PREDICTING IBNYR EVENTS AND DELAYS

I. Continuous Time

WILLIAM S. JEWELL

Engineering Systems Research Center
University of California, Berkely

ABSTRACT

An IBNYR event is one that occurs randomly during some fixed exposure interval and incurs a random delay before it is reported. Both the rate at which such events occur and the parameters of the delay distribution are unknown random quantities. Given the number of events that have been reported during some observation interval, plus various secondary data on the dates of the events, the problem is to estimate the true values of the unknown parameters and to predict the number of events that are still unreported. A full-distributional Bayesian model is used, and it is shown that the amount of secondary data is critical. A recursive procedure calculates the predictive density; however, an explicit formula for the predictive mode can be obtained. The main computational work is the evaluation of an integral involving the prior density of the delay parameters, but this can be simplified in the exponential case using Gammoid approximations.

KEYWORDS

Observations delays; Inccurred But Not Reported (IBNR) models; Bayesian estimation and prediction.

1. INTRODUCTION

An IBNYR (Incurred But Not Yet Reported) claim in insurance is an event whose occurrence in some fixed *exposure interval* is not known until some later date because of random reporting delays. These delays may be administrative in nature, or may be due to the type of the covered contingency, as in the case of occupational illness. With these claims whose existence is not yet known are usually grouped IBNFR (Incurred But Not Fully Reported) claims, whose existence is known but whose cost development is incomplete, as in long-term illnesses or rehabilitation following accidents. Together these claims make up the *IBNR portfolio* for a given exposure year. The correct prediction of the total number of such claims and their ultimate total cost are of critical importance to insurance companies in the continuing process of setting up and modifying their "loss reserves" for each of their policy coverage exposure years. Improper estimation leads to fluctuations in financial results, missed opportunities for loss control, increased regulatory scrutiny, and other problems; thus, there are many pressures for making correct IBNR forecasts and updating them as new information becomes available.

This paper formulates a basic, continuous-time Bayesian model for predicting the *total number* of IBNYR claims arising in a given exposure interval, when only an incomplete number of such claims have been reported by some point in time. In addition to being uncertain about the rate at which events occur, we suppose that the parameters in the distribution that governs the random reporting delays are also uncertain, a priori. For a claim that actually has "surfaced", we permit various cases of *additional information* about occurrence and reporting dates that might be available. We shall see that the problem of predicting the number of as-yet-unreported events cannot be easily separated from the problem of estimating the unknown delay parameter(s). Similar problems arise in other fields, such as survey sampling by mail, and estimating undetected bugs in computer software (JEWELL [1985a] [1985b]).

The IBNR problem has been studied extensively in the actuarial literature, primarily with models where the "developed costs" are reported periodically after the exposure year is over. (STRAUB [1972], KRAMREITER and STRAUB [1973], BÜHLMANN, SCHNIEPER, and STRAUB [1980]. Many other references and a convenient summary through 1980 may be found in VAN EEGHEN [1981]). IBNYR claims are often called "pure IBNR"; other names for IBNFR are: IBN-Enough-R and Reported-But-Not-Settled. The simultaneous availability of several exposure years' data (over varying development intervals) leads to the infamous "IBNR triangle" of data, from which the total ultimate, costs of all exposure years to be forecast simultaneously. BÜHLMANN, SCHNIEPER & STRAUB [1980] first emphasized the additional predictive power available in reporting both quantized counts and costs for the various development years, as have HACHEMEISTER [1980] and NORBERG [1986] in his recent comprehensive model. KAMINSKY [1987] focuses exclusively on count prediction problems and KARLSSON [1974] [1976] considers the growth in mean counts and costs with a known continuous reporting delay process. A recent paper by HESSELAGER & WITTING [1988] introduces unknown quantized delay parameters into a credibility prediction. With these exceptions, one could characterize the field as one in which the solutions are more notable for their ingenuity than for the light they shed on the underlying processes.

We believe the inherent difficulty of estimating even just counts and delays simultaneously has been underrated in these "all-in-one", cost-oriented, discrete-time models. Therefore, in this paper, chosen to examine in great detail only the single exposure-year, continuous-time prediction of unreported events. Later papers will explore the additional complexities introduced by quantized time, multiple data-sources, and simultaneous prediction. Currently, the development of a good model for cost evolution over continuous time appears to require a long-term research effort, one that we believe will use the basic understanding of the event generation and reporting processes developed here, but will require much additional empirical effort to develop an understanding of cost-generating mechanisms and their evolution over time.

For reasons that will become apparent, we believe that the *point estimators* developed in previous papers, either by classical MLE methods or by credibility approximations, can only reveal part of the difficulty in IBNYR estimation. This is why we have adopted an exact, full-distributional Bayesian approach, at least until various approximations become computationally necessary. Admittedly, this approach leaves us open to the criticism that our answers depend upon our prior and model distributional assumptions; as we have remarked before (JEWELL [1980]), this is not a conceptual stumbling block in the actuarial field, as data and experience from related problems often support such assumptions. Anyone who wishes to modify these assumptions can easily implement the necessary changes, thus separating modelling complexity and computational difficulties, which are always open to compromise and tradeoff. Finally, we believe that setting up real IBNR reserves must also include reserves for risk, for which we need a prediction of the spread of final results, not just a point estimator.

We freely admit that this "building-block" model is still far from reality. Additional comments on this point may be found in section 13.

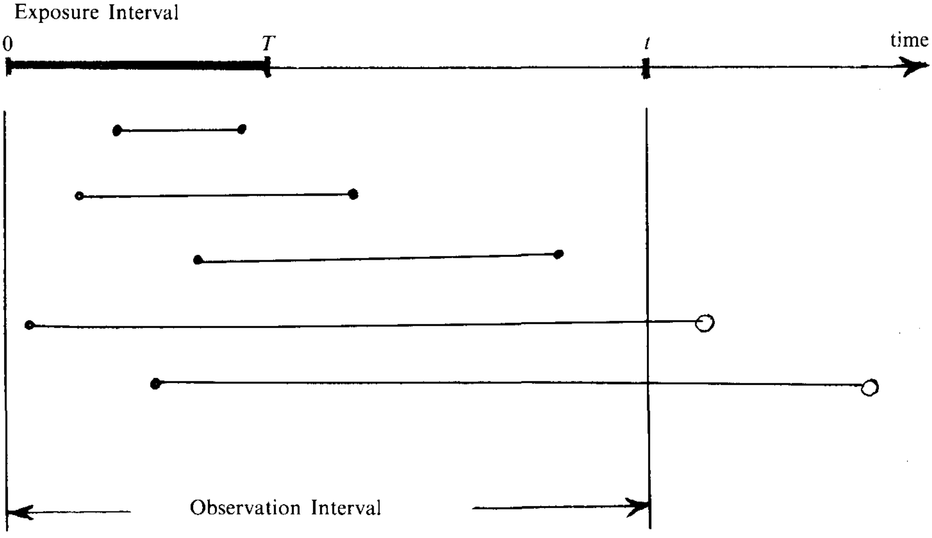
2. THE MODEL

Our basic assumption is that the events of interest are generated by a homogeneous Poisson process with *rate parameter* λ (events/year) over some fixed interval $(0, T]$ (the *exposure interval*). Thus, there are an unknown number, $\tilde{n} = \tilde{n}(T)$, of events at unknown *occurence epochs (accident dates)* $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$, given $\tilde{n} = n$. It follows that $\tilde{n}$ has a Poisson distribution with parameter λT, and that these epochs (with arbitrary numbering) are, a priori, mutually independent rvs, uniformly distributed over $(0, T]$.

Each event j is assumed to have associated with it a positive random *waiting time* (*reporting delay*), $\tilde{w}_j > 0$, such that its *observation epoch* (*reporting date*) is $\tilde{y}_j = \tilde{x}_j + \tilde{w}_j$ $(j = 1, 2, \ldots, n)$. We assume that the $(\tilde{w}_j)$ are iid rvs, with common density $f(w \mid \theta)$ and cdf $F(w \mid \theta)$, where θ is one or more unknown *delay parameter(s)*; both f and F are zero for $w \leq 0$.

Our Bayesian assumption is that $\tilde{\lambda}$ and $\tilde{\theta}$ are random quantities that have a known prior joint density, $p(\lambda, \theta)$. In fact, in this paper we shall assume they are *a priori independent,* with individual prior densities, $p(\lambda)$ and $p(\theta)$, respectively, which we assume can be identified from previous studies of claim frequency and reporting delays. (See discussion in section 13). We *learn* about these parameters through an *experiment* that observes all *reported events* in some *observation interval* $(0, t]$, where $t > 0$ is also continuous. As shown in figure 1 (with $t > t$), this willl lead to an observed number of *reported events,* say $r(t)$, consisting of these events $j = 1, 2, \ldots, n$ for which $\tilde{y}_j \leq t$; the remaining *unreported events,* $\tilde{u}(t) = \tilde{n}(T) - r(t)$ in number, will be those for which $\tilde{y}_j > t$. (Where there is no confusion and t is fixed, we shall write simply $\tilde{u} = \tilde{n} - r$). Section 3 considers various possibilities for reporting *secondary data,* D_j, that might be associated with each observed event j.

FIGURE 1. IBNYR process with $n = 5$, $r = 3$, and $t > T$.



Given the above assumptions, the observed *total data*, $D = \{D_1, D_2, \ldots, D_r\}$, and the prior densities $p(\lambda)$ and $p(\theta)$, the *parameter estimation* problem is to determine the posterior density $p(\lambda, \theta \mid D)$, and the *event prediction* problem is to determine the predictive density $p(u \mid D)$, and hence the distribution for $\tilde{n} = r + \tilde{u}$.

3. OCCURENCE, REPORTING, AND DELAY

Let us examine in more detail the relationship between occurrence and reporting dates, the delays, and the exposure and observation intervals. It can be seen that, given θ, every epoch r.v. pair $(\tilde{x}_j, \tilde{y}_j)$ is statistically independent of every other such pair, with common joint density:

$$p(x, y \mid \theta) = \frac{1}{T} f(y - x \mid \theta), \qquad (0 < x \leq T)\,(x < y < \infty) \tag{3.1}$$

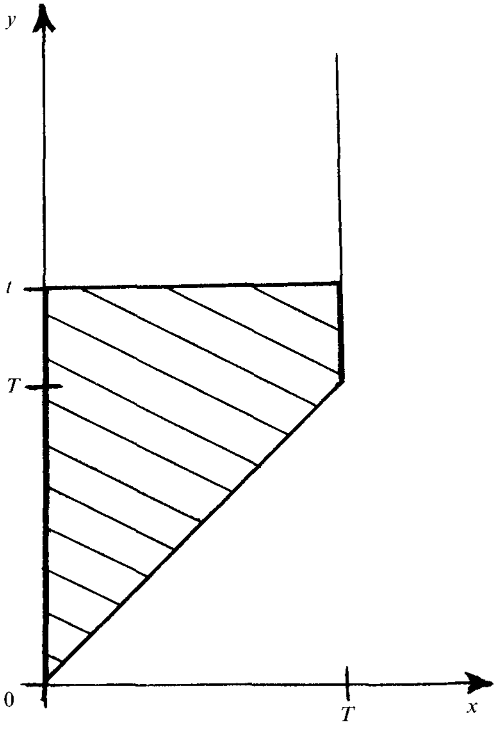
zero otherwise, as shown by the semi-infinite wedge-shaped region in figure 2. Let R_j be the random outcome that event j is reported by time t, i.e., that $(\tilde{x}_j, \tilde{y}_j)$ is a pair for which $\tilde{x}_j < \tilde{y}_j \leq t$. Then, the mixed density $p(x, y, R_j \mid \theta)$ would be (3.1) limited to the cross-hatched area in figure 2. The marginal densities of reported epochs depend upon whether $t < T$ or $t > T$, viz:

$$(3.2)\quad p(y, R_j \mid \theta) = \begin{cases} \dfrac{1}{T} F(y \mid \theta) & (0 < t \le T) \\ \dfrac{1}{T} [F(y \mid \theta) - F(y - T \mid \theta] & (t \ge T) \end{cases}, \quad (0 < y \le t)$$

and

$$(3.3)\qquad p(x, R_j \mid \theta) = \frac{1}{T} F(t - x \mid \theta). \qquad (0 < x \le \min(t, T)).$$

FIGURE 2. Regions of definition of $p(x, y \mid \theta)$ and $p(x, y, R_j \mid \theta)$.



Overall, the probability that a pair $(\tilde{x}_j, \tilde{y}_j)$ will be reported, without regard to the actual dates, is just the probability of the shaded area in figure 2:

$$(3.4)\qquad p(D_j \mid \theta) = \frac{1}{T} \int_{(t-T)^+}^{t} F(w \mid \theta)\, dw,$$

where $u^+ = \max(0, u)$.

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

Now consider again the experiment illustrated in figure 1. When an event j is reported, it is, of course, included in the count $\tilde{r}(t) = r$. There are four possibilities for observing *secondary date*, D_j, about this event, creating individual *secondary data likelihoods*, $p(D_j \mid \theta)$:

Type I Data. Observe Both Occurrence and Reporting Dates (x_j, y_j)

$$(3.1') \qquad p(D_j \mid \theta) = \frac{1}{T} f(y_j - x_j \mid \theta) = \frac{1}{T} f(w_j \mid \theta)$$

(i.e., observing (x_j, y_j) is equivalent to observing only w_j);

Type II Data. Observe Only Reporting Date (y_j)

$$(3.2') \qquad p(D_j \mid \theta) = \frac{1}{T} [F(y_j \mid \theta) - F((y_j - T)^+ \mid \theta)];$$

Type III Data. Observe Only Occurrence Date (x_j)

$$(3.3') \qquad p(D_j \mid \theta) = \frac{1}{T} F(t - x_j \mid \theta);$$

Type IV Data. Observe Event Reported But No Dates

$$(3.4') \qquad p(D_j \mid \theta) = \frac{1}{T} \int_{(t-T)^+}^{t} F(w \mid \theta)\, dw = \Pi(t \mid \theta), \text{ say}.$$

It seem intuitive that decreasing information about θ is provided as we go from Type I to Type IV data; our numerical examples will show that there are strong differences. In practive, of course, there could be a mixture of different types of data from different events. Remember also that t is considered fixed, so that knowing $r = r(t)$ means knowing *one number*; if we know in fact *the curve* $r(s)$ $(0 < s \le t)$, that is tantamount to having Type II data for all events. Finally, note that information of the type "an event has occurred but we have not received the paperwork" would have a likelihood $1 - \Pi(t \mid \theta)$, but be included in the count r!

4. THE DATA LIKELIHOOD

Assume temporarily that $t \ge T$, and suppose that $\tilde{n}(t) = n$. Then the conditional likelihood for the total data D will be:

$$(4.1) \quad p(D \mid \lambda, \theta, n) = \frac{n!}{1!\,1! \ldots 1!\,(n-r)!} \left[\prod_{j=1}^{r} p(D_j \mid \theta) \right] [1 - \Pi(t \mid \theta)]^{n-r}.$$

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

(If any of the D_j were from Type IV, then the multinomial coefficients $(1!\,1!\ldots 1!)$ would be modified here and in (4.2); however, only the ratio $(n!)/(n-r)!$ is of importance in the sequel).

Now, given λ, the distribution of $\tilde{n}(T)$ is Poisson (λT), and forming the product to give $p(D, n \mid \lambda, \theta)$ results in a fortuitous cancellation of $n!$, leaving only terms in $n-r = u$. Marginalizing over all values of $u \geq 0$, we obtain the final *data likelihood*:

$$p(D \mid \lambda, \theta) = \left[\prod_{j=1}^{r} p(D_j \mid \theta)\right] \frac{(\lambda T)^r}{r!} e^{-\lambda T \Pi(t \mid \theta)}. \qquad (t \geq T) \tag{4.2}$$

If $t < T$ the above argument is till correct with regard to $r(t)$ and the (D_j) (there wil be less data with smaller t, on average), but now n represents only the events from $(0, t]$, which have Poisson parameter (λt). Repeating the above analysis, we find that T in (4.2) is simply replaced everywhere by t when $t < T$. For convenience in the sequel we define:

$$\tau = \min(t, T), \tag{4.3}$$

and note that, if we replace T by τ everywhere in (4.2), it will then be correct for any observation interval.

5. MAXIMUM LIKELIHOOD ESTIMATES

It is worthwhile to examine the maximum likelihood point estimators for $\tilde{\lambda}$, $\tilde{\theta}$, and $\tilde{n}$, so that they may be later compared with our Bayesian results.

Assume first that θ and hence the delay distribution are *perfectly known*. From (4.2), we obtain the MLE for $\tilde{\lambda}$:

$$\hat{\lambda} = \frac{r(t)}{\tau \Pi(t \mid \theta)}, \tag{5.1}$$

so that a point estimate for $\tilde{n}(T)$ would be:

$$\hat{n}(T) = \left(\frac{T}{\tau}\right) \frac{r(t)}{\Pi(t \mid \theta)}. \tag{5.2}$$

If $t \geq T$ so that $\tau = T$, (5.2) says simply that a point estimate of the number of events inflates the observed counts by the known factor $\Pi(t \mid \theta)$; if $t < T$, then one must additionally inflate by T/t to take care of the smaller observation interval. Clearly, such estimates will be unrealiable when t is small because of these inflation factors; on the other hand, the estimate will be good when t is large primarily because nearly all events will be reported!

Conversely, suppose that λ is known exactly, but that we wish to estimate a scalar parameter θ in the delay distribution. Let $\mathscr{L}_j(\theta \mid D_j) = \ln p(D_j \mid \theta)$ be the appropriate log-likelihood of secondary data for each reported event. From (4.2), the necessary condition for the MLE of $\tilde{\lambda}$ is:

$$(5.3)\qquad \sum_{j=1}^{r(t)} \frac{\partial \mathscr{L}_j(\theta \mid D_j)}{\partial \theta} = (\lambda\tau)\frac{\partial \Pi(t \mid \theta)}{\partial \theta} \quad (\text{at } \theta = \hat{\theta}).$$

The actual solution depends in a complicated way upon the form of the delay distribution and the different possibilities for secondary data. If no dates are given with each reporting (Type IV), we find the trivial estimate:

$$(5.4)\qquad \Pi(t \mid \hat{\theta}) = \frac{r(t)}{\lambda\tau}.$$

Other secondary data will generally provide a more interesting estimate; for example, for Type I, $\dfrac{\partial \mathscr{L}_j(\theta)}{\partial \theta}$ becomes $\partial \ln f(w_j \mid \theta)/\partial \theta$, thus introducing the samples (w_j).

When both $\tilde{\lambda}$ and $\tilde{\theta}$ are assumed to be unknown parameters, both (5.1) and (5.3) are necessary conditions to determine the joint MLE $(\hat{\lambda}, \hat{\theta})$, i.e., we require the simultaneous solution of:

$$(5.5)\qquad \hat{\lambda}\tau = \frac{r(t)}{\Pi(t \mid \hat{\theta})}; \quad \frac{1}{r(t)} \sum \frac{\partial \mathscr{L}_j(\theta \mid D_j)}{\partial \theta}\bigg|_{\theta=\hat{\theta}} = \frac{\partial \ln \Pi(t \mid \theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}}.$$

Now, if we assume that no dates are reported, we find this second equation is redundant! In other words, with all Type IV data, $\hat{\lambda}$ and $\hat{\theta}$ cannot be determined separately, and there is no estimator $\hat{u}$! Other secondary data will give usable separable estimates, but these are dependable only for large r. For example, with Type I data, if $\bar{w} = (\Sigma w_j / r)$ is sufficient for θ, one can show that the RHS of the second equation in (5.5) is negible when r is very large, and one obtains the usual full-sample MLE from $\Pi f(w_j \mid \theta)$, even though not all events have been observed. We have also tried using the "maximum likelihood predictor" of KAMINSKY [1987] without success.

In short, the MLE approach is not very useful for out model when the observation interval is short, when only a few events have been recorded, or when no dates have been observed.

6. BAYESIAN FORMULATION

In a Bayesian formulation, we must specify our prior information about $\tilde{\lambda}$ and $\tilde{\theta}$, here assumed to be independent, *a priori*. One can, of course, use numerical methods with any empirical priors, but we shall assume analytical priors in attempt to show the general behaviour of our model under reasonable assumptions. A Gamma (a, b) density [1] for $\tilde{\lambda}$ is a convenient model for unimodal infor-

[1] $\tilde{x}$ is Gamma (a, b) means $p(x \mid a, b) = \dfrac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}$ $(x \geq 0)$.

mation, and, in view of the form of (4.2), would be a natural conjugate prior with θ fixed. One can select a and b, for example, from the first two moments, $\mathscr{E}\{\tilde{\lambda}\} = a/b$ and $\mathscr{V}\{\tilde{\lambda}\} = a/b^2$.

Now, given λ, $\tilde{n}(T)$ is Poisson (λT) and independent of $\tilde{\theta}$, so that, prior to the experiment, our opinion is that the number of events generated has a Pascal $(a, T)/(b+T))$ density [2]. In other words, prior to data, the predictive moments are:

$$(6.1) \qquad \mathscr{E}\{\tilde{n}(T)\} = \left(\frac{aT}{b}\right); \qquad \mathscr{V}\{\tilde{n}(T)\} = \left(\frac{aT}{b}\right)\left[1 + \left(\frac{aT}{b}\right)\frac{1}{a}\right].$$

If the prior mean count is held fixed, then a is a shape parameter that can be used to adjust the prior variance, which is naturally always larger than that of a Poisson distribution because of the uncertainty about $\tilde{\lambda}$.

The choice of $p(\theta)$ is more difficult, as θ may enter $f(w \mid \theta)$ and the $p(D_j \mid \theta)$ in a variety of different ways; in fact, θ may stand for a vector of delay parameters that must be estimated! For the moment, we will leave $f(w \mid \theta)$ and $p(\theta)$ arbitrary, and later specialize to particular forms to show typical results.

As the *posterior parameter density,* $p(\lambda, \theta \mid D)$, is not very revealing for any choice of priors, we pass to the central problem of concern, the *prediction* of the *unreported event count,* $\tilde{u}(t) = \tilde{n}(T) - r(t)$. Under the assumptions of our model, if the parameters are given, the reporting delays simply filter the original Poisson process with fixed probabilities; thus, $\tilde{u}(t)$ will also be Poisson with reduced parameters, using the usual decomposition independence arguments. If $t \geq T$, then the parameter will be $\lambda T[1 - \Pi(t \mid \theta)]$. On the other hand, if $t < T$, the unrecorded events in $(0, t]$ have the parameter $\lambda t[1 - \Pi(t \mid \theta)]$, to which must be added the unobservable events in $(t, T]$ with parameter $\lambda(T-t)$, giving a total Poisson parameter for all unreported events generated in $(0, T]$ of $\lambda[T - t\Pi(t \mid \theta)]$. Combining these two different forms for $p(u \mid \lambda, D)$ with appropriate versions of (4.2), we obtain:

$$(6.2) \qquad p(u \mid D) \propto h_\lambda(u \mid D)\, h_\theta(u \mid D),$$

with

$$(6.3) \qquad h_\lambda(u \mid D) = \frac{T^u}{u!} \int \lambda^{r+u} e^{-\lambda T} p(\lambda)\, d\lambda,$$

and

$$(6.4) \qquad h_\theta(u \mid D) = \int \left[\prod_{j=1}^{r} p(D_j \mid \theta)\right] \left[1 - \left(\frac{\tau}{T}\right) \Pi(t \mid \theta)\right]^u p(\theta)\, d\theta,$$

[2] $\tilde{x}$ is Pascal (a, π) means $p(x \mid a, \pi| = \dfrac{\Gamma(a+x)}{\Gamma(a)\,x!}\,(1-\pi)^a \pi^x \; (x = 0, 1, 2, \ldots)$.

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

where $\propto$, "proportional to", indicates that only terms that vary with u need be retained. Note, that there has been a fortuitous cancellation in the term $\exp(-\lambda\tau\Pi(t \mid \theta))$ from the likelihood, so that the predictive density can be represented as the product of two factors:

- one which depends upon $r = r(t)$ and the prior $p(\lambda)$;
- one which depends upon r, the secondary data types and the dates reported, and he prior $p(\theta)$.

This decomposition occurs in other models where one predicts unreported Poisson events (JEWELL [1985a] [1985b]).

With the choice of the Gamma (a, b) prior for $\tilde{\lambda}$, we obtain:

$$h_\lambda(u \mid D) = \frac{\Gamma(a+r+u)}{u!}\left(\frac{T}{b+T}\right)^u, \tag{6.5}$$

that is, of the form of a Pascal $(a+r, T/(b+T))$ distribution. Of course, there is further "shaping" of $p(u \mid D)$ to come from $h_\theta(u \mid D)$.

For later convenience, we note that, with (6.5), the predictive density can be written in recursive form:

$$\frac{p(u+1 \mid D)}{p(u \mid D)} = \left(\frac{a+r+u}{u+1}\right)\left(\frac{T}{b+T}\right)\left[\frac{h_\theta(u+1 \mid D)}{h_\theta(u \mid D)}\right]. \tag{6.6}$$

7. PREDICTION WITH KNOWN DELAY PARAMETERS

As preparation for more complicated cases, we first examine the prediction problem when θ is assumed to be known exactly. Only the term involving $\Pi(t \mid \theta)$ is then significant in $h_\theta(u \mid D)$, and we have:

$$p(u \mid D) = \frac{\Gamma(a+r+u)}{u!}\left[\frac{T-\tau\Pi(t \mid \theta)}{b+T}\right]^u, \tag{7.1}$$

which is a Pascal predictive density, with first two moments:

$$\mathscr{E}\{\tilde{u} \mid D\} = \frac{(a+r)T}{b}\left[\frac{1-(\tau/T)\,\Pi(t \mid \theta)}{1+(\tau/b)\,\Pi(t \mid \theta)}\right]; \tag{7.2}$$

$$\mathscr{V}\{\tilde{u} \mid D\} = \mathscr{E}\{\tilde{u} \mid D\}\left[\frac{b+T}{b+\tau\Pi(t \mid \theta)}\right]. \tag{7.3}$$

With no data ($t = 0$), the moments are identical with (6.1).

If the observation interval is small, (τ/T), $r(t)$, and $\Pi(t \mid \theta)$ will also be small, so that:

$$\mathscr{E}\{\tilde{u} \mid D\} \approx \frac{(a+r)T}{b}\left[1-t\left(\frac{b+T}{bT}\right)\Pi(t \mid \theta)\right], \qquad (t \to 0) \tag{7.4}$$

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

showing that out initial estimate of u is at first increased by the initial reports r, before being diminished by the second-order effects due to increasing t and $\Pi(t \mid \theta)$.

If the observation interval is large, $t > T$ and Π will be near unity, so that:

$$(7.5) \qquad \mathscr{E}\{\tilde{u} \mid D\} \approx \left[\frac{(a+r)\,T}{(b+T)}\right][1-\Pi(t \mid \theta)]\,. \qquad (t \to \infty)$$

The first term is T times the usual *credibility updating*:

$$(7.6) \qquad \mathscr{E}\{\tilde{\lambda} \mid r\} = (1-z)\,(a/b)+z\,(r/T); \qquad z = (T/b+T);$$

of a Poisson process parameter with a Gamma prior, given a number r of ordinary *undelayed* samples from $(0, T]$. This estimate is then diminished by the probability $1-\Pi$ of unreported events ouside $(0, t]$. The first term stabilizes towards the correct value of $\tilde{\lambda}T$ with increasing samples, but it is the second term that makes the predictive mean of $\tilde{\theta}$ decrease with increasing t. Note also that the second term in (7.3) approaches unity with increasing t, so that, in the limit, $\tilde{u}$ is asymptotically (small-mean) Poisson!

8. EXPONENTIAL DELAY LIKELIHOOD FACTORS

We now consider the additional variation due to uncertainty in the delay parameters(s), and the different "learning" effects that occur with various secondary data. For simplicity we use the over-familiar exponential density, $f(w \mid \theta) = \theta \exp(-\theta w)$. However, we expect the phenomena described below to be representative of results obtained with more general delay distributions; only the computational details will differ. A somewhat different approach for Type IV secondary data only is described in Appendix C.

From (3.1), the likelihood for a Type I datum, $D_j = \{x_j, y_j\}$, is:

$$(8.1) \qquad p(D_j \mid \theta) = L_j(\theta \mid D_j) = \frac{1}{T}\,\theta\, e^{-\theta w_j} \qquad (w_j = y_j - x_j)\,,$$

where the new notation $L_j(\theta \mid D_j)$ emphasizes that it is variation in θ that shapes $h_\theta(u \mid D)$ (so that, for example, the term T^{-1} here and below can be delected as uninformative). It can easily be seen that this likelihood is unimodal, with mode $\hat{\theta} = w_j^{-1}$. Data from r such delays would lead to a Gamma-shaped likelihood, peaked at $\hat{\theta} = (\Sigma w_j/r)^{-1}$, with very small "spread" if a is large. Thus, very large amounts of Type I data would force h_θ into a form giving the Pascal predictive density (7.1), with θ replaced by $\hat{\theta}$. In this sence, Type I data has a very strong effect on learning about $\tilde{\theta}$ and in reducing the predictive uncertainty of $\tilde{u}$.

For a Type II datum, $D_j = \{y_j\}$, and (3.2) has two cases:

$$L_j(\theta \mid D_j) = \begin{Bmatrix} \frac{1}{T}[1-e^{-\theta y_j}] & (y_j \leq T) \\ \frac{1}{T}[e^{-\theta(y_j-T)}-e^{-\theta y_j}] & (y_j \geq T) \end{Bmatrix}. \tag{8.2}$$

A small y_j gives a monotone likelihood, leading to a weak shift in θ towards higher values. However, a $y_j > T$ gives again a unimodal L_j with mode at $\hat{\theta}T = -\ln[1-(T/y_j)]$. If $y_j \gg T$, one can show that $\hat{\theta} \approx [y_j-(T/2)]^{-1}$, so the effect is similar to that of Type I data, using a guess of $x_j = (T/2)$; however, (assuming comparable $\hat{\theta}$) one can show that the peak of the likelihood is broader (less "information") for Type II data. Thus, for large amounts of Type II data, and many samples greater than T, the secondary data term in h_θ will also be tighty concentrated around the mode, but less so than if Type I information were available. On the other hand, if most or all of the reporting dates are less than T, then the likelihood will have a very broad peak or no peak at all. In this sense, then, Type II data is not as informative about $\tilde{\theta}$, and hence about $\tilde{u}$, as Type I data.

For a Type III datum, $D_j = \{x_j\}$, (3.3) gives:

$$L_j(\theta \mid D_j) = \frac{1}{T}[1-e^{-\theta(t-x_j)}]. \tag{8.3}$$

Note that this likelihood is monotone, and depends upon the length of the observation period. Because this datum is equivalent to $\{\tilde{w}_j \leq t-x_j\}$ it provides rather weak information about $\tilde{\theta}$, especially as t increases; with many such samples, we shall see that the main effect is to spread out the prior density.

Every Type IV event gives the same likelihood:

$$\Pi(t \mid \theta) = \left(\frac{\tau}{T}\right) - \frac{1}{\theta T}[e^{-\theta(t-T)}-e^{-\theta t}], \tag{8.4}$$

which is also monotome increasing in θ, approaching the asymptote (τ/T) more slowly than any exponential. With many samples, this likehood is very uniformative, and its main effect is to broaden the prior density.

9. COMPUTATIONAL STRATEGIES FOR DELAY INTEGRAL

We now consider various strategies for computing the delay integral (6.4), which, for simplicity, we rewrite as:

$$h_\theta(u \mid D) = \int L(\theta \mid D)\,[K(\theta)]^u\,p(\theta)\,d\theta, \tag{9.1}$$

assuming that the appropriate forms (7.1)-(7.4) are used to calculate $L(\theta \mid D) = \Pi L_j(\theta \mid D_j)$, and the *kernel* $K(\theta) = [1-(\tau/T)\,\Pi(t \mid \theta)]$. The first remark is that (9.1) is a rather easy numerical integration for arbitrary $p(\theta)$, even when there are several θ and many values of u are required. However, this does not give any analytic insight into the shaping of $p(u \mid D)$ from various data types.

Continuing our exponential delay example, we now assume that our prior on the unknown parameter $\tilde{\theta}$ is Gamma (c_0, d_0), that is, our prior opinion is that $\mathscr{E}\{\tilde{\theta}\} = (c_0/d_0)$ and $\mathscr{V}\{\tilde{\theta}\} = (\mathscr{E}\{\tilde{\theta}\})^2/c_0$, and that the density is unimodal, with the mode at $\theta_0 = (c_0-1)/d_0$. This is not only a reasonable prior for unimodal information, but is also conjugate to $L(\theta \mid \mathrm{D})$ for Type I observations.

Our strategy is then to approximate the first two factors in (9.1) by a *Gammoid function*:

$$g(\theta) = (\Delta\theta)^{\Gamma} e^{-\Delta\theta}, \tag{9.2}$$

in the region of the *current* mode of the integrand (which will initially be θ_0, but perhaps modified as we add terms from $L(\theta \mid D)$). This strategy will convert (9.1) into a Gamma integral with a convenient analytic dependence on u. The resulting shape will, of course, be a better approximation to $h_\theta(u \mid \mathscr{D})$, the more precise is our prior knowledge about $\tilde{\theta}$; however, the results are surprisingly good with $c_0 = 3$ or 4 and Type I or II data, for reasons that will become clearer as we proceed. Full details on the Gammoid method will appear in a forthcoming paper.

We now outline this method sequentially, proceeding as if all four data types are present, with the total r being broken down into r_1, r_2, r_3, and r_4 events. It turns out that the Gammoid coefficients Γ and Δ are exactly or approximately linear in r, so that we shall set $\Gamma = r\gamma$ and $\Delta = r\delta$ for each data type, and concentrate on the calculation of the *unit coefficients*, γ and δ. Only basic results are given below; additional formulae and computational details may be found in Appendices A and B.

9.1. Type I Secondary Data

Type I data is the easiest to deal with, as $L(\theta \mid D)$ from (8.2) is *exactly* Gamma. We recommend that the prior coefficients be updated as follows:

$$c_0 \leftarrow c_0 + r_1; \qquad d_0 \leftarrow d_0 + r_1 \bar{w}, \tag{9.3}$$

where $\bar{w}$ is the average of the (w_j) for all Type I data; the *current* mode is then redefined in terms of the new coefficients as $\theta_0 = (c_0-1)/d_0$. (If there is no other secondary data, continue with section 9.5).

9.2. Type II Secondary Data

Data of Type II must be subdivided into two groups: Type IIa consists of the r_{2a} events with $\{y_j \mid y_j \le T\}$, and Type IIb consists of the r_{2b} events with $(y_j \mid y_j > T\}$.

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

Considering the IIb data first, we can show that the likelihood (8.2) for this data is unimodal and well-fitted by a Gamma with unit parameters given by (A.4). To a first approximation, one can take:

$$\text{(9.4)} \qquad \gamma_{2b} \approx 1; \qquad \delta_{2b} \approx \left(\bar{y}^b - \frac{T}{2}\right) - \frac{T}{2}(\theta_0 T),$$

where $\bar{y}^b$ is the average of the r_{2b} Type IIb data values (y_j). Therefore, our strategy with this data is to once again update the coefficients:

$$\text{(9.5)} \qquad c_0 \leftarrow c_0 + r_{2b}\gamma_{2b}; \qquad d_0 \leftarrow d_0 + r_{2b}\delta_{2b},$$

using either the exact approximating coefficients, or (9.4). As before, the *current* mode, θ_0, should be redefined from these new coefficients. From this point on, the Gammoid approximation coefficients usually depend upon θ_0, in a weak way. Therefore, until section 9.6, we recommend keeping θ_0 fixed.

The likelihood factor for Type IIa data is monotone increasing, with no mode. However, we have found that a Gammoid approximation is still locally reasonable. To a good approximation:

$$\text{(9.6)} \qquad \gamma_{2a} \approx 1; \qquad \delta_{2a} \approx \frac{1}{2}\bar{y}^a - \frac{1}{6T} m_2 \theta_0,$$

where $\bar{y}^a$ and m_2 are the first and second moments of the r_{2a} data points (y_j), both small by definition of Type IIa.

9.3. Type III Secondary Data

Type III data is very uninformative, especially for large values of t, with a likelihood is similar to that of Type IIa data, but with all terms in y_j replaced by $t - x_j$. (9.6) still gives an initial approximation:

$$\text{(9.7)} \qquad \gamma_3 \approx 1; \qquad \delta_3 \approx \frac{1}{2}(t - \bar{x}) - \frac{1}{6T} m_2 \theta_0,$$

where m_2 is now the second moment of $(t - x_j)$. Both coefficients become smaller as t increases, reflecting the uninformative nature of the data $\{\tilde{x}_j = x_j \leq \tilde{y}_j \leq t\}$, and it is then necessary to use the exact formulae.

9.4. Type IV Secondary Data

With this minimal information $\{\tilde{y}_j \leq t\}$, (8.4) is monotone increasing, and depends only on r and t. To a rough approximation:

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

$$(9.8)\quad \gamma_4 \approx 1; \qquad \delta_4 \approx \begin{Bmatrix} \frac{1}{3}t & (0 \le t \le T) \\ \frac{1}{3}\left[\frac{t^3-(T-t)^3}{t^2-(T-t)^2}\right] & (t \ge T) \end{Bmatrix}.$$

We have found usually it is necessary to calculate the exact unit coefficients at the current mode. In fact, for large t, δ_4 can become negative, in which case we recommend setting $\delta_4 = 0$, and approximating locally by a polynomial.

9.5. The Kernel $K(\theta)$

As discussed in Appendix B, the kernel is monotone decreasing, much like a negative exponential. Therefore, a reasonable approximating procedure is to set $\gamma_K = 0$, and find δ_K from (B.9) at the current mode. For a quick approximation:

$$(9.9)\qquad \delta_K \approx \begin{Bmatrix} \frac{t^3}{2T^2} & (t \le T) \\ t - \frac{T}{2} & (t \ge T) \end{Bmatrix}.$$

δ_K gives the important dependence of h_θ upon u, since d_0 will be updated by $\Delta_K = \delta_K u$, and c_0 will not change with u.

9.6. Completing the Computations

With all of the above approximations completed, the final coefficients of the Gammoid form $\theta^{c-1} e^{-\theta d}$ representing all factors in (6.4) will be:

$$(9.10)\qquad \begin{aligned} c &= c_0 + \Gamma; \qquad d = d_0 + \Delta + \delta_K u; \\ \Gamma &= r_1 + r_{2a}\gamma_{2a} + r_{2b}\gamma_{2b} + r_3\gamma_3 + r_4\gamma_4; \\ \Delta &= r_1 \bar{w} + r_{2a}\delta_{2a} + r_{2b}\delta_{2b} + r_3\delta_3 + r_4\delta_4. \end{aligned}$$

If desired, one can now make a second pass through all of the approximating formulae using the "final" data-only mode, $\theta_0 = [(c-1)/(d_0+\Delta)]$, to see if there is a significant change in the unit coefficients, and hence in (9.10). In our limited experience, the coefficients will be little modified if the mode of the prior density or of the Type I or Type IIb data likelihood is reasonably concentrated; in other cases, several iterations may be required. The integral of (9.1) is now $\Gamma(c)/d^c$, but only d is informative for u, so we may just as well take:

$$(9.11)\qquad h_\theta(u \mid D) = d^{-c} = (d_0 + \Delta + \delta_K u)^{-(c_0+\Gamma)}.$$

For a quick approximation, one could use the initial terms of all the approximations to compute Δ, δ_k, and r. If all the data is of Type IV and $t > T$, a different approach to calculating h_θ is possible using a Beta prior; details are in Appendix C.

10. CALCULATING THE PREDICTIVE DISTRIBUTION

From (6.6) and (9.11), we obtain finally a recursive relationship for the predictive density:

$$(10.1) \qquad \frac{p(u+1 \mid D)}{p(u \mid D)} = \left(\frac{a+r+u}{u+1}\right)\left(\frac{T}{b+T}\right)\left[\frac{d_0+\Delta+\delta_K u}{d_0+\Delta+\delta_K+\delta_K u}\right]^{c_0+\Gamma},$$

whose values are calculated by setting $p(0 \mid D) = 1$, "bootstrapping" up through "sufficient" values of u, and then renormalizing. Moments and the tail distribution are then obtained numerically. As this recursive method is very efficient, it is easy to explore the full-distributional implications for different parameter and data values.

If one still insists on a *point* estimator for the number of unreported events, the *predictive mode* can be obtained analytically. Let u^* be the (usually non-integral) solution to:

$$(10.2) \qquad u^*+1 = \left(\frac{a+r+u^*}{b+T}\right) T \left[\frac{d_0+\Delta+\delta_K u^*}{d_0+\Delta+\delta_K+\delta_K u^*}\right]^{c_0+\Gamma};$$

this solution always exists, and can be obtained iteratively from (10.2), starting with an arbitrary guess on the RHS; convergence is rapid. The predictive mode, $\hat{u}(D)$, is then *the integer greater than or equal to* u^*.

This type of point estimation is related to an old and well-known formula in population biology, associated with LAPLACE, PETERSEN, and others (JEWELL [1985a]). In section 12, we shall see that (10.2) also has an intersting interpretation in terms of credibility predictors.

Of course, the great advantage of (10.1) is that it provides the *complete* predictive distribution for $\tilde{u}$. As we shall see in the following numerical example, the variance of this distribution remains quite substantial with even a large amount of data. This knowledge is crucial in making a proper risk assessment of IBNR reserves.

11. NUMERICAL EXAMPLE

To illustrate the above theory, we analyzed a numerical example which assumes that our prior knowledge is correct in the means, but is not especially precise. Based on these results, the reader can easily extrapolate to cases where initial knowledge is different from reality, or, conversely, is very accurate.

Specifically, we assumed that $\tilde{\lambda}$ has a Gamma (2, 0.02) prior density, which makes $\mathscr{E}\{\tilde{\lambda}\} = 100$, $\mathscr{V}\{\tilde{\lambda}\} = 5000$; for convenience, we take $T = 1$ year, which makes the mean total rate of IBNYR events 100 per year. This leads to a Pascal $(2, 1.02^{-1})$, density for $\tilde{n}$, with $\mathscr{E}\{\tilde{n}\} = 100$, $\mathscr{V}\{\tilde{n}\} = 5100$, and $\hat{n} = 49$; the 5%, 25%, 75%, and 95% fractiles are: $n_{.05} = 16.5$, $n_{.25} = 47.0$, $n_{.75} = 134.5$, and $n_{.95} = 238.1$, respectively, which is quite a broad range, *a priori*. We assume that $\tilde{\theta}$ has a Gamma (4,6) prior density, so the prior mean delay is $\mathscr{E}\{\tilde{\theta}^{-1}\} = 2.0$ years, and $\mathscr{V}\{\tilde{\theta}^{-1}\} = 2.0$ years^2.

For the purpose of simulation, we further "stacked the deck" by assuming that the true value of the delay parameter was $\theta = 0.5$ per year, and, whatever the true value of λ was, that exactly $n = 100$ IBNYR events were generated during the exposure year. Table 1 shows a few of the simulated values, arranged in order of increasing (y_j), and hence approximately increasing in (w_j). In the 100 samples, the mean delay is 2.35 years, with sample variance 5.35 years, so the coefficient of variation is about right, but the delays are a little long, on average. Figure 3 shows the curves of $\Pi(t \mid \theta)$ and $K(\theta) = [1-(\tau/T)\,\Pi(t \mid \theta)]$ versus t, for the true value $\theta = 0.5$. The ragged curve is the simulated count history for reported events, $r(t)$.

TABLE 1

EXTRACT OF 20 OF THE SIMULATED VALUES FOR NUMERICAL EXAMPLE ($\theta = 0.5$)

x	y	w
.043	.206	.163
.022	.234	.213
.095	.267	.172
.330	.527	.198
.112	.629	.517
....	....	....
....	.....	.....
.570	1.412	.841
.390	1.430	1.040
.600	1.483	.883
.902	1.493	.590
.118	1.558	1.440
....	....	....
....	.....	.....
.269	2.820	2.551
.728	2.823	2.095
.282	2.872	2.590
.055	2.985	2.929
.882	3.055	2.173
....	....	....
....	.....	.....
.036	9.128	9.092
.933	9.408	8.475
.311	9.616	9.305
.349	11.194	10.845
.563	12.967	12.403

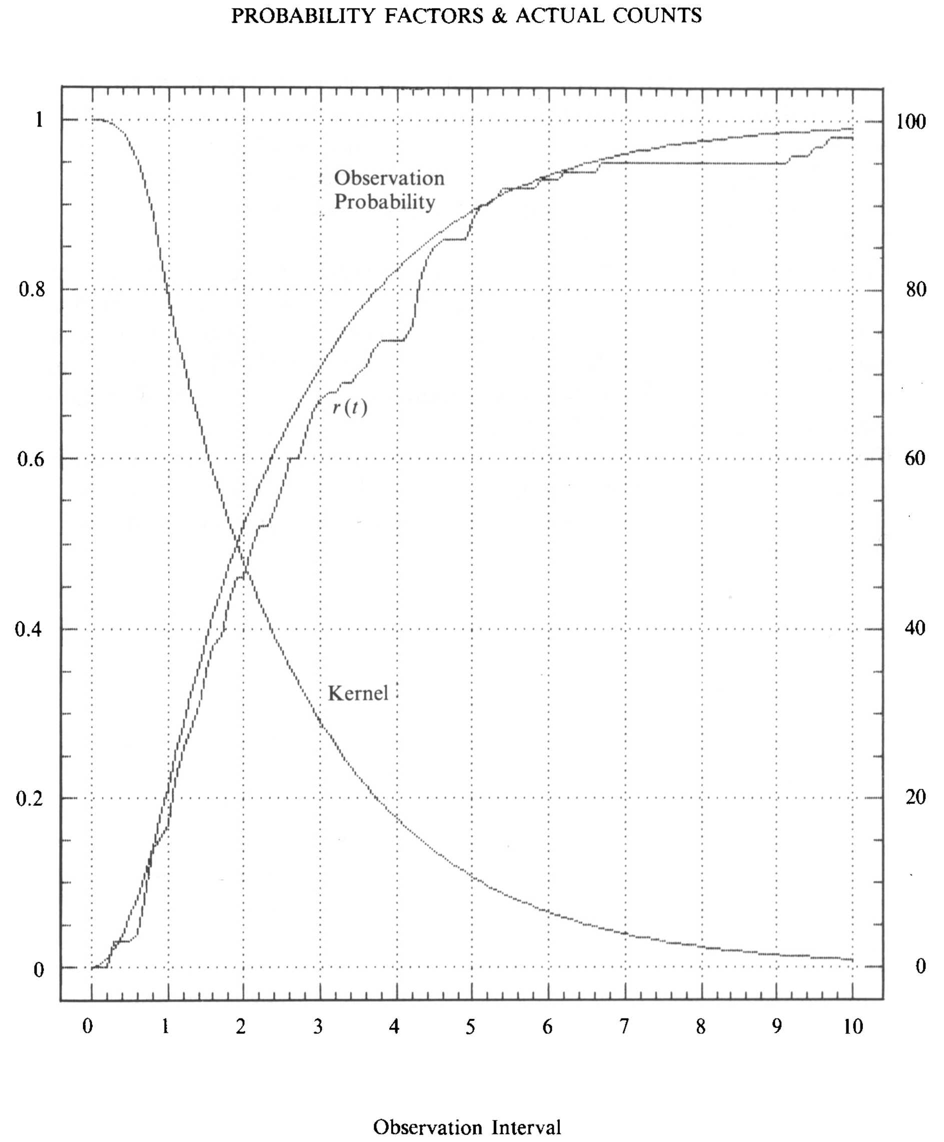


FIGURE 3. Observation Probability, Kernel and simulated count history versus $t(T = 1, \theta = 0.5)$.

11.1. Type I Data Analysis

In the first analysis, we assumed that all data was of Type I, and we examined observation intervals of $t = 0(0.5)10.0$ years (remember $T = 1$ year, and the mean delay is 2.0 years. The results are summarized in figure 4, which shows $r(t)$, $\mathscr{E}\{\tilde{n} \mid D\} = r(t) + \mathscr{E}\{\tilde{u}(t) \mid D\}$, $\hat{n}(D) = \hat{u}(D) + r(t)$, plus the four fractiles of $(\tilde{n} \mid D)$ mentioned previously, all versus t. (Continuous curves are shown for

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

convenience). Of course, these calculations were carried out by first finding the complete predictive densities, $p(u \mid D)$, (over the range [0,1000]) using the appropriate sifted data for the current value of t, and then finding the summary statistics; this took about 10 seconds on a PC-AT for each value of t! All results were translated from predicting $\tilde{u}$ to predicting $\tilde{n}$ for ease in making comparisons.

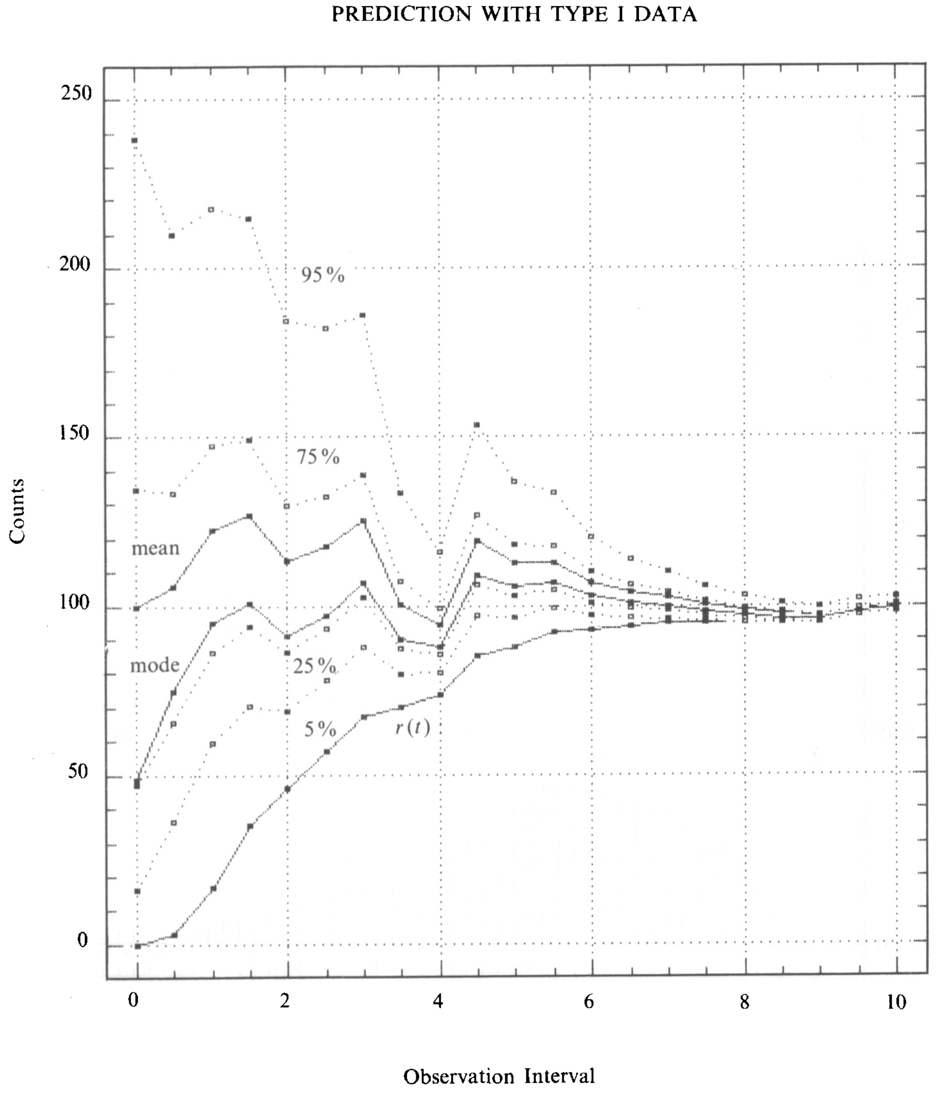


FIGURE 4. Predictive mean, mode, and fractiles versus t for Type I data ($T = 1$).

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

From the figure, it can be seen that the point estimators, the mean and the mode, both wander around the true (and ultimate) value of 100, although, for reasons we do not completely understand, the mode seems to be less "tricked" by intermediate fluctuations in $r(t)$, once the mode has risen from its initial low value of 49 until after, say, $t > T$. It is extremely satisfying to see how the "Bayesian confidence intervals" (predictive quantiles) converge with increasing t, although it must be remembered that much of this is due to the decrease in $1 - \Pi(t \mid \theta)$, and not just the learning due to D!

PREDICTIVE DENSITY $t = 4T$

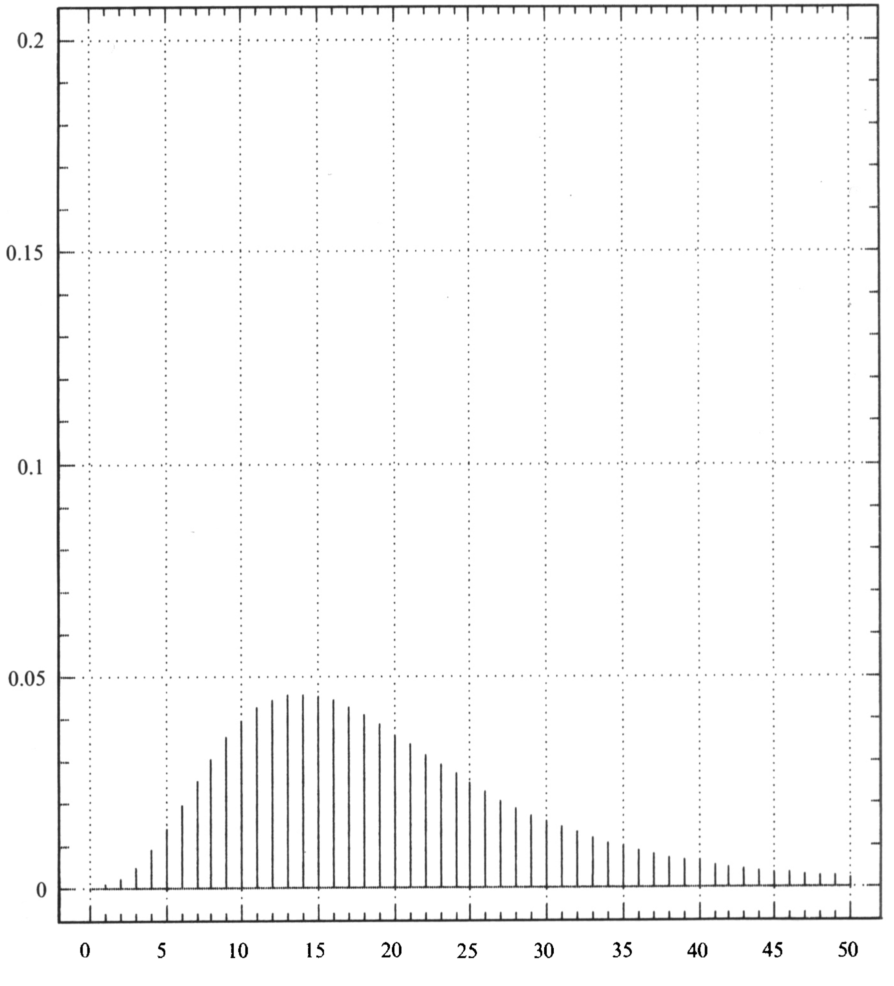


FIGURE 5. Predictive density for Type I data ($t = 4T$).

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

PREDICTIVE DENSITY $t = 8T$

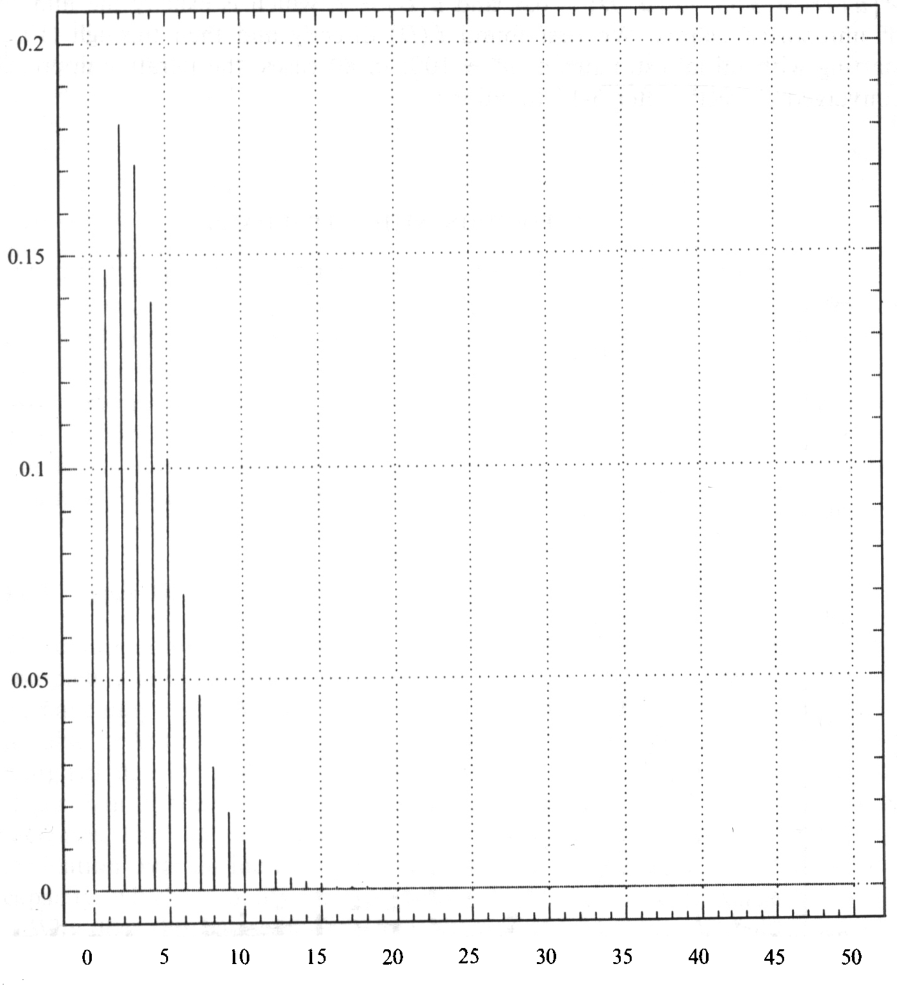


FIGURE 6. Predictive density for Type I data ($t = 8T$).

Specific details for $t = 4$ years ($\Pi = 0.824$) are as follows: the data gave $r = 74$ reported revents, plus secondary information leading to parameters $\Gamma = 74$, $\Delta = 94.509$, so that $c = 78$, $d = 100.509$, and the new (and final) mode was $\theta_0 = 0.7661$, from which $\delta_K = 3.4368$. The resulting $p(u \mid D)$ is shown in figure 5, with $\mathscr{E}\{\tilde{u} \mid D\} = 20.28$, $\mathscr{V}\{\tilde{u} \mid D\} = 143.6$, and $\hat{u}(D) = 14$ (there are, in fact, 22 events outstanding). If we increase the observation to four time constants at $t = 8$ years ($\Pi = 0.976$), there are now $r = 98$ reported events, the

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

parameters are $\Gamma = 95$, $\Delta = 184.436$, so $c = 97$, $d = 190.436$, and the new mode is $\theta_0 = 0.515$, from which $\delta_K = 7.4573$. $p(u \mid D)$ is shown in figure 6, and $\mathscr{E}\{\tilde{u} \mid D\} = 3.55$, $\mathscr{V}\{\tilde{u} \mid D\} = 6.6$, and $\hat{u}(D) = 2$, which is exactly the number of unreported claims. We first found $\hat{u}(D)$ directly and then through (10.2), starting with initial estimates of $u^* = 100$; in all cases, the iterative approach converged correctly after 5-15 iterations.

PREDICTION WITH TYPE II DATA

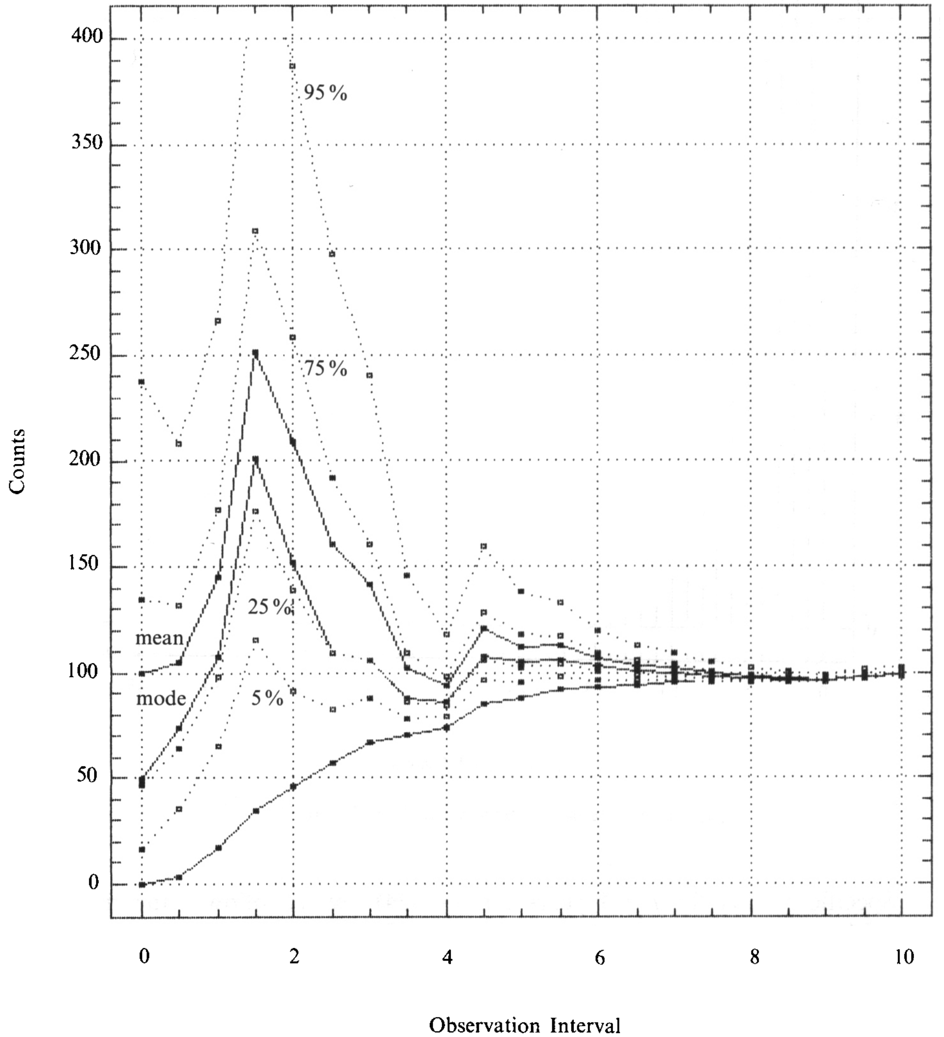


FIGURE 7. Predictive mean, mode, and fractiles versus t for Type II data ($T = 1$).

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

11.2. Type II Data Analysis

In the second analysis, we assumed that all data was of Type II, but otherwise used the same values as above. Figure 7 summarizes the results, which should be compared with figure 4 for Type I data. Roughly speaking, the results are similar for $t \leq T$ and $t > 4T$ (twice the mean delay), but are much more variable in the intermediate region; note particularly how small fluctuations near $t = 0.5$ and $t = 4.5$ "jolt" the predictors more in Type II than in Type I. This poorer behaviour is, of course, due to the missing (x_j) in Type II, which makes the estimation of $\tilde{\theta}$ quite unstable in this region. For $t < T$, there is little learning anyway, and for $t > 4T$, the approximation $w_j \approx y_j - 0.5T$ is good enough.

The main change in computation with Type II data is that it is desirable to iterate a few times to find the correct mode, θ_0. For $t = 4$ years, four iterations stabilized at $\theta_0 = 0.7998$, giving $c = 74.639$, $d = 92.054$, and $\delta_K = 3.4340$, from which $\mathscr{E}\{\tilde{u} \mid D\} = 19.69$, $\mathscr{V}\{\tilde{u} \mid D\} = 183.4$, and $\hat{u}(D) = 12$. For $t = 8$ years, two iterations are enough to give $\theta_0 = 0.5240$, $c = 97.055$, $d = 183.298$, and $\delta_K = 7.4565$, from which $\mathscr{E}\{\tilde{u} \mid D\} = 3.29$, $\mathscr{V}\{\tilde{u} \mid D\} = 6.06$, and $\hat{u}(D) = 2$. The forms of the predictive densities are similar to those shown for Type I data. Again, u^* always converged rapidly to the true answer.

11.3. Types III and IV Data Analysis

The computations with Type III and IV data are much more difficult, and give completely different behaviour than that described above. Considering first that we have only Type IV data (counts only), we obtain the summary results shown in table 2. At first, with t small, we get the modest improvements in the Pascal marginal density that were observed above. However, as soon as t becomes larger than T, there is a steady and dramatic increase in all the predictors as r increases, and our point estimators *grow without bound*! (In fact the need for evaluation over an increasingly wide range of u-values soon exceeds computer capacity, which accounts for the? beside the larger numbers in the table).

Why does this happen? As before, there is at first some instability in finding the current mode, which may require 5 or 10 "assisted" iterations. Then, begin-

TABLE 2

RESULTS FOR TYPE IV DATA VERSUS t ($T = 1$)

t/T	r(t)	θ_0	$\mathscr{E}\{\tilde{u} \mid D\}$	$\hat{u}(D)$	$\mathscr{V}\{\tilde{u} \mid D\}$
0	0	0.000	100.0	49	5100
0.5	3	0.938	102.2	72	2975
1.0	17	1.990	132.6	96	4144
1.1	22	2.178	259.0	209	12290 ?
1.2	26	2.246	450.5	400	23280 ?
1.5	35	2.177	1038	987	53760 ?
2.0	46	1.905	1766	1715	89490 ?
4.0	74	1.237	3370 ?	3320 ?	172300 ?
8.0	95	0.777	4561 ?	4511 ?	232900 ?

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

ning about $t = 2T$, δ_4 becomes negative, and we must change to γ-only modelling, as described in Appendix B. And, admittedly, the Gammoid approximatons for u small are also not as good as in previous cases. But these are second-order effects.

The real reason for the behaviour shown in table 2 is that *there is less and less information in Type IV sampling as t increases*! As r increases with t, the likelihood $[\Pi(t \mid \theta)]^r$ "destabilizes" the prior $p(\theta)$ by diffusing the mode in h_θ, while at the same time h_λ is increasing. This loss of information about $\tilde{\theta}$ and increase in the estimate of $\tilde{\lambda}$ can be seen most clearly in (10.2); there is no technical difficulty in converging to the correct mode, but it is clear from the magnitude of the parameters that the mode must move to larger and larger values as r increases. (But remember we are assuming only that r is known for each t; knowing the history of $r(t)$ would bring us back to Type II).

Besides the lack of information in the likelihood, the behaviour is greatly influenced by our prior certainty about the value of $\tilde{\theta}$. To see this, let us keep $t = 4T$ fixed, and increase both c_0 and d_0 so that the prior mode of $\tilde{\theta}$ (which is the prior mean of $\tilde{\theta}^{-1}$) is kept fixed at its true value of 0.5. As shown in table 3,

TABLE 3

RESULTS FOR TYPE IV DATA AND $t = 4T$, SHOWING EFFECT OF INCREASED PRECISION IN GAMMA PRIOR DENSITY

c_0	d_0	θ_0	$\mathscr{E}\{\tilde{u} \mid D\}$	$\hat{u}(D)$	$\mathscr{V}\{\tilde{u} \mid D\}$
4	6	1.237	3370 ?	3320 ?	172300 ?
8	14	1.036	3008 ?	2958 ?	154300 ?
16	30	0.875	2407 ?	2356 ?	124100 ?
32	62	0.747	1373	1323	73050 ?
40	78	0.714	890	839	49370 ?
50	98	0.683	270.0	212	19760 ?
64	126	0.653	32.9	25	248
128	258	0.588	18.3	17	38
Inf	Inf	0.500	16.0	15	19

as the prior precision increases, the mode of the integrand shrinks slowly towards $\theta = 0.5$ (of course!), and the various predictors are pulled in towards more reasonable numerical values. But notice also that values of say, $c_0 > 60$ are needed to make the values comparable to those obtained with Type I or II data; this is an extraordinary amount of precision, corresponding to a prior standard deviation for $\tilde{\theta}^{-1}$ of less than 0.25 years, when the mean is 2.0 years!

Finally, we can also see what is happening mathematically by examining the details involved in computing the ratio $h(u+1 \mid D)/h(u \mid D)$ in (6.6) (10.1). For $t = 4$ years and the original parameters, we find $\theta_0 = 1.237$ after ten iterations, giving values of $\Gamma = 4.4257$, $\Delta = 0$ (we use polynomial-only approximation), and $\delta_K = 3.3994$. If we compare these with values found previously, we see that it is much easier for the ratio to approach unity more quickly than before. In other words, because the Pascal $\pi = (T/(b+T)) = 1.02^{-1}$ is already very close to one, there is little chance to shape the density downward while it is growing due to increased r. So, while $p(u \mid D)$ is increasing, the ratio is quickly becoming unity, so that the tails of the predictive density must look much like a Pascal $(a+r(t), \pi)$ density. In fact, the means of that Pascal density are 3800 and 4850 for $t = 4T$ and $t = 8T$, respectively, which are comparable to those in table 2.

In contrast, from the analytic form of the shaping ratio, we see that if $(c_0 - 1)/d_0$ is fixed at θ_0, then, as the parameters increase to larger and larger values (with moderate values of u), the ratio approaches $\exp(-\theta_0 \delta_K)$, thus accounting for the convergence shown at the end of table 3. Convergence is also improved with strong prior knowledge about the parameter $\tilde{\theta}$, as this makes the first ratio, $\Pi = (T/(b+T))$ smaller for the same $\mathscr{E}\{\tilde{\theta}\}$.

Turning now to Type III data, we see that similar convergence problems will be encountered because of the shape of the likelihood. Results are analogous to those in table 2. Although the growth is postponed somewhat, the increase in r with t inevitably leads to large increases in the estimators, unless we have very strong prior assumptions.

In summary, we see that not having *at least the date of reporting* of the IBNYR events leads to Bayesian predictions that, while mathematically correct, are operationally useless. This is *not* a result of using Bayesian analysis, but due to a more fundamental problem, namely, that Type III and Type IV data are *uninformative* (some might say, anti-informative) when the priors on $\tilde{\lambda}$ and $\tilde{\theta}$ are not sufficiently precise. In a certain sense, this behaviour is the analogue of the non-existence of MLE's for Type IV data discussed in section 5.

12. INTERPRETATION OF THE PREDICTIVE MODE

There is an interesting interpretation of the predictive mode (10.2) in terms of posterior parameter means that holds even for arbitrary $p(\theta)$ and data types. First note that the predictive mean of u is:

$$(12.1) \qquad \mathscr{E}\{\tilde{u} \mid D\} = \mathscr{E}\left\{\tilde{\lambda} T \left[1 - \frac{\tau}{T} \Pi(t \mid \tilde{\theta})\right] \mid D\right\},$$

and that, because of the factorization (6.2), we might expect the dependence on $\tilde{\lambda}$ and θ to be somehow separable. Recall also that, with a Gamma (a, b) prior on $\tilde{\lambda}$, a measurement of r Poisson events in $(O, T]$ gave in (7.6) a posterior parameter mean, $\mathscr{E}\{\tilde{\lambda} \mid r\} = (a+r)/(b+T)$, in credibility form.

Now, rewrite (10.2) for general $p(\theta)$ as:

$$(12.2) \quad \hat{u}(D) \approx u^* + 1 = \left(\frac{a+r+u^*}{b+T}\right) T \times \left\{\frac{\int \left[1 - \frac{\tau}{T} \Pi(t \mid \theta)\right] [K(\theta)]^{u^*} L(\theta \mid D)\, p(\theta)\, d\theta}{\int [K(\theta)]^{u^*} L(\theta \mid D)\, p(\theta)\, d\theta}\right\}.$$

We see that the first term in brackets is, in fact, $\mathscr{E}\{\tilde{\lambda} \mid r+u^*\}$, the posterior mean for $\tilde{\lambda}$ *under the observation of* $r+u^*$ *samples*! Similarly, the measure $[K(\theta)]^{u^*} L(\theta \mid D)\, p(\theta)$ is essentially $p(\theta \mid D, u^*)$, the density of $\tilde{\theta}$ posterior to the usual data D *plus the "look-ahead" observation of* u^* *events after the observation interval is over*! Thus, the second term on the RHS may be thought of as $\mathscr{E}\{[1-(\tau/T)\,\Pi(t \mid \tilde{\theta})] \mid D, u^*\}$.

We admit that a direct argument that $\hat{u}(D)$ should be approximately like a separated version of (12.1) using anticipatory data (D, u^*) is very slippery indeed. But this type of result for the predictive mode seems to occur over and over in filtered Poisson predictions (JEWELL [1985a] [1985b]).

(12.2) and (6.6) also reveal why simple approximations to h_θ are likely to work well in calculating $p(u \mid D)$. Because only the *ratios* of the integrals are used in the calculations, there is an automatic improvement in the effective accuracy of the approximation. This fact has already been made explicit in more general approaches to Bayesian prediction, see e.g., TIERNEY and KADANE [1986].

13. SUMMARY AND DISCUSSION

The main points of this paper are:

(1) The natural formulation of the IBNYR problem is in continuous time because of the underlying Poisson generation of claims and the continuous nature of reporting delays.

(2) In addition to observing the number of events, r, that are reported during the observation period, it is important to record secondary data consisting of the dates associated with each event in order to improve estimation of the unknown delay parameter $\tilde{\theta}$; the greatest benefit occurs when the exact delays are recorded, and the next best is when reporting dates are observed.

(3) The data likelihood reveals that r is used primarily to estimate the unknown Poisson parameter, $\tilde{\lambda}$, and the secondary data is used primarily to estimate $\tilde{\theta}$; however there is an important coupling term between $\tilde{\lambda}$ and $\Pi(t \mid \tilde{\theta})$, the probability that an event is reported during $(0, t]$. The maximum likelihood estimates of the parameters and of u, the number of events still unreported by time t, are either trivial or non-existent.

(4) Therefore, a Bayesian formulation, with prior densities on $\tilde{\lambda}$ and $\tilde{\theta}$, here assumed *a priori* independent: (i) is a more natural formulation, since prior information about claim rates and reporting delays is always available in practice; and (ii) gives more useful results, since it provides a complete predictive density, $p(u \mid D)$, for any observed data. In fact, emphasizing $p(u \mid D)$, rather than $p(\lambda, \theta \mid D)$, results in a computational simplification, as it eliminates the coupling term in the likelihood and gives $p(u \mid D)$ as the product of two factors that depend upon $p(\lambda)$ and $p(\theta)$, respectively.

(5) The predictive density can easily be calculated for arbitrary priors. With a Gamma (a, b) prior on $\tilde{\lambda}$, the essential work is the calculation of the ratio of two integrals depending upon $p(\theta)$. This ratio can be easily and accurately approximated for all types of secondary data, as shown by an example with

an exponential delay law and a Gamma prior on θ. The numerical computation of $p(u \mid D)$ then proceeds rapidly using a simple recursion, from which the mean, variance, tail distributon, etc., of $\tilde{u}$ can be found. If a quick point estimator is needed, the predictive mode $\hat{u}(D)$ can also be found from a simple iterative formula that always converges rapidly.

(6) A numerical example reveals that there is substantial residual variance in $p(u \mid D)$, even with a large volume of data and consonant prior. This is because, with r large, $\tilde{\theta}$ is estimated as well as it will ever be, especially with good secondary data; $\tilde{u}$ is then approximately Pascal distributed, with mean and variance decreasing as $1-\Pi(t \mid \hat{\theta})$ with increasing time. This effect is due to the underlying assumption of Poisson events, is common to all stochastic IBNR models, and shows the inadequacy of point estimation procedures. On the positive side, availability of the complete density $p(u \mid D)$ enables the direct calculation of risk factors and their incorporation into IBNYR reserves on a sound actuarial basis.

(7) The numerical example also reveals how uninformative and useless are Types III and IV secondary data. Satisfactory stability in estimating the parameters and predicting the unreported events requires the observation of at least the reporting dates, i.e., the time history of $r(t)$.

As mentioned earlier, the model developed here is only a first step on the road to more realistic and formulations. For example, as pointed out by a referee, the assumption of homogeneous process over $(0, T)$ is unnecessarily restrictive, and one could use a rate $\tilde{\lambda} v(t)$, where $v(t)$ is known "volume" of business, and then use operational time. Actually, in the sequel to this paper, we shall develop the modifications necessary when IBNYR reporting occurs only periodically — a quantized form of Type II data. It is in that context that it seems more natural to introduce different volumes or even different random rates for different exposure years. The availability of collateral data from other exposure years leads, in the quantized reporting case, to the possibility of simultaneous learning about all rates and the (common) delay parameter(s). This "IBNR triangle" model will be analyzed in a third report, where we will also attempt to say something about calendar time effects on the delay processes.

The assumption of prior independence of $\tilde{\lambda}$ and $\tilde{\theta}$ is, we realize, a strong one. However, it seems to the author that those who believe they are dependent must have in mind some phenomenon which needs additional modelling — for example, queuing bottlenecks in claims processing. Of course, insurance claims with long delays are usually qualitatively different from rapid filings, but this leads us into cost modelling, which is very difficult.

The author would like to thank two anonymous referees and VALENTIN WÜTHRICH for their comments and criticisms on this paper, many of which have been incorporated. Other suggestions on making this model more realistic and useful are always welcome.

APPENDIX A
GAMMOID APPROXIMATIONS

As discussed in section 9, the strategy in evaluating (9.1) with $p(\theta)$ and possibly a portion of $L(\theta \mid D)$ in Gamma form is to approximate the remainder of the integrand by a *Gammoid function*:

$$g(\theta) = (A\theta)^{\Gamma} e^{-\Delta\theta}, \tag{A.1}$$

in the region of the mode θ_0 of the Gamma part of the integrand; the location of the mode can be recalculated, if necessary. The final integral can then be calculated analytically.

The constant A is usually not of interest in our models. Beginning with the obvious:

$$\frac{d \ln g(\theta)}{d\theta} = \frac{\Gamma}{\theta} - \Delta; \qquad \frac{d^2 \ln g(\theta)}{d\theta^2} = -\frac{\Gamma}{\theta^2}, \tag{A.2}$$

we see that a function $L(\theta)$ can be approximated by (A.1) near θ_0 by using coefficients:

$$\Gamma = -\theta^2 \frac{d^2 \ln L(\theta)}{d\theta^2}\bigg|_{\theta_0}; \qquad \Delta = \frac{\Gamma}{\theta} - \frac{d \ln L(\theta)}{d\theta}\bigg|_{\theta_0}. \tag{A.3}$$

If only a negative exponential approximation is desired, we set $\Gamma = 0$ and find Δ from the first derivative; similarly, for a polynomial—only approximation, we set $\Delta = 0$ in the second formula in (A.3).

The success of the method depends on several factors. First of all, it is desirable to have a concentrated mode to begin with; we have found that even $c_0 = 3$ or 4 in the prior density is adequate. Secondly, if a portion of $L(\theta)$ is already unimodal in the range of interest, we have found it desirable to update the coefficients c_0 and d_0 immediately and to redefine the shifted mode θ_0 for use with the rest of $L(\theta)$, which is locally monotone. Usually, these latter coefficients will be slowly varying in the region of interest (we can make this more precise for our factors) and so the mode does not need continuing redefinition. If desired, after all the Gammoid coefficients have been determined, one can calculate a "final" mode for the integrand, and make one or two more passes to correct the coefficients found from (A.3). In our experience, such iterations lead to minor corrections and usually need to be repeated only a few times; this is essentially because we are only interested in the *ratios* of such integrals, as in (10.2).

More details on Gammoid approximations will appear in a forthcoming paper. Readers interested in the full details of the approximations for the numerical example in section 11 may obtain a copy of the original report from the author.

APPENDIX B
GAMMOID APPROXIMATIONS OF TERMS INVOLVING Π

Analysis of Type IV data $L(\theta)$ and the kernel $K(\theta)$ in (9.1) with exponential delays is simplified if we define the function:

$$\text{(B.1)} \qquad \psi(x) = \frac{1-e^{-x}}{x} = 1 - \frac{x}{2} + \frac{x^2}{6} - \frac{x^3}{24} + \frac{x^2}{120} - + \ldots,$$

and its derivatives:

$$\text{(B.2)} \quad \psi'(x) = -\left[\frac{1-(1+x)\,e^{-x}}{x^2}\right];\ \psi''(x) = \left[\frac{2-(2+2x+x^2)\,e^{-x}}{x^3}\right],$$

which are well-behaved for $x > 0$; for example, $\exp(-x/2) \leq \psi(x) \leq (1+x/2)^{-1}$ in this region.

With Type IV data, $L(\theta) = [\Pi(t \mid \theta)]^r$, and for $t \leq T$, $\Pi(t \mid \theta) = (t/T)(1-\psi(\theta t))$. Then:

$$\text{(B.3)} \qquad \frac{\partial \ln \Pi(t \mid \theta)}{\partial \theta} = \frac{-t\psi'(\theta t)}{1-\psi(\theta t)};$$

$$\frac{\partial^2 \ln \Pi(t \mid \theta)}{\partial \theta^2} = \frac{-t^2\psi''(\theta t)}{1-\psi(\theta t)} - \left[\frac{\partial \ln \Pi(t \mid \theta)}{\partial \theta}\right]^2.$$

When $t > T$, $\Pi(t \mid \theta) = 1-e^{-\theta(t-T)}\psi(\theta T)$, and (B.3) becomes:

$$\text{(B.4)} \qquad \frac{\partial \ln \Pi(t \mid \theta)}{\partial \theta} = \left[\frac{(t-T)\,\psi(\theta T) - T\,\psi'(\theta T)}{e^{\theta(t-T)} - \psi(\theta T)}\right];$$

$$\frac{\partial^2 \ln \Pi(t \mid \theta)}{\partial \theta^2} = \left[\frac{-T^2\psi''(\theta T) + (t-T)\,T\psi'(\theta T)}{e^{\theta(t-T)} - \psi(\theta T)}\right] - \left[\frac{\partial \ln \Pi(t \mid \theta)}{\partial \theta}\right]^2.$$

The gammoid coefficients, γ_4 and δ_4, are then found using (A.3) at the current mode θ_0. For large t, δ_4 can become negative; in this case, we recommend using just a polynomial approximation, with $\delta_4 = 0$ and γ_4 determined from the first derivative in (B.4).

The kernel $K(\theta)$ in (9.1):

$$\text{(B.5)} \quad K(\theta) = \left[1 - \left(\frac{\tau}{T}\right)\Pi(t \mid \theta)\right]$$

$$= \begin{cases} 1 - \left(\dfrac{t}{T}\right)^2 [1-\psi(\theta t)] & (t \leq T) \\ e^{-\theta(t-T)}\,\psi(\theta T) & (t \geq T) \end{cases},$$

is monotonic decreasing in θ, with first logarithmic derivative:

$$\text{(B.6)} \qquad \frac{d\ln K(\theta)}{d\theta} = \begin{Bmatrix} \dfrac{(t^3/T^2)\,\psi'(\theta t)}{K(\theta)} & (t \leq T) \\ -(t-T) + \dfrac{T\psi'(\theta T)}{\psi'(\theta T)} & (t \leq T) \end{Bmatrix}.$$

As both these forms are negative and slowly varying over a wide range of values, in contrast to (B.3) (B.4), it makes little sense to use a full Gammoid approximation, especially since negative values for γ_K may result! Thus, we just approximate by a negative exponential:

$$\text{(B.10)} \qquad \gamma_K = 0; \qquad \delta_K = -\left.\frac{d\ln K(\theta)}{d\theta}\right|_{\theta_0}.$$

This approximation updates the Gamma coefficient d_0 by $\delta_K u$, but does not change c_0.

The Gammoid approximatons presented above can be further refined by the use of additive terms to model the non-zero asymptotes in Types IIa, III, and IV data, or to give a better fit to the long tails of all the factors. However, our limited experience is that the refinements are of second-order effect in modifying the shape of h_θ, especially when the prior parameter density is reasonably informative.

APPENDIX C
TYPE IV DATA ONLY WITH BETA PRIOR

If we have only Type IV information and $t > T$, then:

$$\text{(C.1)} \qquad h_\theta(u \mid D) = \int [\Pi(t \mid \theta)]^r [1 - \Pi(t \mid \theta)]^u \, p(\theta)\, d\theta,$$

which suggests a reparametrization on the r.v. $\tilde{\pi} = \Pi(t \mid \tilde{\theta})$, which, with a Beta prior, $p(\pi)$, would give an analytical integral. The only inconvenience is that, if one truly believes in a Gamma prior on $\tilde{\theta}$, then the transformed density has a rather complex form on (0, 1]. Nevertheless, for a highly peaked Gamma, one could use the Gammoid approximation ideas to approximate $p(\pi)$ by a highly peaked Beta density with equivalent parameters (a_0, β_0); we omit the details.

We would then find $h_\pi(u \mid D) = \Gamma(\beta_0 + u)/\Gamma(a_0 + \beta_0 + r + u)$, changing the shaping factor in (10.2) as follows:

$$\text{(C.2)} \qquad \left[\frac{d_0 + r\delta_4 + \delta_K u}{d_0 + r\delta_4 + \delta_K + \delta_K u}\right]^{c_0 + r + \gamma_4} \to \left[\frac{\beta_0 + u}{a_0 + \beta_0 + r + u}\right].$$

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press

There is no set of parameters which will make these two shapes entirely equivalent, but one might attempt to fit the shapes for $u = 0$ and u very large, say, for fixed t and r. In any case, the shaping for intermediate values of u will be different for h_θ and h_π.

Although somewhat simpler, it is not clear that this approach is "better"; the parametric approach through an assumed from for $f(w \mid \theta)$ seems more "real" to us, as it is difficult to imagine how one could develop a consistent prior $p(\pi)$ for many different values of t. And finally, we must remind the reader of the very poor results obtained in section 11 with Type IV data; we do not expect that this approach will give any improvement for equivalent values of prior precision.

REFERENCES

BÜHLMANN, H., R. SCHNIEPER and E. STRAUB (1980). Claims reserves in casualty insurance based on a probabilistic model. *B. Assoc. Swiss Actuaries* **80**, 21–45.

DE VYLDER, F. (1982). Estimation of IBNYR claims by credibility theory. *Insurance: Math. & Econ.* **1**, 35–40.

HESSELAGER, O. and T. WITTING (1988). A credibility model with random fluctuations in delay probabilities for the prediction of IBNR claims. *ASTIN Bulletin* to appear.

JEWELL, W.S. (1980). Models in insurance: paradigms, puzzles, communications, and revolutions. Transactions, *21st International Congress of Actuaries, Zurich and Lausanne, 19-26 June 1980*, **S**, 87–141.

JEWELL, W.S. (1985a). Bayesian estimation of undetected errors. *Bayesian Statistics 2*, J.M. BERNARDO et al. (Eds.), Elsevier Science Publishers (North-Holland), 1985, 663–672.

JEWELL, W.S. (1985b). Bayesian extension to a basic model of software reliability. *IEEE Trans. Software Eng.* **SE-11**, no. 12, 1465–1471.

HACHEMEISTER, C. (1980). A stochastic model for loss reserving. Transactions, *21st International Congress of Actuaries, Zurich and Lausanne, 19-26 June 1980*, **1**, 185–194.

KARLSSON, J.-E. (1974). A stochastic model for time lag in reporting of claims, *J. Appl. Prob.* **II**, 382-387.

KARLSSON, J.-E. (1976). The expected value of IBNR-claims, *Scand. Actuarial J.* **1976**, 108–110.

KAMINSKY, K.S. (1987). Prediction of IBNR claim counts by modelling the distribution of report lags. *Insurance: Math. & Econ.* **6**, 161–159.

KRAMREITER, H. and E. STRAUB (1973). On the calculation of IBNYR reserves II. *Mitt. Ver. Schweiz. Vers. Math.* **73**. 177–190.

NORBERG, R. (1986). A contribution to modelling of IBNR claims. *Scand. Actuarial J.*, **1986**, 155–203.

STRAUB, E. (1972). On the calculation of IBNR reserves. *Prizewinning papers in the Boleslaw-Monic-Fund Competition 1971*. Nederlandse Reassurantie Groep N.V., Amsterdam, 1972, 123–131.

TIERNEY, L. and J.B. KADANE (1986). Accurate approximations for posterior moments and marginal densities. *Jour. Amer. Stat. Assoc.*, **81**, 82–86.

VAN EEGHEN, J. (1981). Loss reserving methods. *Surveys of actuarial studies no. 1*. Nationale-Nederlanden N.B., Rotterdam, 1981.

WILLIAM S. JEWELL
Engineering Systems Research Center, University of California, Berkeley, CA 94720.

https://doi.org/10.2143/AST.19.1.2014914 Published online by Cambridge University Press