Introduction to Bayesian Loss Development

David R. Clark, FCAS

Abstract

This paper provides an introduction to the use of Bayesian methods for blending prior information with a loss development pattern from a triangle. The methods build upon conjugate forms discussed in earlier literature but introduce the Generalized Dirichlet as a prior, which allows for a significant simplification in calculation. The discussion is mainly restricted to the question of blending observed data with prior beliefs and not on the question of reserve ranges.

The paper is aimed at practicing actuaries seeking an introduction to Bayesian ideas for loss development. The methods will work with a single development triangle analyzed in a spreadsheet.

Keywords. Bayesian loss development, conjugate prior, Generalized Dirichlet

1. INTRODUCTION

The selection of loss development patterns is a critical piece of actuarial analysis for casualty insurance business and it arises in both pricing and reserving. The most common data structure for this analysis is the development triangle. The actuary can estimate a pattern from the triangle but would typically impose judgment in selecting the final pattern based on prior knowledge or external data.

The incorporation of expert judgment and external data makes loss development analysis a natural application for the Bayesian framework. The Bayesian framework provides a way to incorporate this prior knowledge in a systematic way. This paper will provide a very basic model to allow the practicing actuary to begin using the Bayesian ideas.

In many non-insurance applications of Bayesian statistics, the observed data overwhelms the prior distribution, making the exact form of the prior irrelevant. This is not so for insurance, where the data is often sparse or volatile; the prior knowledge can have a great influence on the final results.

We will focus on the narrow problem of selecting a pattern from a loss development triangle (no exposure units or loss ratio information), blending the data in the triangle with prior knowledge. For convenience, this will be done using conjugate forms, which make the calculations trivial to perform. Anyone who knows how to calculate an age-to-age factor will be able to begin doing Bayesian analysis right away.

1.1 Research Context

Bayesian ideas have been part of actuarial thinking for many years, often in the context of credibility theory, which has been called the "cornerstone of actuarial science" (Hickman, 1999). Bayesian methods have previously been introduced in the context of reserving, and have gained more attention recently because of advances in computational algorithms such as Markov Chain Monte Carlo (MCMC) techniques.

The Bayesian approach has been noted for three major advantages:

1) It allows the analyst to incorporate prior knowledge or expertise in a logically coherent way.

2) It can incorporate complex, nonlinear relationships to provide a more realistic model than can be done otherwise.

3) It can incorporate uncertainty in all model parameters and therefore produce realistic reasonable ranges around predicted values.

Prior papers such as Meyers (2015) and Zhang, et al (2012) have generally focused on the problem of estimating ranges around reserve estimates. Authors such as Robbin (1986), Mildenhall (2006), Wüthrich (2007), and England, et al (2012) have also used the Bayesian concepts to illuminate the relationships between traditional models such as chain ladder and Bornhuetter-Ferguson.

While many papers acknowledge that "The Bayesian paradigm offers a formal mechanism for incorporating into one's analysis information not contained in the available data" (Zhang, 2012), it is not always clear how this can be done. Diffuse or noninformative priors are used in much of the literature.

1.2 Objective

In this paper, we present a conjugate Bayesian model applied to a standard loss development triangle. We will assume that the goal of the analyst is to estimate a development pattern using this data to update prior beliefs. Our focus will be on how to organize the prior beliefs about the development pattern into an explicit prior distribution for this blending problem.

By staying in the context of the conjugate[1] models, the blending of prior knowledge with new data can be done with very simple calculations. This allows analyst to begin experimenting with these ideas immediately without the need for special software or programming skills. The hope is that this model will help build intuition in the Bayesian framework and become the stepping stone for expanding to more advanced models.

1.3 Outline

Section 2 of this paper will outline the mathematics of the Bayesian conjugate form for the loss development pattern estimation; this will give all of the theory underlying the approach. Section 3 will provide a numerical example showing how the model can be implemented in practice. Section 4 gives a brief sketch of future research and ways to extend the model into more realistic (and more complex) forms.

2. BACKGROUND AND MATHEMATICS

This section provides all of the mathematics needed to derive the conjugate family for Bayesian loss development. Most of this is not critical for the actuary who is only looking to implement the method, and can be skimmed.

2.1 Bayesian Theory in General

Bayesian theory assumes that an analyst working with a loss development triangle does not start as a "blank slate" with no idea of what a development pattern looks like. Instead, it assumes that the analyst comes with some "prior" expectation and is willing to change that prior belief based on what is observed in the new data.

The theory is derived from Bayes' theorem, which calculates the "inverse probability" of a parameter value θ, based on observed data x.

---

[1] Conjugacy is "the property that the posterior distribution follows the same parametric form as the prior distribution" (Gelman, et al (2013), page 35). This is a technical definition, but the attraction of conjugacy is in the practical implementation and interpretability.

$$f(\theta \mid x) = \frac{f(x \mid \theta) \cdot f(\theta)}{f(x)} = \frac{f(x \mid \theta) \cdot f(\theta)}{\int_{\theta} f(x \mid \theta) \cdot f(\theta) d\theta} \tag{1.1}$$

The major challenge for applying Bayes' theorem in practice is that the parameter θ is usually a vector of multiple parameters. This means that we need to specify a multi-dimensional distribution $f(\theta)$ and also be able to evaluate the multi-dimensional integral in the denominator. This presents a computational challenge.

There have been three main strategies for handling the computation challenge:

1) Use of conjugate priors, allowing closed-form solutions for carefully chosen distributional forms.
2) Linear approximations to the formula (e.g., Bühlmann-Straub)
3) Numerical approximations
    a. Quadrature evaluation of the integral
    b. Simulation-based approached (MCMC)

With greater computer speeds and improved algorithms, the simulation-based methods have allowed for Bayesian methods to be used in many fields. These models are especially useful when we need to evaluate complex models.

The conjugate families are much more useful for introductory purposes because they allow the calculations to be done simply and even manually. It is also very useful to include conjugate forms in some of the components of a simulation model ("conditionally conjugate" parameters) to improve efficiency.

For this paper, we will stay in the conjugate world in order to introduce all of the concepts in the loss development application such that any actuary can implement. If you can calculate an age-to-age factor, then you can do Bayesian analysis!

2.2 The Beta-Binomial Conjugate Relationship

Blending patterns is a multivariate problem, but it is easiest to attack the problem by starting with the univariate case. We begin with the univariate Beta-Binomial case, because it will be the main building block for the loss development application.

The Beta distribution works with a continuous random variable, p, that can be any value between 0 and 1. The density function is given below.

$$f\left(p\right)=\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\cdot\Gamma\left(\beta\right)}\cdot p^{\alpha-1}\cdot\left(1-p\right)^{\beta-1} \tag{2.1}$$

$$E\left(p\right)=\frac{\alpha}{\alpha+\beta} \tag{2.2}$$

The Beta distribution is usefully interpreted as the ratio of gamma random variables. The two gamma random variable have different shape parameters, but share a common scale parameter ϕ, which does not affect the Beta random variable.

$$\begin{gathered} Z_1 \sim Gamma\left(\alpha,\phi\right) \\ Z_2 \sim Gamma\left(\beta,\phi\right) \\ p=\frac{Z_1}{Z_1+Z_2} \end{gathered} \tag{2.3}$$

We can also note that the shape parameters α and β must be positive numbers but they are not restricted to being integer values.

The likelihood function for the observed data x will be assumed to come from a binomial distribution with the probability function below.

$$f\left(x\mid p\right)=\binom{n}{x}\cdot p^{x}\cdot\left(1-p\right)^{n-x} \tag{2.4}$$

The binomial is often interpreted as the number of "successes" observed in a sample of n trials, given a probability of success p. The maximum likelihood estimator of this probability is calculated easily.

$$\hat{p}=\frac{x}{n} \tag{2.5}$$

While the binomial distribution is strictly speaking restricted to integer values, we will make an approximation in this application that the estimator above can include non-integer values for x and/or n when estimating the proportion p.

If the parameter p has a Beta prior distribution as defined above, then we apply Bayes' theorem to revise our distribution based on the observed data.

$$f(p \mid x) = \frac{f(x \mid p) \cdot f(p)}{f(x)} = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x) \cdot \Gamma(\beta+n-x)} \cdot p^{\alpha+x-1} \cdot (1-p)^{\beta+n-x-1} \tag{2.6}$$

The fact that the posterior distribution for p is again a Beta distribution gives us the reason for calling this a "conjugate" form.

The expected value of the proportion can also be written in a linear form.

$$E(p \mid x) = \frac{\alpha+x}{\alpha+\beta+n} = \left(\frac{x}{n}\right) \cdot \left(\frac{n}{\alpha+\beta+n}\right) + \left(\frac{\alpha}{\alpha+\beta}\right) \cdot \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right) \tag{2.7}$$

Alternatively, we can write the updating of parameters in a simple form:

$$\begin{aligned} \alpha^{(1)} &= \alpha^{(0)} + x \\ \beta^{(1)} &= \beta^{(0)} + n - x \end{aligned} \tag{2.8}$$

With this updating formula, we have a very useful way of interpreting the parameters as being "pseudo-data." That is, the prior parameters $\alpha^{(0)}$ and $\beta^{(0)}$ are combined with the new data as though they were previously observed data points. Our prior knowledge is used as though it had been previously observed data.

Koop, et al (2007, page 19) summarize this concept well:

> "Natural conjugate priors have the desirable feature that prior information can be viewed as 'fictitious sample information' in that it is combined with the sample in exactly the same way that additional sample information would be combined. The only difference is that the prior information is 'observed' in the mind of the researcher, not in the real world."

This interpretability is useful when prior knowledge comes in a subjective form. For example, someone may say "I selected the development pattern based upon my twenty years of experience as an actuary." This is still useful in the Bayesian framework but we need to translate twenty years of experience into equivalent dollars of loss development data.

2.3 The Dirichlet-Multinomial Conjugate Relationship

The Dirichlet distribution is a multivariate generalization of the Beta distribution, which allows for a sequence of proportions, $\{p_1, p_2, \cdots, p_k\}$. The probability density function is similar to the Beta distribution except that the random variable is now a vector of percentages. These are interpreted as the incremental percentages of ultimate loss paid or reported in each identified period.

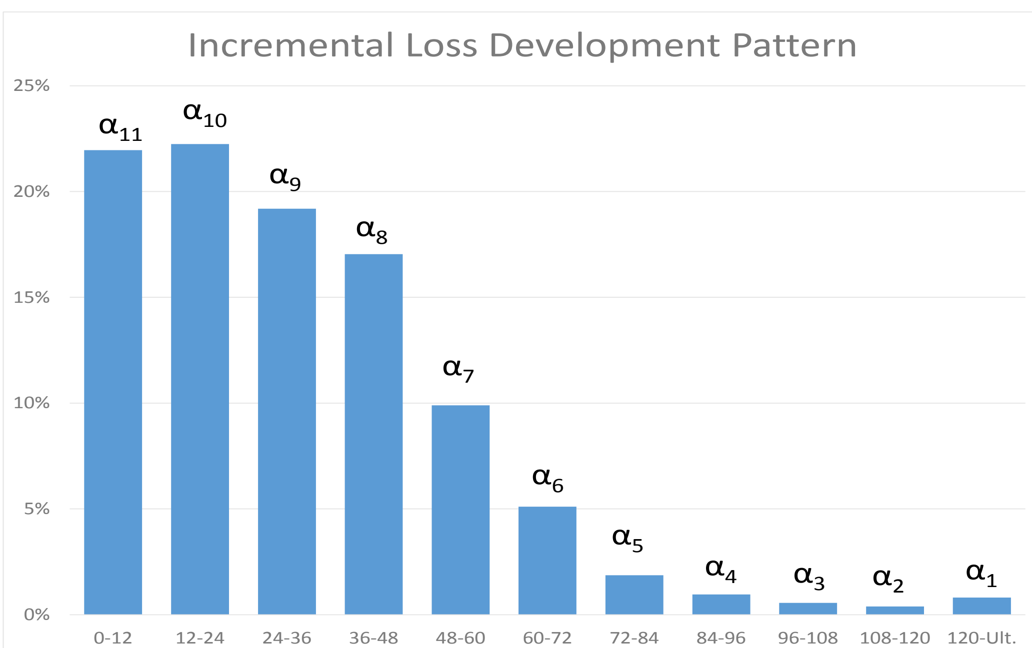
$$f(p) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\cdot\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)} \cdot \prod_{i=1}^{k} p_i^{\alpha_i - 1} \tag{3.1}$$

The expected percent-of-ultimate in each period is proportional to its corresponding alpha.

$$E(p_i) = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \cdots + \alpha_k} \tag{3.2}$$

The sequence of expected percentages produces the expected loss development pattern (either paid or reported). Figure 1 represents the proportion of ultimate loss in each incremental period. This assumption is consistent with Robbin (1986), Hesselager & Witting (1988), de Alba (2002), and Mildenhall (2006).

Figure 1



Similar to the Beta-Binomial model, the Dirichlet is conjugate with a Multinomial distribution, the Multinomial being the multivariate generalization of the Binomial. The parameters are given a similar updating.

$$\begin{aligned} \alpha_1^{(1)} &= \alpha_1^{(0)} + x_1 \\ \alpha_2^{(1)} &= \alpha_2^{(0)} + x_2 \\ &\vdots \\ \alpha_k^{(1)} &= \alpha_k^{(0)} + x_k \end{aligned} \tag{3.3}$$

In this updating formula, the sequence $\{x_1, x_2, \cdots, x_k\}$ is proportional to the observed losses in each development period. It is most convenient to think of these as the shape parameters of gamma random variables, similar to the sequence of $\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$. As such, the new data comes as the incremental losses divided by the common scale parameter ϕ.

The scale parameter ϕ is the variance/mean ratio of the loss data. We will assume that this is a fixed and known quantity, though that assumption can be relaxed later in the work.

If an estimate of the variance/mean ratio is needed, it can be approximated from the data just as is done for the dispersion parameter in a GLM[2], where $C_{t,d}$ is the cumulative loss for year t as of development period d. This is approximately a variance/mean ratio.

$$\phi = \frac{Var(C_{t,d})}{E(C_{t,d})} \approx \frac{1}{n - \#param} \cdot \sum_{t,d} \frac{\left(\left(C_{t,d+1} - C_{t,d}\right) - C_{t,ult} \cdot p_{k-d}\right)^2}{C_{t,ult} \cdot p_{k-d}} \tag{3.4}$$

The major difficulty in the Dirichlet-Multinomial model is that we need to have a complete development pattern from the data in order to perform the updating. This is precisely not the case for loss development; we have a triangle of <u>incomplete</u> patterns. Fortunately, this difficulty is solved via the Generalized Dirichlet distribution.

[2] See for McCullagh & Nelder (1989) as a standard reference.

This is the same concept used in the over-dispersed Poisson (ODP) version of the chain ladder method, as presented in papers such as Renshaw and Verrall (1998). This connection is not accidental, as the binomial model presented here is simply a conditional Poisson model. That is, if X_1 and X_2 are Poisson random variables, then $X_1|X_1 + X_2 = N$ is a binomial random variable. This relationship extends to the over-dispersed and multivariate versions of the distributions.

2.4 The Generalized Dirichlet Distribution

The Generalized Dirichlet distribution was introduced by Connor and Mosimann (1969) in the context of biological science. Wong (1998) further investigated this form and provides the Bayesian updating formulas. Ng, et al (2011) provides more description of this distribution, renaming it the "nested Dirichlet."

Instead of a sequence of model parameters $\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$, we have a parameter set with alphas and betas: $\{\alpha_1, \alpha_2, \cdots, \alpha_{k-1}, \beta_1, \beta_2, \cdots, \beta_{k-1}\}$. Just as α_i was seen to be proportional to incremental loss, the β_i parameter is proportional to cumulative loss. This added flexibility means that we can have different weights for each cumulative development age, making it natural for the development triangle data format.

These parameters generalize the Dirichlet distribution given above. But the random variable $p = \{p_1, p_2, \cdots, p_k\}$, is interpreted exactly the same as before.

$$f(p) = p_k^{\beta_{k-1}-1} \cdot \prod_{i=1}^{k-1} \left[\frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i) \cdot \Gamma(\beta_i)} \cdot p_i^{\alpha_i - 1} \cdot \left(\sum_{j=i}^{k} p_j \right)^{\beta_{i-1} - (\alpha_i + \beta_i)} \right] \tag{4.1}$$

The Generalized Dirichlet has independence[3] between p_1 and $p_2/(1-p_1)$ and between subsequent conditional values $p_3/(1-p_1-p_2)$ and so forth. For the loss development application this implies that all of the age-to-age factors are independent. This independence assumption between age-to-age factors is paralleled in the chain ladder method (Mack, 1993).

The expected incremental losses are given as below. Formulas for all of the moments and co-moments are given in Wong (1998).

$$E(p_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \cdot \prod_{j=1}^{i-1} \frac{\beta_j}{\alpha_j + \beta_j} \quad i = 2, \cdots, k \tag{4.2}$$

The expected incremental values are more easily calculated via a recursive formula.

$$E(p_i) = \frac{\alpha_i}{\alpha_i + \beta_i} \cdot E(p_{i-1}) \cdot \left(\frac{\beta_{i-1}}{\alpha_{i-1}} \right) \quad i = 2, \cdots, k \tag{4.3}$$

---

[3] This property is described as "neutrality" by Connor and Mosimann (1969), and it only holds for the Generalized Dirichlet when the variables are ordered. It is for this reason that we use the notation that the first variable is the tail factor, and then move from right to left up to k as the first (usually age 12) factor. In this order the distribution is a perfect model for development triangle data.

The Dirichlet is a special case when $\beta_j = \alpha_{j+1} + \beta_{j+1}$.

The Bayesian updating formulas are also straightforward.

$$\alpha_j^{(1)} = \alpha_j^{(0)} + x_j \qquad (4.4)$$
$$\beta_j^{(1)} = \beta_j^{(0)} + x_{j+1} + x_{j+2} + \cdots + x_k$$

Using the cumulative losses from the triangle, this is written as shown below. For losses in accident year t as of development period d, the cumulative amount is $C_{t,d}$. The values used for updating the parameters remove the scaling parameter: $x = (C_{t,d+1} - C_{t,d})/\phi$.

$$\alpha_j^{(1)} = \alpha_j^{(0)} + \frac{1}{\phi}\sum_{t=1}^{k}\left(C_{t,d+1} - C_{t,d}\right) \qquad (4.5)$$
$$\beta_j^{(1)} = \beta_j^{(0)} + \frac{1}{\phi}\sum_{t=1}^{k} C_{t,d}$$

The dispersion parameter ϕ acts as a scaling parameter on the loss data from the triangle.

The great advantage of this Generalized Dirichlet is that we can exclude the first p_1 or the first several points and the remaining points are still a Generalized Dirichlet. Further, the relationship of the first increment to the sum of the remaining increments is always a Beta distribution.

$$E\left(p_1\right) = \frac{\alpha_1}{\alpha_1 + \beta_1} \qquad (4.6)$$

This relationship of one period to all the others is exactly what is needed in the calculation of age-to-age link ratios in the chain ladder method. The notation needs to be reversed: for example, count i=1 for last incremental loss oldest period, and $i = k$ for losses in the first year. The model parameters therefore translate very easily into familiar age-to-age factors.

$$ATA_{12-24} = \frac{\alpha_k + \beta_k}{\beta_k} \qquad ATA_{120-ult} = \frac{\alpha_1 + \beta_1}{\beta_1} \qquad (4.7)$$

The age-to-age factor for development period d is calculated from the triangle as shown below. The weighted average age-to-age (ATA) factor should be familiar to most actuaries.

$$ATA_d = \frac{\sum_{t=1}^{k} C_{t,d+1}}{\sum_{t=1}^{k} C_{t,d}} \tag{4.8}$$

The credibility blended numbers are given in a simple form as in formula (4.9) below.

$$ATA_d = \frac{\phi \cdot \left(\alpha_{k-d} + \beta_{k-d}\right) + \sum_{t=1}^{k} C_{t,d+1}}{\phi \cdot \beta_{k-d} + \sum_{t=1}^{k} C_{t,d}} \tag{4.9}$$

Given the model parameters for the Generalized Dirichlet and the scaling parameter ϕ, this credibility blending can be performed in a spreadsheet cell or even on paper. A numerical example of this calculation is given in Section 3.2.

Part of what has made the conjugate form so easy to implement is the assumption of independence between development ages. Unfortunately, the disadvantage of the independence assumption is that ages with little volume will get little credibility weight. There is no consideration of adjacent points, and no more weight assigned if all ages show consistently better (or worse) development than the benchmark. Most notably, the benchmark tail factor will never change based on the client data.

Most users, however, would want some dependence between ages. For example, if all of the age-to-age factors in the client's triangle are below the benchmark, then the benchmark tail should also be reduced. The next section of our paper will provide a way to include such a dependence structure.

2.5 Mixtures of Generalized Dirichlet Distributions

The model above provides a full conjugate Bayesian model that can be easily implemented by an analyst with knowledge of calculating age-to-age factors. The conjugate family is actually a bit more flexible still and allows for further expansion of the prior distributions.

The principle is that a linear combination of conjugate priors will still be a conjugate prior. If the analyst decides that the prior knowledge includes a library of possible development patterns (perhaps slow/medium/fast), then the prior is defined as a weighted average of these priors. The weights $\{w_1, w_2, w_3\}$ act as a discrete mixture distribution.

$$f\left(p\right)=w_1\cdot GD_1\left(p\right)+w_2\cdot GD_2\left(p\right)+w_3\cdot GD_3\left(p\right) \tag{5.1}$$

For each of the individual Generalized Dirichlet distributions $GD_l(p)$, we perform the same updating as outlined in the previous section. In addition, we update the weights in proportion to the likelihood functions for each.

The likelihood functions are the products of the Beta-Binomial functions for each age included in the analysis.

$$f\left(x\right)=\int_0^1 f\left(x\mid p\right)\cdot f\left(p\right)dp=\binom{n}{x}\cdot\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\cdot\Gamma\left(\beta\right)}\cdot\frac{\Gamma\left(\alpha+x\right)\cdot\Gamma\left(\beta+n-x\right)}{\Gamma\left(\alpha+\beta+n\right)} \tag{5.2}$$

For non-integer values of n and x, we can replace $\binom{n}{x}$ with $\frac{\Gamma(n+1)}{\Gamma(n-x+1)\cdot\Gamma(x+1)}$. We may also note that a special case of formula (5.2) is the uniform distribution when $\alpha=\beta=1$, indicating that all values are equally likely.

The updating of the weights is a straight-forward application of Bayes' theorem.

$$w_j^{(1)}=\frac{w_j^{(0)}\cdot f_j\left(x\right)}{w_1^{(0)}\cdot f_1\left(x\right)+w_2^{(0)}\cdot f_2\left(x\right)+w_3^{(0)}\cdot f_3\left(x\right)} \tag{5.3}$$

Section 3.3, below, gives a numerical example illustrating this formula. The ability to adjust the tail factor in the data according to client data is a major practical advantage.

3. NUMERICAL EXAMPLE

3.1 Selecting the Model Parameters

The description of the Bayesian model given in the previous section has flexibility for the analyst to supply a large number of prior parameters. We now discuss how this can be done without making all of these choices arbitrary.

Parodi and Bonche (2010), describe the uncertainty in prior information from two sources:

1. Market heterogeneity – the spread of different risks around some industry average
2. Estimation uncertainty – the industry average, though large, may still be of limited size

We may choose to give the prior distribution more variance depending upon how we evaluate these sources of uncertainty. Nonetheless, we usually have some prior knowledge and are not completely uninformed about external information.

In many application of Bayesian models, the choice of prior is not given much attention because it is assumed that the data will overwhelm the prior assumption anyway. For insurance applications we cannot assume this, and instead want to provide meaningful prior information. The discussion of "noninformative" or "diffuse" priors is therefore just a starting point.

For the Beta or Dirichlet distributions, a standard noninformative prior is to set all of the parameters equal to 1.00. That is, $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 1$; this is sometimes referred to as the Laplace prior. Even more diffuse is the Jeffreys prior with $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 1/k$. In both these cases, the expected pattern would have equal percentages in each period. In the most extreme case, we have $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$; which is an improper prior, sometimes called a Haldane prior, that gives no weight to the prior information and therefore will always result in a posterior expected value equal to the chain ladder calculation.

We would like our prior to have expected values equal to our prior knowledge. In the reserving exercise, this may be equal to the pattern selected in a prior reserve study. In the pricing exercise, the prior pattern may be taken from the expiring pricing or from an average of similar risks.

One approach to setting the sequence of alphas is to make them proportional to the incremental losses in our benchmark pattern. If these are scaled to add up to 1.00 then we have a very wide uncertainty similar to the Jeffreys prior. If the alphas add up to a larger quantity, say 100, then the prior benchmark pattern will be given much more weight. The sequence of betas can be set to make the Generalized Dirichlet equal to a simple Dirichlet: $\beta_j = \alpha_{j+1} + \beta_{j+1}$.

Alternatively, we can set $(\alpha_j + \beta_j)$ as a constant, generally greater than 2, with the α_j and β_j values set to match the ATA factors.

The other key input is the dispersion parameter ϕ, which is defined as the variance/mean of the data in the triangle. A small value of ϕ will result in more weight given to the new data because it implies small process variance.

This dispersion parameter may be estimated empirically from representative triangles, or it can be selected based on other sources for aggregate distributions. For example, in Table M[4] we can approximate the distributions using a Gamma. The expected loss group (ELG) represents the insurance charge at an entry ratio of 1.00. The expected losses for the ELG divided by the Gamma shape parameter is therefore an estimate for the scale ϕ.

Table 1

Theoretical "Table M" (for illustration)

Gamma Shape Parameter	Insurance Charge at Entry=1	Expected Loss Group	Aggregate Loss Size (example)	Implied Variance/Mean
0.5	0.484	48	360,000	720,000
1	0.368	37	1,000,000	1,000,000
1.5	0.308	31	2,000,000	1,333,333
2	0.271	27	3,750,000	1,875,000

For a starting point, we can select a combination of the parameters, such that $\phi \cdot (\alpha_j + \beta_j)$ is constant for all j.

If the prior distribution and scale parameter are calculated from a sample of patterns collected from peer companies, then it may be considered an "empirical Bayes" model. Schmid (2012) and Shi and Hartman (2014) provide models on that basis.

3.2 Numerical Example with One Benchmark Pattern

For an example of the loss development task, we introduce a triangle of cumulative loss payments. This data was taken from a sample of companies collected in the CAS Website, and represents Products Liability loss net of reinsurance. The example is, of course, only for illustration.[5]

The average age-to-age (ATA) factors are calculated as all year weighted averages. The "Col. 1" number represents the sum of losses for a given age, excluding the latest diagonal;

---

[4] Table M is an industry tool for excess-of-aggregate charges for Workers' Compensation. The numbers shown here are not from that source, but were created only to illustrate the concept.

[5] For the interested reader, an Excel file including the example that follows can be provided by the author upon request.

the "Col. 2" number represents the sum of the subsequent column.

Table 2

Sample Triangle = Cumulative Products Liability Paid Losses	12	24	36	48	60	72	84	96
1990	73	262	469	528	536	591	604	606
1991	148	346	391	502	522	514	567	
1992	99	198	219	394	408	430		
1993	118	255	352	412	581			
1994	275	415	645	803				
1995	261	446	637					
1996	130	471						
1997	148							
Col. 1	1,104	1,922	2,076	1,836	1,466	1,105	604	
Col. 2	2,393	2,713	2,639	2,047	1,535	1,171	606	
Avg ATA	2.168	1.412	1.271	1.115	1.047	1.060	1.003	

The average ATA factors are easily calculated by the actuary and, if desired, could be replaced with the values for only including the latest, say, three diagonals.

The ATA factors in the triangle show considerable volatility, so it is desirable to blend the data with other benchmarks to improve the stability.

Table 3

Age-to-Age Factors	12-24	24-36	36-48	48-60	60-72	72-84	84-96
1990	3.589	1.790	1.126	1.015	1.103	1.022	1.003
1991	2.338	1.130	1.284	1.040	0.985	1.103	
1992	2.000	1.106	1.799	1.036	1.054		
1993	2.161	1.380	1.170	1.410			
1994	1.509	1.554	1.245				
1995	1.709	1.428					
1996	3.623						

The table below brings in the prior knowledge. We assume that we know a loss development pattern. This pattern may come from industry sources, peer companies, or prior reserve studies.

We must select the alpha and beta parameters for each age. We can set these such that the expected pattern equals our benchmark: ATA = (Alpha+Beta)/Beta.

The total value of Alpha+Beta is selected to be 4.00 in this example, representing a weakly informative prior. The variance/mean ratio or scale parameter ϕ is selected as 1,000 ($1,000,000 since the original Schedule P units are in thousands). The "Col. 1" and "Col. 2" entries are simply the scale parameter times the Beta and Alpha+Beta parameters of the Generalized Dirichlet.

Table 4

Prior Assumptions

	12	24	36	48	60	72	84	96
LDF	21.950	7.787	3.946	2.512	1.842	1.558	1.415	1.315
ATA	2.819	1.973	1.571	1.364	1.182	1.101	1.076	1.315
Alpha	2.58	1.97	1.45	1.07	0.62	0.37	0.28	0.96
Beta	1.42	2.03	2.55	2.93	3.38	3.63	3.72	3.04
Alpha+Beta	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
Variance/ Mean:		1,000						
Col. 1	1,419	2,027	2,546	2,933	3,383	3,633	3,717	3,042
Col. 2	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000

The blended pattern is simply the addition of the Col. 1 and Col. 2 weights from the triangle and the benchmark pattern (scaled by ϕ).

As noted previously, the conjugate form puts the prior knowledge into a form as though it was prior loss development data. The prior knowledge is added to the data from the new triangle as though we actually had more loss in the weighted-average calculation. The table below makes use of formula (4.9) to blend the patterns.

Table 5

Example of Blending Client and Benchmark Patterns

	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-Ult
ATA from Triangle								
Col. 1	1,104	1,922	2,076	1,836	1,466	1,105	604	-
Col. 2	2,393	2,713	2,639	2,047	1,535	1,171	606	-
ATA	2.168	1.412	1.271	1.115	1.047	1.060	1.003	
Benchmark Pattern								
Col. 1	1,419	2,027	2,546	2,933	3,383	3,633	3,717	3,042
Col. 2	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000
ATA	2.819	1.973	1.571	1.364	1.182	1.101	1.076	1.315
Blended Pattern								
Col. 1	2,523	3,949	4,622	4,769	4,849	4,738	4,321	3,042
Col. 2	6,393	6,713	6,639	6,047	5,535	5,171	4,606	4,000
ATA	2.534	1.700	1.436	1.268	1.141	1.091	1.066	1.315

This calculation can be easily incorporated into reserving studies or pricing work. The values for the alpha, beta and scale parameters in our example are only for illustration; the actuary can sensitivity test values in real examples in order to gain intuition for setting reasonable values.

One limitation in this implementation is that the "tail" factor will always be equal to the benchmark number. This is because we have assumed independence between all ATA factors. This assumption is relaxed in the next section, as more robust priors are used.

3.3 Numerical Example with Library of Benchmark Patterns

The example in section 3.2 assumes that there is a benchmark development pattern and some level of uncertainty around that pattern. It further assumes independence between the ATA factors for each development age.

We can expand the prior assumptions to instead assume that there is not just a single benchmark pattern but rather a library of such patterns. For example, we may assume that there are fast, medium and slow developing businesses, perhaps differing due to settlement strategies or case reserving practices. Each of these patterns has its own Generalized Dirichlet parameters, and there is some prior belief as to the probability of a given triangle being from any member of the library.

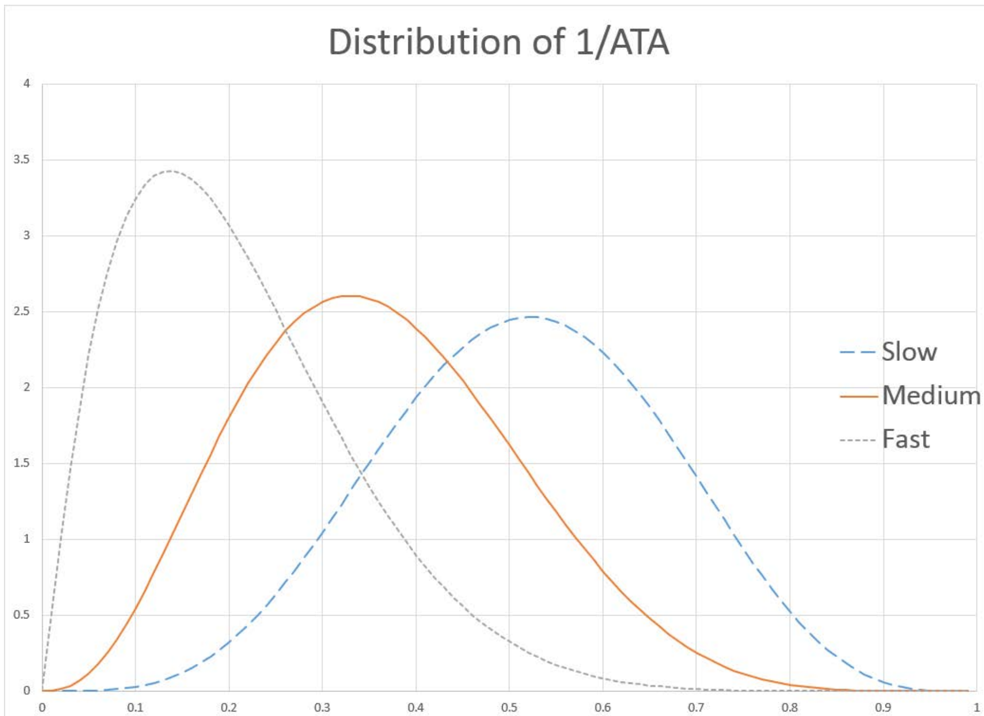
For a reinsurer, this may mean that their client companies' development patterns are naturally clustered into Fast/Medium/Slow groups, but without a perfect way to tell beforehand to which cluster a given client belongs.

Table 6

	Cumulative Loss Development Factors							
	12	24	36	48	60	72	84	96
Fast	14.014	4.930	2.607	1.759	1.406	1.263	1.191	1.155
Medium	21.950	7.787	3.946	2.512	1.842	1.558	1.415	1.315
Slow	49.240	15.860	7.407	4.163	2.706	2.057	1.750	1.567

As we noted earlier, the distribution of 1/ATA always follows a Beta distribution. For each development age, we can make a graph of the density functions for each of the benchmark patterns as a test for reasonableness.

Figure 2



We may have the case that the user has specified three different patterns, with variance within each. The prior mixture weights are assumed to be 1/3 to each of the three benchmark patterns. For Bayesian updating, the same procedure from Section 3.2 is applied for each of these patterns separately.

The mixture weights are then updated using formula (5.3). An example is show for the Fast pattern below, with formula (5.2) calculated as loglikelihood for each development age.

Table 7

Calculaton of Loglikelihood (Fast Pattern)

	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-Ult
Data from Triangle								
Col. 1	1,104	1,922	2,076	1,836	1,466	1,105	604	
Col. 2	2,393	2,713	2,639	2,047	1,535	1,171	606	
Variance/Mean Ratio:		1,000						
N	2.39	2.71	2.64	2.05	1.54	1.17	0.61	
X	1.29	0.79	0.56	0.21	0.07	0.07	0.00	
Benchmark Pattern								
LDF	14.014	4.930	2.607	1.759	1.406	1.263	1.191	1.155
ATA	2.843	1.891	1.482	1.251	1.113	1.060	1.031	1.155
Alpha	6.5	4.7	3.3	2.0	1.0	0.6	0.3	1.3
Beta	3.5	5.3	6.7	8.0	9.0	9.4	9.7	8.7
Alpha+Beta	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0
Loglikelihood	-0.9363	-1.0052	-0.8252	-0.5260	-0.2687	-0.2535	-0.0290	0.0000

This is known as a mixture model and it is still relatively easy to compute because a mixture of conjugate distributions is still a conjugate form. The posterior will again be a discrete mixture of Generalized Dirichlet distributions. Because the data from the triangle generally showed a faster pattern than implied in our benchmark, the weights are revised to shift more weight to the "Fast" curve.

Table 8

Bayesian Updating of Probabilities

	LogLikelihood	Difference in LL	Relative Likelihood	Original Weights	Revised Weights
	A	B=A-Max(A)	C=exp(B)	D	E=C*D/Avg(C)
Slow	-4.61	-0.77	0.464	33.33%	20.41%
Baseline	-4.06	-0.21	0.810	33.33%	35.61%
Fast	-3.84	0.00	1.000	33.33%	43.98%
			0.758	100.00%	100.00%

This use of a mixture of benchmark patterns can be expanded to include as many alternative patterns as desired, though for practical purposes three is sufficient. The major point is simply to illustrate the great flexibility for incorporating prior knowledge.

4. RESULTS AND DISCUSSION

It was the main goal of this paper to provide a Bayesian model that can be implemented quickly for the practicing non-technical actuary. The use of the conjugate form allows that implementation. Once this introductory material has been mastered, it is hoped that actuaries will seek to expand the model to make them more realistic.

4.1 Summary of Conjugate Model

The conjugate model is based on some simplified assumptions for ease of implementation. It is worth remembering some of the assumptions we have required.

1) The variance/mean ratio is assumed to be constant and known (supplied by the analyst)
2) All incremental development should be strictly positive
3) Individual incremental losses are independent

4.2 Extensions of the Model

Some of the ways that we can expand on the simple model are given below. These go beyond the conjugate form and therefore require moving to simulation models. The simple conjugate form may still be a component or special case of these advances.

4.2.1 Parametric versus Nonparametric Models

The use of the Dirichlet or Generalized Dirichlet distribution allows for a pattern with a parameter for each development period. This creates a very flexible shape but requires estimation of many parameters. An alternative is the use of a parametric "growth curve" such as described in Zhang, et al (2012).

A parametric curve creates a much smoother development pattern, which is more constrained because of the fewer parameters. The Dirichlet is sometimes called a nonparametric model because it can follow the data more closely; however, "nonparametric" is a bit of a misnomer because it does not mean "no parameters" but rather potentially "many parameters."

The use of a parametric growth curve can be incorporated in a Bayesian framework, with the prior distribution being on the parameters. This does not fit as neatly into our conjugate form, but can be handled in simulation-based MCMC models.

4.2.2 Including Exposures or Other External Data

The models above assume that the actuary is selecting a loss development pattern from a development triangle, and that the basic assumptions of the chain ladder method apply. For example, that the same pattern is applicable for all accident years.

The Bayesian framework allows us to move beyond this limited data and include other information. We could bring in data such as exposure units (e.g., onlevel premium) or expected loss ratios. This additional information may also have prior distributions reflecting the relative uncertainty in the data. Robbin (1986) and Mildenhall (2006) show that as the relative uncertainties change the results move between familiar methods such as Cape Cod and Bornhuetter-Ferguson.

In addition to exposure or premium information, the model can expand to modify the assumption that all accident years share the same expected development pattern. Meyers (2015) introduces a "changing settlement rate" (CSR) model that includes an interaction term to adjust each accident year.

4.2.3 Calculation of Predictive Distribution

This paper has been focused on getting an estimate of expected ultimate loss that incorporates prior knowledge, and we have not directly discussed the variability around that estimate. However, because all of the analysis presented in this paper has been based on explicit distribution forms, all of the building blocks are in place to calculate ranges around estimated ultimate losses.

The evaluation of variance depends directly upon the scale parameter ϕ, which has been assumed to be fixed and known – in fact supplied by the analyst. For computing ranges we would more generally want this parameter to be considered a random variable with its own prior distribution. The variance should also be considered uncertain in order to evaluate the full uncertainty in the final estimate of ultimate loss.

4. CONCLUSIONS

This paper has presented an introduction to Bayesian loss development and gives an implementation that can be used immediately by any actuary. The use of the Generalized Dirichlet allows simpler computation than presented in earlier papers and allows for calculations that are as direct as the calculation of age-to-age factors. The use of a conjugate form allows an interpretation of prior knowledge in the form of "fictitious" prior loss development. The conjugate form can also be expanded with discrete mixtures to allow greater flexibility in specifying prior knowledge.

It is hoped that this paper will allow more actuaries to experiment with the Bayesian framework and then be comfortable to move to ever more realistic modeling work.

Acknowledgment

The author acknowledges the helpful feedback and discussions with Tho Ngo, Diana Rangelova, Ulrich Riegel, Allen Tan, and Gary Venter. Any errors in the paper remain the fault of the author.

Supplementary Material

The examples given in this paper are easily implements in spreadsheet format. The author can make the examples available upon request.

5. REFERENCES

[1] Bolstad, William M., "Introduction to Bayesian Statistics," second edition, Wiley, 2007.

[2] Connor, Robert J., and James E. Mosimann, "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution," *Journal of the American Statistical Association* 64 (1969) 194-206.

[3] De Alba, Enrique, "Bayesian Estimation of Outstanding Claim Reserves," *North American Actuarial Journal*, 6:4, 1-20.

[4] Dong, Alice X.D. and Jennifer S.K. Chen, "Bayesian analysis of loss reserving using dynamic models with generalized beta distribution," *Insurance: Mathematics and Economics* 53 (2013), 355-365.

[5] England, P.D., R.J. Verrall and M.V. Wüthrich, "Bayesian Overdispersed Poisson Model and the Bornhuetter-Ferguson Method," *Annals of Actuarial Science* 6(2), (2012), 258-283.

[6] Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin, "Bayesian Data Analysis," third edition, CRC Press, 2013.

[7] Hesselager, Ole and Thomas Witting, "A Credibility Model with Random Fluctuations in Delay Probabilities for the Prediction of IBNR Claims," ASTIN Bulletin, Vol. 18, No. 1, 1988.

[8] Hickman, James C., "Credibility Theory: The Cornerstone of Actuarial Science," *North American Actuarial Journal* 1999, Vol. 3, Issue 2.

[9] Koop, Gary, Dale J. Pirier, and Justin L. Tobias, "Bayesian Econometric Methods," Cambridge University Press, 2007.

[10] Mack, Thomas, "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates," ASTIN Bulletin, 1993, Vol. 23, No. 2, 213-225.

[11] McCullagh, Peter and J.A. Nelder, "Generalized Linear Models" (second edition), Chapman & Hall/CRC Press, 1989.

[12] Meyers, Glenn G., "Stochastic Loss Reserving Using Bayesian MCMC Models," *CAS Monograph Series*, Number 1, 2015.

[13] Mildenhall Stephen J., "A Multivariate Bayesian Claim Count Development Model with Closed Form Posterior and Predictive Distributions," *CAS Forum* Winter 2006, 451-493.

[14] Ng, Kai Wang, Guo-Liang Tian and Man-Lai Tang, "Dirichlet and Related Distributions: Theory, Methods and Application," Wiley, 2011.

[15] Parodi, Pietro and Stephane Bonche, "Uncertainty-Based Credibility and its Applications," *Variance*, 2010 Vol. 04, No. 1, 18-29.

[16] Renshaw, Arthur E., and Richard Verrall, "A Stochastic Model Underlying the Chain-Ladder Technique," *British Actuarial Journal* 4, 1998, 903-923.

[17] Robbin, Ira, "A Bayesian Credibility Formula for IBNR Counts," *PCAS* 1986, 129-164.

[18] Schmid, Frank, "A Total Credibility Approach to Pool Reserving," *Casualty Actuarial Society E-Forum*, Summer 2012.

[19] Shi, Peng and Brian M. Hartman, "Credibility in Loss Reserving," *Casualty Actuarial Society E-Forum*, Summer 2014-Volume 2.

[20] Verrall, R.J., "Bayes and Empirical Bayes Estimation for the Chain Ladder Model," ASTIN Bulletin, 1990, Vol. 20, No. 2, 217-243.

[21] Wong, Tzu-Tsung, "Generalized Dirichlet Distribution in Bayesian Analysis," Applied Mathematics and Computation 97 (1998), 165-181.

[22] Wüthrich, Mario V., "Using a Bayesian Approach for Claims Reserving," *Variance*, 2007 Vol. 01, No. 2, 292-301.

[23] Zhang, Yanwei, Vanja Dukic, James Guszcza, "A Bayesian Non-Linear Model for Forecasting Insurance Loss Payments," *Journal of the Royal Statistical Society* 2012, Vol. 175, Issue 2, 637-656.

Abbreviations and Mathematical Notation

ATA	Age-to-Age factor , or "link ratio"
LDF	Cumulative Loss Development Factor

p	A variable representing a portion between 0 and 1. It is a parameter (number of successes) in the Binomial distribution or the random variable itself for Beta distribution. In the univariate distributions (Binomial, Beta) it is written without a subscript; in the multivariate cases (Multinomial, Dirichlet) it is written with a subscript.
ϕ	Scale Parameter, or variance-to-mean ratio of aggregate loss
α, β	Shape parameters of Gamma, Beta and Generalized Dirichlet distributions
$C_{t,d}$	Cumulative losses for accident year t as of development age d

Biography of the Author

David R. Clark is a senior actuary with Munich Reinsurance, working in the Actuarial Research and Modeling area. He received the 2015 Ronald Bornhuetter Prize with co-author Diana Rangelova for the Non-Technical Call Paper "Accident Year / Development Year Interactions."