DISCUSSION BY STUART A. KLUGMAN, PH.D.

While actuaries have had a Bayesian view of the world for decades, the adoption of methods that adhere strictly to the principles of modern Bayesian analysis has been slow. In his paper, Glenn Meyers shows that for a particular problem such an approach is not only feasible, but easy to complete. I am delighted that he has continued to take up the Bayesian cause, and with this note, I hope to provide just two extensions. One is to demonstrate that Meyers employed an approximation that was not needed for the particular prior distribution. The other is to provide an example that will confirm that his suggestions are indeed not limited to the Pareto distribution nor to one-parameter distributions.

To be fair to Meyers, and to continue his promotion of Bayesian methods as a practical solution to estimation problems, I will employ his definition of "practical:" that solutions can be obtained via simple spreadsheet calculations.

1. EXACT BAYESIAN CALCULATIONS

There have been a number of reasons for the slow adoption of exact Bayesian methods. One excellent discussion is Efron [3]. Aside from philosophical issues, there is a major computational one. Begin by defining the customary Bayesian estimation problem:

$$x = \text{data}$$

$$\theta = \text{parameter}$$

$$p(\theta) = \text{prior density}$$

$$f(x|\theta) = \text{model density}$$

$$f(\theta|x) = \text{posterior density}$$

$$t(\theta) = \text{quantity of interest}$$

$f(t|x)$ = posterior density of the quantity of interest.

In Meyers's paper, $\theta = q$, and t is the layer average severity, while the model density is the likelihood function.

One standard Bayesian estimate of a parameter is the posterior mean. For continuous models, the formula is

$$E(\theta|x) = \frac{\int \theta f(x|\theta)p(\theta)d\theta}{\int f(x|\theta)p(\theta)d\theta}. \tag{1.1}$$

The estimate of the quantity of interest is

$$E[t(\theta)|x] = \frac{\int t(\theta)f(x|\theta)p(\theta)d\theta}{\int f(x|\theta)p(\theta)d\theta}. \tag{1.2}$$

Thus any Bayesian estimation problem using the posterior mean reduces to evaluating a (possibly) multi-dimensional integral. The number and efficiency of methods to do so have greatly increased in the past decade. Four methods (extensions of one-dimensional numerical integration methods, Gauss-Hermite, Tierney-Kadane, Monte Carlo, empirical Bayes) are outlined in Klugman [4]. Recently two additional methods have been developed: the Gibbs sampler (Casella and George [2]) and sampling-resampling (Smith and Gelfand [5]). All of the methods require a large number of calculations and clearly do not meet our present standard of being spreadsheet-friendly.

Meyers offers the only alternative that requires a limited amount of calculation: Replace the customary continuous prior distribution with a discrete one. The integrals then become sums and are easy to calculate. The question that remains is whether additional approximations are needed in order to complete the posterior calculations.

2. EXACT CALCULATIONS FOR THE SINGLE PARAMETER PARETO DISTRIBUTION

For the specific problem addressed by Meyers we have

$$p(q) \propto q^{\alpha-1}e^{-\beta q} \qquad (2.1)$$

$$f(x|q) = k^{nq}q^{n}(\Pi x_i)^{-q-1}. \qquad (2.2)$$

The prior distribution is a Gamma distribution when α and β are both positive, and reduces to Meyers's noninformative prior when they are both zero. The posterior distribution is

$$f(q|x) \propto k^{nq}q^{n+\alpha-1}(\Pi x_i)^{-q-1}e^{-\beta q} \propto q^{n+\alpha-1}e^{-(\gamma+\beta)q} \qquad (2.3)$$

where $\gamma = -n\ln(k) + \Sigma\ln(x_i)$. This is just another Gamma distribution and so the posterior mean is

$$E(q|x) = \frac{\gamma}{\gamma+\beta}\hat{q} + \frac{\beta}{\gamma+\beta}\frac{\alpha}{\beta}, \qquad (2.4)$$

the usual weighted average of the maximum likelihood estimator and the prior mean. When $\alpha = \beta = 0$, the posterior mean is $\hat{q}$ as in Meyers (so once again the "WYSIWYG" estimator is obtained) but without resorting to the Normal approximation. The posterior variance is

$$\text{Var}(q|x) = \frac{n+\alpha}{(\gamma+\beta)^2}. \qquad (2.5)$$

With $\alpha = \beta = 0$ it is $\hat{q}^2/n$, also in agreement with Meyers.

The difference comes when other features are desired or when numerical approximations are needed. The other feature desired in Meyers's paper is the layer average severity. The required integral for the posterior mean of the layer average severity is

$$\mathrm{E}(t|q) = \int_0^\infty \frac{k^q}{q-1}\left(\frac{1}{R^{q-1}} - \frac{1}{L^{q-1}}\right)\frac{(\gamma+\beta)^{n+\alpha}}{\Gamma(n+\alpha)}\, q^{n+\alpha-1}e^{-(\gamma+\beta)q}dq\,, \quad (2.6)$$

which cannot be integrated analytically. A simple discretization (the composite trapezoid rule) should approximate this integral. For Meyers's example, the values are $n = 100$, $\gamma = 57.143$, $R = 1{,}000{,}000$, $L = 5{,}000{,}000$, $k = 100{,}000$, $\alpha = 0$, and $\beta = 0$. The approximate integral, evaluating q every 0.05 from 0 to 3, produced a posterior mean of 18,971 and a posterior standard deviation of 9,989. These cannot be compared with Meyers's paper as he did not solve this example.

The above calculations took advantage of the fact that the Gamma prior distribution turned out to be conjugate for the single parameter Pareto likelihood. (That is, the posterior turned out to have the same density type as the prior. The major advantage is that the constants needed to make Equation 2.3 an equality can be found without integrating.) This will seldom be the case for actuarial examples. To continue Meyers's example, we can use the prior that appears in his Exhibit 1. Exhibit 1 of this discussion provides the equivalent results using Meyers's discrete prior but retaining the exact likelihood function. The results are similar to the exact calculation done previously with a posterior mean of 18,972 and standard deviation of 9,989. These numbers are similar to those obtained by Meyers, but were not expected to be exactly the same.

3. EXTENSIONS TO MULTI-PARAMETER PROBLEMS

Avoiding the Normal approximation for the distribution of maximum likelihood estimators—in fact, avoiding maximum likelihood estimators of q altogether— may make extensions of Meyers's analysis easier. The major difficulty is that any sums must now be taken over a relatively large number of values. This is because, for example, two-dimensional approximate integration requires the square of the number of function evaluations as compared with a similar one-dimensional approximation. Rather than produce general formulas, the example used previously is extended to the case of a Lognormal distribution. Suppose a sample of size 100 was taken and the sufficient statistics (the only numbers other than the parameter values needed to compute the likelihood function) were

$$\sum \ln(x_i) = 1000, \text{ and } \sum [\ln(x_i)]^2 = 10{,}300.$$

Thus, the maximum likelihood estimates of the Lognormal parameters are $\hat{\mu} = 10$, and $\hat{\sigma}^2 = 3$. This leads to a maximum likelihood estimate of the layer average severity of 48,770. The noninformative prior distribution selected is the standard one (Berger [1], pp. 83-87) for the normal distribution: $p(\mu,\sigma) \propto 1/\sigma$. This implies a uniform (over the entire real line) prior on μ that is independent of the prior on σ. Possible values were restricted to the range $8 \le \mu \le 12$ and $1.0 \le \sigma \le 2.5$. The other relevant functions are:

$$f(x|\mu,\sigma) \propto \sigma^{-n} e^{-\frac{1}{2}\sum\left(\frac{\ln x_i - \mu}{\sigma}\right)^2}$$

$$= \sigma^{-n} e^{-\frac{10{,}300 - 2{,}000\mu + 100\mu^2}{2\sigma^2}}, \qquad (3.1)$$

and

$$t(\mu,\sigma) = e^{\mu+\sigma^2}\left[\Phi\left(\frac{\ln(5{,}000{,}000)-\mu}{\sigma} - \sigma\right) - \Phi\left(\frac{\ln(1{,}000{,}000)-\mu}{\sigma} - \sigma\right)\right]$$

$$+ 4{,}000{,}000 - (5{,}000{,}000)\,\Phi\left(\frac{\ln 5{,}000{,}000 - \mu}{\sigma}\right)$$

$$+ 1{,}000{,}000\,\Phi\left(\frac{\ln(1{,}000{,}000)-\mu}{\sigma}\right). \qquad (3.2)$$

For the calculations, the ranges on μ and σ were split into 30 equally spaced intervals. This led to 961 function evaluations, of which 13 are displayed in Exhibit 2. The relevant posterior quantities appear at the end of the exhibit.

4. CONCLUSIONS

Through his paper, Glenn Meyers has reminded us that Bayesian calcuations can be relatively simple, and that they provide quantities of great interest to actuaries (mainly the standard deviation, and perhaps the complete distribution of the quantity to be estimated). This discussion points out that there may be simpler ways to do the calculations and that two-dimensional calculations are indeed feasible as Meyers indicated.

REFERENCES

[1] Berger, James O., *Bayesian Inference in Statistical Analysis,* Second edition, New York, Springer-Verlag, 1985.

[2] Casella, George and Edward I. George, "Explaining the Gibbs Sampler," *The American Statistician*, 46, 1992, pp. 167-174.

[3] Efron, Bradley, "Why Isn't Everyone a Bayesian?," *The American Statistician*, 40, 1986, pp. 1-11.

[4] Klugman, Stuart A., *Bayesian Statistics in Actuarial Science with an Emphasis on Credibility*, Boston, Kluwer, 1992.

[5] Smith, Adrian F.M. and Alan E. Gelfand, "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46, 1992, pp. 84-88.

EXHIBIT 1

SINGLE PARAMETER PARETO—POSTERIOR ANALYSIS

q	t	b(post)	B(postcdf)	$q*b$	$q*q*b$	$t*b$	$t*t*b$
1.000	160944	0.0000	0.0000	0.00000	0.00000	0.07	10512
1.050	137822	0.0000	0.0000	0.00000	0.00000	0.40	554478
1.100	118085	0.0000	0.0000	0.00002	0.00002	1.98	233847
1.150	101229	0.0001	0.0001	0.00009	0.00010	7.95	804496
1.200	86826	0.0003	0.0004	0.00037	0.00044	26.46	2297361
1.250	74512	0.0010	0.0014	0.00125	0.00156	74.21	5529579
1.300	63979	0.0028	0.0042	0.00361	0.00469	177.72	11370313
1.350	54964	0.0067	0.0109	0.00903	0.01220	367.78	20214671
1.400	47245	0.0141	0.0249	0.01970	0.02758	664.74	31405343
1.450	40631	0.0261	0.0510	0.03781	0.05482	1059.42	43045108
1.500	34961	0.0430	0.0940	0.06443	0.09664	1501.62	52498447
1.550	30099	0.0634	0.1573	0.09824	0.15226	1907.56	57414828
1.600	25926	0.0844	0.2417	0.13498	0.21596	2187.06	56700811
1.650	22343	0.1019	0.3436	0.16819	0.27752	2277.50	50885512
1.700	19265	0.1125	0.4561	0.19120	0.32502	2166.60	41739159
1.750	16619	0.1139	0.5700	0.19931	0.34880	1892.83	31457522
1.800	14344	0.1064	0.6764	0.19149	0.34468	1525.98	21889173
1.850	12387	0.0921	0.7684	0.17030	0.31506	1140.29	14124672
1.900	10702	0.0741	0.8425	0.14079	0.26750	793.00	8486676
1.950	9251	0.0557	0.8982	0.10860	0.21177	515.19	4765789
2.000	8000	0.0392	0.9374	0.07844	0.15688	313.75	2510000
2.050	6922	0.0260	0.9634	0.05322	0.10910	179.69	1243787
2.100	5992	0.0162	0.9796	0.03402	0.07145	97.07	581624
2.150	5189	0.0096	0.9892	0.02055	0.04419	49.60	257374
2.200	4496	0.0053	0.9945	0.01176	0.02587	24.03	108049
2.250	3897	0.0028	0.9973	0.00639	0.01438	11.07	43138
2.300	3380	0.0014	0.9988	0.00331	0.00760	4.86	16415
2.350	2932	0.0007	0.9995	0.00163	0.00383	2.03	5966
2.400	2545	0.0003	0.9998	0.00077	0.00185	0.82	2075
2.450	2210	0.0001	0.9999	0.00035	0.00085	0.31	692
2.500	1920	0.0001	1.0000	0.00017	0.00038	0.12	222

	mle	post mean	post sd	low 95	high 95
For q	1.750	1.7500	0.1749	1.400	2.100
For E(x)	16,619	18,972	9,989	5,992	47,245

EXHIBIT 2

Lognormal Posterior Analysis

μ	σ	prior	t	model	posterior
8	1	1	0	9.930E-153	3.173E-108
8	1.75	.57143	320	1.155E-74	2.108E-30
8	2.5	.4	14,390	7.683E-65	9.818E-21
9.867	1.75	.57143	12,154	1.989E-46	3.631E-2
10	1	1	20	7.175E-66	2.292E-21
10	1.7	.58824	12,174	2.593E-46	4.874E-2
10	1.75	.57143	15,172	2.658E-46	4.854E-2
10	1.8	.55556	18,616	2.326E-46	4.128E-2
10	2.5	.4	114,619	6.066E-51	7.753E-7
10.133	1.75	.57143	18,845	1.989E-46	3.631E-2
12	1	1	20,411	9.930E-153	3.173E-108
12	1.75	.57143	234,939	1.155E-65	2.108E-30
12	2.5	.4	560,653	7.683E-65	9.818E-21

$E(\mu)$	10.000
$StdDev(\mu)$	.17586
$E(\sigma)$	1.7541
$StdDev(\sigma)$	.12609
$E(t)$	17,452
$StdDev(t)$	10,811

Note: The entries in the "model" column are the evaluation of Equation 3.1. The entries in the "posterior" column are the entries in the "model" column divided by σ and then multiplied by a constant to make them sum to one.