p(\eta \mid \zeta, I_1)$ is improper? The usual procedure would be to replace '=' with ' \propto ' above, in the hope that the end result will be a proper distribution, and then normalize when we are done. The problem, however, is that *by definition* the integral on the third line above diverges. Thus, the above derivation of $p(z \mid \zeta)$ is invalid when using an improper prior over η . In such circumstances, Jaynes tells us that we must consider the sequence of proper priors whose limit our improper prior represents, go back and do our analysis with these proper priors, and see if our result converges to some well-defined limit. If it does converge, we take this limit as our result for the improper prior. If it does not converge, then we cannot use the improper prior.

In this case we get the same answer at every point in whatever sequence of proper priors we might choose, and so we have our well-defined limit. Thus we can justify (3) even for improper priors.

At this point Jaynes, again following DSZ, postulates another Bayesian B_2 who does not know of the existence of η or y ; his knowledge is limited to ζ , z , the prior $\pi_1(\zeta)$, and the marginal sampling distribution (3). B_2 then applies Bayes' rule to derive (PTLOS 15.70)

$$p(\zeta \mid z, I_2) \propto \frac{\pi_1(\zeta)c^{-\zeta}}{[Q(\zeta, z)]^n}, \quad (4)$$

where I_2 represents the state of information of B_2 .

Suppose now that B_1 assigns an improper prior for η : $\pi_2(\eta) \propto \eta^{-k}$, $0 < \eta < \infty$. Applying Bayes' Rule with $p(\zeta, \eta \mid I_1) = \pi_1(\zeta)\pi_2(\eta)$ and the sampling

distribution (3) to obtain $p(\zeta, \eta | y, z, I_1)$, then integrating out η , Jaynes and DSZ obtain the marginal posterior for ζ (PTLOS 15.72)

$$p(\zeta | y, z, I_1) \propto \frac{\pi_1(\zeta)c^{-\zeta}}{[Q(\zeta, z)]^{n-k+1}} \quad (5)$$

and thereby conclude, since (5) has no dependence on y , that

$$p(\zeta | z, I_1) \propto \frac{\pi_1(\zeta)c^{-\zeta}}{[Q(\zeta, z)]^{n-k+1}}. \quad (6)$$

(4) and (6) are clearly different posterior distributions for ζ . Jaynes asserts that (6) is in fact correct for B_1 , and attributes the difference with (4) to the different prior information of B_1 and B_2 .

To see that Jaynes's analysis is incorrect, it suffices to note that

$$\begin{aligned} p(z | \zeta, I_1) &= p(z | \zeta, I_2) \\ p(\zeta | I_1) &= p(\zeta | I_2); \end{aligned}$$

B_1 is also entitled to apply Bayes' Rule to the above, and doing so, obtains the same answer as B_2 :

$$p(\zeta | z, I_1) \propto \frac{\pi_1(\zeta)c^{-\zeta}}{[Q(\zeta, z)]^n}. \quad (7)$$

Since this contradicts (6), we do indeed have a paradox to resolve.

We resolve the paradox as follows:

- We show that the step from (5) to (6) is invalid: taking this step involves integrating $p(y | z, I_1)$, which itself is improper, and hence the integral diverges. Thus we cannot obtain (6).
- We take the conservative approach of defining the sequence of proper priors whose limit $\pi_2(\eta)$ represents, then deriving $p(\zeta | y, z)$ for each of these proper priors; this gives us (7), and not (6), as the limit.

2 An overlooked impropriety

As we did with (3), let us carefully go through the steps we would have to take to derive (6) from (5):

$$\begin{aligned} p(\zeta | z, I_1) &= \int_0^\infty dy p(\zeta, y | z, I_1) \\ &= \int_0^\infty dy p(\zeta | y, z, I_1)p(y | z, I_1) \\ &= p(\zeta | y_0, z, I_1) \int_0^\infty dy p(y | z, I_1) \\ &= p(\zeta | y_0, z, I_1) \end{aligned}$$

for any $0 < y_0 < \infty$, since $p(\zeta | y, z, I_1)$ has no dependence on y . As in the discussion of (3), the above derivation fails if $p(y | z, I_1)$ is improper, as the integral on the second-to-last line diverges. We now show that $p(y | z, I_1)$ is, indeed, improper, given the improper prior over η .

Applying (1) and the improper prior $\pi_2(\eta) \propto \eta^{-k}$, we obtain

$$\begin{aligned}
p(y | z, I_1) &\propto p(y, z | I_1) \\
&\propto \sum_{\zeta} \int_0^{\infty} d\eta p(y, z | \zeta, \eta, I_1) \pi_1(\zeta) \eta^{-k} \\
&\propto y^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} \int_0^{\infty} d\eta \eta^{n-k} \exp(-\eta y Q(\zeta, z)) \\
&= y^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} \frac{\Gamma(n-k+1)}{(yQ(\zeta, z))^{n-k+1}} \\
&\propto y^{k-2},
\end{aligned}$$

which is improper. (In the above we used the equality $\int_0^{\infty} x^{\alpha-1} e^{-\beta x} dx = \Gamma(\alpha)/\beta^{\alpha}$, which should be familiar as the normalization constant for the gamma distribution.) Thus we cannot obtain (6) from (5), and the paradox disappears.

I believe this to be the heart of the Marginalization Paradox: the implicit use of a divergent integral when an improper prior for a parameter leads to an improper distribution for a data variable.

3 Take it to the limit

The divergent integral in Section 2 tells us that we need to return to basics; that is, we need to repeat the analysis with a sequence of proper priors, and see if the results tend to some well-defined limit. Following Jaynes, we look at proper priors of the form

$$\pi_2(\eta) \propto \eta^a \exp(-b\eta) \tag{8}$$

for $b > 0$ and $a > -1$; our previous improper prior corresponds to $a = -k$ and $b \rightarrow 0$. Let us write $I_1(b)$ for the state of information corresponding to use of this prior for a specific value b .

Corresponding to (5) we now have

$$p(\zeta | y, z, I_1(b)) \propto \frac{\pi_1(\zeta) c^{-\zeta}}{(b + yQ(\zeta, z))^{n+a+1}}, \tag{9}$$

which is just (PTLOS 15.76). We must be careful here, though. We are going to multiply by $p(y | z, I_1(b))$ then integrate out y in order to obtain $p(\zeta | z, I_1(b))$. In using \propto instead of $=$ we have dropped factors that do not depend on ζ , but may depend on y ; we need the y -dependent factors to get the correct result. So

we shall put the normalization constant back in:

$$p(\zeta | y, z, I_1(b)) = \frac{\pi_1(\zeta)c^{-\zeta}(b + yQ(\zeta, z))^{-(n+a+1)}}{\sum_{\zeta'} \pi_1(\zeta')c^{-\zeta'}(b + yQ(\zeta', z))^{-(n+a+1)}}. \quad (10)$$

Now let us again use (1) to derive the marginal distribution of y conditional on z , only this time for a proper prior:

$$\begin{aligned} p(y | z, I_1(b)) &\propto p(y, z | I_1(b)) \\ &= \sum_{\zeta} \pi_1(\zeta) \int_0^{\infty} d\eta p(\eta | I_1(b)) p(y, z | \zeta, \eta, I_1(b)) \\ &\propto y^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} \int_0^{\infty} d\eta \eta^{n+a} \exp(-\eta(b + yQ(\zeta, z))) \\ &\propto y^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} (b + yQ(\zeta, z))^{-(n+a+1)} \end{aligned} \quad (11)$$

The proportionality constant depends only on z and b ; in particular, it doesn't depend on ζ .

We now combine (11) with (10) to obtain

$$\begin{aligned} p(\zeta | z, I_1(b)) &= \int_0^{\infty} dy p(\zeta | y, z, I_1(b)) p(y | z, I_1(b)) \\ &\propto \pi_1(\zeta) c^{-\zeta} \int_0^{\infty} dy y^{n-1} (b + yQ(\zeta, z))^{-(n+a+1)}. \end{aligned} \quad (12)$$

To finish the derivation, we need the following facts, obtained by use of Mathematica:

$$\int \frac{y^r}{(b + uy)^m} dy = \frac{y^{r+1}}{b^m(r+1)} {}_2F_1(r+1, m, r+2, -uy/b) \quad (13)$$

$${}_2F_1(\alpha, \beta, \gamma, s) \equiv \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1} (1-t)^{\gamma-\beta-1} (1-ts)^{-\alpha} dt \quad (14)$$

From (14) we find

$$\begin{aligned} {}_2F_1(\alpha, \beta, \gamma, 0) &= \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1} (1-t)^{\gamma-\beta-1} dt \\ &= \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \frac{\Gamma(\beta)\Gamma(\gamma-\beta)}{\Gamma(\gamma)} \\ &= 1; \end{aligned} \quad (15)$$

furthermore, as $y \rightarrow \infty$, we have

$$\begin{aligned} &y^{r+1} f(r+1, m, r+2, -uy/b) \\ &= \kappa \int_0^1 t^{m-1} (1-t)^{r-m+1} y^{r+1} (1 + yut/b)^{-(r+1)} dt \end{aligned}$$

$$\begin{aligned}
&= \kappa \int_0^1 t^{m-1} (1-t)^{r-m+1} (y^{-1} + ut/b)^{-(r+1)} dt \\
&\rightarrow \kappa \left(\frac{b}{u}\right)^{r+1} \int_0^1 t^{m-r-2} (1-t)^{r-m+1} dt \\
&= \kappa \left(\frac{b}{u}\right)^{r+1} \Gamma(m-r-1) \Gamma(r-m+2) \\
&= \frac{\Gamma(r+2) \Gamma(m-r-1)}{\Gamma(m)} \left(\frac{b}{u}\right)^{r+1} \tag{16}
\end{aligned}$$

where $\kappa \equiv \Gamma(r+2)/(\Gamma(m)\Gamma(r-m+2))$. Combining (13), (15), and (16), we obtain

$$\int_0^\infty y^r (b+uy)^{-m} dy = (r+1)^{-1} b^{-m} \frac{\Gamma(r+2) \Gamma(m-r-1)}{\Gamma(m)} \left(\frac{b}{u}\right)^{r+1}$$

Substituting this into (12), with $r = n-1$, $u = Q(\zeta, z)$, and $m = n+a+1$, we obtain

$$p(\zeta | z, I_1(b)) \propto \pi_1(\zeta) c^{-\zeta} [Q(\zeta, z)]^{-n},$$

which accords with (7). The above holds regardless of b , so it is also the limit as $b \rightarrow 0$. Thus (7), and not (6), is the correct solution for B_1 .

4 What went wrong: non-uniform convergence

The reader may still be puzzled as to how $p(\zeta | z, I_1)$ could differ from $p(\zeta | y, z, I_1)$ if the latter has no dependence on y . To clarify what is happening we look again at what happens with the proper priors (8) as $b \rightarrow 0$ and we approach the improper prior $\propto \eta^a$.

The problem is one of nonuniform convergence. Comparing (5) and (9), we have

$$\frac{p(\zeta) c^{-\zeta}}{(b+yQ(\zeta, z))^{n+a+1}} \approx \frac{p(\zeta) c^{-\zeta}}{Q(\zeta, z)^{n+a+1}} y^{-n-a-1}$$

only for $y \gg b/Q(\zeta, z)$. For any *fixed* value of y , $p(\zeta | z, I_1(b))$ does indeed tend to (5) as $b \rightarrow 0$; nonetheless, for *all* values of b , no matter how small, there remain values of y for which $p(\zeta | y, z, I_1(b))$ differs greatly from (5). Is this important? Since the only way to get rid of the conditioning on y is to multiply by $p(y | z, I_1(b))$ and then integrate out y , to obtain (6) we need to have

$$p(y \leq Cb/Q(\zeta, z) | z, I_1(b)) \rightarrow 0,$$

for any ζ, z , and constant $C > 0$, as $b \rightarrow 0$.

In Figure 1 we plot $p(u | z, I_1(b))$, where $u \equiv y\bar{Q}/b$ and \bar{Q} is the average of $Q(\zeta, z)$, $1 \leq \zeta \leq n$. We chose $a = n = 10$, $c = 1.5$, and randomly generated the vector z as follows: x_i , $1 \leq i \leq n/2$, were drawn from an exponential distribution with mean 1; x_i , $n/2 < i \leq n$, were drawn from an exponential

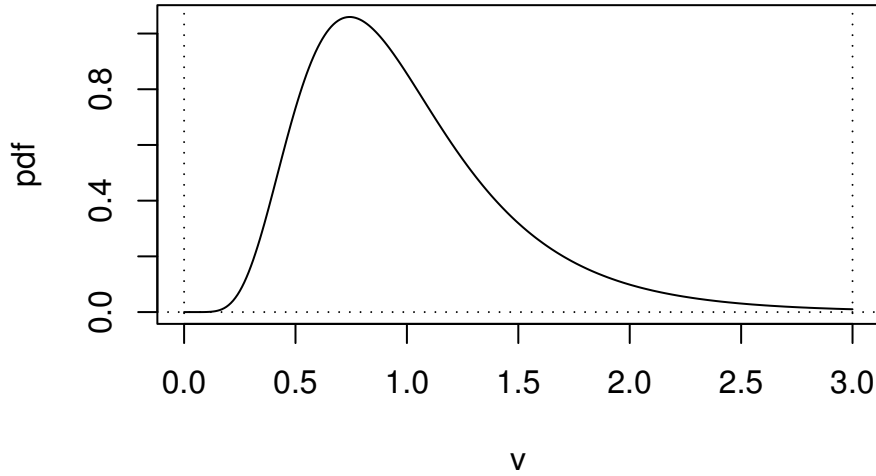


Figure 1: $p(u | z, I_1(b))$, where $u \equiv y\bar{Q}/b$

distribution with mean $1/c$; and $z_i \equiv x_i/x_1$. The ratios $Q(\zeta, z)/\bar{Q}$ range from 0.819 to 1.167; thus $u \leq 1$ corresponds roughly to $y \leq b/Q(\zeta, z)$. We see then that in this case there is substantial probability mass in the region $y \leq b/Q(\zeta, z)$ for all ζ , which is precisely the region in which (5) differs substantially from (9).

We have not specified b , because $p(u | z, I_1(b))$ has no dependence on b ! Thus as $b \rightarrow 0$ we find that, for any ζ , $P(u \leq C\bar{Q}/Q(\zeta, z) | z, I_1(b))$ remains constant, and hence so does $P(y \leq Cb/Q(\zeta, z) | z, I_1(b))$. To see this, note from (11) and the definition of u that

$$\begin{aligned} p(u | z, I_1(b)) &\propto u^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} (b + bu Q(\zeta, z)/\bar{Q})^{-(n+a+1)} \\ &\propto u^{n-1} \sum_{\zeta} \pi_1(\zeta) c^{-\zeta} (1 + u Q(\zeta, z)/\bar{Q})^{-(n+a+1)}, \end{aligned}$$

which has no dependence on b .

5 DSZ Example #5

Jaynes analyzes a second specific MP example in PTLOS section 15.9.1; again, the apparent paradox can be traced to an overlooked improper distribution, and the paradox is resolved by noting that the integral we must take to obtain

the paradox diverges. In this case we have iid variables x_i , with a normal distribution for each x_i with mean μ and variance σ^2 . We then define $\zeta = \mu/\sigma$, $\nu = \sigma$, $y = r$, and $z = R$, where

$$\begin{aligned} R^2 &\equiv n(\bar{x}^2 + s^2) \\ r &\equiv n\bar{x}/R \\ \bar{x} &\equiv n^{-1} \sum_{i=1}^n x_i \\ s^2 &\equiv n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned}$$

The paradox arises when we use an improper prior of the form

$$p(\mu, \sigma) \propto \sigma^{-\gamma-1},$$

$\gamma > 0$, taken as the limit of the proper prior

$$p(\mu, \sigma) \propto \sigma^{-\gamma-1} \exp(-\beta/\sigma - \alpha\mu^2)$$

as $\alpha, \beta \rightarrow 0$.

Similarly to the first example, $p(\zeta | R)$ is derived by first finding $p(\zeta | r, R) \propto h(\zeta, R)$, noting that this has no dependence on r , and concluding that $p(\zeta | R) \propto h(\zeta, R)$ also. As before, this last step is invalid, and therefore the paradox cannot be created, because $p(R | r)$ is improper when we use the improper prior for μ and σ . We now present the proof that $p(R | r)$ is improper.

Theorem: Let $x_i \sim \mathcal{N}(\mu, \sigma)$ for $1 \leq i \leq n$, and define $S^2 = \sum_i (x_i - \bar{x})^2$. Then

- \bar{x} and S^2 are independent given μ and σ ;
- $\bar{x} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$;
- $S^2 \sigma^{-2} \sim \chi^2(n-1)$.

Proof: This is a standard result in statistics; see, for example, M. G. Bulmer, *Principles of Statistics*.

We have $s^2 = S^2/n$; define also $u = S^2 \sigma^{-2} = n\sigma^{-2} s^2$. Then $du = n\sigma^{-2} ds^2$, and combining this with the above theorem we have

$$\begin{aligned} p(\bar{x} | \mu, \sigma) &\propto \sigma^{-1} \exp\left(-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right) \\ p(u | \mu, \sigma) &\propto u^{(n-3)/2} e^{-u/2} \\ p(s^2 | \mu, \sigma) &\propto \sigma^{-n+1} s^{n-3} \exp\left(-\frac{n}{2\sigma^2} s^2\right) \\ p(\bar{x}, s^2 | \mu, \sigma) &\propto \sigma^{-n} s^{n-3} \exp\left(-\frac{n}{2\sigma^2} (s^2 + (\bar{x} - \mu)^2)\right). \end{aligned}$$

We will later remove the conditioning on μ and σ by multiplying by the priors on these variables and integrating them out. For this reason we explicitly show any

factors dependent on μ or σ , rather than absorbing them into the proportionality constants above.

We now want to do a change of variables from \bar{x} and s^2 to R . To compute the Jacobian, we first find

$$\begin{aligned}
\frac{\partial R}{\partial \bar{x}} &= \frac{1}{2} (n(\bar{x}^2 + s^2))^{-1/2} \cdot 2n\bar{x} \\
&= n\bar{x}R^{-1} \\
\frac{\partial R}{\partial s^2} &= \frac{1}{2} (n(\bar{x}^2 + s^2))^{-1/2} \cdot n \\
&= \frac{1}{2} nR^{-1} \\
\frac{\partial r}{\partial \bar{x}} &= \frac{Rn - n\bar{x}(n\bar{x}R^{-1})}{R^2} \\
&= \frac{n}{R} - \frac{n^2\bar{x}^2}{R^3} \\
\frac{\partial r}{\partial s^2} &= -n\bar{x}R^{-2} \cdot \left(\frac{1}{2}nR^{-1}\right) \\
&= -\frac{n^2\bar{x}}{2R^3}
\end{aligned}$$

Then the Jacobian is the absolute value of

$$\begin{aligned}
&\frac{\partial R}{\partial \bar{x}} \frac{\partial r}{\partial s^2} - \frac{\partial R}{\partial s^2} \frac{\partial r}{\partial \bar{x}} \\
&= n\bar{x}R^{-1} \cdot \left(-\frac{1}{2}n^2\bar{x}R^{-3}\right) - \frac{1}{2}nR^{-1} \cdot (nR^{-1} - n^2\bar{x}^2R^{-3}) \\
&= -\frac{1}{2}n^3\bar{x}^2R^{-4} - \frac{1}{2}n^2R^{-2} + \frac{1}{2}n^3\bar{x}^2R^{-4} \\
&= -\frac{1}{2}n^2R^{-2},
\end{aligned}$$

giving $d\bar{x} ds^2 = 2n^{-2}R^2 dR dr$. We now combine this with the fact that

$$\begin{aligned}
\bar{x} &= n^{-1}rR \\
s &= (n^{-1}R^2 - \bar{x}^2)^{1/2} \\
&= (n^{-1}R^2 - n^{-2}r^2R^2)^{1/2} \\
&= n^{-1}R(n - r^2)^{1/2}
\end{aligned}$$

to obtain

$$\begin{aligned}
p(R, r | \mu, \sigma) &\propto R^2 p(\bar{x}, s^2 | \mu, \sigma) \\
&\propto R^2 \sigma^{-n} s^{n-3} \exp\left(-\frac{n}{2\sigma^2}(s^2 + \bar{x}^2 + \mu^2 - 2\bar{x}\mu)\right) \\
&\propto R^2 \sigma^{-n} s^{n-3} \exp\left(-\frac{1}{2\sigma^2}(R^2 - 2\mu r R + n\mu^2)\right) \\
&\propto R^{n-1} \sigma^{-n} v(r) \exp\left(-\frac{1}{2\sigma^2}(R^2 - 2\mu r R + n\mu^2)\right)
\end{aligned}$$

where $v(r) \equiv (n - r^2)^{(n-3)/2}$.

Now we remove the μ dependence by multiplying $p(R, r | \mu, \sigma)$ by the prior over μ and then integrating out μ :

$$\begin{aligned} f(r, R, \mu, \sigma) &\equiv \exp\left(-\frac{1}{2\sigma^2}(R^2 - 2\mu r R + n\mu^2)\right) p(\mu) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}(R^2 - 2\mu r R + (n + 2\sigma^2\alpha)\mu^2)\right) \end{aligned}$$

Let $\lambda \equiv n + 2\sigma^2\alpha$ and $m \equiv rR/\lambda$; then

$$f(r, R, \mu, \sigma) \propto \exp\left(-\frac{1}{2\sigma^2}(R^2(1 - r^2/\lambda) + \lambda(\mu - m)^2)\right)$$

and

$$\int_{-\infty}^{\infty} f(r, R, \mu, \sigma) d\mu \propto \lambda^{-1/2} \sigma \exp\left(-\frac{R^2}{2\sigma^2}(1 - r^2/\lambda)\right)$$

Now, we are interested in what happens as $\alpha, \beta \rightarrow 0$ simultaneously. As $\beta \rightarrow 0$ we see that the prior over σ goes to $p(\sigma) \propto \sigma^{-\gamma-1}$, and as we approach this limit it becomes highly probable that σ is very near zero. (Recall that $\gamma > 0$ is fixed.) Thus as $\alpha, \beta \rightarrow 0$ we can assume that $\alpha\sigma^2$ is very small. Then

$$\lambda^{-1} \approx n^{-1}(1 - 2n^{-1}\alpha\sigma^2),$$

hence

$$1 - r^2\lambda^{-1} \approx 1 - n^{-1}r^2 + 2n^{-2}r^2\alpha\sigma^2,$$

and so

$$-\frac{R^2}{2\sigma^2}\left(1 - \frac{r^2}{\lambda}\right) \approx -\frac{R^2}{2\sigma^2}\left(1 - \frac{r^2}{n}\right) - \frac{\alpha r^2 R^2}{n^2}.$$

Likewise, we have

$$\lambda^{-1/2} \approx n^{-1/2}(1 - n^{-1}\alpha\sigma^2).$$

So we have

$$\int_{-\infty}^{\infty} f(r, R, \mu, \sigma) d\mu \propto \sigma \exp\left(-\frac{R^2}{2\sigma^2}w(r)\right) g(\alpha, \sigma) h(\alpha, r, R),$$

where, for $\alpha, \beta \ll 1$,

$$\begin{aligned} w(r) &\equiv 1 - r^2/n \\ g(\alpha, \sigma) &\approx (1 - n^{-1}\alpha\sigma^2) \\ h(\alpha, r, R) &\approx \exp(-n^{-2}\alpha r^2 R^2). \end{aligned}$$

(Note that $-\sqrt{n} \leq r \leq \sqrt{n}$ and so $0 \leq w(r) \leq 1$.)

This gives us

$$p(R, r | \sigma) \propto R^{n-1} v(r) h(\alpha, r, R) g(\alpha, \sigma) \sigma^{-n+1} \exp\left(-\frac{R^2}{2\sigma^2}w(r)\right).$$

Now we remove the σ dependence by multiplying by the prior over σ and then integrating out σ :

$$\begin{aligned}
F(\alpha, r, R, \sigma) &\equiv p(\sigma)g(\alpha, \sigma)\sigma^{-n+1} \exp\left(-\frac{R^2}{2\sigma^2}w(r)\right) \\
&\approx F_1(r, R, \sigma) + F_2(\alpha, r, R, \sigma) \\
F_1(r, R, \sigma) &\equiv p(\sigma)\sigma^{-n+1} \exp\left(-\frac{R^2}{2\sigma^2}w(r)\right) \\
&= Z^{-1}\sigma^{-n-\gamma} \exp\left(-\frac{\beta}{\sigma} - \frac{R^2}{2\sigma^2}w(r)\right) \\
F_2(\alpha, r, R, \sigma) &\equiv -n^{-1}\alpha\sigma^2 F_1(r, R, \sigma)
\end{aligned}$$

where Z is the normalization constant for the prior over σ .

We now note that

$$1 > \exp(-\beta/\sigma) > \exp(-\beta) \exp(-\beta/\sigma^2)$$

(Proof: $1 < \sigma + 1/\sigma$, hence $1/\sigma < 1 + 1/\sigma^2$, hence $-\beta/\sigma > -\beta - \beta/\sigma^2$.) This means that

$$\begin{aligned}
&Z^{-1}\sigma^{-n-\gamma} \exp\left(-\frac{R^2}{2\sigma^2}w(r)\right) \\
&> F_1(r, R, \sigma) \\
&> e^{-\beta} Z^{-1}\sigma^{-n-\gamma} \exp\left(-\frac{R^2}{2\sigma^2}w(r) - \frac{\beta}{\sigma^2}\right).
\end{aligned}$$

Applying the change of variables $\rho = \sigma^2$, with $d\rho = 2\sigma d\sigma$, we find

$$\begin{aligned}
&(2Z)^{-1}\rho^{-(n+\gamma+1)/2} \exp\left(-\frac{R^2}{2\rho}w(r)\right) d\rho \\
&> F_1(r, R, \sigma) d\sigma \\
&> (2Ze^\beta)^{-1}\rho^{-(n+\gamma+1)/2} \exp\left(-\frac{R^2}{2\rho}w(r) - \frac{\beta}{\rho}\right) d\rho.
\end{aligned}$$

We've now matched the form of an inverted-gamma distribution for ρ , so the integrals are just the corresponding normalization constants:

$$\begin{aligned}
&\frac{\Gamma((n+\gamma-1)/2)}{2Z(R^2w(r)/2)^{(n+\gamma-1)/2}} \\
&> \int_0^\infty F_1(r, R, \sigma) d\sigma \\
&> \frac{\Gamma((n+\gamma-1)/2)}{2Ze^\beta(R^2w(r)/2 + \beta)^{(n+\gamma-1)/2}}.
\end{aligned}$$

Likewise, we find for F_2 that

$$-\frac{\alpha\Gamma((n+\gamma+1)/2)}{2nZ(R^2w(r)/2)^{(n+\gamma+1)/2}}$$

$$\begin{aligned}
&< \int_0^\infty F_2(\alpha, r, R, \sigma) d\sigma \\
&< -\frac{\alpha\Gamma((n+\gamma+1)/2)}{2nZe^{\beta(R^2w(r)/2+\beta)^{(n+\gamma+1)/2}}}.
\end{aligned}$$

We see that as $\alpha, \beta \rightarrow 0$, if $w(r) \neq 0$ then

$$\int_0^\infty F(\alpha, r, R, \sigma) d\sigma \rightarrow C_1(R^2w(r))^{-(n+\gamma-1)/2}$$

for some constant C_1 . Then

$$\begin{aligned}
p(R | r) &\propto p(R, r) \\
&\propto v(r)R^{n-1}h(\alpha, r, R) \int_0^\infty F(\alpha, r, R, \sigma) d\sigma \\
&\rightarrow C_1v(r)R^{n-1}(R^2w(r))^{-(n+\gamma-1)/2} \\
&\propto R^{-\gamma}
\end{aligned}$$

which is improper.

What about the case $w(r) = 0$? This occurs when $r = \pm\sqrt{n}$, that is to say, when all the x_i are identical. As previously mentioned, as $\beta \rightarrow 0$ the prior for σ becomes concentrated near 0, and so the x_i tend to group closely together in the prior predictive distribution. Thus, the prior predictive distribution for $w(r)$ is $\delta(w(r))$, making $w(r) = 0$ a case of interest.

For $w(r) = 0$ we obtain

$$F_1(r, R, \sigma) = Z^{-1}\sigma^{-n-\gamma} \exp(-\beta/\sigma).$$

Thus $F(\alpha, r, R, \sigma)$ has no dependence on r or R , nor does its integral over σ , and so we find

$$\begin{aligned}
p(R | r) &\propto p(R, r) \\
&\propto v(r)R^{n-1}h(\alpha, r, R) \int F(\alpha, r, R, \sigma) d\sigma \\
&\propto R^{n-1}h(\alpha, r, R) \\
&\rightarrow R^{n-1}
\end{aligned}$$

which is also improper.

6 Conclusion

What have we learned from this? Certainly, we see that the use of improper priors can be a tricky business. However, avoiding improper priors altogether may be an extreme solution. In a sense, the usual rule for dealing with improper priors, more-or-less explicitly laid out in PTLOS, still protects us from the Marginalization Paradox. That rule has two parts:

- Make sure you carefully follow the rules of probability theory; don't skip steps that seem obvious or unimportant.
- If you ever find yourself trying to evaluate a divergent integral, stop immediately; you must abandon the improper prior and use a proper prior. (Optionally, you may do the analysis for a sequence of proper priors and see if the solutions tend to a well-defined limit.)

This note does not, of course, constitute a proof that the above simple rule will always protect one from inconsistencies in the use of improper priors. It seems clear that some foundational work remains to be done on the safe use of improper priors. In particular, the following questions need to be answered:

- What, exactly, do we mean when we say that an improper prior is the limit of a particular sequence of proper priors? This is a trickier question than one might think; in particular, it appears that no reasonable definition can be given without reference to some class of likelihood functions with which the prior is to be combined.
- If π_i , $1 \leq i < \infty$, are a sequence of proper priors whose "limit" is the improper prior π_∞ , what rules for use of improper priors will guarantee that a solution obtained using π_∞ is the well-defined limit of the sequence of solutions obtained by use of the priors π_i ?

References

- [1] 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.
- [2] Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*, Cambridge University Press.