Bayes Statistics for Mixed Erlang-models

Erhard Kremer

Department Mathematik

Bundesstraße 55, 20144 Hamburg, Germany

E-mail: Kremer@math.uni-hamburg.de

Abstract

Bayesian techniques are usually no standard techniques of Mathematical Statistics. Nevertheless they are applied in different fields of research and practice, e.g. in insurance premium rating. Many theoretical results were derived about Bayesian-statistical methods. Applied to more special model assumptions they give often handy techniques for application. The present paper gives under spezialized conditions such theory-based techniques, that have not been given like that in the literature up to now.

Keywords: Bayes Statistics, Mixed Erlang Model, Estimator, Test, Parameter-Estimators

1. Introduction

Bayesian ideas are already comparably old. Already in the 18th century Thomas Bayes and Pierre-Simon Laplace developed first conclusions on the Bayes-calculus. Nowadays Bayes-theory or Bayes-Statistics is a highly developed part of Mathematical Statistics. The significant difference to usual Mathematical Statistics is the assumption that the parameter of the statistical model is a realization of a random variable. The distribution of that random variable is called a-priori-distribution. Usually one interprets the a-priori-distribution as something like a previous information on the value of the parameter of the model. But note that sometimes one has a more concrete explanation for such an a-priori-distribution. For example in the actuarial field it is the concretely given distribution of the socalled risk parameter in a collective of insurance risks. The classical approach of Bayes-Statistics consists in calculating a socalled a-posteriori-distribution out of the a-priori-distribution by integrating the data of a statistical sample. With that a-posteriori distribution one derives optimal statistical decision rules then. These decision rules are called Bayes-rules.

A lot of research on the Bayes-rules was done already. For surveys on most important things see the books of Berger (1985), Gosh et.al. (2003) and Robert (1994). An elegant introduction (in German) was given by the author (see Kremer (2005a)).

Recently the author noted a certain gap in the field of Bayes research. The closing of that gap is contents of the following paper, that is quite elegant and closed in its results.

2. The context

Let be given the sample $X = (X_1, ..., X_n)$ as random variables:

$$X_i : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B})$$

and the random-variable θ (on also $(\Omega, \mathcal{A}, P)$ with values in a set Θ), describing as realisation ϑ the parameter of the underlying stochastic model. Denote with

$$P_\vartheta^X := P^{X|\theta=\vartheta}$$

the conditional probability (measure) of the X given the realisation ϑ of θ. With this notation the probability law P^θ is the **a-priori-distribution** (of θ) and the contitional distribution of θ given $X = x$, the socalled **a-posteriori-distribution** (of θ), just in symbol: $P^{\theta|X=x}$. The last one can be computed (usually) with the socalled **Bayes-Theorem** (see on this e.g. Kremer (2005a), Theorem 2.1, on page 30 there).

For all that follows assume that the $X_1, ..., X_n$ are **i.i.d. given θ**, what means in details:

(A) $P^{X|\theta=\vartheta} = \bigotimes_{i=1}^n P^{X_i|\theta=\vartheta}$, $\forall \vartheta$

(B) $P^{X_i|\theta=\vartheta} = P^{X_j|\theta=\vartheta}$, $\forall \vartheta,\ \forall i \neq j$.

Usually one assumes in addition that the $P^{X_i|\theta=\vartheta}$ $(\vartheta \in \theta)$ are dominated by a σ-finite measure μ, giving the μ-density $f_\vartheta^{X_i}$ of the conditional distribution of X_i given ϑ, with symbol:

$$P_\vartheta^{X_i} := P^{X_i|\theta=\vartheta} .$$

So far for the general notations. The new parts of the paper are specialized to the additional model assumptions:

(C) θ has an **Erlang(a, b)-distribution** with (given) fixed $a \in \mathbb{N}$ and parameter $b \in (0, \infty)$, meaning P^{θ} has the Lebesgue-density:

$$f_b(\vartheta) = \frac{b^a}{\Gamma(a)} \cdot \vartheta^{a-1} \cdot \exp(-b\vartheta)$$

on $\vartheta \in (0, \infty)$.

(D) $P_{\vartheta}^{X_i}$ is an **Erlang(α, ϑ)-distribution** with (given) fixed $\alpha \in \mathbb{N}$ and parameter $\vartheta \in (0, \infty)$, meaning that one has (with μ as Lebesgue-measure):

$$f_{\vartheta}^{X_i}(y) = \frac{\vartheta^{\alpha}}{\Gamma(\alpha)} \cdot y^{\alpha-1} \cdot \exp(-\vartheta \cdot y)$$

on $y \in (0, \infty)$.

Note that for $n = 1$ the model given by (A)-(D) is related to the model in Johnson & Kotz (1970), section 8.2.

3. Basic results

For giving Bayes-rules under the model defined through (A)-(C), one can take certain general theorems in Kremer (2005a). These shall be listed up adequately in the following:

Result 1: The assumptions (A), (B) with $\vartheta \in \theta = (a_1, b_1) \subset \mathbb{R}$. Suppose that one has:

(a) P^{θ} is dominated by the Lebesgue-measure with density of type:

$$f(\vartheta) = [C(\vartheta)]^m \cdot \frac{\exp(-x_0 \cdot \vartheta)}{D(m, x_0)} ,$$

where $m \geq 0$ is fixed, $C(\cdot)$ is a positive function, $x_0 \in (b_1, b_2) \subset \mathbb{R}$ is a parameter and $D(m, x_0)$ is the norming constant such that:

$$\int_{a_1}^{a_2} f(\vartheta)\, d\vartheta = 1 .$$

(b) $P_{\vartheta}^{X_i}$ is dominated by the Lebesgue-measure and has the density:

$$f_{\vartheta}^{X_i}(y) = C(\vartheta) \cdot \exp(-\vartheta \cdot y) \cdot h(y)$$

with the function $C(\cdot)$ of (a) and where $h(\cdot)$ is an adequate, nonnegative, measurable function.

Then the a-posteriori-distribution $P^{\theta|X=x}$ has the Lebesgue density according to the formula:

$$f^{\theta|X=x}(\vartheta) = [C(\vartheta)]^{n+m} \cdot \frac{\exp\left(-\left(x_0 + \sum_{i=1}^n x_i\right)\vartheta\right)}{D\left(n + m, x_0 + \sum_{i=1}^n x_i\right)} \tag{1}$$

where $x = (x_1, \ldots, x_n)$ and $D\left(n + m, x_0 + \sum_{i=1}^n x_i\right)$ is again the norming constant like in (a). □

This result is Theorem 2.14 in Kremer (2005a). It will be applied lateron to models according assumptions (A)-(D).

For the next let the function γ on θ be defined according:

$$\begin{aligned} \gamma(\vartheta) &= E(X_i|\theta = \vartheta) \\ &= \int y\, P_{\vartheta}^{X_i}(dy), \end{aligned} \tag{2}$$

and consider the problem of estimating $\gamma(\vartheta)$, based on (the data) X.

The corresponding optimal estimator (in the context of the above Bayes-model (see all in front of (C)in part 2.)) is the so called **Bayes-estimator**. For its general definition see part 4.1 in Kremer (2005a). Here one gets more special:

Result 2:

Take the complete context of Result 1 with in addition:

$$f(b) - f(a) = 0.$$

The Bayes-estimator is given according to:

$$\hat{\gamma}(x) = \frac{x_0 + \sum_{i=1}^n x_i}{m + n}$$

with $x = (x_1, \ldots, x_n)$. □

This result is just Theorem 4.6 in Kremer (2005a). Also it will be applied lateron.

Finally let Θ be split up into two disjoint sets H and K:

$$\Theta = H + K. \tag{3}$$

A decision (based on X) shall be made between the **hypotheses**:

$$\vartheta \in H$$

and the **alternative**:

$$\vartheta \in K.$$

The Bayes-rule for this (so called) testing problem is called **Bayes-Test**. For details on it see again Kremer (2005a), part 3.1.

Result 3:

Take the assumptions (A),(B) of section 2 and assume the existence of μ-densities $f_\vartheta^{X_i}(\cdot)$ of $P_\vartheta^{X_i}$. Suppose $f_\vartheta^{X_i}(y_i)$ is measurable in (ϑ, y_i). Then one has as Bayes-Test:

$$\varphi(x) = \begin{cases} 1, & \text{when } P^{\theta|X=x}(K) > \frac{c_1}{c_1+c_2} \\ 0, & \text{otherwise,} \end{cases}$$

with $x = (x_1, \ldots, x_n)$. □

Here c_1, c_2 are certain weighting constants in the so called Bayes-risk (see Theorem 3.2 in Kremer (2005a)). They are for defining a certain loss-function (see Kremer (2005a), page 53). Most simple is to take $c_1 = c_2 = 1$.

The proof of a more generalized version of Result 3 can be found in Kremer (2005a), pages 55-57 (note, the above φ is given in the remark on page 57).

4. Bayes-estimator

Take the Bayes-context of section 2 with (A),(B) **and more special (C),(D) in addition**. One has:

Theorem 1:

The Bayes-estimator of $\gamma(\vartheta)$ (according to (2)) is given by:

$$\hat{\gamma}(x) = \alpha \cdot \left(\frac{b + \sum_{i=1}^n x_i}{a + \alpha \cdot n - 1} \right)$$

with (the sample) $x = (x_1, \ldots, x_n)$, when one has: $\alpha \neq 1$.

Remark 1:

Note that the excluded case $\alpha = 1$ is just the (classical) exponential distribution, what (in usual applications, like nonlife **insurance rating**) is of not such great importance.

But also note, that this exponential case is already done in (Kremer 2005a), example 2.11 amd example 4.5 case 4.

Remark 2:

From the properties of the Erlang(α, ϑ)-distribution one knows that:

$$\int_0^\infty y \, f_\vartheta^{X_i}(y) \, dy = \frac{\alpha}{\vartheta},$$

what means:

$$\boxed{\gamma(\vartheta) = \frac{\alpha}{\vartheta}.} \qquad \square \tag{4}$$

Proof of Theorem 1:

The $f_\vartheta^{X_i}$ of (D) is of type (b) of section 3. One has:

$$C(\vartheta) = \frac{\vartheta^\alpha}{\Gamma(\vartheta)}, \quad h(y) = y^{\alpha-1}.$$

Also f_b of (C) is of the Type (a) of section 3. Here one has:

$$x_0 = b, \qquad m = \frac{a-1}{\alpha}. \tag{5}$$

Obviously the conditions of result 1 are given with:

$$a_1 = 0, \quad a_2 = \infty, \quad b_1 = 0, \quad b_2 = \infty.$$

Since:

$$f_b(0) = 0, \quad f_b(\infty) = 0$$

all conditions of result 2 are given. Inserting there into the formula of $\hat{\gamma}(x)$ the special choices (5), one arrives at the result of theorem 1. □

Of certain interest are:

Remark 3:

From formula (1) and (4) one concludes easily that the a-posteriori-distribution $P^{\vartheta|x=x}$ with $x = (x_1, \ldots, x_n)$ is just the

$$\text{Erlang}(a + \alpha, b + \sum_{i=1}^{n} x_i),$$

what is for $n = 1$ in agreement with one result in table 3.2 in Robert (1994). □

Remark 4:

Example 4.9 in Robert (1994) is related to the above theorem. In the special situation $n = 1$ one has for the Bayes-estimator $\delta_1^{\pi}(X)$ in Robert (1994) and the $\hat{\gamma}(X)$ of Theorem 1:

$$\hat{\gamma}(X) = \alpha \cdot \delta_1^{\pi}(X).$$

The factor α can be easily explained. Robert (1994) takes the loss function:

$$L(\vartheta, \delta) = \left(\delta - \frac{1}{\vartheta}\right)^2,$$

whereas the above is based on:

$$L(\vartheta, \delta) = \left(\delta - \frac{\alpha}{\vartheta}\right)^2.$$

□

5. Bayes-Test

Suppose again that the Bayes-context of section 2 with (A), (B) **and** with in addition (C), (D) is given. Considered shall be an adequate testing problem now, more concretely the hypotheses:

$$H = [\vartheta_0, \infty)$$

against the alternative:

$$K = (0, \vartheta_0),$$

where ϑ_0 is fixed and given. That this type is most nearlying, one concludes from (4).

Note on this:

Usually most nearlying is $H = (0, \vartheta_0)$ against $K = (\vartheta_0, \infty)$. But since in (4) one has ϑ^{-1}, above H and K are the most right choice.

One has:

Theorem 2:

The Bayes-test for H against K is given by:

$$\varphi(x) = \begin{cases} 1, & \text{when } \bar{x}_n > \frac{1}{n} \cdot (k(a, \alpha, \vartheta_0) - b) \\ 0, & \text{otherwise,} \end{cases}$$

when $x = (x_1, \ldots, x_n)$ and with:

$$\bar{x}_n = (\frac{1}{n}) \cdot \sum_{i=1}^{n} x_i$$

$$k(a, \alpha, \vartheta_0) = F_{\vartheta_0}^{-1}\left(\frac{c_1}{c_1 + c_2}\right),$$

where $F_{\vartheta_0}^{-1}$ is the inverse of the (strictly monotone increasing) function:

$$F_{\vartheta_0}(z) = 1 - \exp(-\vartheta_0 \cdot z) \cdot \sum_{l=1}^{a_*-1} \frac{(\vartheta_0 \cdot z)^l}{l!},$$

with:

$$a_* = a + \alpha \ .$$

Proof:

According to Remark 2 one has:

$$P^{\vartheta|X=x}(K) = F_{\vartheta_0}(b_*)$$

with

$$b_* = b + \sum_{i=1}^{n} x_i \ ,$$

since the Erlang(a_*, b_*)-distribution has the distribution function value $F_y(b_*)$ at the point y. As a consequence the condition:

$$P^{\vartheta|X=x}(K) > \frac{c_1}{c_1 + c_2}$$

is equivalent with:

$$b_* > F_{\vartheta_0}^{-1}\left(\frac{c_1}{c_1 + c_2}\right)$$

and that again with

$$\bar{x}_n > \frac{1}{n} \cdot (k(a, \alpha, \vartheta_0) - b) \ .$$

□

Remark 5:

Note that F_{ϑ_0} is the distribution function of the Erlang(a_*, ϑ_0)-distribution. This means, that one can compute the $k(a, \alpha, \vartheta_0)$ just as γ-fractile (with $\gamma = \frac{c_2}{c_1+c_2}$) of the Erlang$(a_*, \vartheta_0)$- distribution. □

Note that the above Bayes-test of theorem 2 can not be found in Berger (1994) (not even for the special case $n = 1$!) or somewhere else.

Finally a fine:

Application:

Also in Bayes-statistics one thinks about certain confidence regions on the parameter ϑ. A quickest introduction into that area can be found again in Kremer (2005a) (see part 4.4 there). There one derives confidence regions from the Bayes-tests. Denote with $C(x)$ a confidence region for the parameter ϑ based on (the sample) $x = (x_1, ..., x_n)$. According to formula (3.7) in Kremer (2005a) it has greatest sense to take:

$$C(x) = \{\vartheta : \varphi_\vartheta(x) = 0\}$$

where φ_ϑ is the test of Theorem 2 with choice ϑ instead of ϑ_0. Obviously:

$$C(x) = \left\{\vartheta : \bar{x}_n \leq \frac{1}{n}(k(a, \alpha, \vartheta) - b)\right\}$$

what can be rewritten as:

$$C(x) = \left\{\vartheta : F_\vartheta\left(b + \sum_{i=1}^{n} x_i\right) \leq c\right\}$$

with

$$c = c_1/(c_1 + c_2)$$

(compare proof of Theorem 2).

For practical applications one needs an adequate value for c. Certainly one wants that the confidence region holds in a certain confidence level, say $(1-\delta)$ with a fixed, chosen $\delta \in (0,1)$ (small, e.g. 0.05). Consequently one has the condition for choosing c:

$$P^{\theta|X=x}(C(x)) \geq 1-\delta \ , \tag{6}$$

where one can replace "'$\geq$"' through "'$=$"'. According to remark 2 one has $P^{\theta|X=x}$. It is Erlang $(a+\alpha, b+\sum_{i=1}^{n} x_i)$. But first rewrite $C(x)$ into a more nice form.

Take the notation:

$$G_{b_*}(\vartheta) := F_\vartheta\left(b+\sum_{i=1}^{n} x_i\right)$$

with

$$b_* = b+\sum_{i=1}^{n} x_i \ .$$

Since also $G_{b_*}(\cdot)$ is strictly increasing, also its inverse $G_{b_*}^{-1}$ exists what gives:

$$C(x) = \{\vartheta : \vartheta \leq G_{b_*}^{-1}(c)\} \ .$$

The condition (6) (with "'$\geq$"' replaced by "'$=$"') means as a consequence that $G_{b_*}^{-1}(c)$ must be the δ-fractile $u_\delta(x)$ of $P^{\theta|X=x}$. Alltogether one has as **Bayesian-confidence-region for ϑ**:

$$C(x) = (0, u_\delta(x)] \ , \tag{7}$$

where $u_\delta(x)$ is the δ-fractile of the Erlang$(a+\alpha, b+\sum_{i=1}^{n} x_i)$-distribution.

$C(x)$ is a **HPD δ-credible region** in the sense of Robert (1994) (see definition 5.7 there). But note, that the above results (especially (7)) are **not** given by Robert (1994) (and others). □

6. Parameter-estimation

For application of the results of the section 4 and 5 one needs to know the parameter b of the a-priori-distribution (remember $a \in \mathbb{N}$, $\alpha \in \mathbb{R}$ were assumed to be given (and known)).

Since b is not given, one needs an estimator for b. For deriving such an estimator suppose, that one has k replications of the X, say:

$$X_j = (X_{j1}, ..., X_{jn}) \ , \quad j = 1, ..., k \ .$$

with for each the random variable θ_j $(=\theta)$ of the parameter ϑ_j $(=\vartheta)$ $(j = 1, ..., k)$.

Assume for the following:

a) $\theta_1, ..., \theta_k$ are identically distributed.

b) (X_j, θ_j), $j = 1, ..., k$ are independent.

c) $X_{j1}, ..., X_{jn}$ are i.i.d. given θ_j (for $j = 1, ..., k$)
(defined analogy to (A), (B)).

Obviously one has the context of section 6.2 and 6.3 in Kremer (2005a).

In addition assume that θ_j is distributed like the θ in (C) (for all $j = 1, ..., k$) and that:

$$P_\vartheta^{X_{ji}} := P^{X_{ji}|\theta_j=\vartheta}$$

is that $P_\vartheta^{X_{ji}}$ of (D) (for all $j = 1, ..., k$ and $i = 1, ..., n$).

According to section 6.2 in Kremer (2005a) one gets the socalled **moment-estimator** for the b as the solution $\hat{b}$ of

$$E_{\hat{b}}(E(X_{ji}|\theta_j)) = \bar{X}_{\cdot\cdot} \tag{8}$$

with:

$$\bar{X}_{\cdot\cdot} := \frac{1}{k\cdot n}\cdot\sum_{j=1}^{k}\sum_{i=1}^{n} X_{ji}$$

and where the outer integral $E_b(\cdot)$ is standing for the integration over $\theta_j = \vartheta_j$ according to the Erlang(a, b)-law. According to (4) one has

$$E(X_{ji}|\theta_j) = \frac{a}{\theta_j} ,$$

and since:

$$E(\theta_j^{-1}) = \frac{b}{a-1}$$

it follows as equation for $\hat{b}$ from (8):

$$\alpha \cdot \frac{\hat{b}}{a-1} = X_{\cdot\cdot} \quad .$$

Consequently one has as moment-estimator $\hat{b}_{ME}$ for b simply:

$$\hat{b}_{ME} = \frac{a-1}{\alpha} \cdot \bar{X}_{\cdot\cdot} \quad .$$

Finally one also likes to know, how to calculate the **maximum-likelihood-estimator for** b in the above given Bayes-context. According to section 6.3 the $\hat{b}$ is a solution of the equation:

$$\sum_{j=1}^{k} \frac{d}{db} \left(ln \int \left(\prod_{i=1}^{n} f_{\vartheta}^{X_i}(X_{ji}) \right) P_b(d\vartheta) \right) \Big|_{b=\hat{b}} = 0 , \tag{9}$$

where P_b is the a-piori-distribution with parameter b (a is known!).

Inserting all densities (according to (C) and (D)), one gets after routine calculations:

$$\frac{d}{db} \left(ln \int \left(\prod_{i=1}^{n} f_{\vartheta}^{X_i}(X_{ji}) \right) P_b(d\vartheta) \right) = \frac{a}{b} - (a + n \cdot \alpha) \cdot \left(b + \sum_{i=1}^{n} X_{ji} \right)^{-1}$$

As consequence, (9) gives as equation for $\hat{b}$:

$$k \cdot \frac{a}{\hat{b}} = (a + n \cdot \alpha) \cdot \sum_{j=1}^{k} \frac{1}{\hat{b} + \sum_{i=1}^{n} X_{ji}} .$$

With further modifications one arrives at the final result, that the maximum-likelihood-estimator $\hat{b}_{ML}$ of b is given as:

$$\hat{b}_{ML} = n \cdot \hat{B} ,$$

where $\hat{B}$ is (the) solution of the equation:

$$\hat{B} = \frac{a}{a + n \cdot \alpha} \cdot \frac{1}{S(\hat{B})} \tag{10}$$

with the $S(\cdot)$ according:

$$S(B) = \frac{1}{k} \cdot \sum_{j=1}^{k} \frac{1}{B + \bar{X}_{j\cdot}} ,$$

where:

$$\bar{X}_{j\cdot} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_{ji} .$$

Clearly the solution $\hat{B}$ of (10) has to be calculated in practical application with an adequate method of numerical mathematics.

Certainly the practitioner might prefer the $\hat{b}_{ME}$ to the $\hat{b}_{ML}$. But note, according to certain general theoretical investigations of Asymptotic Statistics, also the more unhandy $\hat{b}_{ML}$ has its sense.

7. Final remarks

Note, that the author's roots go back to non-parametric statistics. He made elegant research on Bahadur efficiencies of rank tests (see e.g. Kremer (1979), (1981)). At the begin of the 80th he changed more to mathematical risk theory. In that field he brought a lot of new research in Bayes-techniques e.g. in premium rating (see e.g. Kremer (1982), (2005b)). Much work he made in mathematical stochastics of reinsurance (see e.g. Kremer (2004), (2008)).

References

Berger, J.O. (1985). *Statistical decision theory and Bayesian Analysis*. Springer.

Gosh, J.K. & Ramamoorthi, R.V. (2003). *Bayesian noparametric statistics*. Springer.

Johnson, N.I. and Kotz, S. (1970). *Continuous distributions* - 1. John Wiley.

Kremer, E. (1979). Approximate and local Bahadur efficiency of linear rank tests of the two sample problem. *Annals of Statistics.*

Kremer, E. (1981). Local comparison of rank tests for the independence problem. *Journal of Multivariate Statistics.*

Kremer, E. (1982). Credibility theory for some evolutionary models. *Scandinavian Actuarial Journal.* Kremer, E. (2004). Einf*ü*hrung in die Risikotheorie der verallgemeinerten H*ö*chstschadenr*ü*ckversicherung. *Verlag Versicherungswirtschaft.*

Kremer, E. (2005a). Einf*ü*hrung in die Mathematik der Bayes-Statistik f*ü*r Aktuare und Andere. *Logos Verlag.*

Kremer, E. (2005b). Credibility for the regression model with autoregressive error terms. *Blàtter der deutschen Gesellschaft für Versicherungsmathematik.*

Kremer, E. (2008). Einf*ü*hrung in die Risikotheorie des Stoploss-Vertrags. *Der Andere Verlag.*

Robert, C.P. (1994). The Bayesian Choice. *Springer.*