ACTUARIAL RESEARCH CLEARING HOUSE
1993 VOL. 1

A Mixed Lognormal Estimator of a Risk Distribution

JACQUES CARRIERE

Department of Actuarial & Management Sciences

The University of Manitoba

Abstract: Consider a portfolio of insurance policies where the mean frequency of claims for each policy may vary. This heterogeneity in the portfolio may be modeled with a risk distribution function $F(\lambda)$ that mixes the mean frequency λ. Using the observed claim frequencies of this portfolio, we present a continuous nonparametric estimator of the risk distribution $F(\lambda)$ that reproduces some of the empirical moments and converges uniformly. The estimator that we investigate is a mixture of lognormal distributions whose parameters are calculated by considering the determinants of certain moment matrices.

Keywords: Risk distribution, moment matrices, mixtures of distributions, uniform consistency.

Abbreviated Title: Lognormal Estimator

1. Introduction

Suppose that the number of claims N for a policy can be modeled with the Poisson probability density function (pdf)

$$p(n|\lambda) = \frac{e^{-\lambda}\ \lambda^n}{n!} \tag{1.1}$$

where $\lambda > 0$ is the mean frequency and $n = 0,1,2,\ldots$ In a heterogeneous population the mean frequency is distributed according to some unknown distribution $F(\lambda)$. We will assume throughout the discussion that the risk distribution $F(\lambda)$ is continuous and that $F(0)=0$. Moreover, we will assume that the risk distribution is uniquely determined by its moments.

Suppose we observe the frequencies N_i for $i=1,\ldots,T$ where T is the number of policies in some insurance portfolio and $N_1, N_2, \ldots$ are independent and identically distributed random variables with a common pdf equal to

$$p(n) = \int_{(0,\infty)} p(n|\lambda)\ dF(\lambda) \tag{1.2}$$

where $p(n|\lambda)$ is given in (1.1). The moments $m_k = E(\lambda^k)$ for $k=1,2,\ldots$ of $F(\lambda)$ can be estimated with the empirical moments

$$\hat{m}_k = \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1)\ \hat{p}(n) \tag{1.3}$$

where

$$\hat{p}(n) = \frac{1}{T} \sum_{i=1}^{T} \mathrm{I}(N_i = n) \tag{1.4}$$

is an empirical pdf. Note that $\hat{m}_k$ is an unbiased and almost surely consistent estimator of m_k. The nonparametric estimator of $F(\lambda)$ that we will investigate will have the form

$$\tilde{F}(\lambda) = \sum_{j=1}^{\rho} \pi_j \; L_j(\lambda) \tag{1.5}$$

where

$$L_j(\lambda) = \int_0^\lambda (2\pi)^{-1/2} \, (y\sigma)^{-1} \exp\Bigl(-(\log_e(y/\beta_j))^2 \Big/ 2\sigma^2\Bigr) \, dy. \tag{1.6}$$

Note that we are not suggesting that the underlying true distribution is a mixture but we are simply using a mixture as an estimator. Later we will show that the estimator $\tilde{F}(\lambda)$ reproduces the empirical moments $\hat{m}_k$ for $k=1,\ldots,2\rho-1$. We will also show that decreasing σ makes $\tilde{F}(\lambda)$ look like a step function and so σ may be viewed as a smoothing parameter.

Let us review the current literature on estimators for $F(\lambda)$. Hossack, Pollard and Zehnwirth (1983) gave an asymptotically consistent estimator of $F(\lambda)$ under the assumption that it belongs to a Gamma class of distributions. Willmot (1987) also gave a consistent estimator when $F(\lambda)$ belongs to an Inverse-Gaussian class of distributions. Obviously, these estimators will be asymptotically biased if the true distribution is not in these parametric classes. Lindsay (1989) constructed a discrete estimator that reproduces some of the empirical moments and is consistent when $F(\lambda)$ is uniquely determined by its moments and $F(0)=0$. Our continuous estimator $\tilde{F}(\lambda)$ will be a generalization of Lindsay's result.

2. A Moment Problem

Consider the empirical moments $\hat{m}=\{\hat{m}_k : k=0,1,\ldots\}$ where $\hat{m}_k$ is defined in (1.3). Assume T is large enough so that $\hat{p}(0)\neq 1$. Note that there always exists $n_0\geq 1$ such that $\hat{p}(n)=0$ for all $n\geq n_0$. This means that $\hat{m}_k=0$ for all $n\geq n_0$ and so there does not exist a distribution function $\hat{F}(\lambda)$ such that $\hat{F}(0)=0$ and $\hat{m}_k=\int_0^\infty \lambda^k d\hat{F}(\lambda)$ for all $k=0,1,\ldots$ But Lindsay (1989) proves that there exists a discrete distribution $\hat{F}(\lambda)$ with $\rho\geq 1$ atoms such that $\hat{F}(0)=0$ and $\hat{m}_k=\int_0^\infty \lambda^k d\hat{F}(\lambda)$ for all $k=0,1,\ldots 2\rho-1$. In this section we will present various definitions and we will summarize Lindsay's result. Using this foundation, we will then present our estimator and discuss its asymptotic properties.

Using the sequence $\hat{m}$, define $M_0=\{1\}$, $M_0^s(\hat{m})=\{\hat{m}_1\}$ and for $k=1,2,\ldots$ define the moment matrix as

$$M_k(\hat{m}) = \begin{bmatrix} 1 & \hat{m}_1 & \cdots & \hat{m}_k \\ \hat{m}_1 & \hat{m}_2 & \cdots & \hat{m}_{k+1} \\ \vdots & \vdots & & \vdots \\ \hat{m}_k & \hat{m}_{k+1} & \cdots & \hat{m}_{2k} \end{bmatrix} \tag{2.1}$$

and the shifted moment matrix as

$$M_k^s(\hat{m}) = \begin{bmatrix} \hat{m}_1 & \hat{m}_2 & \cdots & \hat{m}_{k+1} \\ \hat{m}_2 & \hat{m}_3 & \cdots & \hat{m}_{k+2} \\ \vdots & \vdots & & \vdots \\ \hat{m}_{k+1} & \hat{m}_{k+2} & \cdots & \hat{m}_{2k+1} \end{bmatrix} . \tag{2.2}$$

Using $M_k(\hat{\underline{m}})$ and $M_k^*(\hat{\underline{m}})$ for $k=0,1,\ldots$ we define $\rho(\hat{\underline{m}})$ as follows

$$\rho(\hat{\underline{m}}) = 1 + \sup\Big\{k : \det(M_i(\hat{\underline{m}}))>0 \text{ and } \det(M_i^*(\hat{\underline{m}}))>0 \ \forall i=0,\ldots,k\Big\}. \tag{2.3}$$

For the ensuing discussion we will assume that $\rho \le \rho(\hat{\underline{m}})$. Now consider the following polynomial of degree ρ,

$$P(t\,;\,\hat{\underline{m}}) = \det\begin{bmatrix} 1 & \hat{m}_1 & \cdots & \hat{m}_{\rho-1} & 1 \\ \hat{m}_1 & \hat{m}_2 & \cdots & \hat{m}_\rho & t \\ \vdots & \vdots & & \vdots & \vdots \\ \hat{m}_\rho & \hat{m}_{\rho+1} & \cdots & \hat{m}_{2\rho-1} & t^\rho \end{bmatrix}. \tag{2.4}$$

Let $r_j(\hat{\underline{m}})$ for $j=1,\ldots,\rho$ denote the roots of $P(t\,;\,\hat{\underline{m}})$. Next calculate $\underline{\tau}(\hat{\underline{m}})=(\tau_1,\ldots,\tau_\rho)^T$ as follows

$$\begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_\rho \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_\rho \\ \vdots & \vdots & & \vdots \\ r_1^{\rho-1} & r_2^{\rho-1} & \cdots & r_\rho^{\rho-1} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \hat{m}_1 \\ \vdots \\ \hat{m}_{\rho-1} \end{bmatrix} \tag{2.5}$$

whenever the inverse matrix exists. We now give Lindsay's result.

Lemma 1. If $\rho \le \rho(\hat{\underline{m}})$ then there exists a unique distribution $\hat{F}(\lambda)$ with ρ distinct atoms of mass $\pi_j(\hat{\underline{m}})>0$ at $r_j(\hat{\underline{m}})>0$ for $j=1,\ldots,\rho$ whose moments are equal to $\hat{m}_k$ for $k=1,\ldots,2\rho-1$.□

Lindsay's result is related to the Stieltjes moment problem, that is discussed in Shohat and Tamarkin (1943). The proof for the following lemma is based on the results presented there.

Lemma 2. a) A distribution $F(\lambda)$ with a moment sequence $\underline{m}=\{m_k : k=0,1\ldots\}$ has ρ distinct atoms of mass $\pi_j>0$ at $r_j>0$ for $j=1,\ldots,\rho$ if and only if $\det(M_i(\underline{m}))>0$, $\det(M_i^*(\underline{m}))>0 \;\forall i=0,\ldots,\rho-1$ and $\det(M_i(\underline{m}))=0$, $\det(M_i^*(\underline{m}))=0 \;\forall i=\rho,\rho+1,\ldots$

b) If $F(\lambda)$ is continuous and $F(0)=0$ then $\det(M_i(\underline{m}))>0$, $\det(M_i^*(\underline{m}))>0 \;\forall i=0,1,\ldots$□

3. The Mixed Lognormal Estimator

In this section, we will show how to calculate the parameters σ, ρ, $\pi_1,\ldots,\pi_\rho$ and $\beta_1,\ldots,\beta_\rho$ for our mixed lognormal estimator. Define $\underline{m}^*=\{m_k^* : k=0,1,\ldots\}$ where

$$m_k^* = \hat{m}_k \times \exp\Big\{-k^2\sigma^2/2\Big\}. \tag{3.1}$$

Note that $m_k^* \to \hat{m}_k$, $\det(M_i(\underline{m}^*))\to\det(M_i(\hat{\underline{m}}))$ and $\det(M_i^*(\underline{m}^*))\to\det(M_i^*(\hat{\underline{m}}))$ as $\sigma\to 0$. Therefore, if $\rho\le\rho(\hat{\underline{m}})$ then there exists $\sigma_0>0$ such that $\rho\le\rho(\underline{m}^*)$ for all $\sigma\le\sigma_0$. In the ensuing discussion we will always assume that $\sigma\le\sigma_0$.

The parameters $\beta_j = r_j(\underline{m}^*) > 0$ for $j=1,\ldots,\rho$ are equal to the distinct positive roots of the polynomial $P(t\ ;\ \underline{m}^*)$. Finally, the parameters $0 < \pi_j < 1$ for $j=1,\ldots,\rho$ are equal to $\underline{\pi} = \underline{\tau}(\underline{m}^*)$. Let us prove that calculating the parameters in this manner yields an estimator $\tilde{F}(\lambda)$ that reproduces the empirical moments. Define $\tilde{\underline{m}} = \{\tilde{m}_k : k=0,1,\ldots\}$ where

$$\tilde{m}_k = \int_0^\infty \lambda^k d\tilde{F}(\lambda)\ . \tag{3.2}$$

According to Lemma 1, the distribution with mass of π_j at the points β_j has moments equal to m_k^* for $k=1\ldots,2\rho-1$. Therefore $\tilde{m}_k = \sum_{j=1}^{\rho} \pi_j \beta_j^k \exp\{k^2\sigma^2/2\} = \exp\{k^2\sigma^2/2\} \times m_k^* = \hat{m}_k$.

Now, let us show that σ is a smoothing parameter by proving that $\tilde{F}(\lambda) \xrightarrow{d} \hat{F}(\lambda)$ as $\sigma \to 0$. First, note that $P(t;\ \underline{m}^*) \to P(t;\ \hat{\underline{m}})$ as $\sigma \to 0$. This implies that $\beta_j = r_j(\underline{m}^*) \to r_j(\hat{\underline{m}})$ as $\sigma \to 0$ because $r_j(\hat{\underline{m}})$ is a root of the continuous polynomial $P(t;\ \hat{\underline{m}})$. In turn this implies that $\underline{\pi} = \underline{\tau}(\underline{m}^*) \to \underline{\tau}(\hat{\underline{m}})$ as $\sigma \to 0$ because $\underline{\tau}(\hat{\underline{m}})$ is a continuous function in $\hat{\underline{m}}$. Therefore $\tilde{m}_k = \sum_{j=1}^{\rho} \pi_j \beta_j^k \exp\{k^2\sigma^2/2\} \to \sum_{j=1}^{\rho} \tau_j(\hat{\underline{m}}) \Big(r_j(\hat{\underline{m}}) \Big)^k$ as $\sigma \to 0$ for $k=1,2,\ldots$ In other words, all the moments of $\tilde{F}(\lambda)$ converge to the moments $\hat{F}(\lambda)$. According to a theorem by Frechet and Shohat that is given in Serfling (1980), p. 17, this is a sufficient condition for convergence in distribution. We now summarize our results.

<u>Theorem 3a</u>) If $k=1,\ldots 2\rho-1$ then $\tilde{m}_k = \hat{m}_k$ and
b) if $\sigma \to 0$ then $\tilde{F}(\lambda) \xrightarrow{d} \hat{F}(\lambda)$. □

Let us apply the result to some motor vehicle data given in Johnson and Hey (1971). In this data we find that T=421,240 and that

$$\hat{p}(0) = .879337 \qquad \hat{p}(1) = .110495$$
$$\hat{p}(2) = .009341 \qquad \hat{p}(3) = .000753 \tag{3.3}$$
$$\hat{p}(4) = .000066 \qquad \hat{p}(5) = .000007.$$

Using (1.3) we find that

$$\hat{m}_1 = .131735 \qquad \hat{m}_2 = .024132$$
$$\hat{m}_3 = .006522 \qquad \hat{m}_4 = .002424 \tag{3.4}$$
$$\hat{m}_5 = .000840 \qquad \hat{m}_6 = .000000.$$

Using this data we constructed two graphs. Figure 1 shows a plot of the pdf of the mixed lognormal estimator and the pdf of the inverse-gaussian estimator given in Willmot (1987). The smoothing parameter for the mixed lognormal estimator was σ=.45 and we found that increasing it slightly yielded an estimator that was almost identical to the inverse-gaussian estimator. Figure 2 compares the pdf of the mixed lognormal estimator when the smoothing parameter σ is equal to .15 and .45. The parameter values for the plot with σ=.15 were $\rho=2$, β_1=.3704, β_2=.1048 and π_1=.0958, π_2=.9042. These graphs and all the necessary calculations were made with the statistical computing language called GAUSS.

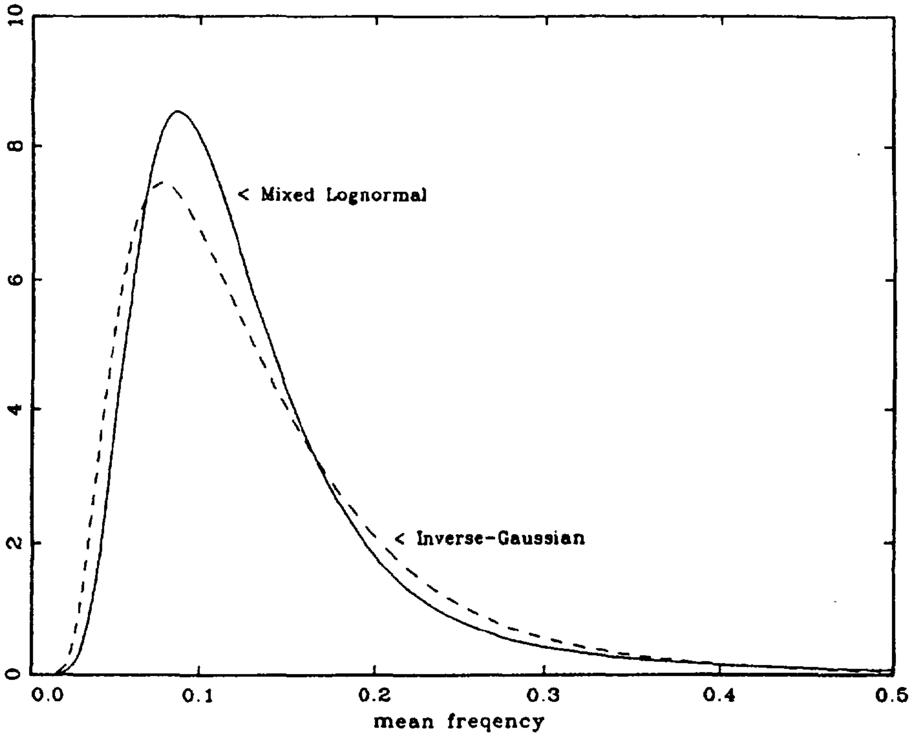

Figure 1. A comparison of the mixed lognormal and inverse-gaussian estimators.

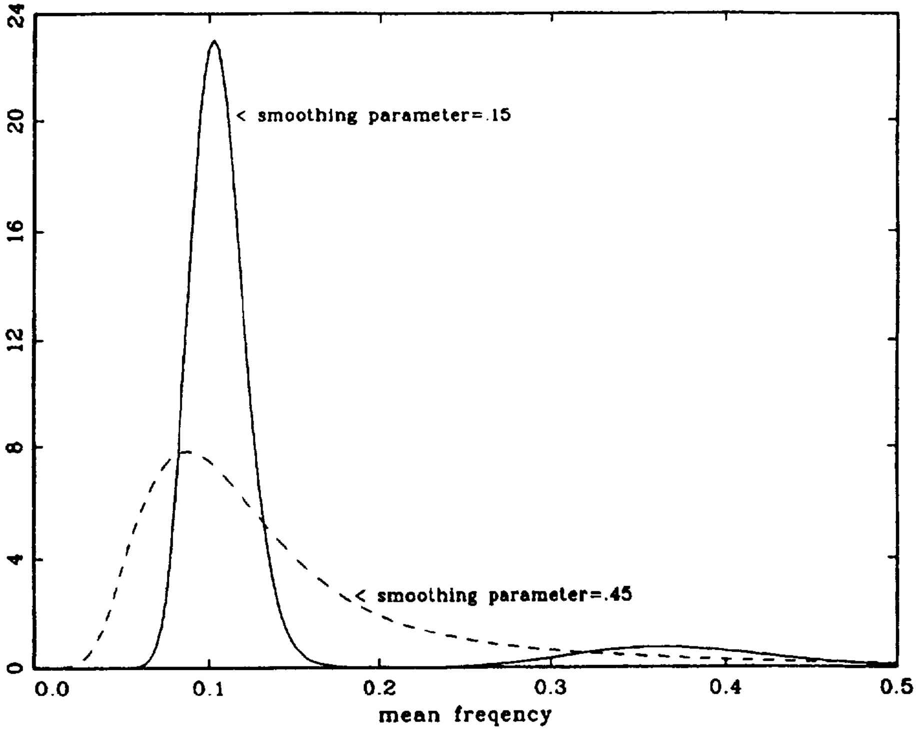

Figure 2. A comparison of the mixed lognormal estimator when $\sigma = .15$ and $\sigma = .45$.

4. Asymptotic Results

Using Kolmogorov's strong law of large numbers we know that $\hat{m}_k \xrightarrow{a.s.} m_k$ as $T \to \infty$ for $k=1,2,\ldots$ Therefore $\det(M_k(\hat{\underset{\sim}{m}})) \xrightarrow{a.s.} \det(M_k(\underset{\sim}{m}))$ and $\det(M_k^1(\hat{\underset{\sim}{m}})) \xrightarrow{a.s.} \det(M_k^1(\underset{\sim}{m}))$ as $T \to \infty$ because the determinants are continuous functions of the moments. Lemma 2 states that if $F(0)=0$ and $F(\lambda)$ is continuous then $\det(M_k(\underset{\sim}{m}))>0$ and $\det(M_k^1(\underset{\sim}{m}))>0$ for all $k=0,1,\ldots$ Therefore, $\rho(\hat{\underset{\sim}{m}}) \xrightarrow{a.s.} \rho(\underset{\sim}{m})=\infty$ as $T \to \infty$. We summarize the result as follows.

<u>Lemma 4.</u> If $T \to \infty$ then $\rho(\hat{\underset{\sim}{m}}) \xrightarrow{a.s.} \infty$. □

We will now give some asymptotic results for our mixed lognormal estimator $\bar{F}(\lambda)$. To prove these results we will use certain approximation theorems found in Serfling (1980). We will assume throughout that $\rho=\rho(\hat{\underset{\sim}{m}})$.

<u>Theorem 5.</u> a) If $k=1,2,\ldots$ and $T \to \infty$ then $\bar{m}_k \xrightarrow{a.s.} m_k$.

b) Let $g(\lambda)$ be a bounded and continuous function for all $\lambda>0$,

then $\int_0^\infty g(\lambda)\ d\bar{F}(\lambda) \xrightarrow{a.s.} \int_0^\infty g(\lambda)\ dF(\lambda)$ as $T \to \infty$.

c) If $T \to \infty$, then $\sup_{\lambda>0} |\bar{F}(\lambda)-F(\lambda)| \xrightarrow{a.s.} 0$.

d) Let $r>0$. then $E|\bar{F}(\lambda)-F(\lambda)|^r \to 0$ and $E(\bar{F}(\lambda))^r \to (F(\lambda))^r$ as $T \to \infty$.

Proof: a) Using Lemma 4 we find that for a fixed $k=1,2,\ldots$ there exists T_k such that for all

$T > T_k$ we must have $k < 2\rho - 1$. Therfore, by Theorem 3 $\tilde{m}_k = \hat{m}_k$. Finally, by

Kolmogorov's strong law of large numbers $\tilde{m}_k \xrightarrow{a.s.} m_k$ as $T \to \infty$.

b) This result follows immediately after applying a theorem by Frechet and Shohat

that is given in Serfling (1980), p. 17.

c) This follows immediately after applying Polya's theorem in Serfling (1980), p. 18.

d) This follows immediately from standard theorems in Serfling (1980). pp. 11-15. □

References

Hossack. I.B., Pollard, J.H. and Zehnwirth, B. (1983). *Introductory statistics with applications in general insurance*. Cambridge University Press, New York.

Johnson, P.D. and Hey, G.B. (1971). Statistical Studies in Motor Insurance. *Journal of the Institute of Actuaries*, Vol. 97. The Alden Press, Oxford, Great Britain.

Lindsay, B.G. (1989). Moment Matrices: Applications in Mixtures. *The Annals of Statistics*, Vol. 17, No. 2, pp. 722-740.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shohat, J.A. and Tamarkin, J.D. (1943). *The Problem of Moments. Mathematical Surveys, Number* 1. American Mathematical Society, New York.

Willmot G.E. (1987). The Poisson-Inverse Gaussian Distribution as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, pp. 113-127.