

EXPLORATIONS BY JOHN A. MAJOR, ASA

# Bayesian Dragons: A Cautionary Note

“Here Be Dragons”

— Ancient map label for unexplored regions

This is the story of how I discovered moment-busting monsters lurking in the unexplored regions of Bayesian predictive distributions. Don't get me wrong. I'm a believer. I used to be a skeptic, but recently I saw the light. Bayesian methods, and Markov Chain Monte Carlo (MCMC) technology in particular, now have a special place in my toolbox. But on my journey to the light, I stumbled upon some interesting facts that, it seems, are not widely appreciated.

## Background

The task set out before me was to reproduce the fit of the leveled chain ladder (LCL) model to the “Illustrative Insurer” ultimate incurred losses in Meyers [2015]. Readers unfamiliar with Bayesian methods are urged to read Meyers' monograph for background and literature references. The LCL model on this data consists of 29 parameters and the likelihood function is still simple enough to be able to program directly. Numerical methods sufficed to find the maximum likelihood estimate (MLE) of the parameters.

The parameters, of course, are not known precisely; they are simply estimated and there is some uncertainty around them. One way to express this uncertainty is to generate a *predictive distribution*. In effect, one mixes the various possible lognormal distributions

implied by a range of plausible model parameter values.

My challenge was to get estimates close to Meyers' using methods faster than MCMC. My weapon of choice was importance sampling [Rubinstein & Kroese, 2011]. It failed miserably, but in an interesting way. No matter how many samples I drew (and I went into tens of millions), the estimator of the mean ultimate loss would not stabilize. The central limit theorem (CLT) seemed not to apply.

This got me thinking ... when does the CLT not apply? One failure mode is when the target random variable does not have a finite variance. Could it be that the variance of the predictive ultimate loss in the LCL model did not exist?

Eventually, I traced my problems to the curse of dimensionality. Even a 29-dimensional problem, modest though it might be compared to the thousands or even millions of dimensions in some industrial-strength statistical models, was too much for importance sampling.

Yet my initial speculation continued to haunt me. Eventually, I found out I was right ... sort of.

## Bayesian Dragons in a Simple Model

Consider the following simple model, basically a one-cell triangle. There is one accident year with parameter  $\mu$  and one development period with parameter  $\varpi$ . The ultimate loss is a random variable  $Z$

$= \exp(X)$  where  $X$  has a normal distribution with mean  $\mu$  and variance  $\varpi$ .

Given particular parameter values, the expected value of  $Z$  can be readily calculated as  $\exp(\mu + \varpi/2)$ .

Say we impose a flat prior over  $(-\infty, \infty)$  on  $\mu$  and another flat prior over  $(0, \infty)$  on  $\varpi$ . Say also we have  $n > 1$  independent and identically distributed observations  $z_i$ . The likelihood can be written down easily; it is the independent multivariate normal formula. With flat priors, the posterior distribution is the same formula, with a restriction that  $\varpi > 0$  and with a normalizing constant:

$$Post(\mu, \varpi) = K_1 \cdot \left(\frac{1}{\sqrt{2 \cdot \pi \cdot \varpi}}\right)^n \cdot \exp\left[-\frac{1}{2 \cdot \varpi} \cdot \sum_i (z_i - \mu)^2\right] \quad [1]$$

Say the data exhibit mean  $M$  and variance  $V > 0$  (the case  $V = 0$  leads to a singular solution at  $\mu = M$ ). Rewrite [1] as

$$Post(\mu, \varpi) = K_1 \cdot \left(\frac{1}{\sqrt{2 \cdot \pi \cdot \varpi}}\right)^n \cdot \exp\left[-\frac{1}{2 \cdot \varpi} \cdot \left(\sum_i (z_i - M)^2 + n \cdot (\mu - M)^2\right)\right] \quad [2]$$

$$= K_2 \cdot \varpi^{-\frac{n}{2}} \cdot \exp\left(-\frac{n \cdot V}{2 \cdot \varpi}\right) \cdot \exp\left[-\frac{n \cdot (\mu - M)^2}{2 \cdot \varpi}\right]$$

where the  $K_i$  are constants not involving the parameters. This is the product of an inverse gamma in  $\varpi$ , and, conditional on  $\varpi$ , a normal in  $\mu$ . Therefore, the predictive distribution is proper and  $K_2 > 0$ .

**Proposition:** Assume the model described above with the posterior distribution given by [2] with  $V > 0$ . The predictive mean loss is given by the fol-

lowing expression:

$$P = K_3 \cdot \int_0^\infty \left[ \frac{\sigma^{-1}}{\sigma} \cdot \exp\left(-\frac{n \cdot V}{2 \cdot \sigma}\right) \right] \cdot \exp\left(\sigma \cdot \frac{n+1}{2 \cdot n}\right) d\sigma \quad [3]$$

and this integral diverges. Therefore the predictive mean does not exist. (So neither does the variance.)

The formal proof of this, to appear in another paper, is omitted here. Notice that the expression [3] can be recognized as the expectation of  $\exp(k^* \varpi)$  where  $\varpi$  is distributed as an inverse gamma. Such expectations do not exist for the inverse gamma. It should also be pointed out that with the usual conjugate reference priors, the predictive distribution of a lognormal random variable is log- $t$ , and the log- $t$  has no finite mean.

What if a  $1/\varpi$  prior were used instead of a flat prior on  $\varpi$ ? The same result obtains; it doesn't help. (Thanks to Gary Venter for raising this question.)

It is instructive to examine the attempt to calculate expression [3] numerically.

Take, for example, the concrete values  $n=11$  and  $V=1$  (M doesn't matter up to a constant).

The inverse gamma posterior density of  $\varpi$  (square brackets in [3]), up to a constant, is shown in Figure 1.

One might think that truncating  $\varpi$  at, say, 15, would suffice to estimate the value of [3]. After all, this encompasses roughly 99.95% of the probability. However, see Figure 2. This shows what the integrand of [3] looks like. Note: this *includes* the probability density factor.

Evidently, at 15, one is only beginning to see the exponential rise in the integrand. And the choice of where to truncate, in this example, has a material impact on the calculated value, as an inspection of the cumulative integrand in Figure 3 shows.

### LCL's Potential Dragon

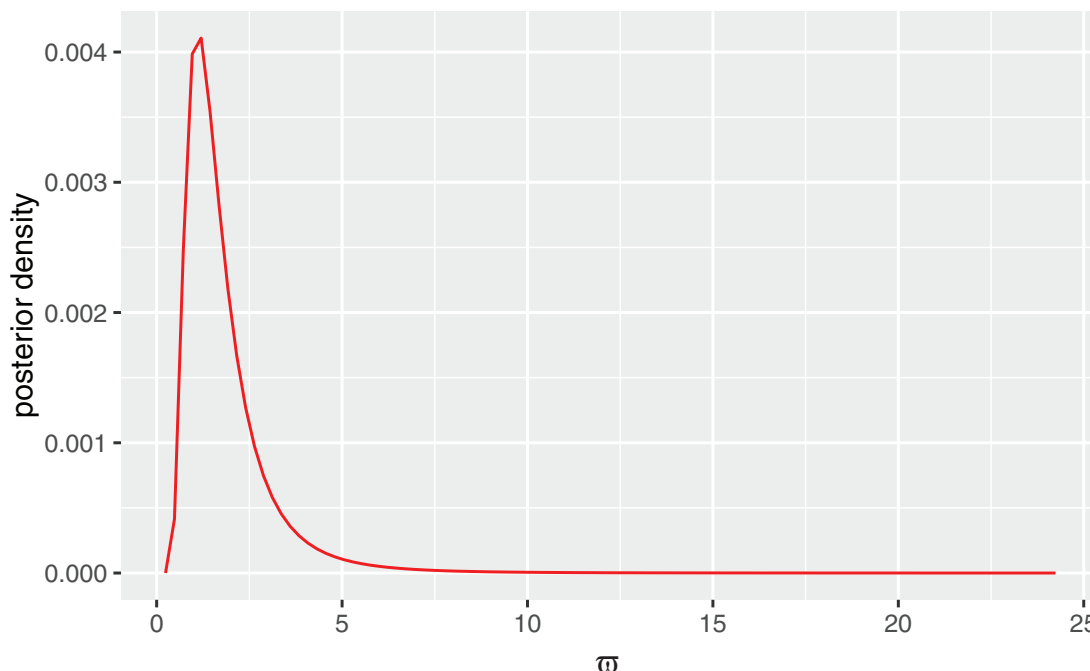
The leveled chain ladder is different from this simple model in several respects. In particular, the priors on the variances are bounded. This is an important difference, and it guarantees that the predictive mean does indeed exist. If they were not bounded, would the LCL have the same problem?

Figure 4 shows a "profile" of the posterior density  $f(\varpi_N)$  of the LCL as a function of the last development period variance  $\varpi_N$ , with all other parameters held constant at their MLE values.

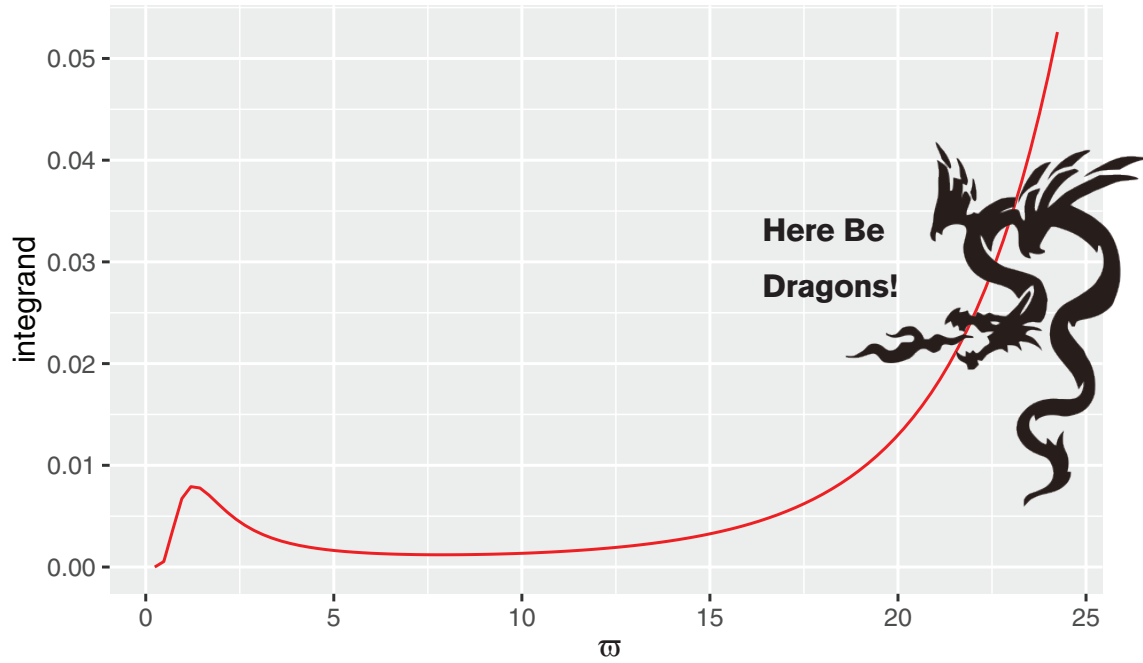
By the time  $\varpi = 4e-05$  is reached, over 99.99% of the cumulative posterior probability has been covered. Yet this is still far from the truncation boundary of  $\varpi = 1$  that Meyers used.

Figure 5 shows the integrand, i.e., the product of posterior density  $f(\varpi_N)$  and conditional expected loss for the final accident year  $E[Y_N | \varpi_N]$ .

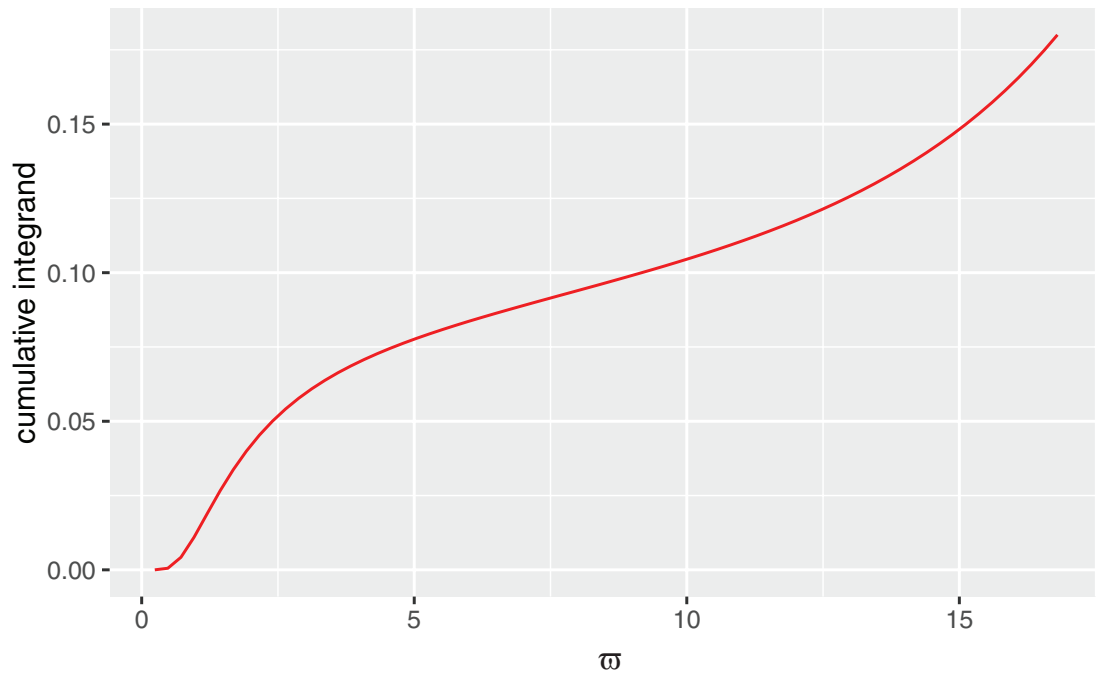
Figure 1: Density of inverse gamma distribution.



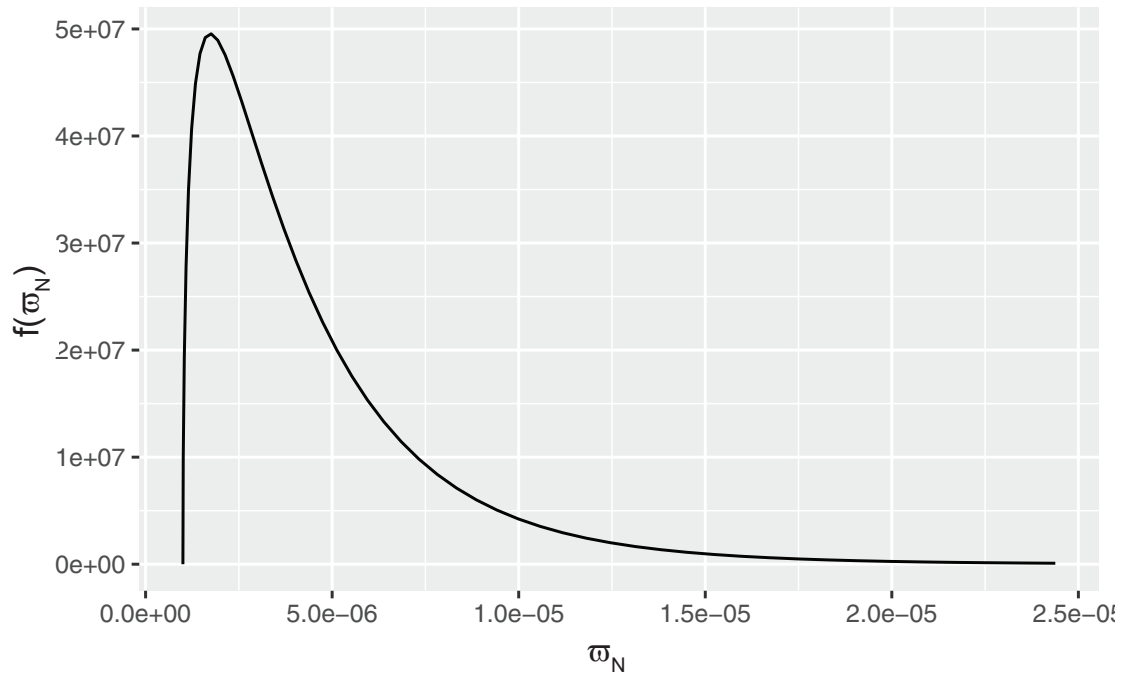
**Figure 2: Integrand for predictive mean ultimate loss.**



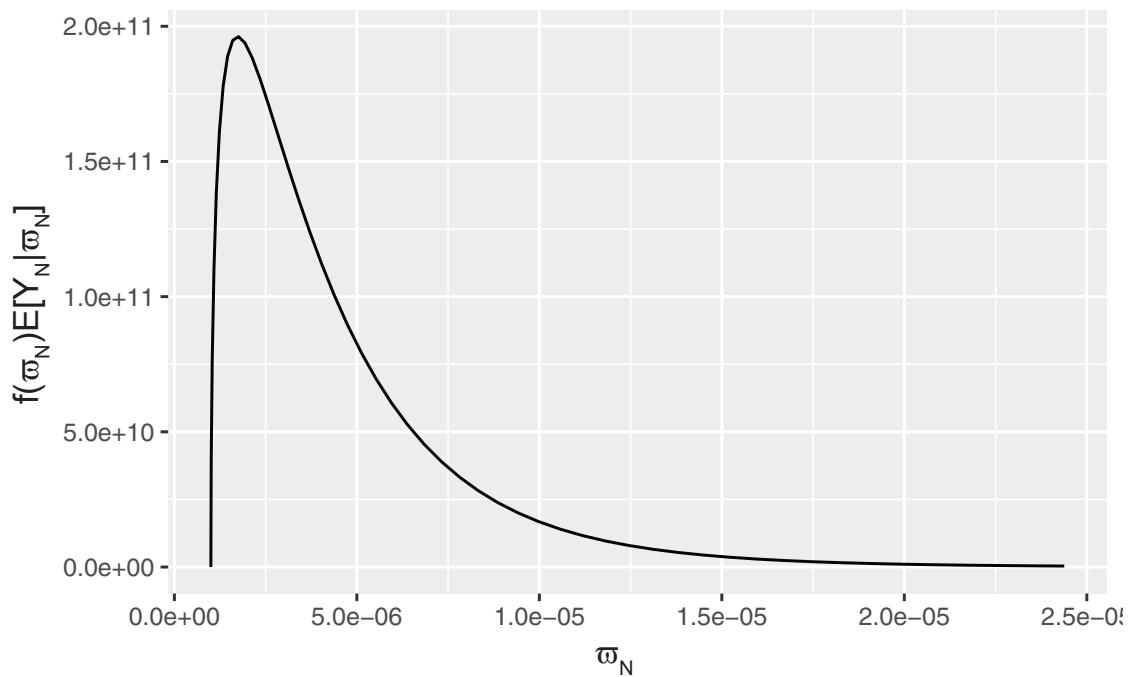
**Figure 3: Cumulative sum of the integrand of Figure 2.**



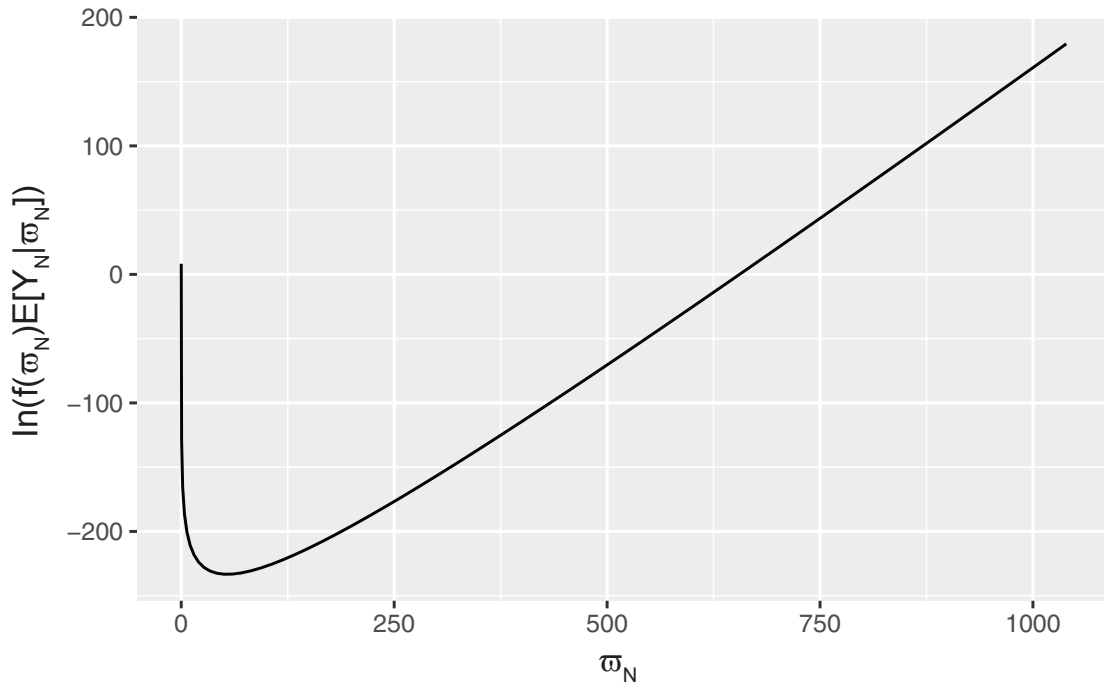
**Figure 4: Posterior density profile (up to scale factor) as function of last DP variance.**



**Figure 5: Contribution to predictive mean ultimate loss.**



**Figure 6: Log contribution to predictive mean ultimate loss.**



This graph is nearly isomorphic to Figure 4 because the conditional expected loss only varies from \$3959.63 to \$3959.68 over the range shown.

When the scale is taken out to  $\varpi = 1000$ , an absurdly improbable value, the divergence of the predictive mean can be seen. This is shown in Figure 6, with the y-axis rendered as the log integrand.

If  $\varpi$  were not capped and the predictive mean were computed by integrating over the full (all-parameters) posterior, the calculation would indeed diverge. However, it is unlikely that any MCMC application would ever reach the extreme, and extremely improbable,  $\varpi$  values required. The dragons live in unexplored territory.

Unlike the behavior seen in Figure 3, there is no material difference here between bounding  $\varpi$  at 0.01, 1, or even 500. The contribution to the predictive mean consists of a small pond at  $\varpi < .00003$  and an infinite ocean at  $\varpi > 500$ .

This suggests that when capping  $\varpi$  by design, the choice of limit is not material — in this case. Lack of materiality is true of the particular example in Meyers’ monograph. We can’t know a priori how other triangle data might behave.

**A Call for Caution**

This argues for caution, or at least circumspection, when applying numerical Bayesian methods to problems involving the lognormal. Moments may not exist, or may be made arbitrary by model design choices. Quantiles, on the other hand, should be well-behaved.

I don’t mean to pick on the fantastic work of Dr. Meyers; his just happens to be what led me to this line of inquiry. I have seen at least one blog explaining MCMC that went right ahead and calculated the predictive mean of a lognormal without even considering the question of the existence of the target. In a more formal venue, the seminal paper by De

Alba [2002] seems to exhibit this same lack of consideration.

The literature on stochastic loss reserving is large and growing, and a significant portion of it addresses Bayesian methods. How many other triangle models out there are affected by this phenomenon? I don’t know, but the question is worth addressing.

**References**

De Alba, Enrique. “Bayesian estimation of outstanding claim reserves.” *North American Actuarial Journal* 6.4 (2002): 1-20.

Meyers, Glenn. “Stochastic Loss Reserving Using Bayesian MCMC Models.” *CAS Monograph Series 1* (2015).

Rubinstein, Reuven Y., and Dirk P. Kroese. *Simulation and the Monte Carlo method*. Vol. 707. John Wiley & Sons, 2011. ●