“Parametric empirical Bayes estimation of the net premium with right censored data”

AUTHORS	Mostafa Mashayekhi
ARTICLE INFO	Mostafa Mashayekhi (2011). Parametric empirical Bayes estimation of the net premium with right censored data. *Insurance Markets and Companies*, *2*(1)
RELEASED ON	Monday, 04 April 2011
JOURNAL	"Insurance Markets and Companies"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES
0

NUMBER OF FIGURES
0

NUMBER OF TABLES
0

© The author(s) 2021. This publication is an open access article.

BUSINESS PERSPECTIVES

businessperspectives.org

Mostafa Mashayekhi (USA)

Parametric empirical Bayes estimation of the net premium with right censored data

Abstract

This paper considers an empirical Bayes estimation of the net single premium for one-year insurance policies with right-censored observations of the times of claim causing events. It is assumed there are m independent classes of insureds with $n_i + 1$ members in the i^{th} class ($i \in \{1,\ldots, m\}$) and n_i of these members provide the claims experience. Members of the i^{th} class are assumed to be subject to the same constant hazard rate θ_i, during the policy period, with $\theta_1,\ldots,\theta_m$ being the values of independent unobservable random variables with common gamma distribution with unknown parameters α and β. The force of interest is assumed to be a deterministic function of time. The claim sizes are allowed to be random. The authors obtains sufficient conditions for asymptotic optimality of empirical Bayes estimators of the net single premium, under the squared error loss function, and provide an easy to compute example of estimators that satisfy the sufficient conditions.

Keywords: parametric empirical Bayes, asymptotically optimal, right censored data, net single premium.

Introduction and preliminaries

The assumption of a constant hazard rate, or constant force of mortality, is frequently adopted by actuaries in the process of determining premiums for life insurance policies. This assumption is also present in models of risk assessment in casualty insurance that use a Poisson process for the number of claims in a given period of time. In this paper we adopt the assumption of a constant hazard rate in an empirical Bayes approach to estimation of the net single premium for coverage against an event that cannot happen to a policyholder more than once during the policy period. The examples of such events include a death, loss of a limb, catching a certain incurable disease, or loss of a certain unique property which cannot be replaced, etc.

Consider the net single premium for a one-year insurance policy that pays one dollar if a certain event occurs within the next year when the following assumption is satisfied.

Assumption 1. (1) The force of interest, $\delta(t)$ at time t, is a deterministic function of time. (2) The time to the event has a constant hazard rate, θ, during the policy period.

The net single premium satisfying the equivalence principle for this insurance policy, for the case when the value of θ is known, is denoted throughout this paper by $\pi(\theta)$. With Y denoting the time to the event, we have:

$$\pi(\theta) = E_\theta[e^{-\int_0^Y \delta(t)dt} 1_{[Y<1]}] =$$

$$= \int_0^1 (e^{-\int_0^y \delta(t)dt})\theta e^{-\theta y} dy. \tag{1}$$

In general, let B denote the amount of the claim random variable. Note that since there is no claim if the event does not happen during the policy period, B and Y are not independent, and $B = B1_{[Y<1]}$.

Suppose if there is a claim its size is independent of the time of occurrence of the event under every value of θ. Then given the value of $1_{[Y<1]}$ the random variables B and Y are conditionally independent and with E_θ denoting the expectation with respect to the joint distribution of B and Y under θ, the conditional expectation $E_\theta[B \mid Y<1]$ does not depend on θ. Suppose this conditional expectation is finite and let $\mu = E_\theta[B \mid Y<1]$.

The net single premium satisfying the equivalence principle, in the above case, is given by:

$$\begin{aligned} & E_\theta[Be^{-\int_0^Y \delta(t)dt} 1_{[Y<1]}] = \\ & = E_\theta[E_\theta(Be^{-\int_0^Y \delta(t)dt} 1_{[Y<1]} \mid 1_{[Y<1]})] = \\ & = E_\theta\{E_\theta[B \mid 1_{[Y<1]}]E_\theta[e^{-\int_0^Y \delta(t)dt} 1_{[Y<1]} \mid 1_{[Y<1]}]\} = \\ & = (1-e^{-\theta})\mu\pi(\theta)(1-e^{-\theta})^{-1} = \mu\pi(\theta). \end{aligned} \tag{2}$$

In many cases data for separate policies have been collected by observing small cohorts. In such cases an actuary can improve the quality of his/her estimates by borrowing strength from similar data. A popular method that is often used for this purpose is the empirical Bayes credibility method. The estimates, produced by the empirical Bayes credibility

© Mostafa Mashayekhi, 2011.

method, are linear empirical Bayes estimates under the squared error loss function. In this paper by "Bayes estimator" we mean unrestricted Bayes estimator and use "credibility estimator" to refer to a restricted linear Bayes estimator.

The use of an empirical Bayes credibility estimator instead of an empirical Bayes estimator is often justified by the simplicity of computation of the empirical Bayes credibility estimators. It is argued (see Bühlmann and Gisler, 2005) that in general the Bayes estimator does not have a closed form representation and, therefore, the value of the Bayes estimator can only be computed by numerical procedures. The argument about the difficulty in the computation of a Bayes solution extends to the empirical Bayes method because every empirical Bayes procedure either involves computation of a Bayes solution versus some estimate of the unknown prior distribution or some direct estimation of the Bayes solution versus the unknown prior. Another advantage of using an empirical Bayes credibility estimator is that an empirical Bayes credibility estimator can be completely non-parametric and, therefore, more robust relative to the empirical Bayes approach. For computation of an empirical Bayes credibility estimate, one does not need to assume any functional forms for the model distribution (i.e., the conditional distribution of an observation given the parameter) or the prior distribution. An empirical Bayes estimator, on the other hand, cannot be found without a parametric assumption about the functional form of the model distribution. Supporters of parametric inference would argue, however, that the extra precision gained by a reasonable parametric model can outweigh the extra robustness of a non-parametric model.

In addition to the above arguments, in favor of the credibility estimators, a famous theorem of Jewell (1974; 1975) asserts that under certain regularity conditions the credibility estimator is equal to the Bayes estimator when the model distribution belongs to a single parameter exponential family of distributions consisting of distributions with density of the form

$$p(x \mid \theta) = \frac{a(x)e^{-\theta x}}{c(\theta)},$$

if the natural conjugate distribution is used for the prior. A rigorous treatment of Jewell's theorem with its extension from 1-dimensional parameter case to the *d*-dimensional case and a proof for the converse of Jewell's theorem are given in Diaconis and Ylvisaker (1979). The converse of Jewell's theorem asserts that under the stated conditions the Bayes estimator is equal to the credibility estimator only if the conjugate prior is used.

However, as we will show in the model that we are considering asymptotically optimal empirical Bayes estimates are not hard to compute and the credibility solution does not coincide with the Bayes solution. We will give sufficient conditions for asymptotic optimality of the empirical Bayes estimators in our model and provide an easy to compute example which satisfies the sufficient conditions.

Let $Y_1,...,Y_{n+1}$ denote the times to the claim-causing event for $n+1$ policies, with $Y_1,...,Y_n$ corresponding to past policies, and suppose that given θ the times $Y_1,...,Y_{n+1}$ are independent and have a constant hazard rate equal to θ for the first year. In practice an insurer is not in a position to observe $Y_1,...,Y_n$ and the available observations are values of censored times.

Suppose it is only possible to observe the type 1 right censored random variables $X_1,...,X_n$, where for each $i \in \{1,...,n\}$,

$$X_i = Y_i 1_{[Y_i < 1]} + 1_{[Y_i \geq 1]}.$$

Let

$$D(s) = e^{-\int_0^s \delta(t)dt} 1_{[s < 1]}. \qquad (3)$$

Then in the case when the $(n+1)^{th}$ policy pays one dollar at the time of the event, the random present value of the payoff due to the $(n+1)^{th}$ policy is equal to $D(X_{n+1})$. Since the value of $\delta(t)$ changes over time the random variables $D(X_1),...,D(X_n)$ are not the discounted payoffs at time of issue, of the first n policies. However, since the random variables $D(X_1),...,D(X_{n+1})$ satisfy the assumptions of Bühlmann's credibility model (c.f., Bühlmann, 1967; or for example Klugman et al., 1998, pp. 436-437) and the information generated by $X_1,...,X_n$ is the same as the information generated by $D(X_1),...,D(X_n)$, these random variables may be used in the Bühlmann formula to obtain the credibility estimate of the net single premium for the $(n+1)^{th}$ policy. With Z denoting the credibility factor, the credibility estimator of the net single premium for the $(n+1)^{th}$ policy is equal to:

$$Z[n^{-1}\sum_{i=1}^{n} D(X_i)] + (1-Z)\ E[\pi(\theta)]. \qquad (4)$$

Note that the credibility estimator given in equation (4) is not equal to the Bayes estimator because the conditional distribution of $D(X_1)$ given θ and the assumed gamma distribution for θ do not satisfy the conditions of Jewell's theorem. Indeed, it is easy to check that for the simple case, where $\delta(t)=0$, and hence the conditional distribution of $D(X_1)$ given θ is Bernoulli with probability of success equal to $1-e^{-\theta}$, the gamma distribution is not the conjugate prior. Therefore, by the Diaconis and Ylvisaker (1979) converse of Jewell's theorem the credibility estimator cannot be equal to the Bayes estimator in this case.

Theorem 1 in Section 1 gives the Bayes estimator of the net single premium for the $(n+1)^{th}$ policy for the case when the claim size is constant, and the prior distribution is gamma. Theorem 2 gives a generalization of Theorem 1 for the case when the claim size is random. In Section 2 we consider empirical Bayes estimation of the net single premium and provide sufficient conditions for asymptotic optimality of empirical Bayes estimators in Theorem 3. Examples of estimators that satisfy the sufficient conditions for asymptotic optimality are provided in Section 3.

1. The Bayes estimator

The following theorem gives the unrestricted Bayes estimator of the net single premium for one-year policies with constant claim sizes based on type 1 right censored data.

Theorem 1. Let $X_1,...,X_{n+1},\theta$ be random variables such that θ is gamma (α,β) and given θ the random variables $X_1,...,X_{n+1}$ are $i.i.d.$ with distribution function

$$F_\theta(x)=(1-e^{-\theta x})1_{[0<x<1]}+1_{[x\geq 1]}. \tag{5}$$

Then with $\pi(\theta)$ defined in equation (1) and $D(s)$ defined in equation (3):

$$E[D(X_{n+1})\mid X_1,...,X_n]=E[\pi(\theta)\mid X_1,...,X_n]= \\ =E^*[\pi(\theta)],$$

where with $\alpha^*=\alpha+\sum_{i=1}^{n}1_{[X_i<1]}$ and $\beta^*=\beta+\sum_{i=1}^{n}X_i$, the symbol E^* denotes expectation under gamma (α^*,β^*) distribution for θ.

Proof. Let λ denote the uniform probability measure on $[0, 1]$ and let ϑ be the probability measure that gives mass 1 to $\{1\}$ and let η be the measure defined by $\eta=\lambda+\vartheta$. Then, given θ, each X_i has density (Radon Nikodym derivative) with respect to η given by:

$$f_\theta(x)=\theta^{1_{[x<1]}}e^{-\theta x}1_{[x>0]}.$$

Let g denote the density of θ and, for $k>1$ let h_k denote the marginal density of $X_1,...,X_k$ with respect to η^k. Then:

$$E[D(X_{n+1})\mid X_1=x_1,...,X_n=x_n]= \\ =\int_{[0,1)}D(x_{n+1})\left(\frac{h_{n+1}(x_1,...,x_{n+1})}{h_n(x_1,...,x_n)}\right)d\eta(x_{n+1})= \\ =\int_0^1 D(x_{n+1})\left(\frac{\int_0^\infty f_\theta(x_{n+1}).\theta^{\sum_{i=1}^{n}1_{[x_i<1]}}e^{-\theta\sum_{i=1}^{n}x_i}g(\theta)d(\theta)}{h_n(x_1,...,x_n)}\right)dx_{n+1}. \tag{6}$$

Let $g(\theta\mid x_1,...,x_n)$ denote the conditional density of θ given $X_1=x_1,...,X_n=x_n$ and note that

$$g(\theta\mid x_1,...,x_n)=\frac{\theta^{\sum_{i=1}^{n}1_{[x_i<1]}}e^{-\theta\sum_{i=1}^{n}x_i}g(\theta)}{h_n(x_1,...,x_n)}.$$

By the Tonelli theorem the right hand side of equation (6) is equal to

$$\int_0^\infty(\int_0^1 D(x_{n+1})f_\theta(x_{n+1})dx_{n+1})g(\theta\mid x_1,...,x_n)d\theta= \\ =\int_0^\infty \pi(\theta)g(\theta\mid x_1,...,x_n)d\theta= \\ =E\big(\pi(\theta)\mid X_1=x_1,...X_n=x_n\big).$$

Observe that $g(\theta\mid x_1,...x_n)$ is proportional to $\theta^{(\alpha+\sum_{i=1}^{n}1_{[x_i<1]})-1}e^{-(\beta+\sum_{i=1}^{n}x_i)\theta}$. Therefore, it is a gamma $(\alpha+\sum_{i=1}^{n}1_{[x_i<1]},\ \beta+\sum_{i=1}^{n}x_i$) density.

For the case when the claim sizes are random, let $B_1,...,B_{n+1}$ denote the random claim amounts of $(n+1)$ policies and suppose the following assumption is satisfied.

Assumption 2. (1) Conditional on the information that there is a claim for the $(n+1)^{th}$ policy, the size of the claim is independent of the time of occurrence of the event. (2) The dependence of $B_1,...B_{n+1}$ is only through the dependence of $X_1,...X_{n+1}$, so that the conditional distribution of (B_{n+1},X_{n+1}) given $(B_1,X_1),...,(B_n,X_n)$ is the same as the conditional distribution of (B_{n+1},X_{n+1}) given $X_1,...X_n$, and the conditional distribution of B_{n+1} given $X_1,...,X_{n+1}$ is the same as the conditional distribution of B_{n+1} given X_{n+1}. (3) $E[B_i \mid X_i < 1] = \mu$, and $Var[B_i \mid X_i < 1] = \sigma^2 < \infty$ for every $i \in \{1,...,n\}$.

Let E denote the expectation under the joint distribution of all of the random variables involved. Then by (1) and (3) of Assumption 2

$$E[B_{n+1} \mid X_{n+1}] = 1_{[X_{n+1}<1]} E[B_{n+1} \mid 1_{[X_{n+1}<1]}] = \\ = \mu 1_{[X_{n+1}<1]}.$$

Note that the Bayes estimator of the net single premium for the $(n+1)^{th}$ policy under the squared error loss function is $E[B_{n+1}D(X_{n+1}) \mid (B_1,X_1),...,(B_n,X_n)]$. The following theorem gives the Bayes estimator of the net single premium for the case when the claim sizes are random.

Theorem 2. Let $X_1,...,X_{n+1},\theta$ be as in Theorem 1 and let $B_1,...,B_{n+1}$ be such that $(B_1X_1),...,(B_{n+1}X_{n+1}),\theta$ satisfy Assumption 2. Then with E^* as in Theorem 1,

$$E[B_{n+1}D(X_{n+1}) \mid (B_1,X_1),...,(B_n,X_n)] =$$

$$= E[\mu\pi(\theta) \mid X_1,...,X_n] = \mu E^*[\pi(\theta)].$$

Proof. By Assumption 2 we have:

$$E[B_{n+1} \mid X_1,...,X_n,X_{n+1}] = \\ = E[B_{n+1} \mid X_{n+1}] = \mu 1_{[X_{n+1}<1]} \quad (7)$$

and

$$E[B_{n+1}D(X_{n+1}) \mid (B_1,X_1),...,(B_n,X_n)] = \\ = E[B_{n+1}D(X_{n+1}) \mid X_1,...,X_n].$$

Since the information generated by $X_1,...,X_n$ is contained in the information generated by $X_1,...,X_n,X_{n+1}$, we have (see Theorem 34.4 of Billingsley, 1986):

$$E[B_{n+1}D(X_{n+1}) \mid X_1,...,X_n] = \\ E[E[B_{n+1}D(X_{n+1}) \mid X_1,...,X_{n+1} \mid X_1,...,X_n].$$

Since $D(X_{n+1})$ is a Borel function of $X_1,...,X_{n+1}$,

$$E[B_{n+1}D(X_{n+1}) \mid X_1,...,X_{n+1}] = \\ = D(X_{n+1})E[B_{n+1} \mid X_1,...,X_{n+1}]. \quad (8)$$

By equation (7) it follows that the right hand side of equation (8) is equal to $D(X_{n+1})\mu 1_{[X_{n+1}<1]} = \mu D(X_{n+1})$

Therefore, an application of Theorem 1 completes the proof.

The computation of the Bayes estimator (with known prior parameters) and, as we will see, the empirical Bayes estimators (based on estimates of the prior parameters) of the net single premium involves computation of integrals of the form

$$\int_0^\infty \pi(\theta)g(\theta)d\theta,$$

where g is a gamma (a,b) density for some $a > 0$ and $b > 0$.

To approximate such an integral with error less than an arbitrary positive ε, note that with $0 < c < d < \infty$ we can write:

$$\int_0^\infty \pi(\theta)g(\theta) = \int_0^c \pi(\theta)g(\theta) + \\ + \int_c^d \pi(\theta)g(\theta) + \int_d^\infty \pi(\theta)g(\theta). \quad (9)$$

Since $\pi(\theta) < 1$, the first term on the right hand side of equation (9) is less than $(\Gamma(a))^{-1}b^a)c^{a-1}$. Hence, we can choose c such that the first term is less than $\varepsilon/3$.

By Markov inequality the third term on the right hand side of equation (9) is less than or equal to $a(bd)^{-1}$. Hence, we can choose d such that the third term is less than $\varepsilon/3$.

When the integrand has continuous second derivative on $[c,d]$, we can use numerical integration, for example the Composite Midpoint rule, to approximate the second term with error less than $\varepsilon/3$.

For example it is easy to see that when the force of interest is a constant δ, we have $\pi(\theta) = = (\theta+\delta)^{-1}\theta[1-e^{-(\delta+\theta)}]$, and the integrand has continuous second derivative.

2. Asymptotically optimal empirical Bayes estimators

In the empirical Bayes approach pioneered by Robbins (1955) it is assumed there are m independent random pairs $(\theta_1,V_1),...,(\theta_m,V_m)$ such that $V_1,...,V_m$ are observable and distribution of V_i depends on the parameter θ_i. The parameters $\theta_1,...,\theta_m$ are independent unobservable random variables with a common distribution G. There is a non-negative loss function L and the task is to find decision rules $t_m(.) = t_m(V_1,...,V_m,.)$ that are asymptotically optimal in the sense that

$$E[L(t_m(V_m),\theta_m)] - \min_t E[L(t(V_m),\theta_m] \longrightarrow 0$$

as $m \longrightarrow \infty$.

In the non-parametric empirical Bayes approach introduced by Robbins (1955), G is completely unknown. In the parametric empirical Bayes approach that was later explored by Efron and Morris (1973, 1975), the prior distribution G belongs to a parametric family of distributions.

The model we are considering for an empirical Bayes approach is formally described in the following assumption.

Assumption 3. With $V_i = (B_{i1}, X_{i1}, ..., B_{in_{i+1}}, X_{in_{i+1}})$, the random pairs $(\theta_1, V_1), ..., (\theta_m, V_m)$ are independent and for each i, conditional on θ_i, the random variables $X_{i1}, ..., X_{in_{i+1}}$ are independent with distribution function

$$F_{\theta_i}(x) = (1 - e^{-\theta_i x})1_{[0<x<1]} + 1_{[x\geq 1]}.$$

The random variables $\theta_1, ..., \theta_m$ have common gamma distribution with unknown parameters α and β, and for each i Assumption 2 is satisfied with $B_{i1}, ..., B_{in_{i+1}}$ in place of $B_1, ..., B_n$ and $X_{i1}, ..., X_{in_{i+1}}$ in place of $X_1, ..., X_n$.

This problem is slightly different from the problem introduced by Robbins (1955) in the sense that the component problems are not identical when the n_i's are not equal. However, if τ is a Borel function, one can consider an empirical Bayes estimator $\hat{\tau}_i^{EB}$ of $\tau(\theta_i)$, $i \in \{1, ..., m\}$, asymptotically optimal under squared error loss if with $\hat{\tau}_i^{B}$ denoting the Bayes estimator of $\tau(\theta_i)$,

$$E[\hat{\tau}_i^{EB} - \tau(\theta_i)]^2 - E[\hat{\tau}_i^{B} - \tau(\theta_i)]^2 \longrightarrow 0$$

as $m \longrightarrow \infty$.

Asymptotic optimality for empirical Bayes forecasts may be defined similarly.

Theorem 3 below provides sufficient conditions for asymptotic optimality of the empirical Bayes estimators of the net single premium. Lemma 1 and Lemma 2 are used in the proof of Theorem 3. The proofs of these lemmas are deferred to the Appendix.

In the rest of this paper all incompletely described limits are as $m \longrightarrow \infty$ through positive integers.

Lemma 1. Suppose for some $K < \infty$:

- $0 < a \leq a_m \leq a + K$, and $0 < a \leq \hat{a}_m \leq a + K$;
- $0 < b < b_m < b + K$, and $0 < b < \hat{b}_m < b + K$;
- $\hat{a}_m - a_m \longrightarrow 0$, and $\hat{b}_m - b_m \longrightarrow 0$.

Then

$$\int_0^\infty | y^{\hat{a}_m - 1} e^{-\hat{b}_m y} - y^{a_m - 1} e^{-b_m y} | dy \longrightarrow 0.$$

Lemma 2. Suppose a_m, $\hat{a}_m$, b_m, and $\hat{b}_m$ are as in Lemma 1. Let g_m denote the gamma density with parameters a_m and b_m, and let $\hat{g}_m$ denote the gamma density with parameters $\hat{a}_m$ and $\hat{b}_m$. Then

$$\int_0^\infty | \hat{g}_m(x) - g_m(x) | dx \longrightarrow 0.$$

Theorem 3. Suppose $\sup_{i\geq 1} n_i < \infty$, and $\alpha \in (0, N_1]$ and $\beta \in (0, N_2]$, and $\mu \in (0, N_3]$, where N_1, N_2, and N_3 are known numbers. Let $0 \leq \hat{\alpha} \leq N_1$, and $0 \leq \hat{\beta} \leq N_2$, and $0 \leq \hat{\mu} \leq N_3$ be such that

- $\hat{\alpha} \xrightarrow{P} \alpha$;
- $\hat{\beta} \xrightarrow{P} \beta$;
- $\hat{\mu} \xrightarrow{P} \mu$.

Let $\hat{\pi}_i^{B}$ denote the Bayes estimator of $\pi(\theta_i)$, and let $\hat{\pi}_i^{EB}$ be defined by

$$\hat{\pi}_i^{EB} = \int_0^\infty \pi(\theta) \hat{g}_i^*(\theta) d\theta,$$

where with $\hat{\alpha}_i^* = \hat{\alpha} + \sum_{j=1}^{n_i} 1_{[X_{ij}<1]}$ and $\hat{\beta}_i^* = \hat{\beta} + \sum_{j=1}^{n_i} X_{ij}$, $\hat{g}_i^*$ denotes the gamma $(\hat{\alpha}_i^*, \hat{\beta}_i^*)$ density . Then for each $i \in \{1, ..., m\}$,

$$E[\hat{\mu}\hat{\pi}_i^{EB} - \mu\pi(\theta_i)]^2 - E[\mu\hat{\pi}_i^{B} - \mu\pi(\theta_i)]^2 =$$
$$= E[\hat{\mu}\hat{\pi}_i^{EB} - \mu D(X_{in_i+1})]^2 -$$
$$- E[\mu\hat{\pi}_i^{B} - \mu D(X_{in_i+1})]^2 \longrightarrow 0.$$

Proof. Let g_i^* denote the gamma density with parameters $\alpha_i^* = \alpha + \sum_{j=1}^{n_i} 1_{[X_{ij}<1]}$ and $\beta_i^* = \beta + \sum_{j=1}^{n_i} X_{ij}$. Observe that $| \hat{\pi}_i^{EB} - \hat{\pi}_i^{B} | \leq \int_0^\infty | \hat{g}_i^*(\theta) - g_i^*(\theta) | d\theta$ since $\pi(\theta_i) \leq 1$.

Let $a = \alpha/2$, $b = \beta/2$, and $K = a + b + \sup_{i\geq 1} n_i$. Then $a < \alpha_i^* < a + K$ and $b < \beta_i^* < b + K$ for every $i \in \{1, ..., m\}$. Note that if $a_m \longrightarrow \alpha$ and $b_m \longrightarrow \beta$ then there is M such that for all

$m > M$ we have $a \le a_m + \sum_{j=1}^{n_i} 1_{[X_{ij}<1]} \le a + K$ and $b \le b_m + \sum_{j=1}^{n_i} X_{ij} \le b + K$. Use the fact (see Billingsley, 1986, Theorem 20.5) that $Z_n \xrightarrow{P} Z$ if and only if every subsequence of Z_n has a further subsequence that converges to Z with probability 1. Then by Lemma 2 it follows that $|\hat{\pi}_i^{EB} - \hat{\pi}_i^B| \xrightarrow{P} 0$. Since $\hat{\pi}_i^{EB}$ and $\hat{\pi}_i^B$ are both in $(0,1]$, by the Bounded Convergence Theorem it follows that:

$$E(\hat{\pi}_i^{EB} - \hat{\pi}_i^B)^2 \longrightarrow 0. \tag{10}$$

Since $\pi(\theta_i)$ and $D(X_{in_i+1})$ are both in $(0,1]$, by Lemma 2.1 of Singh (1979):

$$E[\hat{\mu}\hat{\pi}_i^{EB} - \mu\pi(\theta_i)]^2 - E[\mu\hat{\pi}_i^B - \mu\pi(\theta_i)]^2 =$$
$$= E[\hat{\mu}\hat{\pi}_i^{EB} - \mu D(X_{in_i+1})]^2 - E[\mu\hat{\pi}_i^B - \mu D(X_{in_i+1})]^2 =$$
$$= E(\hat{\mu}\hat{\pi}_i^{EB} - \mu\hat{\pi}_i^B)^2.$$

Let $\|\ \|$ denote the L_2 norm. By the triangle inequality and the fact that $0 < \hat{\pi}_i^{EB} < 1$,

$$\|\hat{\mu}\hat{\pi}_i^{EB} - \mu\hat{\pi}_i^B\| \le \|\hat{\mu} - \mu\| + \mu\|\hat{\pi}_i^{EB} - \hat{\pi}_i^B\| \longrightarrow 0$$

by the assumed property of $\hat{\mu}$ and equation (10) and the Bounded Convergence Theorem.

3. Estimation of the structural parameters

For $t > 0$, let

$$\bar{\delta}(t) = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} 1_{[X_{ij}\ge t]}. \tag{11}$$

Let $t_1 = 1$, and $t_2 = 0.5$, and for $k = 1,2$ let

$$\bar{U}_k = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} (X_{ij}1_{[X_{ij}<t_k]} + t_k 1_{[X_{ij}\ge t_k]}). \tag{12}$$

With $\varsigma = \bar{\delta}(1)\bar{U}_2 - 0.5\bar{\delta}(0.5)\bar{U}_1$, and $\xi = (1-\bar{\delta}(1))\bar{U}_2 - (1-\bar{\delta}(0.5)\bar{U}_1$, let

$$\tilde{\beta} = \begin{cases} \dfrac{\varsigma}{\xi} & \text{if } \xi \ne 0 \\ N_2 & \text{otherwise} \end{cases} \tag{13}$$

Let

$$\tilde{\alpha} = \bar{U}_1^{-1}[\tilde{\beta}(1-\bar{\delta}(1) - \bar{\delta}1)] + 1, \tag{14}$$

and

$$\tilde{\mu} = (1-\bar{\delta})^{-1} m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} B_{ij}. \tag{15}$$

The following proposition gives estimators of the structural parameters α, and β, and μ that satisfy the conditions of Theorem 3 under the assumption that $\alpha > 1$.

Proposition 1. Suppose $\alpha \in (1, N_1]$, and $\beta \in (0, N_2]$, and $\mu \in (0, N_3]$. Let $\tilde{\alpha}$ and $\tilde{\beta}$, and $\tilde{\mu}$ be as defined in equations (14), (13) and (15), respectively. Let

$$\hat{\alpha} = \tilde{\alpha}1_{[0<\tilde{\alpha}\le N_1]} + N_1 1_{[\tilde{\alpha}>N_1]},$$
$$\hat{\beta} = \tilde{\beta}1_{[0<\tilde{\beta}\le N_2]} + N_2 1_{[\tilde{\beta}>N_2]},$$

and

$$\hat{\mu} = \tilde{\mu}1_{[\tilde{\mu}\le N_3]} + N_3 1_{[\tilde{\mu}>N_3]}.$$

Then $\hat{\alpha}$ and $\hat{\beta}$, and $\hat{\mu}$ satisfy the conditions of Theorem 3.

Proof. Since $n_i^{-1}\sum_{j=1}^{n_i}[X_{ij} \ge 1] \le 1$, we have:

$$Var[\bar{\delta}(t)] \le m^{-2}\sum_{j=1}^{m} 1 = m^{-1} \longrightarrow 0.$$

Similarly,

$$Var(\bar{U}_k) \le m^{-1}[\max(t_k,1)]^2 \longrightarrow 0.$$

Therefore, by Chebychev's inequality

$$\bar{\delta}(t_k) \xrightarrow{P} E[\bar{\delta}(t_k] = P[X_{ij} > t_k] = [(\beta + t_k)^{-1}\beta]^{\alpha}$$

and

$$\bar{U}_k \xrightarrow{P} E(\bar{U}_k) = (\alpha - 1)[\beta - (\beta + t_k)(E(\bar{\delta}(t_k))].$$

Using the fact that convergence in probability is preserved under continuous transformations, it is easily seen that $\tilde{\beta} \xrightarrow{P} \beta$ and $\tilde{\alpha} \xrightarrow{P} \alpha$.

Hence, $\hat{\alpha} \xrightarrow{P} \alpha$ and $\hat{\beta} \xrightarrow{P} \beta$.

Let $q = P[X_{ij} < 1]$. Observe that by the moment inequality $(n_i^{-1}\sum_{j=1}^{n_i} B_{ij})^2 \le n_i^{-1}\sum_{j=1}^{n_i} B_{ij}^2$. Therefore,

$$Var(n_i^{-1}\sum_{j=1}^{n_i} B_{ij}) \le E(n_i^{-1}\sum_{j=1}^{n_i} B_{ij})^2 \le E(n_i^{-1}\sum_{j=1}^{n_i} B_{ij}^2) =$$
$$= n_i^{-1}\sum_{j=1}^{n_i} EE[B_{ij}^2 \mid 1_{[X_{ij}<1]}] = q(\sigma^2 + \mu^2).$$

Therefore,

$$Var(\bar{B}) \le m^{-1}q(\sigma^2 + \mu^2) \longrightarrow 0.$$

Since,

$$E(\overline{B}) = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} E[E(B_{ij} \mid 1_{[X_{ij}=1]}] = q\mu$$

by Chebychev's inequality it follows that $\overline{B} \xrightarrow{P} q\mu$. Hence, $(1-\overline{\delta}(1))^{-1}\overline{B} \xrightarrow{P} \mu$.

Concluding remarks

It is easy to check that $\pi(\theta)$ is a continuous function of θ. Also the distribution of the censored random variables given in equation (5) satisfies an identifiability assumption (see Mashayekhi, 2002) that makes it possible to obtain asymptotically optimal non-parametric empirical Bayes estimators of continuous functions of θ under the assumption that the range of θ_i's is compact. We considered a parametric empirical Bayes approach in this paper because the parametric approach is more attractive to many statisticians (see Morris, 1983) since it is considered to be closer to Bayesian. Also the parametric approach produces estimates that are much easier to compute.

Acknowledgment

I would like to thank the anonymous referees and the editorial board for helpful comments which led to a number of improvements in the presentation of this paper.

References

1. Billingsley, P. (1986). *Probability and Measure*, Wiley, New York.
2. Bühlmann, H. (1967). Experience rating and credibility, *The ASTIN Bulletin*, 4, pp. 119-207.
3. Bühlmann, H., and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*, Springer.
4. Diaconis, P., and Ylvisaker, D. (1979). Conjugate priors for exponential families, *Ann. Statist.*, 7, pp. 269-281.
5. Efron, B., and Morris, C. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach, *Journal of the American Statistical Association*, 68, pp. 117-130.
6. Efron, B., and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations, *Journal of the American Statistical Association*, 70, pp. 311-319.
7. Fabian, Vaclav and Hannan, James (1985). Introduction to Probability and Mathematical Statistics, John Willey & Sons.
8. Jewell, W.S. (1974). Credible means are exact for exponential families, *The ASTIN Bulletin*, 8, pp. 77-90.
9. Jewell, W.S. (1975). Regularity conditions for exact credibility. *The ASTIN Bulletin*, 8, pp. 336-341.
10. Klugman, S., Panjer, H., and Willmot G. (1998). Loss Models from Data to Decisions, John Willey & Sons.
11. Mashayekhi, Mostafa (2002). Compound estimation of a monotone sequence, *Statistics and Probability Letters*, 60, pp. 7-15
12. Morris, C. (1983). Parametric empirical Bayes inference: theory and applications (with discussion), *Journal of the American Statistical Association*, 78, pp. 47-65.
13. Robbins, Herbert (1955). An empirical Bayes approach to statistics, *Proc. Thrd Berkely Symp. Math. Statist. Prob.*, 1, pp. 157-164.
14. Robbins, Herbert (1964). The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.*, 35, pp. 1-20.
15. Robbins, Herbert (1983). Some thoughts on empirical Bayes estimation, *Ann. Statist.*, 11, pp. 713-723.
16. Singh, R.S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate, *Ann. Statist.*, 7, pp. 890-902.

Appendix

Proof of Lemma 1.

$$\int_0^\infty | y^{\hat{a}_m-1}e^{-\hat{b}_m y} - y^{a_m-1}e^{-b_m y} | \, dy = \int_0^\infty | y^{\hat{a}_m-a_m}e^{-(\hat{b}_m-b_m)y} - 1 | \, y^{a_m-1}e^{-b_m y} dy . \qquad (16)$$

Clearly the integrand on the right hand side of equation (16) converges to zero as $m \longrightarrow \infty$.

Let $y>0$. Observe that $u(x) = y^x$ has negative derivative and hence is decreasing when $0<y<1$, and has non-negative derivative and hence non-decreasing when $y>1$. Thus,

$$| y^{\hat{a}_m-1}e^{-\hat{b}_m y} - y^{a_m-1}e^{-b_m y} | \le y^{\hat{a}_m-1}e^{-\hat{b}_m y} + y^{a_m-1}e^{-b_m y} \le 2y^{a-1}e^{-by}1_{[y<1]} + 2y^{a+K-1}e^{-by}1_{[y\ge 1]} .$$

Hence, the assertion of the lemma follows by the Lebesgue Dominated Convergence Theorem.

Proof of Lemma 2. Let $r_m = [\Gamma(a_m)]^{-1}\Gamma(\hat{a}_m)\hat{b}_m^{-\hat{a}_m} b_m^{a_m}$. We first show that $r_m \longrightarrow 1$.

Since $\Gamma(a_m) \geq \int_0^1 y^{a_m-1}e^{-y}dy \geq e^{-1}\int_0^1 y^{a_m-1}dy = e^{-1}a_m^{-1} \geq e^{-1}(a+K)$ and by Lemma 1 $\Gamma(\hat{a}_m)-\Gamma(a_m)\longrightarrow 0$ we have $|[\Gamma(a_m)]^{-1}\Gamma(\hat{a}_m)-1| \leq e(a+K)^{-1}\,|\Gamma(\hat{a}_m)-\Gamma(a_m)|\longrightarrow 0$.

Since $0<b<\hat{b}_m<b+K$ we have $\hat{b}_m^{-1}b_m \longrightarrow 1$, and hence $\log(\hat{b}_m^{-\hat{a}_m}b_m^{a_m}) = (a_m-\hat{a}_m)\log\hat{b}_m +$ $+\,a_m\log(b_m/\hat{b}_m)\longrightarrow 0$. Therefore $\hat{b}_m^{-\hat{a}_m}b_m^{a_m}\longrightarrow 1$. Hence, $r_m \longrightarrow 1$.

Since $\hat{a}_m$ and $\hat{b}_m$ are bounded there is $0<M<\infty$ such that $[\Gamma(\hat{a}_m)]^{-1}\hat{b}_m^{\hat{a}_m}<M$. By the triangle inequality it follows that

$$\int_0^\infty |\,\hat{g}_m(x)-g_m(x)\,|\,dx \leq M\int_0^\infty |\,y^{\hat{a}_m-1}e^{-\hat{b}_m y}-y^{a_m-1}e^{-b_m y}\,|\,dy + M\,|1-r_m|\int_0^\infty y^{a_m-1}e^{-b_m y}dy. \quad (17)$$

The first term on the right hand side of equation (17) converges to zero by Lemma 1 and the second term converges to zero because $y^{a_m-1}e^{-b_m y} \leq (y^{a-1}1_{[y<1]}+y^{a+K-1}1_{[y\geq 1]})e^{-by}$.