Bayes Convolution

Edwin R. van den Heuvel¹ and Chris A.J. Klaassen²

¹Institute for Business and Industrial Statistics, IBIS UvA BV, University of Amsterdam, The Netherlands. ²IBIS UvA B.V. and Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands. E-mail: chrisk@wins.uva.nl

Summary

A general convolution theorem within a Bayesian framework is presented. Consider estimation of the Euclidean parameter $\theta$ by an estimator $T$ within a parametric model. Let $W$ be a prior distribution for $\theta$ and define $G$ as the $W$-average of the distribution of $T - \theta$ under $\theta$. In some cases, for any estimator $T$ the distribution $G$ can be written as a convolution $G = K \ast L$ with $K$ a distribution depending only on the model, i.e. on $W$ and the distributions under $\theta$ of the observations. In such a Bayes convolution result optimal estimators exist, satisfying $G = K$. For location models we show that finite sample Bayes convolution results hold in the normal, loggamma and exponential case. Under regularity conditions we prove that normal and loggamma are the only smooth location cases. We also discuss relations with classical convolution theorems.

Key words: Bayes; Convolution; Loggamma.

1 Introduction

 Much of mathematical statistics is focussed on constructing optimal or efficient estimators. The impact of the convolution theorem has been enormous here, because it formulates a natural and elegant optimality criterion for estimators. Let $X$ be random with distribution $P_\theta$, $\theta \in \Theta \subset \mathbb{R}^k$. Based on $X$ we want to estimate the parameter $\theta$ optimally. A way to prove optimality of a candidate optimal estimator $T_0 = t_0(X)$ is both to find a bound on the performance of all (relevant) estimators $T = t(X)$ and to show that $T_0$ attains this bound. Several classical results fit into this framework.

A well-known lower bound follows from the (Fréchet-)Cramér-Rao inequality, also called the information inequality. It is a lower bound on the mean square error of (linear combinations of the components of) estimators $T = t(X)$ in terms of the Fisher information matrix $I(\theta)$; see Fréchet (1943), Rao (1945) and Cramér (1946, Sections 32.2 and 32.3). There are many generalizations of this inequality; we mention Bhattacharyya (1946) and Chapman & Robbins (1951). For unbiased estimators of $\theta$ with dimension $k = 1$, the Fréchet-Cramér-Rao inequality reduces to an inequality for the variance of $T$. It has been shown by Wijsman (1973) that equality may hold if and only if $\{P_\theta : \theta \in \Theta\}$ is an exponential family.

In the Bayesian approach to estimation, a prior distribution $W$ on the parameter space $\Theta$ is chosen. Given a loss function $\ell(\cdot, \cdot)$, one considers the resulting minimum Bayes risk

$$
\inf_{t(\cdot)} \int_{\mathbb{R}^k} E_\theta \ell(t(X), \theta) \, dW(\theta). \tag{1.1}
$$

With regard to optimality this infimum may be viewed as a lower bound, and one derives Bayes estimators attaining this bound. We refer to Berger (1985) for a comprehensive treatise.

The distribution of the error $T - \theta$ weighed by prior $W$ on $\Theta$ is at least as spread out as a
distribution determined by the distributions \( P_\theta \) and \( W \). This so-called spread inequality for one-dimensional parameters \( k = 1 \) is discussed in Klaassen (1989) and van den Heuvel & Klaassen (1997). For the symmetric location model the lower bounds mentioned here simplify and still other lower bounds may be proved; see Klaassen (1981). It may be shown that the lower bound in the spread inequality is attained for normal, exponential and loggamma distributions with \( \theta \) as location parameter.

If \( X \) is one observation from a location model with parameter \( \theta \in \mathbb{R}^k \), then the error \( T - \theta \) of any equivariant estimator \( T = t(X) \) of \( \theta \) based on \( X \) has a distribution that may be viewed as a convolution; namely the convolution of the distribution of \( X - \theta \) itself and some other distribution. This convolution result (Boll, 1955) represents a bound on the performance of equivariant estimators with \( T = t(X) = X \) as the optimal estimator; we will mention some consequences of the convolution structure in terms of loss functions after Theorem 2.1 below.

The main contribution of the present paper is a Bayesian approach to the convolution result of Boll (1955). It is a finite sample theorem (Theorem 2.1) which we illustrate now by one of the simplest possible estimation problems, namely of one observation in the one-dimensional normal location model. So, let \( X \) be normally distributed with mean \( \theta \) and variance unity; \( X \sim N(\theta, 1) \). As prior \( W \) we choose \( N(0, \sigma^2) \) and we introduce the random variable \( \theta \) with realizations \( \theta \) and distribution \( W \). For every estimator \( T = t(X) \) we may write

\[
T(X) - \theta = \left\{ \frac{\sigma^2}{1 + \sigma^2} X - \theta \right\} + \left\{ t(X) - \frac{\sigma^2}{1 + \sigma^2} X \right\}.
\]

The crux is that the two terms at the right hand side are independent. This may be seen easily by noting that the first term and \( X \) are jointly normal and uncorrelated. Indeed, we have

\[
\text{Cov} \left( \frac{\sigma^2}{1 + \sigma^2} X - \theta, X \right) = \frac{\sigma^2}{1 + \sigma^2} EX^2 - E\theta X = \frac{\sigma^2}{1 + \sigma^2} (1 + E\theta^2) - E\theta^2 = 0.
\]

This independence in (1.2) shows that the distribution of \( T - \theta \) is a convolution and that the optimal estimator is \( t(X) = \sigma^2(1 + \sigma^2)^{-1} X \).

In asymptotic statistics, specific estimation problems are approximated by simple location estimation problems. More formally, a specific estimation problem is viewed as a member of a sequence of experiments and the resulting limit experiment is studied. Typically, the sequence has the LAN-property (LAN = Local Asymptotic Normality, introduced by Le Cam (1960)), which means that the limit experiment is a normal (or Gaussian) shift experiment. In this limit experiment all the above optimality bounds and criteria may be applied. However, we will focus on convolution results. Equivariance leads to the celebrated so-called Hájek–Le Cam convolution theorem, which implies that under LAN optimal (sequences of) estimators are asymptotically normal with covariance matrix \( I^{-1} \theta \). Such estimators are called asymptotically efficient, or efficient for short.

The drawback of this classical convolution theorem due to Hájek (1970), is that the class of competing estimators is restricted to regular estimators. Regularity means asymptotic equivariance here. Our Bayesian convolution theorem does not put restrictions on the class of estimators. Therefore, it leads to asymptotic Bayes convolution theorems that are not restricted to regular estimators; see Chapter 3 of van den Heuvel (1996).

Hájek (1970) proved his asymptotic convolution theorem via averaging by uniform priors. This Bayesian aspect of the proof of the classical convolution theorem is exploited in the present paper and leads to our main result, the finite sample Bayes convolution theorem of Section 2. Sections 3 through 5 will discuss normal, exponential and loggamma location models, where this convolution theorem holds. Section 6 will show that Bayes convolution for smooth location models holds if and only if \( P_\theta \) is normal or loggamma. Section 7 will contain more details about the relation between the classical convolution theorem and our Bayes convolution theorem than could be given in this
introductory Section 1.

2 Bayes Convolution

Suppose we have a random element \( X \) drawn via a distribution from a parametric family \( P = \{ P_\theta : \theta \in \Theta \} \), \( \Theta \subset \mathbb{R}^k \), on a measurable space \( (X, A) \). On the basis of this sample we want to estimate the parameter \( \theta \). In a Bayesian set-up we choose a weight function or prior \( W \) on \( \mathbb{R}^k \), putting its mass in \( \theta \), and we consider the average distribution function

\[
G(y) = \int_{\mathbb{R}^k} P_\theta(T - \theta \leq y) dW(\theta), \quad y \in \mathbb{R}^k, \tag{2.1}
\]

where \( T \equiv t(X) \) is an estimator of \( \theta \).

The following simple observation tells us that for general dimension \( k \) the average distribution \( G \) is the convolution of a distribution, which depends on \( P \) and \( W \), but which does not depend on the estimator \( T \), and another distribution.

Let \( \psi : X \to \mathbb{R}^k \) be a measurable function such that \( \psi(X) - \vartheta \) and \( X \) are independent, where \( \vartheta \) has distribution \( W \). Then \( \psi(X) - \vartheta \) and \( t(X) - \psi(X) \) are also independent. Since \( T - \vartheta \) may be written as

\[
T - \vartheta = t(X) - \vartheta = \{ \psi(X) - \vartheta \} + \{ t(X) - \psi(X) \}, \tag{2.2}
\]

we may conclude that \( G(\cdot) = P(T - \vartheta \leq \cdot) \) is a convolution of \( K(\cdot) = P(\vartheta(X) - \vartheta \leq \cdot) \), the distribution of \( \vartheta(X) - \vartheta \), which indeed does not depend on \( T \). Consequently, there exists a distribution \( L \) with

\[
G = K * L. \tag{2.3}
\]

We will call this equality (2.3) a Bayes Convolution Theorem. Furthermore, we will call \( T = \psi(X) \) the best estimator in the sense of (2.3), since this choice makes \( L \) degenerate at 0. Summarizing, we have the following observation.

**Theorem 2.1. Bayes Convolution.** Let \( \vartheta \) be a random variable taking values in \( \Theta \subset \mathbb{R}^k \) and let the conditional distribution of \( X \) given \( \vartheta = \theta \) be \( P_\theta \) on \( (X, A) \). If the measurable function \( \psi : X \to \mathbb{R}^k \) is such that \( \psi(X) - \vartheta \) and \( X \) are independent, then the distribution \( G(\cdot) = P(T - \vartheta \leq \cdot) \) of \( T - \vartheta \) is the convolution of the distribution \( K(\cdot) = P(\psi(X) - \vartheta \leq \cdot) \) and some other distribution \( L \), i.e. (2.3) holds.

If such a \( \psi \) exists, the best estimator with respect to Bayes risk is \( \psi(X) + c \), where \( c \in \mathbb{R}^k \) may depend on the loss function. Indeed,

\[
\inf_T E\ell(T - \vartheta) = \inf_c E\ell(\psi(X) - \vartheta + c) \tag{2.4}
\]

holds for all convex loss functions \( \ell \) in view of the following application of the conditional Jensen inequality

\[
E\ell(T - \vartheta) \geq E\ell(E(\psi(X) - \vartheta + T - \psi(X) | \psi(X) - \vartheta)) = E\ell(\psi(X) - \vartheta + E(T - \psi(X))) \geq \inf_c E\ell(\psi(X) - \vartheta + c). \tag{2.5}
\]

If for \( k = 1 \) the distribution of \( \psi(X) - \vartheta \) is strongly unimodal, then equality (2.4) holds for all loss functions which are decreasing-increasing, and again the best estimator is determined by \( \psi \) and does not depend on the loss function apart from a shift over \( c \). This may be seen as follows. First, note that the convolution (2.3) with \( K \) strongly unimodal, implies that \( G \) is at least as spread out as \( K \), i.e. that the quantiles of \( G \) are at least as far apart as those of \( K \); see Droste & Wefelmeyer (1985)
and Klaassen (1985). Subsequently, if there exists a \( u_0 \in [0, 1] \) with \( G^{-1}(u_0) = 0 \) then this spread property yields

\[
E\ell(T - \vartheta) = \int_0^1 \ell(G^{-1}(u) - G^{-1}(u_0))du \geq \int_0^1 \ell(K^{-1}(u) - K^{-1}(u_0))du
\]

\[
= E\ell(\psi(X) - \vartheta - K^{-1}(u_0)).
\]

Finally, note that (2.6) may be adapted to the case where 0 is not a quantile of \( G \); cf. the proof of Theorem 2.1 of Klaassen (1984).

In view of (2.4), (2.5) and (2.6) it seems interesting to investigate the existence of combinations of \( P \), \( W \) and \( \psi \) such that the distribution \( G \) from (2.1) has a Bayes convolution structure (2.3). To this end we restrict attention to location parameters \( \vartheta \). In Sections 3, 4 and 5 respectively, we will show that normal, exponential and loggamma location families yield such convolution structures. Furthermore, we will prove in Section 6 that for "smooth" distributions the normal and loggamma are the only two generating location families yielding a Bayes convolution structure. In Section 6 we will also indicate relationships between our Bayes Convolution Theorem 2.1 and the classical convolution theorem for location parameters.

### 3 Normal Location Models

In the model of Section 2 with \( \vartheta \in \Theta = \mathbb{R}^k \) and \( k \) general, let there exist an \( m \times k \) matrix \( A \) and an \( m \times m \) matrix \( \chi \) such that conditionally on \( \vartheta = \theta \) the random \( m \)-vector \( X \) is normally distributed with mean vector \( A\theta \) and nonsingular covariance matrix \( \chi \). We will assume that the \( k \times k \) matrix \( A^T A \) is nonsingular. Then \( \theta \) is identifiable via \( (A^T A)^{-1}A^T E(X | \vartheta = \theta) = \theta \). As prior on \( \mathbb{R}^k \) we take the normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). Consequently, the \((m + k)\)-vector \((X^T, \vartheta^T)^T\) is normally distributed,

\[
\left( \begin{array}{c} X \\ \vartheta \end{array} \right) \sim \mathcal{N}\left( \left( \begin{array}{c} A\mu \\ \mu \end{array} \right), \left( \begin{array}{cc} A\Sigma A^T + \chi & A\Sigma \\ \Sigma & \Sigma \end{array} \right) \right).
\]

With covariance as inner product in the Hilbert space of mean zero random variables with finite variance, the componentwise projection of \( \vartheta - \mu \) onto the linear subspace spanned by the components of \( X - A\mu \) equals

\[
\Pi(\vartheta - \mu | [X - A\mu]) = E(\vartheta - \mu)(X - A\mu)^T [E((X - A\mu)(X - A\mu)^T)]^{-1} (X - A\mu) = \Sigma A^T (A\Sigma A^T + \chi)^{-1} (X - A\mu).
\]

Consequently, all components of \( \vartheta - \mu - \Pi(\vartheta - \mu | [X - A\mu]) \) are orthogonal to and hence uncorrelated with all components of \( X - A\mu \). In view of the joint normality this means that

\[
\psi(X) - \vartheta = \mu + \Pi(\vartheta - \mu | [X - A\mu]) - \vartheta = \mu + \Sigma A^T (A\Sigma A^T + \chi)^{-1} (X - A\mu) - \vartheta
\]

and \( X - A\mu \) are independent random vectors. We may conclude that a Bayes convolution result (2.3) holds and that the convolution lower bound \( K \) is the distribution of \( \psi(X) - \vartheta \) from (3.3), which is normal with mean vector 0 and covariance matrix

\[
\Sigma_K = \Sigma - \Sigma A^T (A\Sigma A^T + \chi)^{-1} A\Sigma = (A^T A)^{-1} A^T A\Sigma A^T + \chi)^{-1} A\Sigma.
\]

We may formulate this Bayes convolution result as follows.
THEOREM 3.1. In the above normal location model,
\[ T = t(X) = \psi(X) = \mu + \Sigma A^T (A \Sigma A^T + \chi)^{-1} (X - A \mu) \]  
\hspace{0.5cm} (3.5)
is the best estimator for \( \theta \) in the sense of (2.3).

Of course, the case of \( n \) i.i.d. observations \( X_1, \ldots, X_n \) from a normal distribution with mean vector \( \theta \) and nonsingular \( k \times k \) covariance matrix \( \Omega \) is of greatest interest. It fits into our framework by the choices \( m = nk, \ X^T = (X_1^T, \ldots, X_n^T), \ A^T = (I \ldots I) \) with \( I \) the \( k \times k \) identity matrix, and \( \chi \) the \( m \times m \) block diagonal matrix with all \( k \times k \) blocks on the diagonal equal to \( \Omega \). We will use a matrix inversion formula which for \( n = 2 \) reads as follows:
\[ \begin{pmatrix} \Omega + \Sigma & \Sigma \\ \Sigma & \Omega + \Sigma \end{pmatrix}^{-1} = \begin{pmatrix} \Omega^{-1} - A & -A \\ -A & \Omega^{-1} - A \end{pmatrix} \]  
\hspace{0.5cm} (3.6)
with
\[ A = (\Omega + 2 \Sigma)^{-1} \Sigma \Omega^{-1} = (\Omega + 2 \Sigma)^{-1} (\Sigma + 2 \Sigma \Omega^{-1} \Sigma) (\Omega + 2 \Sigma)^{-1} = \Omega^{-1} \Sigma (\Omega + 2 \Sigma)^{-1}. \]  
\hspace{0.5cm} (3.7)
Then, \( T = \psi(X) \) from (3.5) may be written as
\[ \psi(X) = \mu + \Omega (n \Sigma + \Omega)^{-1} \Sigma \Omega^{-1} \sum_{i=1}^{n} (X_i - \mu), \]  
\hspace{0.5cm} (3.8)
whereas the covariance matrix of \( \psi(X) - \theta \) from (3.4) reads as
\[ \Sigma_K = \Omega (n \Sigma + \Omega)^{-1} \Sigma. \]  
\hspace{0.5cm} (3.9)

With \( \Sigma = \sigma^2 B, \ B \) nonsingular, and with \( \sigma \) converging to infinity, our normal prior converges to the uninformative prior and the best estimator becomes the sample mean,
\[ \lim_{\sigma \to \infty} \psi(X) = \frac{1}{n} \sum_{i=1}^{n} X_i, \]  
\hspace{0.5cm} (3.10)
with covariance matrix
\[ \lim_{\sigma \to \infty} \Sigma_K = n^{-1} \Omega. \]  
\hspace{0.5cm} (3.11)
Finally, consider a translation equivariant estimator \( T = t(X) \) of \( \theta \), i.e.
\[ t(x_1 + a, \ldots, x_n + a) = t(x_1, \ldots, x_n) + a, \]  
\hspace{0.5cm} (3.12)
for all \( a, x_1, \ldots, x_n \in \mathbb{R}^k \). Then the average distribution of \( T \) from (2.1) reduces to
\[ G(y) = P_0(T \leq y), \hspace{0.5cm} y \in \mathbb{R}^k, \]  
\hspace{0.5cm} (3.13)
and does not depend on the prior distribution \( W \). Consequently, the preceding limit procedure with \( \sigma \to \infty \) yields
\[ G = \mathcal{N}(0, n^{-1} \Omega) \ast L \]  
\hspace{0.5cm} (3.14)
for some \( L \) and the sample mean is the best translation equivariant estimator of \( \theta \).

4 Exponential Location Models

Consider exponential densities with a location parameter,
\[ f_\lambda(x - \theta) = \lambda e^{-\lambda(x - \theta)} 1_{(0, \infty)}(x - \theta), \hspace{0.5cm} x \in \mathbb{R}. \]  
\hspace{0.5cm} (4.1)
We will study estimators of $\theta \in \Theta = \mathbb{R}$ based on $X = (X_1, \ldots, X_n)^T$, where the $X_i$ have a density from (4.1), but they have to be neither independent nor identically distributed. As a weight function or prior $W$ on $\Theta$ we will take the distribution with density

$$w(\theta) = \mu e^{(\theta-a)}1_{(-\infty,a)}(\theta), \quad \theta \in \mathbb{R}.$$  

(4.2)

Note that this means that $\theta$ from Section 1 has the structure $\theta = a - Y$ with $Y$ an exponential random variable with density $f_\mu(x)$ from (4.1).

For a start let $X_1, \ldots, X_n$ be independent and let $X_i$ have density $f_{\lambda_i}(\cdot - \theta)$. We will show that

$$\psi(X) = \min\{X_1, \ldots, X_n, \theta\}$$  

(4.3)

leads to the Bayes Convolution Theorem 2.1 for this situation. Note that this includes the i.i.d. case with $\lambda_1 = \cdots = \lambda_n$. Our approach will be based on the following lemma.

**Lemma 4.1.** Let $Y$ and $Z$ be independent random variables with exponential densities $f_\lambda$ and $f_\mu$, respectively. Then $Y \wedge Z$ and $Z - Y$ are independent.

**Proof:** Straightforward computation shows,

$$P(Y \wedge Z \geq u, Z - Y \geq v) = \begin{cases} e^{-(\lambda+\mu)u} \cdot \frac{1}{\lambda+\mu} e^{-\mu v}, & v \geq 0 \\ e^{-(\lambda+\mu)u} \cdot \left(1 - \frac{\mu}{\lambda+\mu} e^{\lambda v}\right), & v \leq 0. \end{cases}$$

Q.E.D.

In order to prove (2.3) for (4.3) under (4.1) and (4.2), we have to show independence of $\psi(X) - \theta$ and $X = (X_1, \ldots, X_n)^T$, i.e.

$$\psi(X) - \theta \perp X.$$  

(4.4)

Consequently, it suffices to prove independence of $\psi(X) - \theta$ and $X_i$, $i = 1, \ldots, n$. Write $X_i = \theta + Y_i = a + Y_i - Y$, where $Y_i$ has density $f_{\lambda_i}(\cdot)$ and $Y_1, \ldots, Y_n, Y$ are independent. By the Lemma, $Y_i \wedge Y$ and $X_i = a + Y_i - Y$ are independent. In view of

$$\psi(X) - \theta = \min\{\theta + Y_1, \ldots, \theta + Y_n, a\} - \theta = \min\{Y_1, \ldots, Y_n, Y\}$$

(4.5)

and the independence of $Y_1, \ldots, Y_n, Y$, this yields (4.4) and hence (2.3) for this situation.

The next situation we will study is the fatal shock model of Proschan & Sullo (1976) for dependent exponentially distributed random variables. To formulate this model we take $Y_0, Y_1, \ldots, Y_n, Y$ to be independent exponentially distributed random variables with location parameter $0$ as before. Again, we take $\theta = a - Y$, but we replace $X_i = \theta + Y_i$ by

$$X_i = \theta + Y_i \wedge Y_0, \quad i = 1, \ldots, n.$$  

(4.6)

The random variables $X_i - \theta = Y_i \wedge Y_0$ all have an exponential distribution, but they are not independent. Let $Y_i$ be the waiting time till part $i$ of some equipment fails by its internal causes and let $Y_0$ be the time till a fatal shock occurs that ruins all parts $i = 1, \ldots, n$ of the equipment. Then $X_i - \theta = Y_i \wedge Y_0$ is the survival time of part $i$, and this interpretation explains why this is called the fatal shock model.

In this model too, the statistic (4.3) satisfies (4.4) and hence yields a Bayes Convolution Theorem 2.1. To verify this, first note

$$\psi(X) - \theta = \min\{\theta + Y_0, \theta + Y_1, \ldots, \theta + Y_n, a\} - \theta = \min\{Y_0, Y_1, \ldots, Y_n, Y\}$$

(4.7)
and

\[ X_i = a + \min\{Y_0 - Y, Y_i - Y\}. \quad (4.8) \]

Therefore, it suffices to show

\[ \min\{Y_0, Y_1, \ldots, Y_n\} \perp Y_i - Y, \ i = 0, 1, \ldots, n. \quad (4.9) \]

However, we have

\[ Y_j \perp Y_i - Y, \quad j \neq i, \quad (4.10) \]

and by Lemma 3.1

\[ Y_i \perp Y_i - Y, \quad i = 0, 1, \ldots, n. \quad (4.11) \]

Together, (4.10) and (4.11) yield (4.9) and hence (4.4).

A generalization of these models for multivariate exponential distributions is given by Esary & Marshall (1974). Fix a collection \( \mathcal{J} \) of subsets of \( \{1, \ldots, n\} \). For each \( J \in \mathcal{J} \), let \( Y_J \) be an exponential random variable with density \( f_{\lambda_J}(\cdot) \), and let all \( Y_J, J \in \mathcal{J}, \) and \( Y \) be independent. The exponential random variables \( X_i \) are defined now by

\[ X_i = \theta + \min\{Y_J : i \in J \in \mathcal{J}\}, \quad i = 1, \ldots, n. \quad (4.12) \]

If \( \mathcal{J} = \{\{1\}, \ldots, \{n\}\} \) is the collection of singletons, then \( X_i = \theta + Y_{\{i\}}, \ i = 1, \ldots, n, \) describes our first case of independent exponentials. With \( \mathcal{J} = \{\{i\}, \ldots, \{n\}, \{1, \ldots, n\}\} \) we obtain the fatal shock model (4.6).

Again, \( \psi(X) \) from (4.3) leads to the Bayes Convolution Theorem 2.1. Indeed,

\[ \psi(X) - \theta = \min_{J \in \mathcal{J}} Y_J \perp Y \quad (4.13) \]

is independent of

\[ X_i = a + \min\{Y_J - Y : i \in J \in \mathcal{J}\} \quad (4.14) \]

in view of Lemma 3.1 and the independence of \( Y, Y_J, J \in \mathcal{J}. \)

We may formulate these Bayes Convolution results as follows.

**THEOREM 4.1.** In all of the above exponential location models

\[ T = t(X) = \psi(X) = \min\{X_1, \ldots, X_n, a\} \quad (4.15) \]

is the optimal estimator for \( \theta \) in the sense of (2.3).

Recall that the chosen prior implies \( \theta \leq a \), and note that \( T \leq a \), a.s. Finite sample optimality of the smallest observation has been studied for more general densities in Klaassen (1995). Furthermore, note that for translation equivariant (cf. (3.12)) estimators of \( \theta \), (3.13) holds again, and that the limit procedure with \( a \to \infty \) shows that the sample minimum is the best translation equivariant estimator of \( \theta \).

To conclude this Section we note that model (4.12) includes the case of \( m \) i.i.d. copies of \( k \)-dimensional exponentials of the type (4.12) itself. Just take \( n = mk \) and let \( \mathcal{J} \) consist of isomorphic subsets of \( \{1, \ldots, k\}, \{k + 1, \ldots, 2k\}, \ldots, \{(m - 1)k + 1, \ldots, mk\}, \) respectively.
5 Loggamma Location Models

Let the random variable $Y_{\alpha, \beta}$ have a gamma distribution with shape parameter $\alpha$, scale parameter $\beta$ and density

$$g_{\alpha, \beta}(y) = \frac{1}{\Gamma(\alpha) \beta^\alpha} y^{\alpha-1} e^{-\gamma/y/\beta}, \quad y > 0. \tag{5.1}$$

Then $\gamma^{-1} \log Y_{\alpha, \beta}$ has a so-called loggamma distribution with parameters $\alpha > 0$, $\beta > 0$, and $\gamma \neq 0$, and with density

$$f_{\alpha, \beta, \gamma}(x) = \frac{|\gamma|}{\Gamma(\alpha) \beta^\alpha} \exp\left\{ -\frac{1}{\beta} e^{\gamma x} + \alpha \gamma x \right\}, \quad x \in \mathbb{R}. \tag{5.2}$$

With the shape parameter $\alpha$ held fixed, this is a location-scale family with scale parameter $\gamma^{-1} \log \beta$. Density (5.2) is a generalization of the type I extreme value density $\exp(-e^{-x} - x), x \in \mathbb{R}$. For simplicity we will restrict $\gamma$ to be positive and we reparametrize to location parameter $\theta = \gamma^{-1} \log \beta$. Keeping $\alpha$ and $\gamma$ positive and fixed we obtain the location family

$$f_{\alpha, \gamma}(x, \theta) = \frac{\gamma}{\Gamma(\alpha)} \exp\{ -e^{\gamma(x-\theta)} + \alpha \gamma (x - \theta)\}, \quad x \in \mathbb{R}. \tag{5.3}$$

Theorem 5.1. Let $X_1, \ldots, X_n$ be independent random variables and let $X_i$ have density $f_{\alpha, \gamma}(x_i - \theta)$, $x_i \in \mathbb{R}$, $i = 1, \ldots, n$. Furthermore, let the prior $W$ on $\mathbb{R} = \Theta \ni \theta$ be described by a random variable $\Theta = a - Y$ with $Y$ a loggamma random variable with density $f_{\alpha, \gamma}(y, 0)$, $y \in \mathbb{R}$, and with $a$ being a fixed constant. With

$$\psi(X) = \frac{1}{\gamma} \log \left\{ \sum_{i=1}^n e^{\gamma X_i} + e^{\gamma a} \right\} \tag{5.4}$$

the Bayes Convolution Theorem 2.1 holds here.

Proof: We will write $X_i = \theta + Y_i$ with $Y_1, \ldots, Y_n$, $Y$ independent with densities $f_{\alpha_1, \gamma}$, $\ldots$, $f_{\alpha_n, \gamma}$, $f_{\alpha, \gamma}$, respectively. Note that

$$X_i = a + Y_i - Y, \quad i = 1, \ldots, n,$$

and

$$\psi(X) - \theta = \frac{1}{\gamma} \log \left\{ \sum_{i=1}^n e^{\gamma Y_i} + e^{\gamma Y} \right\}$$

hold. Consequently, the independence of $\psi(X) - \theta$ and $X$, which is the only condition we need for (2.3), holds if

$$e^{\gamma Y_i} + e^{\gamma Y} \perp Y_i - Y, \quad i = 1, \ldots, n.$$

This is the case as is implied by the Lemma below. Q.E.D.

Lemma 5.1. Let $Y_1$ and $Y_2$ be independent random variables with loggamma densities $f_{\alpha_1, \gamma}$ and $f_{\alpha_2, \gamma}$, respectively (cf. (5.3)). Then

$$e^{\gamma Y_1} + e^{\gamma Y_2}$$

are independent.

Proof: Since $\gamma Y_i$ has density $f_{\alpha_i, \gamma}$, we will assume $\gamma = 1$ without loss of generality. Define
Then \( Y_1 - Y_2 = \log(Z_1/Z_2) \) holds, and we see that it suffices to show that 
\( Z_1 + Z_2 \) and \( Z_1/Z_2 \) are independent, whenever \( Z_1 \) and \( Z_2 \) are independent gamma random variables with densities \( g_{\alpha_1,1} \) and \( g_{\alpha_2,1} \), respectively (cf. (5.1)). Straightforward computation shows,

\[
\frac{\partial^2}{\partial v \partial u} P(Z_1 + Z_2 > u, Z_1/Z_2 > v) = \frac{\partial^2}{\partial v \partial u} \int_{u-z_i}^{z_i/u} \int_{u-z_i}^{z_1} g_{\alpha_1,1}(z_1)g_{\alpha_2,1}(z_2)dz_2dz_1
\]

\[
= -\frac{\partial}{\partial v} \int_{u-z_i}^{z_i/u} g_{\alpha_1,1}(z_1)g_{\alpha_2,1}(u-z_1)dz_1
\]

\[
= g_{\alpha_1,1} \left( \frac{uv}{1+v} \right) g_{\alpha_2,1} \left( \frac{u}{1+v} \right) \frac{u}{(1+v)^2}
\]

\[
= \left\{ \frac{1}{\Gamma(\alpha_1 + \alpha_2)} u^{\alpha_1 + \alpha_2 - 1} e^{-u} \right\} \left\{ \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1 - 1}(1+v)^{-\alpha_1 - \alpha_2} \right\}, u > 0, v > 0,
\]

and the joint density factorizes; cf. Example II.3.(a) of Feller (1971). Q.E.D.

Note that for translation equivariant estimators of \( \theta \), (3.13) holds again, and that a limit procedure with \( u \to -\infty \) shows that the statistic

\[
T = t(X) = \frac{1}{\gamma} \log \left\{ \sum_{i=1}^{n} e^{yX_i} \right\}
\]

is the best translation equivariant estimator of \( \theta \). We conclude this Section by noting that the normal model treated in Section 3, is a limiting case of (5.3) by the choices \( \alpha = y^{-2} \), \( \theta = \mu + 2y^{-1} \log \gamma \) and \( y \downarrow 0 \).

6 Necessity of Loggamma

The Bayes Convolution Theorem 2.1 is derived from independence of \( \psi(X) - \theta \) and \( X \). This independence implies that the conditional distribution of \( \psi(X) \) given \( X = x \) does not depend on \( x \). In other words, the conditional distribution of \( \psi(X) \) given \( X = x \) has location parameter \( \psi(x) \), but does not depend on \( x \) otherwise. We will derive a consequence of this for the location model with \( \theta \in \Theta = \mathbb{R} \) and \( X = (X_1, \ldots, X_n)^T \). Here \( X_1, \ldots, X_n \) are independent and \( X_i \) has positive density \( f_i(x_i - \theta), x_i \in \mathbb{R}, i = 1, \ldots, n \), with respect to Lebesgue measure.

**Theorem 6.1.** Let \( \log f_i \) be twice continuously differentiable and bounded from above, \( i = 1, \ldots, n \). Let the prior distribution \( W \) have log-differentiable positive density \( w \) on \( \mathbb{R} \). If there exists a function \( \psi : \mathbb{R}^n \to \mathbb{R} \) such that \( \psi(X) - \theta \) and \( X \) are independent then each \( f_i \) is either a normal or a loggamma density.

**Remark 6.1.** As noted at the end of the preceding Section 5 the normal distribution may be viewed as a limit case of the loggamma distributions. See Ferguson (1962) for this and for other characterizations of loggamma distributions.
Proof: As noted above the existence of $\psi$ implies that there exists a density $g$ such that the conditional density of $\theta$ given $X = x$ satisfies

$$
\frac{\left( \prod_{i=1}^{n} f_i(x_i - \theta) \right) w(\theta)}{\int_{-\infty}^{\infty} \left( \prod_{i=1}^{n} f_i(x_i - \xi) \right) w(\xi) d\xi} = g(\theta - \psi(x)), \ \theta \in \mathbb{R}, \ x \in \mathbb{R}^n.
$$

(6.1)

Denoting the denominator, i.e. the marginal density of $X$, by $f(x)$, we see that (6.1) may be rewritten as

$$
\sum_{i=1}^{n} \log f_i(x_i - \theta) + \log w(\theta) = \log g(\theta - \psi(x)) + \log f(x), \ \theta \in \mathbb{R}, \ x \in \mathbb{R}^n.
$$

(6.2)

Differentiation with respect to $\theta$ and $x_j$ of the left-hand side of (6.2) and hence of the right-hand side is possible and yields

$$
\sum_{i=1}^{n} \frac{f_i'}{f_i}(x_i - \theta) + \frac{w'}{w}(\theta) = \frac{g'}{g}(\theta - \psi(x)), \ \theta \in \mathbb{R}, \ x \in \mathbb{R}^n,
$$

(6.3)

and

$$
\frac{f_j'}{f_j}(x_j - \theta) = -\frac{g'}{g}(\theta - \psi(x)) \frac{\partial \psi(x)}{\partial x_j} + \frac{\partial}{\partial x_j} \log f(x), \ \theta \in \mathbb{R}, \ x \in \mathbb{R}^n.
$$

(6.4)

Eliminating $g'/g$ from (6.3) and (6.4) we arrive at the equation

$$
\left(1 - \frac{\partial \psi(x)}{\partial x_j}\right) \frac{f_j'}{f_j}(x_j - \theta) = \left\{ -\frac{w'}{w}(\theta) + \sum_{i=1}^{n} \frac{f_i'}{f_i}(x_i - \theta) \right\} \frac{\partial \psi(x)}{\partial x_j}
$$

$$
+ \frac{\partial}{\partial x_j} \log f(x), \ \theta \in \mathbb{R}, \ x \in \mathbb{R}^n.
$$

(6.5)

Fix $j$. If $\partial \psi(x)/\partial x_j = 1$ holds at some $x = x_0 \in \mathbb{R}^n$, then (6.5) may be rewritten at this $x_0$ as

$$
\frac{w'}{w}(\theta) - \sum_{i=1}^{n} \frac{f_i'}{f_i}(x_0i - \theta) = \frac{\partial}{\partial x_j} \log f(x_0) = c(x_0), \ \theta \in \mathbb{R}.
$$

(6.6)

Integration shows that there exists a constant $d(x_0)$ such that

$$
\log w(\theta) + \sum_{i=1}^{n} \log f_i(x_0i - \theta) = c(x_0)\theta + d(x_0), \ \theta \in \mathbb{R},
$$

(6.7)

or

$$
w(\theta) \prod_{i=1}^{n} f_i(x_0i - \theta) = \exp(c(x_0)\theta + d(x_0)), \ \theta \in \mathbb{R}.
$$

(6.8)

The boundedness of $f_i$, $i = 1, \ldots, n$, implies that the left-hand side of (6.8) is integrable with respect to $\theta$ over $\mathbb{R}$. However, there do not exist constants $c(x_0)$ and $d(x_0)$ such that the right-hand side is integrable.
Consequently, (6.5) may be rewritten as

\[
\frac{f'_j(x_j - \theta)}{f_j} = \left\{ -\frac{w'}{w}(\theta) + \sum_{i \neq j} \frac{f'_i(x_i - \theta)}{f_i} \right\} \frac{\partial \psi(x)/\partial x_j}{1 - \partial \psi(x)/\partial x_j} \\
+ \frac{\partial \log f(x)/\partial x_j}{1 - \partial \psi(x)/\partial x_j}, \quad \theta \in \mathbb{R}, \ x \in \mathbb{R}^n.
\]

(6.9)

We will consider \( x_i \) fixed for \( i \neq j \). By Lemma 6.1 below, (6.9) implies that the score function \( f'_i/f_j \) is either linear or exponential up to a constant. But this means that the density \( f_j \) itself is either normal or loggamma. Q.E.D.

**Lemma 6.1.** Let \( \chi : \mathbb{R} \to \mathbb{R} \) be such that there exist functions \( f, g, \) and \( h \) such that

\[
\chi(x + y) = f(x) + g(x)h(y), \quad x, \ y \in \mathbb{R}.
\]

(6.10)

If \( \chi \) is continuously differentiable, then there exist constants \( b, c \) and \( \lambda \) such that either

\[
\chi(x) = bx + c.
\]

(6.11)
or

\[
\chi(x) = be^{\lambda x} + c, \quad x \in \mathbb{R}.
\]

(6.12)

**Proof:** Differentiation of (6.10) by \( y \) yields existence of the derivative \( h' \) of \( h \) and

\[
\chi'(x + y) = g(x)h'(y), \quad x, \ y \in \mathbb{R}.
\]

(6.13)

However, this implies that \( \chi' \) is exponential or a constant and hence the result. Indeed, the classical argument runs as follows. Equation (6.13) implies

\[
\chi'(x) = g(x)h'(0) = g(0)h'(x),
\]

and hence

\[
\chi'(x)/(g(0)h'(0)) = g(x)/g(0) = h'(x)/h'(0)
\]

(without loss of generality we may assume \( g(0)h'(0) \neq 0 \), since \( g(0)h'(0) = 0 \) implies that \( \chi \) is constant). With \( \Omega(x) = \log |\chi'(x)/(g(0)h'(0))| \), this yields \( \Omega(x + y) = \Omega(x) + \Omega(y) \) and hence \( \Omega(kx) = k\Omega(x) \). Consequently,

\[
\Omega(\frac{p}{q}) = \frac{1}{q} \Omega(q \cdot \frac{p}{q}) = \frac{1}{q} \Omega(p) = \frac{p}{q} \Omega(1).
\]

Since the set \( \mathbb{Q} \) of the rationals is dense in \( \mathbb{R} \), the continuity of \( \Omega \) implies its linearity. Q.E.D.

### 7 Classical Convolution Theorems

The classical convolution theorem has been formulated and proved by Hájek (1970). He assumed Local Asymptotic Normality (LAN) of the log-likelihood ratio, which means that the original experiment can be approximated asymptotically by a normal shift experiment, i.e. a normal location model. The covariance matrix of the corresponding normal distributions is related to the Fisher information matrix. The LAN-terminology stems from Le Cam (1960), whereas the theory of limits of experiments has a milestone in Le Cam (1972). It had been shown by Boll (1955), that the distribution of an equivariant estimator of the location parameter, based on one observation in a location model, is the convolution of the distribution of the observation and another distribution. Asymptotic equivariance has been formulated by Hájek (1970) as regularity. His convolution theorem states then that under LAN the limit distribution of a regular sequence of estimators is the convolution
of a normal distribution and another one, where the normal distribution has mean vector 0 and covariance matrix the inverse of the Fisher information matrix.

To prove this, Hájek uses averaging by uniform priors with increasing support. The same device is used by Strasser (1985), Chapter 38, p. 191, to present the convolution result of Boll (1955), and by Le Cam (1986), Section 8.3, pp. 125-128, to prove a lifting theorem that yields Hájek's convolution theorem. An alternative proof via analytic continuation was suggested by Bickel, see Theorem 3.1 of Roussas (1972) or Theorem 2.3.1 of Bickel et al. (1993). In his Chapter 2, Van der Vaart (1988), as cited in Theorem 5.6.3, pp. 85, 86, of Le Cam & Yang (1990), formulated and proved another convolution theorem under LAN, without the condition of regularity. Instead, he averaged over the shift parameter and proved a Bayes convolution theorem for the normal case. From this normal Bayes convolution theorem, one can easily derive the convolution theorem of Hájek. In view of his regularity assumption one restricts the class of estimators to equivariant estimators. The distribution $G$ in (2.1) reduces to (3.13) then, and it does not depend on the prior. However, the distribution $K$ in the Bayes convolution does, since $\psi$ is in general not shift equivariant. By changing the prior to an uninformative prior the distribution $G$ does not change, but the distribution $K$ becomes most spread out. This device has been presented already after formula (3.9). In passing we note that formula (2.2) for the LAN situation has been given already on p. 86 of Le Cam & Yang (1990).

If we start with a sequence of experiments that is asymptotically a normal, exponential or loggamma shift experiment, then the above argument can be applied to obtain asymptotic convolution theorems from the Bayes convolution results as presented in Sections 3, 4 and 5. The normal case leads to the classical theorem of Hájek (1970) as we have just seen, whereas the exponential convolution theorem is due to Ibragimov & Has'minskii (1981), Theorem V.5.2, p. 278.

The step from Bayes convolution to Hájek convolution is made by replacing the Bayes prior by an uninformative, improper prior. This is done via a limit procedure. This device may be applied directly to the argument leading from (2.2) to (2.3) for location models. It leads to the result of Boll (1955), mentioned above, as follows. Let $w_\sigma(\cdot) = \sigma^{-1} w(\sigma^{-1} \cdot)$ be the density of a Bayes prior on $\mathbb{R}^k$, and let $\theta_\sigma$ have density $w_\sigma$. We take $w$ bounded, continuous at 0, and fixed, and we let $\sigma$ tend to infinity. Let $X$ be a random vector with location parameter $\theta$ and density $f(\cdot - \theta)$ with $f$ fixed. We choose $\psi$ the identity and $T = t(X)$ an arbitrary translation equivariant estimator of $\theta$. To prove (2.3) with $K$ the distribution with density $f$ it suffices to show that $X - \psi$ and $T - X$ are independent asymptotically, as $\sigma \to \infty$. Indeed, the conditional distribution of $X - \psi$ given $X = x$ does not depend on $x$ in the limit and has density $f$:

$$
\lim_{\sigma \to \infty} P(X - \theta_\sigma \leq z \mid X = x) = \lim_{\sigma \to \infty} \frac{\int_{\mathbb{R}^k} \mathbf{1}_{\{x - \theta \leq 1\}} f(x - \theta) \frac{1}{\sigma} w(\frac{\psi}{\sigma}) d\theta}{\int_{\mathbb{R}^k} f(x - \theta) \frac{1}{\sigma} w(\frac{\theta}{\sigma}) d\theta}
$$

$$
= \lim_{\sigma \to \infty} \frac{\int_{\mathbb{R}^k} \mathbf{1}_{\{y \leq 1\}} f(y) w(\frac{y - x}{\sigma}) dy}{\int_{\mathbb{R}^k} f(y) w(\frac{y}{\sigma}) dy} = \int_{\mathbb{R}^k} \mathbf{1}_{\{y \leq 1\}} f(y) dy, \quad z \in \mathbb{R}^k.
$$

Acknowledgement

We would like to thank editor Vijay Nair, an associate editor and two referees for their suggestions for improved presentation.

References

Bayes Convolution


Résumé

Nous considérerons l’estimation d’un paramètre euclidien \( \theta \) par un estimateur \( T \) dans un modèle paramétrique. Soit \( W \) une distribution a priori pour \( \theta \) et définirons \( G \) par la \( W \)-moyenne de la distribution de \( T - \theta \) sous \( \theta \). Il y a des situations dans lesquelles la distribution \( G \) est une convolution \( G = K \ast L \) pour tout estimateur \( T \), où \( K \) est une distribution qui dépend seulement du modèle, c.-à-d. de \( W \) et des distributions sous \( \theta \) des observations. En cas de ce résultat de convolution bayésienne, des estimateurs optimaux existent qui satisfont à \( G = K \). Pour des modèles de location nous prouvons, qu’il y a des résultats de convolution bayésienne à échantillons finis dans les cas normaux, log-gamma et exponentiels. Sous des conditions de régularité nous prouvons, que les seules situations polies de location sont des normales et log-gamma. Nous discutons aussi des relations avec les théorèmes classiques de convolution.

[Received December 1997, accepted April 1999]