Journal of
Statistical Theory and Practice
Volume 3, No. 1, March 2009

The Imprecise Logit-Normal Model and its Application to Estimating Hazard Functions

Miķelis Bickis, *Department of Mathematics & Statistics,*
University of Saskatchewan, Saskatoon SK, S7N 5E6, Canada.
Email: bickis@math.usak.ca

Received: November 23, 2007 *Revised:* March 30, 2008

Abstract

Given data on inter-arrival times, the imprecise Dirichlet model can be used to determine upper and lower values on the survival function. Similar bounds on the hazard function can be quite irregular without some structural assumptions. To address this problem, a family of prior distributions for a binomial success probability is contructed by assuming that the logit of the probability has a normal distribution. Posterior distributions so defined form a three-dimensional exponential family of which the beta family is a limiting case. This family is extended to the multivariate case, which provides for the inclusion of prior information about autocorrelation in the parameters. By restricting the hyperparameters to a suitably chosen subset, this model is proposed as an alternative to the usual imprecise Dirichlet model of Walley, having the advantage of providing smoother estimates of the hazard function. The methods are applied to data on inter-occurrence times of pandemic influenza.

AMS Subject Classification: 62F15; 62G05; 62N02; 62E17.

Key-words: Hazard function; Imprecise inference; Autocorrelation.

1. Introduction

Consider a renewal process, i.e., a sequence $T_1, T_2, \ldots$ of i.i.d. non-negative random variables representing the time intervals between a sequence of incidents, with $S_n = \sum_{i=1}^{n} T_i$ being the time of occurrence of the nth incident. After observing the process for some time, we are required to estimate the probability that the next incident is imminent. The information on which we can base this estimate is the pattern of previously observed incidents, including the waiting time since the most recent incident. We simplify the discussion by discretizing time into intervals, in which case "imminent" means occuring during the next interval. The continuous time case can always be approximated in this fashion.

* 1559-8608/09-1/$5 + $1pp – see inside front cover

If at time s we have observed $n-1$ incidents, then the problem at hand can be expressed mathematically as determining

$$\Pr\{S_n \le s+1 \mid S_{n-1} \le s < S_n; S_1, \ldots, S_{n-1}\}. \tag{1.1}$$

Since the T_i's are i.i.d., the above quantity is equal to

$$h_t = \Pr\{T_n \le t+1 \mid T_n > t\} \tag{1.2}$$

where $t = s - S_{n-1}$. This *discrete hazard function* does not depend on n and characterizes the distribution of T_i, since the *survival function*

$$S(t) = \Pr\{T_i > t\} = \prod_{j=1}^{t} (1 - h_j). \tag{1.3}$$

In continuous time the hazard function is defined as

$$\lambda(t) = \lim_{\varepsilon \downarrow 0} \frac{\Pr\{T_n \le t+\varepsilon \mid T_n > t\}}{\varepsilon}$$

in which case

$$h_t = 1 - e^{-\int_t^{t+1} \lambda(s)\, ds} \approx \lambda(t)$$

if both λ and its derivative are small. This paper focusses on the discrete case, and in view of this approximation will simply use *hazard* to refer to the discrete hazard function (1.2).

The stochastic independence of the inter-occurrence times is conditional on the knowledge of S. If S is known, then finding the probability (1.1) is mere computation—no inference is required. In practice, however, S would not be known. A frequentist approach would make inferences about S and compute the required probability from the inferred distribution, without acknowledging any probabilities about S itself.

In a Bayesian framework, we would consider the distribution of T_i to depend on a parameter θ having a prior distribution π, and write

$$h_t(\theta) = \Pr\{T_n \le t+1 \mid T_n > t; \theta\}$$

and then take its expectation with respect to the posterior distribution of θ to get the predictive probability

$$h_t = \int \Pr\{T_n \le t+1 \mid T_n > t; \theta\}\, d\pi(\theta | T_1, \ldots, T_{n-1}).$$

One could let θ index the distribution of T_i as a member of a parametric family, but here we will take a non-parametric approach, and use $\boldsymbol{\theta}$ to represent the complete vector $\{h_t, t \ge 1\}$.

The choice of prior has little influence on the predictive probability after a large number of observations but could make a large difference while there are still only a few previous incidents. Tools of imprecise probability can be used in presenting the uncertainty in the predictive probabilities.

Inference with imprecise probabilities was put on a solid foundation by Walley (1991), who also proposed the imprecise Dirichlet model (Walley, 1996) as a practical tool for implementing such inference. In the final section of this latter paper, Walley points out how the imprecise Dirichlet model can also be used for computing upper and lower CDF's based on a sequence of i.i.d. random variables. Coolen (1997) showed how this model could be generalized to take account of censored observations. (In the present context, the data have precisely one censored observation, since we have observed T_i for $i < n$ but know only that $T_n > t$.) A totally non-parametric approach based on *a priori* exchangeability of the waiting times was adopted by Coolen and Yan (2004). They showed how minimal assumptions about non-informative censoring allowed one to incorporate partial probability information from censored observations into the predictive distribution of the next waiting time. Their paper also includes a procedure for computing upper and lower survival functions.

The imprecise Dirichlet model is characterized by respecting certain invariance requirements which make sense for many situations, but not necessarily for the present problem. The time intervals are not generally exchangeable, nor are they subject to arbitrary agglomeration, as might be the case with balls of various colours. There is prior information in the order of the time intervals, including the expectation that the hazards would not fluctuate wildly from interval to interval. Reasonable estimates of the hazard function should be reasonably smooth. Lambert and Eilers (2005) use Bayesian methods to estimate hazard functions which are explicitly smoothed through the use of penalized splines. An alternative approach, which we explore in what follows, is to model prior dependence between close intervals, so that the posterior expectation in an interval is also influenced by data from nearby intervals. The methods of Coolen and Yan, by contrast, deliberately avoid any structural assumptions beyond exchangeability and non-informative censoring.

Although we ultimately do not use the imprecise Dirichlet model, we summarize its essential properties in the next section, since we use these concepts as a prototype for developing other semi-parametric imprecise inference models. In Section 3 we present a model that exploits the conditional independence of incidents among intervals. Unfortunately, this model does not address the need for smooth hazard estimates, which requires an explicit modelling of prior dependence across time. Flexible dependence structures are easier to model with a normal distribution than with the beta family, and so we propose in Section 4 the logit-normal model as an alternative to the beta family, and examine its properties. The adaptation of this family to imprecise inference, by analogy to the imprecise Dirichlet model, is discussed in Section 5, and this idea is extended to the multivariate case in the section that follows. Finally, we show an application to estimating the hazard function for influenza pandemics, which has been presented previously (Bickis and Bickis, 2007).

2. Imprecise Dirichlet Model

Walley (1996) proposed the imprecise Dirichlet model as a general procedure for making imprecise inferences. If $\mathbf{y}$ is an observed vector having a multinomial distribution with probabilities $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_k)$ (where $\sum_i \theta_i = 1$), the conjugate prior distribution on $\boldsymbol{\theta}$ is Dirichlet with hyperparameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)$

$$d\pi(\boldsymbol{\theta}) = K(\boldsymbol{\beta}) \prod_{j=0}^{k} \theta_j^{\beta_j}, \quad \theta_j > 0, \quad \sum_j \theta_j = 1,$$

$K(\boldsymbol{\beta})$ being a normalizing constant. The posterior distribution of $\boldsymbol{\theta}$ is again Dirichlet, with the hyperparameters $\boldsymbol{\beta}$ updated to $\boldsymbol{\beta}+\mathbf{y}$. The posterior expectation of $\boldsymbol{\theta}$, i.e., the vector of predictive probabilities of the next observation, is then

$$\mathrm{E}(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\boldsymbol{\beta}+\mathbf{y}}{\sum_j(\beta_j+y_j)} . \tag{2.1}$$

To make the model imprecise, the hyperparameter $\boldsymbol{\beta}$ is re-expressed as $\boldsymbol{\beta} = \nu\boldsymbol{\alpha}$ where $\sum_j \alpha_j = 1$ and $\nu > 0$ is a scalar. Upper and lower predictive probabilities are then obtained by fixing ν and taking the infimum and supremum, respectively, of (2.1) over the simplex $\alpha_j > 0, \sum_j \alpha_j = 1$.

The Dirichlet family is closed under agglomeration, with the marginal distribution's hyperparameters being just the sums of the hyperparameters corresponding to the agglomerated categories. Walley pointed out (Walley, 1996, p. 32) how this feature can be used to define upper and lower CDF's from an imprecise Dirichlet model in which the categories are intervals in which the values of a random variable may fall. Similarly, a set of independent observations of waiting times would provide upper and lower values of the survival function (1.3), with censoring handled by the method of Coolen (1997).

Upper and lower survival functions, however, do not provide upper and lower probabilities of imminent occurrence, for the transformation from survival function to hazard does not preserve order. On the other hand, upper and lower hazards *will* transform to upper and lower survival functions.

3. Product Beta Model

Let θ_t be the conditional probability of an incident in the tth interval since the last incident, given that there have been none in the intervening $t-1$ intervals. If among the preceding $n-1$ waiting times there have been m_{t-1} cases in which the waiting time has been at least $t-1$ and n_t cases in which the waiting time has been precisely t, then the likelihood function for this observation will be $(1-\theta_t)^{m_t}\theta_t^{n_t}$, where $m_t = m_{t-1} - n_t$ in the absence of censoring. If observations are censored at time $t-1$ then $m_t < m_{t-1} - n_t$. In the present context, only the current observation is censored, so that $m_{t+1} = m_t - n_{t+1} - 1$. Incidents in different intervals will be conditionally independent (given waiting), so the likelihood function of the entire sequence will be

$$\prod_{t=1}^{k}(1-\theta_t)^{m_t}\theta_t^{n_t}, \tag{3.1}$$

where k can be taken to be $\max_i T_i$.

The conjugate prior to this product binomial likelihood is a product beta distribution. We can then use, for each interval, an imprecise beta prior with hyperparameters $\alpha_t\nu$ and $(1-\alpha_t)\nu$. The upper and lower predictive hazards, (i.e., the upper and lower posterior expectations of θ_t) then become

$$\widehat{\overline{h}}_t = (n_t+\nu)/(n_t+m_t+\nu) \quad \text{and} \tag{3.2}$$

$$\underline{\widehat{h}}_t = n_t/(n_t + m_t + \nu), \tag{3.3}$$

respectively. The upper and lower survivor functions can then be computed from (1.3) as

$$\widehat{\overline{S}}_t = \prod_{j=1}^{t} \left(1 - \frac{n_j}{n_j + m_j + \nu}\right), \quad \text{and} \tag{3.4}$$

$$\widehat{\underline{S}}_t = \prod_{j=1}^{t} \left(1 - \frac{n_j + \nu}{n_j + m_j + \nu}\right). \tag{3.5}$$

If the precise current waiting time t has not been previously experienced (so that no interval has both censoring and incidence) then $\widehat{\overline{S}}_t$ is the same as the estimate obtained using the generalization by Coolen (1997) of the imprecise Dirichlet model. It converges to the Kaplan-Meier estimator as $\nu \to 0$. The lower survivor function $\widehat{\underline{S}}_t$, however, can be much lower than that obtained by Coolen for the reason that (3.2) is positive in every interval even if no incidents have been observed.

Although one might be reluctant to postulate a precise prior distribution on the parameter $\boldsymbol{\theta}$, it is nonetheless reasonable to assume that θ_t does not change dramatically with t. Our model is a discretization of a continuous-time process, and one would expect that the continuous hazard be a smooth function of time. We can model this prior information by incorporating an autocorrelation into the family of priors, rather than assuming the *a priori* independence of the product beta distribution. Modelling dependence among beta distributions is rather messy, and does not seem amenable to a stationary autoregressive structure such as one might expect of a hazard function. In the next section we propose an alternative prior family, which extends the two-dimensional beta family to three dimensions. This will be extended to the multivariate case in Section 6, which will allow for the definition of *a priori* autocorrelation.

4. Logit-Normal Model

Consider an observed random variable Y having a binomial distribution with n trials and success probability θ. Instead of using a beta prior, suppose that the logit of θ, $\omega = \log\big(\theta/(1-\theta)\big)$ has a prior normal distribution with mean μ and variance σ^2.

Since $\theta = e^{\omega}/(1+e^{\omega})$, the posterior density of ω given an observation $Y = y$ is proportional to

$$e^{-(\omega-\mu)^2/(2\sigma^2)} \frac{e^{\omega y}}{(1+e^{\omega})^n} \propto \frac{\exp\left[-\big(\omega - (\mu + \sigma^2 y)\big)^2/(2\sigma^2)\right]}{(1+e^{\omega})^n}. \tag{4.1}$$

In terms of the parameter θ, the posterior density transforms to

$$\pi(\theta|y) = K \exp\left[-\frac{\left(\log\left(\frac{\theta}{1-\theta}\right) - (\mu + \sigma^2 y)\right)^2}{2\sigma^2}\right] \frac{(1-\theta)^{n-1}}{\theta}, \tag{4.2}$$

where K is a normalizing constant. The constant K appears not to be analytically tractable, but can be computed using numerical or Monte Carlo integration.

An approximation of (4.1) can be obtained by linearization. Expanding $\theta = (1+e^{-\omega})^{-1}$ in a Taylor series around μ, we get

$$\theta \approx \frac{1}{1+e^{-\mu}} + \frac{e^{-\mu}}{(1+e^{-\mu})^2}(\omega-\mu)\ . \tag{4.3}$$

Putting $\bar{\theta} = (1+e^{-\mu})^{-1}$, we then get, under the prior distribution,

$$\mathrm{E}(Y \mid \omega) = n\theta \approx n\bar{\theta} + n\bar{\theta}(1-\bar{\theta})(\omega-\mu) \tag{4.4}$$

and

$$\begin{aligned} \mathrm{Var}(Y \mid \omega) &= n\theta(1-\theta) \\ &\approx n\bar{\theta}(1-\bar{\theta})\big(1+(1-\bar{\theta})(\omega-\mu)\big)\big(1-\bar{\theta}(\omega-\mu)\big) \\ &\approx n\bar{\theta}(1-\bar{\theta}) \end{aligned} \tag{4.5}$$

if ω is close to μ.

Thus the joint distribution of (ω, Y) can be approximated as a bivariate normal with mean $(\mu, n\bar{\theta})$ and dispersion matrix

$$\begin{bmatrix} \sigma^2 & n\bar{\theta}(1-\bar{\theta})\sigma^2 \\ n\bar{\theta}(1-\bar{\theta})\sigma^2 & n\bar{\theta}(1-\bar{\theta}) + n^2\bar{\theta}^2(1-\bar{\theta})^2\sigma^2 \end{bmatrix},$$

from which it follows that the posterior distribution of ω given $Y = y$ is approximately normal with

$$\begin{aligned} \mathrm{E}(\omega \mid y) &= \mu + \frac{\sigma^2(y-n\bar{\theta})}{1+n\bar{\theta}\left(1-\bar{\theta}\right)\sigma^2} \\ \mathrm{Var}(\omega \mid y) &= \frac{\sigma^2}{1+n\bar{\theta}\left(1-\bar{\theta}\right)\sigma^2}. \end{aligned} \tag{4.6}$$

More precisely, the posterior distributions (4.2) can be seen to form an exponential family (e.g. Schervish, 1995, Chapter 2). The log density of the prior distribution can be written

$$M(\theta, \boldsymbol{\phi}) = \phi_1 \left[\log\left(\frac{\theta}{1-\theta}\right)\right]^2 + \phi_2 \log\theta + \phi_3 \log(1-\theta) + C(\boldsymbol{\phi}) \tag{4.7}$$

where C is a normalizing constant and the hyperparameters ϕ_1, ϕ_2, ϕ_3 are

$$\begin{aligned} \phi_1 &= -1/(2\sigma^2) \\ \phi_2 &= \mu/\sigma^2 - 1 \\ \phi_3 &= -\mu/\sigma^2 - 1. \end{aligned} \tag{4.8}$$

Thus in canonical form, this is a three-dimensional exponential family with basis functions $\big(\log(\theta/(1-\theta))\big)^2$, $\log\theta$, and $\log(1-\theta)$. When $\phi_1 = 0$, i.e., the limiting case when

$\sigma^2 \to \infty$, we have the beta family. We call the family (4.7) the extended logit-normal family. It can be considered as the convex hull of the union of the logit-normal and beta families, although the natural (hyper)parameter space is actually bigger:

$$\begin{aligned} \phi_1 &\leq 0 \\ \phi_2, \phi_3 &> -1 \quad \text{if } \phi_1 = 0 \\ -\infty < \phi_2, \phi_3 &< \infty \quad \text{otherwise.} \end{aligned}$$

It is easy to see that after observing a binomial random variable y, the posterior distribution is obtained by updating the hyperparameters:

$$\begin{aligned} \phi_1 &\mapsto \phi_1 \\ \phi_2 &\mapsto \phi_2 + y \\ \phi_3 &\mapsto \phi_3 + n - y. \end{aligned} \tag{4.9}$$

For $\phi_1 < 0$, the densities all approach zero at the endpoints 0 and 1, whereas vertical asymptotes will obtain for $\phi_1 = 0$ and $\phi_2 < 0$ or $\phi_3 < 0$. The mode of the density is the solution of

$$\log\left(\frac{\theta}{1-\theta}\right) = \frac{(\phi_2 + \phi_3)\theta - \phi_2}{2\phi_1}, \tag{4.10}$$

which in the case of the prior distribution becomes.

$$\log\left(\frac{\theta}{1-\theta}\right) = \mu + (2\theta - 1)\sigma^2. \tag{4.11}$$

The prior density will be bimodal if $\sigma^2 > 2$ (i.e., $\phi_1 > -\frac{1}{4}$) and if

$$|\mu| < \log(\sigma^2 - \sigma\sqrt{\sigma^2 - 2} - 1) + \sigma\sqrt{\sigma^2 - 2}. \tag{4.12}$$

5. Imprecise Logit-Normal Model

The imprecise Dirichlet model is based on identifying a subfamily (indeed, an exponential subfamily) of the conjugate family such that the expectations of the parameters over the family cover the entire interval $(0,1)$. Bayesian updating of this subfamily shifts it such that the expectations with respect to the posterior distribution cover only a subinterval, giving posterior predictive probabilities that, while imprecise, are no longer vacuous. In order to apply the logit-normal model to imprecise inference, we need to find a corresponding subfamily.

There does not appear to be a closed form solution to the expected value of the extended logit-normal family (although Johnson (1949) did present an explicit but horrendous expression for the mean of the logit-normal family). The relation (4.10), however can be used for putting bounds on the mode. The right side represents a linear function in θ with slope $(\phi_2 + \phi_3)/(2\phi_1)$, whereas the left side is the inverse of the logistic curve. There will be a unique mode if the slope is negative, whereas two modes may occur for positive slopes, as pointed out above. A set of hyperparameters defines a set of lines in the plane whose intersection with the logistic curve defines the set of modes.

Using the relations (4.8) and (4.9), the right side of (4.10) becomes

$$\mu + (y-1)\sigma^2 - (n-2)\sigma^2\theta\,,$$

giving a negative slope, provided that $n > 2$. This line intersects the θ axis at

$$\theta_0 = \frac{y-1}{n-2} + \frac{\mu}{(n-2)\sigma^2}. \tag{5.1}$$

If the slope is negative and θ_0 is greater than $\frac{1}{2}$, the mode must be contained in $(\frac{1}{2}, \theta_0)$. Similarly, if θ_0 is less than $\frac{1}{2}$, the mode must be in $(\theta_0, \frac{1}{2})$. (see Figure 5.1.) Thus, if Φ_0 is a family of prior hyperparameters, the posterior modes must lie between

$$\inf_{(\mu,\sigma^2)\in\Phi_0} \left\{ \frac{y-1}{n-2} + \frac{\mu}{(n-2)\sigma^2} \right\} \quad \text{and} \quad \sup_{(\mu,\sigma^2)\in\Phi_0} \left\{ \frac{y-1}{n-2} + \frac{\mu}{(n-2)\sigma^2} \right\}, \tag{5.2}$$

provided that the family includes a wide enough range of hyperparameters that both $\theta_0 < \frac{1}{2}$ and $\theta_0 > \frac{1}{2}$ occur.

It is convenient to construct Φ_0 to be a one-dimensional subfamily, defined by a single hyperparameter μ. It remains then to express σ^2 as a function of μ in such a way that the upper and lower prior expectations of θ enclose the whole interval $(0,1)$, but the posterior expectations lie in a proper subinterval.

If Φ_0 is defined such that the bounds in (5.2) are properly contained in $(0,1)$, the posterior modes will also be constrained to the subinterval. The mean of a beta distribution is closer to $\frac{1}{2}$ then the mode. Numerical computations suggest that the same is true for the extended normal-logit family, although no rigorous proof has been found. If this conjecture is true, then the bounds (5.2) will also ensure that the posterior upper and lower expectations of θ will be contained in a proper subinterval of $(0,1)$. This will be achieved if in Φ_0, σ^2 grows at a rate at least as fast as μ. We also want to avoid a σ^2 that is too small, for in that case the prior will dominate the likelihood, stretching the intervals of imprecision towards such prior means.

On the other hand, to make the prior probabilities stretch from 0 to 1, we need that Φ_0 include all values of $\mu \in (-\infty, \infty)$. Moreover, μ must move away from the origin faster than σ, since otherwise a non-vanishing probability will remain on the other side of origin, which will prevent the expectation from going to the extreme.

These considerations lead us to propose defining

$$\Phi_0 = \left\{(\mu, \sigma^2) : \sigma = \sigma_0 + \tau|\mu|^{1/2},\ -\infty < \mu < \infty\right\}, \tag{5.3}$$

where $\sigma_0 > 0$ and $\tau < 1$.

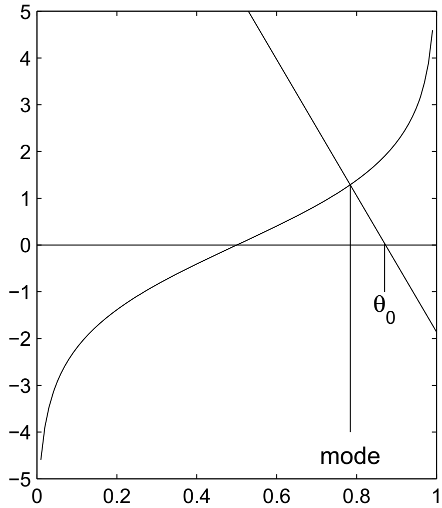

Figure 5.1. Finding the mode of the posterior density. The curve represents the logit function $\omega = \log(\theta/(1-\theta))$. The line is $\omega = \mu + (y-1)\sigma^2 - (n-2)\sigma^2\theta$.

6. Multivariate Logit-Normal Model

The logit-normal model generalizes quite readily to the multivariate situation. Let $\mathbf{y}$ be a k-dimensional product binomial random vector with $\mathbf{n}$ trials, and success probabilities $\boldsymbol{\theta}$. Letting $\mathbf{R}$ be a fixed correlation matrix, we let the vector of logits $\boldsymbol{\omega} = \log(\boldsymbol{\theta}/(1-\boldsymbol{\theta}))$ have a prior multivariate normal distribution with mean vector $\mu\mathbf{1}$ and dispersion $\sigma^2\mathbf{R}$, again with a two-dimensional hyperparameter (μ, σ^2).

In canonical form, the log density of the posterior becomes

$$M(\boldsymbol{\theta}, \boldsymbol{\phi}) = \phi_1 \boldsymbol{\omega}'\mathbf{R}^{-1}\boldsymbol{\omega} + \boldsymbol{\phi}_2 \log\boldsymbol{\theta} + \boldsymbol{\phi}_3 \log(1-\boldsymbol{\theta}) + C(\phi_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3), \tag{6.1}$$

where

$$\begin{aligned} \phi_1 &= -1/(2\sigma^2) \\ \boldsymbol{\phi}_2 &= \frac{\mu}{\sigma^2}\mathbf{R}^{-1} + \mathbf{y} - \mathbf{1} \\ \boldsymbol{\phi}_3 &= \frac{\mu}{\sigma^2}\mathbf{R}^{-1} + \mathbf{n} - \mathbf{y} - \mathbf{1}. \end{aligned} \tag{6.2}$$

The posterior densities thus form a $2k+1$ dimensional exponential family.

The multivariate posterior distribution can also approximated by linearization, analogously to (4.6):

$$\begin{aligned} &\mathrm{E}(\boldsymbol{\omega} \mid \mathbf{y}) = \mu\mathbf{1} + \sigma^2\mathbf{R}(\mathbf{I} + \sigma^2\boldsymbol{\Theta}\mathbf{R})^{-1}(\mathbf{y} - \bar{\theta}\mathbf{n}) \\ &\mathrm{Var}(\boldsymbol{\omega} \mid \mathbf{y}) = \sigma^2\mathbf{R}\big(\mathbf{I} - \sigma^2(\mathbf{I} + \sigma^2\boldsymbol{\Theta}\mathbf{R})^{-1}\boldsymbol{\Theta}\mathbf{R}\big), \end{aligned} \tag{6.3}$$

where $\bar{\theta} = (1+e^{-\mu})^{-1}$ as before, and $\boldsymbol{\Theta} = \bar{\theta}(1-\bar{\theta})\mathrm{diag}(\mathbf{n})$.

This approximation is not expected to be particularly good unless σ^2 is small. However, it may be a good proposal distribution for computing the posterior distribution by Markov chain Monte-Carlo. In the present context, however, we do not need the complete posterior distribution, but only its expectation. This can be readily computed by importance sampling (Robert, 2007), as described in the application below.

This multivariate logit-normal model can now be used to compute imprecise posterior hazards, as implied by a sequence of waiting times to previously observed incidents, as described in the introduction. As in Section 3, we let θ_i be the conditional probabity of incidence in the ith interval, given no incidence in the preceding $i-1$ intervals. We assume a prior distribution of the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ to be multivariate logit-normal, choosing $\mathbf{R}$ to have an autocorrelation structure, so that we have high prior density for the values of neighbouring hazards to be close, regardless of the values of the (variable) hyperparameters. This correlation will tend to smooth out the posterior estimates. Then we compute the range of the posterior estimates as the hyperparameters (μ, σ^2) vary over the subhyperparameter space defined in (5.3). The extremes of these posterior expectations will then give us imprecise predictive probabilities.

7. An Application

Bickis and Bickis (2007) applied this model to derive upper and lower probabilities for the imminent recurrence of pandemic influenza. Patterson (1987) reported that pandemic influenza has been recorded in the following years: 1729, 1732, 1781, 1788, 1830, 1833, 1836, 1889, 1899, 1918, 1957, 1968, and 1977. Considering the most remote year 1729 as the starting point, this gives waiting times of 3, 49, 7, 42, 3, 3, 53, 10, 19, 39, 11, and 9 years as well as the censored 30 year pandemic-free period to the time of writing. We consider these times as realizations of the T_i's, and use the theory of the preceding section to determine upper and lower hazards, given the data.

Defining the hyperparameters of the prior space as in (5.3) with $\sigma_0 = 8/3$ and $\tau = 1/2$, posterior distributions of the discrete hazard function were computed by importance sampling. For a range of values of μ, a Gaussian first-order autoregressive process was simulated, with autocorrelation 0.99 and variance as defined in (5.3). The values of this process were interpreted as the vector of logits $\boldsymbol{\omega}$ in (6.1), which were then used to compute the likelihood of the data according to (3.1). This procedure was iterated 1000 times, averaging the simulated θ's weighted by the likelihoods, giving an estimated predictive probability of recurrence after each waiting time. The entire process was repeated for a range of μ's ranging from -8 to 2.

The values of the hyperparameters were chosen heuristically. Hyperparameters of $\mu = 0$ and $\sigma = 8/3$ in the univariate case give a distribution of θ that closely resembles a beta distribution with parameters $(\frac{1}{2}, \frac{1}{2})$, i.e., in the language of Section 2, with an imprecision parameter $\nu = 1$, as recommended by Walley (1996). The value of σ thus attempts to maintain a level of imprecision comparable to that used for the imprecise Dirichlet model. The autocorrelation coefficient was chosen to achieve an appreciable, but not excessive, level of smoothing. Experimentation with the autocorrelation demonstrated that values much lower than 0.99 still provided rather ragged estimates, whereas at 0.999, time-dependent features of the hazard function were obscured. (See Figure 7.1) (It should be stressed that

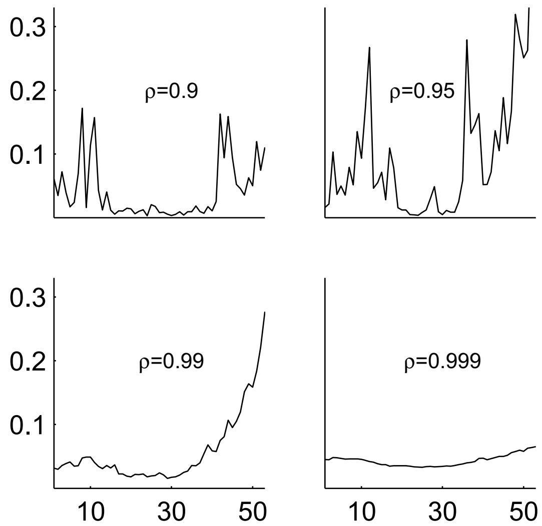

Figure 7.1. Monte Carlo estimates of posterior hazards using $\mu = -3$ and various autocorrelations ρ. Each path is the average of 1000 simulations.

the autocorrelation is between prior probabilities of adjacent inter-pandemic years. We are not modelling a correlation from one pandemic to the next.)

In theory, μ should range from $-\infty$ to ∞, but of course a simulation is restricted to finite bounds. The centre of the range $(-8, 2)$ corresponds to the logit of 0.05, which is approximately the average hazard level as estimated by maximum likelihood. Experimentation with ranges of μ suggested that the posterior expectations would not change much by going outside this range.

The results are shown in Figure 7.2, where the points corresponding to a particular value of μ have been connected. The bundle of paths then approximates the intervals between upper and lower predictive probabilities.

The same simulation can also be used to estimate the survival function of the process. For each step of the simulation, the survival function was computed as in (1.3) and again averaged with likelihood weights. The corresponding graph is show in Figure 7.3. This imprecise estimate of the survival function is superimposed in Figure 7.4 on the upper and lower survival functions determined according to the method of Coolen and Yan (2004). While the smooth estimate displays a somewhat greater degree of imprecision, the two methods give fairly similar patterns. In particular, the structural and parametric assumptions in our model have not imposed notably more precision into the estimates than are obtained by the minimalist assumptions of Coolen and Yan.

8. Conclusions

In this paper we have presented a model for imprecise inference that is particularly suitable in a multivariate situation where a prior dependence structure is natural. The model is based on an exponential family of priors in which we can define a subfamily of hyperpa-

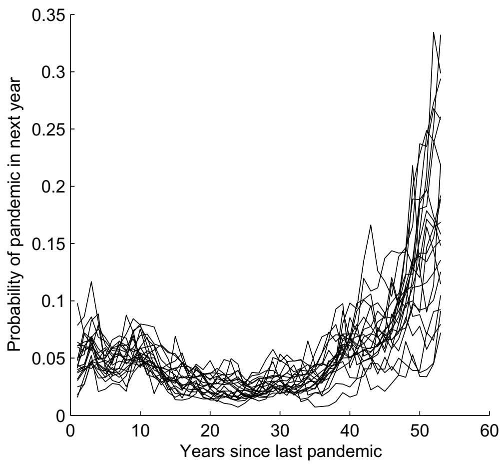

Figure 7.2. Sampled hazard functions from autocorrelated imprecise posterior. Each path is the average of 1000 simulations.

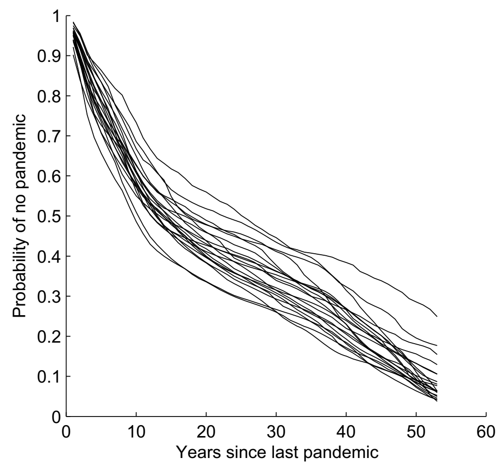

Figure 7.3. Sampled survival functions from autocorrelated imprecise posterior. Each path is the average of 1000 simulations.

rameters which will lead to vacuous prior expectations, but posterior expectations having less imprecision. Such a subfamily is an essential property of the commonly used imprecise Dirichlet model, but is not exclusive to it. We note in passing that any useful imprecise inference model requires prior information of *some* sort, since Walley (1991) established that a totally vacuous prior necessarily leads to a vacuous posterior.

The model appears to have satisfactory performance in the example presented, but more works remains to be done. In particular, one would like a rigorous proof that the extended logit-normal family shares the property of the beta family that the means are closer to $\frac{1}{2}$ than the modes. This property is required to unequivocally establish that the proposed sub-hyperparameter space has the desired property for imprecise inference. Unlike the imprecise

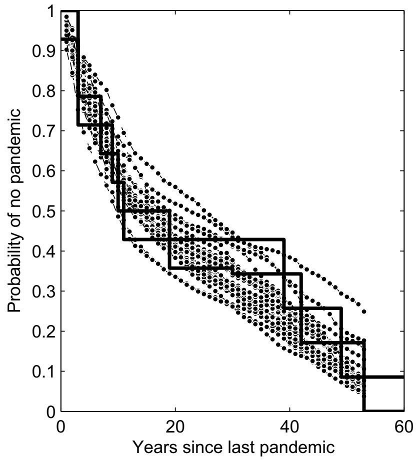

Figure 7.4. Autocorrelated imprecise posterior (dotted lines) compared with upper and lower survival functions of Coolen and Yan (solid lines)

Dirichlet model, the logit-normal family is not analytically tractable, so efficient computations are important. In the one-dimensional case, numerical integration works fairly well except for cases with large values of σ^2, where standard integration algorithms have difficulty dealing with the sharp peaks near the endpoints. Some attempts at computing the posterior distribution using Markov chain Monte Carlo have been implemented, but the performance of this method depends crucially on a suitable proposal distribution. If only posterior expectations are required, then importance sampling, as used in the example, is a fairly quick and accurate procedure. More details on these questions will be discussed in future papers.

References

Bickis, M., Bickis, U., 2007. Predicting the next pandemic: An exercise in imprecise hazards. In *ISIPTA'07 – Fifth International Symposium on Imprecise Probability: Theories and Applications*, de Cooman, G., Vejnarová, J., Zaffalon M. (editors), 41–46, Prague, Czech Republic.

Coolen, F.P.A., 1997. An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, 56, 61–68.

Coolen, F.P.A., Yan, K.J., 2004. Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, 126, 25–54.

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149–176.

Lambert, P., Eilers, P.H.C., 2005. Bayesian proportional hazards model with time-varying regression coefficients: A penalized Poisson regression approach. *Statistics in Medicine*, 24, 3977–3989.

Patterson, K.D., 1987. *Pandemic Influenza, 1700–1900 : A Study in Historical Epidemiology*. Rowman and Littlefield, Totowa, NJ.

Robert, C., 2007. *The Bayesian Choice: From Decision Theoretic Foundations to Computational Implementation*. Springer, New York.

Schervish, M.J., 1995. *Theory of Statistics*. Springer, New York.

Walley, P., 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.

Walley, P., 1996. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society*, 58B, 3–57.