Exact multidimensional credibility

Exact Multidimensional Credibility

By William S. Jewell

University of California, Berkeley

Introduction

Credibility theory is the name given by American actuaries to a linearized Bayesian forecast of the mean observation. In modern notation, let ξ be a discrete or continuous random variable (the *risk*), depending upon a parameter θ (the *risk parameter*) through a *likelihood* density $p(\cdot \mid \theta)$; a *prior density* $u(\theta)$ is assumed known. If n independent risk samples, $\underline{x} = \{\xi_t = x_t ; t = 1, 2, \ldots, n\}$ (the *experience data*), are drawn, then the Bayesian forecast of the mean risk next period (the *experience-rated fair premium*) is just the conditional mean:

$$E\{\xi_{n+1} \mid \underline{x}\} = \iint yp(y \mid \theta)\, dy \frac{\prod_{t=1}^{n} p(x_t \mid \theta)}{\int \prod_{t=1}^{n} p(x_t \mid \Phi)\, u(\Phi)\, d\Phi} u(\theta)\, d\theta. \tag{1}$$

Practically speaking, this expression can only be evaluated with the aid of a computer, or by using natural conjugate prior families [4] [8] of likelihood and prior to carry out the updating.

Based on practical arguments, American actuaries in the 1920's proposed forecasting the mean risk through the *credibility formula*:

$$E\{\xi_{n+1} \mid \underline{x}\} \approx f(\underline{x}) = (1 - Z) \cdot m + Z\left(\frac{1}{n} \sum_{t=1}^{n} x_t\right), \tag{2}$$

$$Z = \frac{n}{N + n}.$$

Here, $m = E_\theta m(\theta) = E_\theta E\{\xi \mid \theta\}$ is the prior mean (the *manual fair premium*), and N is a time constant chosen heuristically. Z is called the *credibility factor*, which increases to unity with increasing data. The interesting feature of (2), as contrasted with (1), is the linear dependence of $f(\underline{x})$ on the data, through the *sample mean* $\frac{1}{n}\Sigma x_t$. The credibility method of experience rating has worked well in the insurance industry for over 40 years.

In the 1950's, Bailey and Mayerson showed that (2) was, in fact, the exact Bayesian result (1) for the Beta-Binomial, Gamma-Poisson, and Normal-Normal conjugate prior likelihood families. Bühlmann then showed, in 1967, that $f(\underline{x})$ was the minimum least-squares linear estimator for arbitrary priors and likelihood provided that $N = E_\theta V\{\xi \mid \theta\} / V_\theta E\{\xi \mid \theta\}$. A fuller historical discussion may be found in [7].

In [8], the author showed that $f(\underline{x})$ was also exact for the *simple* (or linear) *exponential family* (with natural parameterization):

$$p(x \mid \theta) = \frac{a(x)e^{-\theta x}}{c(\theta)}, \; (x \in X) \tag{3}$$

provided that the *natural conjugate prior*,

$$u(\theta) \propto [c(\theta)]^{-n_0} e^{\theta x_0}, \; (\theta \in \Theta) \tag{4}$$

is used over the complete parameter space Θ, and that $u(\theta) \to 0$ at both ends of the range [9]. It turns out that the hyperparameter n_0 is the time constant N of Bühlmann, while $x_0 = m \cdot N$ [8]. Credibility also holds, in an extended sense, for exponential families with different sufficient statistics.

Multidimensional Credibility

There are many applications in which Bayesian analysis of more than one random variable is of interest. For instance, in the collective risk model of casualty insurance, forecasts are needed of the moments of the three dependent random variables {number of claims; total cost of claims; average per-claim cost}.

In the sequel, we shall use certain lower case letters (such as ξ, x, θ) as column vectors. Upper case letters without subscript (X, Ξ, N, E, D) are matrices, with $\Xi_{\cdot t}$ being the vector which is the t^{th} column of Ξ, etc. [3].

Now consider a multivariate extension to (1), in which there is a p-dimensional risk variable ξ and an $p \times n$ data matrix $\Xi = X$, or $\{\xi_{it} = x_{it}; i = 1, 2, \ldots, p; t = 1, 2, \ldots, n\}$, of n independent vector samples, given θ, a vector-valued risk parameter. Define the indexed vectors $\xi_t = \Xi_{.t}$ and $x_t = X_{.t}$. An obvious modification to (1) will give the multidimensional forecast of a selected component s, $E\{\xi_{s,n+1} \mid X\}$ of the next observation, or a vector forecast $E\{\xi_{n+1} \mid X\}$. In [7], the author gives the multidimensional credibility (linear least-squares) result corresponding to (2). Let:

$$m_i(\theta) = E\{\xi_i \mid \theta\} \quad ; m_i = E_\theta m_i(\theta)$$

$$C_{ij}(\theta) = C\{\xi_i; \xi_j \mid \theta\}; E_{ij} = E_\theta C_{ij}(\theta) \qquad (i,j = 1, 2, \ldots, p) \tag{5}$$

$$D_{ij} = C_\theta\{m_i(\theta) \quad ; m_j(\theta)\}$$

and let $m(\theta)$, m, $C(\theta)$, E, and D be the corresponding vector and matrix quantities. (Here, as in the rest of the paper, we assume all indicated moments exist.) E and D being covariance matrices, they must be symmetric and non-negative definite. We shall additionally require D to be positive definite, so that D^{-1} exists (see [3], page 106); otherwise, with probability one $m(\theta)$ lies in some hyperplane for all θ, and at least one dimension of the forecast problem is uninteresting.

Now define the square matrices N, Z:

$$N = ED^{-1}, Z(E + nD) = nD. \tag{6}$$

From the results given in [7], one can then show that the vector forecast function $f(X) = E\{\Xi_{.(n+1)} \mid X\}$ is given by the following generalization of (2):

$$f(X) = (I - Z)m + Z\bar{x} \tag{7}$$

where I is the identity matrix, and $\bar{x}$ is the vector of sample means. The *credibility matrix* Z satisfies formulae analogous to the one-dimensional result:

$$Z = n(N + nI)^{-1}; (I - Z) = \frac{1}{n}(ZN) = \frac{1}{n}(NZ). \tag{8}$$

If the eigenvalues of N are $\{v_i\}$, then those of Z are $\{n/(n+v_i)\}$; one can show that in the nondegenerate case $\lim_{n\to\infty} Z = I$. In other words, in the multidimensional case, the initial forecast (no data) is the prior mean m; successive forecasts utilize linear mixtures of *all* sample means in varying proportions, but then, ultimately, each component of the risk is estimated only through its own sample mean, as $n \to \infty$. Specific examples are given in [7].

The purpose of this paper is to show in what sense the multidimensional credibility formula (7) is exact, that is, to find the multidimensional families for which the Bayesian conditional mean is linear in the data X. Contrary to expectations, it will turn out that the simple one-dimensional credibility formula holds for each dimension of many multivariate distributions and a "natural" conjugate prior. In order to require the full multidimensional generalization (7), we shall have to consider special likelihood families, and enrich the associated priors.

Multidimensional Exponential Family

The *linear multivariate exponential family* with natural parameterization θ has a likelihood density:

$$p(x \mid \theta) = \frac{a(x)\, e^{-\theta' x}}{c(\theta)} \qquad (x \in X). \tag{9}$$

The normalization factor, $c(\theta)$, is determined by

$$c(\theta) = \iiint_{x \in X} a(x)\, e^{-\theta' x}\, dx \qquad (\theta \in \Theta), \tag{10}$$

and the complete parameter space Θ consists of all points in R^p for which (10) is finite; Θ is known to be convex. Studies of distributions in this family have been carried out by Bildikar and Patil [2] (see also [10]); it includes, for example, the multinational with known precision. It is also known that this is the only family which, subject to certain regularity conditions and a sample space X independent of θ' has the sample mean vextor $\bar{x}$ as a sufficient statistic [4]. The family is characterized by certain relationships between its moments [2]

$$m_i(\theta) = \frac{-\dfrac{\partial c(\theta)}{\partial \theta_i}}{c(\theta)}; \; C_{ij}(\theta) = \frac{\dfrac{\partial^2 c(\theta)}{\partial \theta_i \, \partial \theta_j}}{c(\theta)} = -\frac{\partial m_i(\theta)}{\partial \theta_j} = -\frac{\partial m_j(\theta)}{\partial \theta_i}. \tag{11}$$

Credibility for a Simple Prior

Natural conjugate priors for specific multivariate distributions have been developed by several authors [1] [10]; however, there seems to be no discussion in the literature of priors conjugate to (9). Following (4), we shall first assume the (scalar function) prior:

$$u(\theta) \propto [c(\theta)]^{-n_0} e^{-\theta' x_0} \qquad \theta \in \Theta \tag{12}$$

where the scalar n_0 and the vector $x_0 = \{x_{i0};\, i = 1, 2, \ldots, p\}$ are hyperparameters. Since the likelihood of n independent vector samples X governed by (9) is proportional to $[c(\theta)]^{-n} \exp\left\{-\sum_{i=1}^{p} \theta_i \sum_{t=1}^{n} x_{it}\right\}$, it follows that the prior (12) is *closed under sampling*, i.e., $u(\theta \mid X)$ is of the same form as (12), with the hyperparameters updated by:

$$\begin{aligned} n_0 &\leftarrow n_0 + n \\ x_0 &\leftarrow x_0 + \sum_{t=1}^{n} x_t . \end{aligned} \tag{13}$$

Following the method used in [8], we assume that $u(\theta) \equiv 0$ everywhere on the boundary of Θ to relate n_0 and x_0 to moments of the mixed density through (5) and (11). First

$$\frac{\partial u(\theta)}{\partial \theta_i} \propto [n_0 \cdot m_i(\theta) - x_{i0}] \cdot u(\theta), \qquad (i = 1, 2, \ldots, p) \tag{14}$$

so that

$$\iiint_{\Theta} \frac{\partial u(\theta)}{\partial \theta_i}\, d\theta = n_0 \cdot E_\theta m_i(\theta) - x_{i0} .$$

But, by definition of Θ, the first integral is zero, so that

$$m_i = E_\theta m_i(\theta) = \frac{x_{i0}}{n_0}. \tag{15}$$

Further,

$$\frac{\partial^2 u(\theta)}{\partial\theta_i\,\partial\theta_j} \propto \left[-n_0 C_{ij}(\theta) + \left(n_0 m_i(\theta) - x_{i0}\right)\left(n_0 m_j(\theta) - x_{j0}\right)\right] u(\theta), \tag{16}$$

and under the assumption that $\iiint\limits_{\Theta} \frac{\partial^2 u}{\partial\theta_i\,\partial\theta_j}\, d\theta \equiv 0$ we find

$$E_{ij} = E_\theta\, C_{ij}(\theta) = n_0\, C_\theta\{m_i(\theta)\,;\, m_j(\theta)\} = n_0 D_{ij} \tag{17}$$

for all i, j. Thus N is a diagonal matrix, $n_0 I$.
Since $E\{\xi_{s,n+1} \mid X\} = E_\theta\, m_s(\theta)$, it follows from (13) and (15) that

$$E\{\xi_{s,n+1} \mid X\} = \frac{x_{s0} + \sum_{t=1}^{n} x_{st}}{n_0 + n} = \frac{n_0 \cdot m_s}{n_0 + n} + \frac{n}{n_0 + n}\left(\frac{1}{n}\sum_{t=1}^{n} x_{st}\right). \tag{18}$$

Thus, with (12) as natural conjugate prior, we reach the embarrassing conclusion that there is only a single time constant n_0 operating in the multivariate case, Z is a diagonal matrix with entries $n/(n_0 + n)$, and each component is predicted by a one-dimensional credibility formula (2)! We shall refer to this degenerate result as a *self-dimensional* credibility forecast.
The condition $u(\theta) \equiv 0$ on the boundary of Θ is satisfied in most problems of practical interest, and is, fortunately, easily checked. An examination of this condition in the one-dimensional case may be found in [9].

Enriched Priors for Linearly Dependent Exponential Families

In order to require full multidimensional credibility, additional hyperparameters and functions of θ must be used to enrich the prior; however, these functions must be chosen in a special way so as to obtain a linear mixture of means.
A clue to the correct enrichment can be gotten by imagining first that there was a vector of *independent* risks η, with a vector of natural parameters Φ, *all of whose marginal distributions were in the exponential family*. The joint density of $\eta = y$, analogous to (9), would be of the form:

$$p(y \mid \Phi) = \frac{\prod b_i(y_i)\, e^{-\Phi' y}}{\prod d_i(\Phi_i)}\,;\; d_i(\Phi_i) = \int e^{-\Phi_i y_i} b_i(y_i)\, dy_i \qquad (i = 1, \ldots, p) \tag{19}$$

for appropriate functions $b_i(\cdot)$ and $d_i(\cdot)$. Here y_i and Φ_i are components of y and Φ, respectively. Because of the assumed independence, the overall normalization $d(\Phi)$ is a product of normalization factors.

An appropriate conjugate prior, $v(\Phi)$, would also be a product of individual priors:

$$v(\Phi) \propto \prod \left[d_i(\Phi_i)\right]^{-n_{i0}} e^{-\Phi' y_0} \tag{20}$$

but with hyperparameters n_{i0} and $y_0 = \{y_{i0};\ i = 1, 2, \ldots, p\}$ for each component. Once again, the argument leading to (15) would give independent credibility forecasts for each component, with the difference that each component, η_i, would have *its own time constant*, n_{i0}. Or in vector terminology, (7) applies with N a special diagonal matrix N_0:

$$N_0 = \mathrm{diag}\,\{n_{10}, n_{20}, \ldots, n_{p0}\}. \tag{21}$$

Now consider again the general multivariate exponential-type density (9), and suppose there exists an invertible linear transformation, defined by a square matrix A and a vector k, that

$$\xi = A\eta + k \qquad |A| \neq 0 \tag{22}$$

causes $a(Ay + k)$ to factor *into the product of independent components* $b_i(y_i)$ in each variable y_i $(i = 1, 2, \ldots, p)$. We will call such a likelihood a *linearly-dependent multivariate exponential family* (LDMEF) distribution.

Defining the vector of parameters θ by

$$\Phi = A'\theta, \tag{23}$$

and using definition (19) of the $d_i(\cdot)$, the normalization factor $c(\theta)$ can be decomposed into:

$$c(\theta) = \prod_{i=1}^{p} d\left(\sum_h \theta_h A_{hi}\right) \cdot e^{-k_i \theta_i} \tag{24}$$

where A_{hi} and k_i are the elements of A and k, respectively. Then in place of (12), we define the *enriched prior for the LDMEF family* as:

$$u(\theta) \propto \prod_{i=1}^{p} \left[d_i\left(\sum_h \theta_h A_{hi}\right)\right]^{-n_{i0}} e^{-\theta' x_0}. \tag{25}$$

[Note that the translation factors k_i which effect the factorization of $a(x)$ can be absorbed into the definition of the x_{i0} in the prior; for the most part, we shall neglect k in the sequel.] The parameter updating becomes:

$$N_0 \leftarrow N_0 + nI\,;\; x_0 \leftarrow x_0 + \sum_{t=1}^{n} x_t = x_0 + n \cdot \bar{x}\,. \tag{26}$$

Theorem:
If $p(x \mid \theta)$ belongs to the *linearly-dependent multivariate exponential family* (LDMEF) as defined above, and the *enriched natural conjugate prior* (25) is used with hyperparameters $n_{i0} > 0$ $(i = 1, 2, \ldots, p)$, the full multi-dimensional credible forecast (7), (8) is used to predict $E\{\xi_{n+1} \mid X\} = f(X)$ with:

$$N = AN_0A^{-1} = ED^{-1}\,. \tag{27}$$

Add: providing that $u(\theta)$ and $\partial u(\theta)/\partial\theta$ one zero everywhere on the boundary of Θ.

Proof:
Neglect k, and assume $x = Ay$ factors $a(x)$ into $\Pi b_i\,(y_i)$. Directly from the enriched prior (25):

$$\frac{\partial}{\partial\theta_j}u(\theta) = \left[\sum_k A_{jk}n_{0k}\left(-\frac{\dfrac{dd_k\,(\Phi_k)}{d\Phi_k}}{d(\Phi_k)}\right) - x_{j0}\right]\cdot u(\theta) \qquad (j = 1, 2, \ldots, p)\,. \tag{28}$$

Now from (19), the term in parentheses is just $E\{\eta_{kt} \mid \Phi\}$, and $E\{\xi \mid \theta\} = m(\theta) = A \cdot E\{\eta \mid \Phi\}$. Using definition (21) of N_0, we rewrite (28) in vector notation as:

$$\frac{\partial}{\partial\,\theta}u(\theta) = \left[A \cdot N_0 \cdot A^{-1} \cdot m(\theta) - x_0\right] \cdot u(\theta)\,. \tag{29}$$

Since we assume $u(\theta) \equiv 0$ everywhere on the boundary of Θ [see (14), (15)], the expectation of the term in brackets vanishes, and

$$N \cdot m = x_0\,, \tag{30}$$

with N defined by (27), and updated by (26).
Using (11), the matrix of second derivatives of $u\,(\theta)$ is [see (16), (17)]:

$$\frac{\partial^2}{\partial\theta\partial\theta'} u(\theta) = \{-N \cdot C(\theta) + [N \cdot m(\theta) - x_0]\,[N \cdot m(\theta) - x_0]'\} \cdot u(\theta) \quad (31)$$

and under the assumption that the various gradients of $u(\theta)$ are zero everywhere on the boundary of Θ, the expectation of the term in braces vanishes, and from (5):

$$N \cdot E = C_0\{N \cdot m(\theta)\} = N \cdot D \cdot N'. \quad (32)$$

Now E and D are symmetric, so $E = ND$. Since D^{-1} is assumed to exist, $N = ED^{-1}$.
The desired forecast is $f(X) = E_{0|X} m(\theta)$, so from the updating (26) and from (29):

$$E_{0|X}\left[A(N^0 + nI)A^{-1}m(\theta) - (x_0 + n\bar{x})\right] = 0$$

or

$$(N + nI)f(X) = N \cdot m + n\bar{x},$$

which is the multidimensional forecast (7), (8). Q.E.D.

The key point to requiring full credibility is the factorization of $a(x)$ [and hence $c(\theta)$] after a linear transformation of variables. If such a transformation only factors out less than p mutually independent random variables η_i, then this limits the amount of enrichment possible in the prior, leading to a degenerate form of N. Details are left to the reader.

Multinormal with Random Mean and Known Precision

As an example, consider the nonsingular p-dimensional multivariate normal with unknown mean vector μ, and a known (symmetric) precision matrix W (covariance W^{-1}). In the usual notation, the likelihood is:

$$p(x \mid \mu) = (2\pi)^{-\frac{p}{2}} \mid W \mid^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)'\,W\,(x-\mu)\right\}. \quad (33)$$

This is clearly in the exponential family, with $\theta = -W\mu$, and normalizing factor

$$c(\theta) = \left(2\pi\right)^{\frac{p}{2}} | W |^{-\frac{1}{2}} \exp\left\{\frac{1}{2}\theta' W^{-1}\theta\right\}. \tag{34}$$

If we take the simple conjugate prior (12), we would have:

$$u(\theta) \propto \exp\left\{-\frac{1}{2}\theta' \left(n_0 W^{-1}\right)\theta - x_0'\theta\right\}, \tag{35}$$

or upon completing the square and returning to a traditional notation

$$u(\mu) \propto \exp\left\{-\frac{1}{2}\left(\mu - \left(\frac{1}{n_0}\right)x_0\right)' (n_0 W)\left(\mu - \left(\frac{1}{n_0}\right)x_0\right)\right\}, \tag{36}$$

that is, a multinormal with mean $\left(\frac{1}{n_0}\right)x_0$ and precision $n_0 W$.

Since the mean and covariance of the likelihood are:

$$m(\mu) = \mu;\ C(\mu) = W^{-1},$$

it follows that

$$m = \left(\frac{1}{n_0}\right)x_0;\ E = EC(\mu) = W^{-1};\ D = (n_0 W)^{-1}.$$

Thus, we get

$$N = ED^{-1};\ x_0 = N\cdot m,$$

only with $N = n_0 I$, i.e., the one-time-constant credibility forecast (18) holds. Enriching the prior amounts to replacing $n_0 W^{-1}$ in (35) by NW^{-1}. The new prior is then:

$$u(\mu) \propto \exp\left\{-\frac{1}{2}\left(\mu - N^{-1}x_0\right)' (WN)\left(\mu - N^{-1}x_0\right)\right\} \tag{37}$$

i.e., a multinormal with mean $N^{-1}x_0$ and precision WN. Clearly,

$$m = N^{-1}x_0;\ E = W^{-1};\ D = (WN)^{-1}$$

so that $x_0 = N \cdot m$; N is not an arbitrary matrix of hyperparameters, but must satisfy $N = ED^{-1}$, where E and D are the symmetric matrices of prior mean covariance and covariance of the means, respectively.
The reader may easily verify that the updating of the prior is governed by (26), so that the full multidimensional forecast (7) (8) applies. As a by-product, the posterior distribution of μ is multinormal, with a mean given by the same credibility formula, and with updated precision $(N + nI)W = D^{-1} + nE^{-1} = D^{-1}(I + nN^{-1})$. The latter results are well known (see, for example, [4], page 175).

Credible Means for the General Multiexponential Family

In principle, there is no difficulty extending the above arguments to generalized multiexponential families of the form:

$$p(x \mid \theta) = \frac{a(x)\exp\left(-\theta' f(x)\right)}{c(\theta)} \quad (x \in X) \tag{38}$$

by following the approach given in [8]. Here, the various functions $f(x) = \{f_i(x);\ i = 1, 2, \ldots, p\}$ are sufficient statistics, and credibility updates their mean values.
The normalization is, as usual,

$$c(\theta) = \int_X a(x)\exp\left(-\theta' f(x)\right)\, dx \tag{39}$$

for θ in the natural parameter space Θ for which (39) is finite.
Generalized mean vectors

$$m(\theta) = E_{\xi \mid \theta} f(\theta) = -\frac{\partial}{\partial \theta} \ln c(\theta) \tag{40}$$

and covariance matrices

$$C(\theta) = C_{\xi|\theta}\{f(\xi)\} = + \frac{\partial^2}{\partial\theta\theta'} \ln c(\theta), \tag{41}$$

as defined in the obvious manner, and assumed to exist.
If the simple natural conjugate prior

$$u(\theta) \propto \left[c(\theta) \right]^{-n_0} \exp\left(-\sum f_0' \theta\right) \tag{42}$$

is assumed, we get self-dimensional forecasts in the form

$$E\{f_s(\xi_{n+1}) \mid X\} = \frac{f_{0s} + \sum_t f_s(x_t)}{n_0 + n}, \tag{43}$$

in an obvious analogy to (18) above and to (36) in [8].
Clearly one could get various enrichments of (36) in special cases where the $f_i(x)$ factored under various linear transformations of x, and thus get a multi-dimensional credibility forecast. Rather than attempting to extend the theory any more, we must examine special cases where this factorization is possible.

Multinormal with Random Mean and Precision

Simple Prior

If we now consider the p-dimensional multinormal in which the precision matrix ω is also random (and symmetric with probability 1), the likelihood is:

$$p(x \mid \mu, \omega) = \left(2\pi\right)^{-\frac{p}{2}} |\omega|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu)'\,\omega(x-\mu)\right\}. \tag{44}$$

This is in the generalized multiexponential family, sometimes called the quadratic exponential family,

$$p(x \mid \theta, \omega) = \frac{a(x) \exp\left\{-\theta' x - \frac{1}{2} x' \omega x\right\}}{c(\theta, \omega)} \tag{45}$$

if we make the identification $\theta = -\omega\mu$, $a(x) \equiv 1$, and

$$c(\theta, \omega) = (2\pi)^{\frac{p}{2}} \mid \omega \mid^{-\frac{1}{2}} \exp\left\{\frac{1}{2} \theta' \omega^{-1} \theta\right\}. \tag{46}$$

We extend the idea of a simple natural conjugate prior (12) by adding a symmetric matrix $Q_0 = \{Q_{ij0};\, i, j = 1, 2, \ldots, p\}$ of $\frac{1}{2}p(p+1)$ hyperparameters (for the random variables ω) to the usual $p + 1$ hyperparameters n_0 and x_0:

$$u(\theta, \omega) \propto \left[c(\theta, \omega)\right]^{-n_0} \exp\left\{-\theta' x_0 - \frac{1}{2} tr\,(\omega Q_0)\right\}. \tag{47}$$

The Jacobian of the transformation is $\mid \omega \mid$, so in traditional notation

$$u(\mu, \omega) \propto \mid \omega \mid^{\frac{1}{2}(n_0 + 2)} \exp\left\{-\frac{1}{2} \mu' (n_0 \omega) \mu + \mu' \omega x_0 - \frac{1}{2} tr\,(\omega Q_0)\right\} \tag{48}$$

which can be seen to be a *Normal-Wishart* distribution by factoring $u(\mu, \omega)$ into $u(\mu \mid \omega) \cdot u(\omega)$.

For the conditional distribution of the mean, we take:

$$u(\mu \mid \omega) \propto \mid \omega \mid^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mu - \frac{x_0}{n_0}\right)' (n_0 \omega) \left(\mu - \frac{x_0}{n_0}\right)\right\} \tag{49}$$

that is, a *multinormal* distribution, with mean vector x_0 / n_0 and precision matrix $n_0 \omega$.
This leaves, after some algebra,

$$u(\omega) \propto \mid \omega \mid^{\frac{1}{2}(n_0 + 1)} \exp\left\{-\frac{1}{2} tr\,(\omega U_0)\right\}, \tag{50}$$

where

$$U_0 = Q_0 - (x_0 x_0' / n_0) \tag{51}$$

which can be recognized as a *Wishart* distribution with $\alpha = n_0 + p + 2$ degrees of freedom, and precision-parameter matrix U_0. For future reference, the covariance of the likelihood, $\Sigma = \omega^{-1}$, has mean [10]:

$$E\{\omega^{-1}\} = \frac{1}{\alpha - p - 1} U_0 = U_0 / (n_0 + 1). \tag{52}$$

Formulae for the covariance of the covariance are also available [12].
The marginal distribution of μ, after some algebra (see, e.g., [4]), is then:

$$u(\mu) \propto \left| 1 + n_0 \left(\mu - \frac{x_0}{n_0}\right)' U_0^{-1} \left(\mu - \frac{x_0}{n_0}\right) \right|^{-\frac{1}{2}(n_0 + p + 3)}, \tag{53}$$

which is a *multivariate Student-t* distribution, with $\beta = n_0 + 3$ degrees of freedom, location vector x_0/n_0, and precision-parameter matrix $T = n_0(n_0 + 3) U_0^{-1}$. For future reference:

$$E\{\mu\} = x_0/n_0, \tag{54}$$

$$C\{\mu\} = \frac{\beta}{\beta - 2} T^{-1} = U_0/n_0 (n_0 + 1). \tag{55}$$

In this case, we can also get an explicit form for the mixed (predictive) density, $p(x) = \int p(x \mid \mu, \omega) u(\mu, \omega) \, d\mu \, d\omega$, by retracing some of the above steps, as in pp. 178–180 of [4]. We find that an arbitrary vector ξ (no data) is distributed as *multivariate Student-t*, with $\beta = n_0 + 3$ degrees of freedom, location vector x_0/n_0, and precision-parameter matrix $T = n_0(n_0 + 3)(n_0 + 1)^{-1} U_0^{-1}$. Updating of the parameters can be shown to follow:

$$n_0 \leftarrow n_0 + n \tag{56}$$

$$x_0 \leftarrow x_0 + \sum_{t=1}^{n} x_t \tag{57}$$

$$Q_0 \leftarrow Q_0 + \sum_{t=1}^{n} x_t x_t'. \tag{58}$$

The relationship with the moments defined in (5) is then:

$$m = x_0 / n_0 , \tag{59}$$

$$E = U_0 / (n_0 + 1) , \tag{60}$$

$$D = U_0 / n_0 (n_0 + 1) , \tag{61}$$

so that, as expected, the Normal-Wishart prior (48) is of the simple type, with $N = n_0 I$, giving a self-dimensional forecast (18) of the mean.
Since $C\{\xi\} = D + E = U_0 / n_0$, it follows that a simple forecast of the covariance can also be given:

$$C\{\xi_{n+1} \mid X\} = (1 - Z)\, C\{\xi\} + Z \left[\frac{1}{n} \sum_{t=1}^{n} (x_t - \overline{x})(x_t - \overline{x})' \right] + Z(1 - Z)(m - \overline{x})(m - \overline{x})' , \tag{62}$$

where, of course, Z is a scalar, equal to $n/(n_0 + n)$. This should be compared with Equation (41) of [8].
The prior (48) was discovered by Ando and Kaufman [1], and its "thinness" seems to be well known in the literature. The usual criticism is that one cannot set both the means and covariances of both μ and $\Sigma = \omega^{-1}$ independently. From our point of view, the limitation is that $E = n_0 D$, so that the two covariance components of observations from the collective cannot be chosen independently.
Actually, the prior of Ando and Kaufman is slightly more general than (48), being multiplied by an additional factor

$$| \omega |^{\frac{1}{2}(\alpha - p - n_0 - 2)} ,$$

with α arbitrary, but greater than $p - 1$. Thus, their joint prior begins with a term

$$| \omega |^{\frac{1}{2}(\alpha - p)} .$$

This leads to the following changes in the marginal priors, as described above:

$u(\mu \mid \omega)$ — no change;
$u(\omega)$ — α degrees of freedom;
$u(\mu)$ — $\beta = \alpha - p + 1$ degrees of freedom, and precision-parameter matrix $T = n_0(\alpha - p + 1)U_0^{-1}$;
$p(x)$ — $\beta = \alpha - p + 1$ degrees of freedom, and precision-parameter matrix $T = n_0(\alpha - p + 1)(n_0 + 1)^{-1}U_0^{-1}$.

This changes (60) and (61) to:

$$E = U_0/(\alpha - p + 1) \tag{60'}$$

$$D = U_0/n_0(\alpha - p + 1), \tag{61'}$$

but clearly does not affect the updating, the forecasting, or the fact that $E = n_0 D$. In this sense, α is an invariant nuisance hyperparameter which only scales the variance of the observations, independent of the mean.
In some unpublished work [11] [12], Kaufmann further enriches this prior by multiplying the Wishart distribution $u(\omega)$ by arbitrary powers of the products of determinants of principal minors of ω, thus introducing $p - 1$ additional hyperparameters, and eliminating some of the objections to the Ando-Kaufman prior. However, the formulae are quite complicated, and do not appear relevant to the credibility problem.

Enriched Prior

To obtain a more general multinormal prior which will require full credibility, we follow a previous argument, and start with a vector of *independent* risks η, distributed as a (degenerate) multinormal, with mean vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)'$, and a *diagonal* precision matrix $\pi = \text{diag}\{\pi_1, \pi_2, \ldots, \pi_p\}$. The joint density of $\eta = y$ can be expressed as a product of independent one-dimensional normals, or as:

$$p(y \mid \lambda, \pi) \propto |\pi|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - \lambda)' \pi (y - \lambda)\right\} \tag{63}$$

remembering the diagonal form of π, and not confusing it with 3.14159 ....

For each component $i = 1, 2, \ldots, p$, we take a one-dimensional prior distribution of the type suggested by (20) and (48):

$$v_i(\lambda_i, \pi_i) \propto \left(\pi_i\right)^{\frac{1}{2}(n_{0i}+2)} \exp\left\{-\frac{1}{2}\lambda_i^2(n_{i0}\pi_i) + \lambda_i\pi_i y_{i0} - \frac{1}{2}R_{i0}\pi_i\right\} \tag{64}$$

where there are three hyperparameters $\{n_{i0}, y_{i0}, R_{i0}\}$ for each dimension. This is clearly a one dimensional *Normal-Gamma*, or *Normal-Chi Squared* [8], and the marginals and moments follow from the last section, using n_{i0} instead of n_0, y_{i0} instead of x_0, R_{i0} instead of Q_0, and $p = 1$. Each η_i is predicted independently by a credibility formula with *different* time constant n_{i0}.

The joint density $v(\lambda, \pi) = \prod_{i=1}^{p} v_i(\lambda_i, \pi_i)$ can be written in matrix form as:

$$v(\lambda, \pi) \propto \left[\prod_{i=1}^{p}\left(\pi_i^{\frac{1}{2}(n_{i0}+2)}\right)\right] \exp\left\{-\frac{1}{2}\lambda'(\pi N_0)\lambda + \lambda'\pi y_0 - \frac{1}{2}tr(\pi R_0)\right\}. \tag{65}$$

N_0 is given in (21), $R_0 = \text{diag}\{R_{10}, R_{20}, \ldots, R_{p0}\}$, and $y_0 = (y_{10}, y_{20}, \ldots, y_{p0})'$.

We then obtain a general linearly-dependent multinormal by transforming the random variables as follows:

$$x = Ay; \quad \mu = A\lambda; \quad \omega^{-1} = A\pi^{-1}A^{-1}, \tag{66}$$

where A is an invertible $p \times p$ matrix, otherwise arbitrary. μ and ω are now the vector mean and matrix precision of ξ, with likelihood (44).

Changing the prior (65) to the variables (μ, ω) can be most easily accomplished by defining new hyperparameters:

$$x_0 = Ay_0; \quad N = AN_0A^{-1}; \quad Q_0 = AR_0A'. \tag{67}$$

The first two relations are those of the section on LDMEF distributions; the last transformation follows from the identity $tr(A'B) = tr(BA)$.

The enriched-multinormal prior is then:

$$u(\mu,\omega) \propto \left[\prod_{i=1}^{p}\left(\sum_{j,k} A_{ji}\omega_{jk}A_{ki}\right)^{\frac{1}{2}(n_{i0}+2)}\right] \exp\left\{-\frac{1}{2}[\mu'(\omega N)\mu] + \mu'\omega x_0 - \frac{1}{2}tr(\omega Q_0)\right\} \tag{68}$$

which, when compared with (48), has a generalized product term in front, and a precision factor ωN in place of $n_0\omega$. We shall call this prior the *linearly-dependent Normal-Wishart* (LDNW) distribution.
The factorization can be carried out as before, using the easily proven identities $u'Av = tr\,(Avu')$ (u, v arbitrary vectors), and $\omega N = N'\omega'$. We get:

$$u(\mu \mid \omega) \propto |\omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mu - N^{-1}x_0\right)' (\omega N)\left(\mu - N^{-1}x_0\right)\right\} \tag{69}$$

and

$$u(\omega) \propto \left[\prod_i \left(\sum_{j,k} A_{ji}\omega_{jk}A_{ki}\right)^{\frac{1}{2}(n_{i0}+1)}\right] \exp\left\{-\frac{1}{2}\,tr\,(\omega U_0)\right\} \tag{70}$$

where now

$$U_0 = Q_0 - x_0\left(N^{-1}x_0\right)' = AV_0A' \tag{71}$$

and V_0 is the diagonal matrix

$$V_0 = R_0 - y_0y_0'N_0^{-1}. \tag{72}$$

The conditional distribution of the mean vector, given ω, is thus *multinormal*, with mean vector $N^{-1}x_0$, and precision matrix ωN.
The marginal distribution of the precision (70) is a new generalization of a Wishart distribution obtained by a linear mixture of one-dimensional Gamma or Chi Squared distributions, each with a different degree of freedom. We shall call (70) the *linearly-dependent Wishart* (LDW) distribution. Because of the third relation (66), the expected covariance has simple form:

$$E\{\omega^{-1}\} = (N + \mathrm{I})^{-1}U_0. \tag{73}$$

To obtain the marginal distribution of the mean vector, one begins with the individual Student-t distributions, with different degrees of freedom for each dimension:

$$v(\lambda_i) \propto \left\{1 + (n_{i0}/V_{ii0})\,[\lambda_i - (y_{i0}/n_{i0})]^2\right\}^{-\frac{1}{2}(n_{i0}+4)} \tag{74}$$

and then mixes the product of these marginals according to $\mu = A\lambda$. It seems difficult to get the resulting distribution into matrix form because of the different exponents. The moments, however, follow easily from the definitions:

$$E\{\lambda\} = N^{-1}x_0 \tag{75}$$

$$C\{\lambda\} = (N+I)^{-1}N^{-1}U_0 = N^{-1}(N+I)^{-1}U_0 . \tag{76}$$

For completeness, we could call this a *linearly-dependent Student-t* (LDST) distribution. Generalizations of this and more complicated types have been studied in conjunction with multivariate analysis; see, for example, [5].

By reasoning similar to that described in the last section, the predictive density for the vector ξ is also LDST, with a density similar to that which will be obtained from (74) by mixing, but with n_{i0} / V_{ii0} replaced by $n_{i0} / (n_{i0}+1) V_{ii0}$.

From the above, we then see that moments of ξ with the extended LDNW prior (68) will be:

$$m = N^{-1}x_0 \tag{77}$$

$$E = (N+I)^{-1}U_0 \tag{78}$$

$$D = N^{-1}(N+I)U_0 \tag{79}$$

from which we see that $E = ND$, and the general credibility forecast (7) of the mean will be exact, since the updating will follow

$$N \leftarrow N + nI \tag{80}$$

and (57) (58).

Since $C\{\xi\} = N^{-1}U_0$, the general credibility forecast of the covariance can be gotten from examining the updating of the hyperparameters. After some algebra, and liberal use of symmetry, we find

$$C\{\xi_{n+1} \mid x\} = (I-Z)\,C\{\xi\} + Z\left[\frac{1}{n}\sum_{t=1}^{n}(x_t-\overline{x})(x_t-\overline{x})'\right] + (I-Z)(m-\overline{x})(\overline{m}-\overline{x})'Z' \tag{81}$$

where now Z is the general credibility matrix function (8).

We believe that the enriched LDNW prior (68) is a new multinormal prior, and has not been given before in the literature. It clearly answers many of the objections to the Ando and Kaufman prior, having a total of $3p + p^2$ independent hyperparameters. Although the marginals cannot be written in neat matrix form, the moments are trivially obtained through linear transformations. It can, of course, be slightly generalized by making the degrees of freedom of the LDW distributions into general values a_i; we leave the details to the reader.

Summary

What we have attempted to show in this paper is that exact credibility can be required in the multidimensional case only in certain special cases. This is because, for the multidimensional exponential family of likelihoods, it is difficult to construct a conjugate prior which is rich enough to specify both components E and D of the collective covariance independently. In certain cases, such as the LDMEF family and the multinormal, full-dimensional credibility can be exact because the prior can be enriched because certain factorizations are possible. In other cases, one will either have to accept the limitations of the simple prior, develop special formulae for the prior of interest, or else use multidimensional credibility as an approximation, then valid for any prior.

References

[1] *Ando, A.* and *G. Kaufmann:* "Bayesian Analysis of the Independent Multinormal Process – Neither Mean Nor Precision Known", *Journal of the American Statistical Association*, Vol. 60, pp. 347–358 (1965).

[2] *Bildikar, S.* and *G. P. Patil:* "Multivariate Exponential-Type Distributions", *Annals of Mathematical Statistics,* Vol. 39, No. 4, pp. 1316–1326 (1968).

[3] *Bodewig, E.:* "Matrix Calculus", North-Holland, 452 pp. (1959).

[4] *DeGroot, M. H.:* "Optimal Statistical Decisions", McGraw-Hill, 489 pp. (1970).

[5] *Dickey, J. M.:* "Three Multidimensional-Integral Identities with Bayesian Applications", *Annals of Mathematical Statistics*, Vol. 39, No. 5, pp. 1615–1627.

[6] *Ferguson, T. S.:* "Mathematical Statistics: A Decision Theoretic Approach", Academic Press, 396 pp. (1967).

[7] *Jewell, W. S.:* "Multidimensional Credibility", ORC 73-7, Operations Research Center, University of California, Berkeley (April 1973); to appear in *Journal of Risk and Insurance.*

[8] *Jewell, W. S.:* "Credible Means Are Exact Bayesian for Simple Exponential Families", ORC 73-21, Operations Research Center, University of California, Berkeley (October 1973); to appear in *ASTIN Bulletin.*

[9] *Jewell, W. S.:* "Regularity Conditions for Exact Credibility", ORC 74-22, Operations Research Center, University of California, Berkeley (July 1974).

[10] *Johnson, N. L.* and *S. Kotz:* "Distributions in Statistics: Continuous Multivariate Distributions", John Wiley and Sons, 333 pp. (1972).

[11] *Kaufman, G. M.:* "Some Extensions of Natural Conjugate Analysis of the Multivariate Normal and Normal Regression Processes", unpublished (1965).

[12] *Kaufman, G. M.:* "Some Bayesian Moment Formulae", Discussion Paper No. 6710, CORE, Louvain (April 1967).

Zusammenfassung

In einer früheren Arbeit wurde die Exaktheit von Credibilityformeln bewiesen für eindimensionale Versicherungen der Exponentialfamilie und dazugehörige konjugierte A-priori-Verteilungen. Die vorliegende Arbeit befasst sich mit dem mehrdimensionalen Fall und den Bedingungen, unter welchen die übliche Credibilityformel gleich dem exakten Aposteriori-Erwartungswert ist. Nimmt man mehrdimensionale Exponentialverteilungen und deren einfachste Konjugierte als Strukturfunktionen, so ist der zugehörige Aposteriori-Erwartungswert zwar linear, aber trivial, indem nur Beobachtungen der zu schätzenden Komponente in der Schätzfunktion vorkommen. Sind dagegen in einem praktisch wichtigen Spezialfall auch gewisse allgemeinere a priori Verteilungen zugelassen, so erhält man echte mehrdimensionale Credibilityformeln. Dasselbe Vorgehen liefert auch neue verallgemeinerte konjugierte Verteilungen zur mehrdimensionalen Normalverteilung mit unbekanntem Mittelwert und unbekannter Streuung.

Résumé

Dans un travail antérieur, il a été démontré l'exactitude de formules de crédibilité en ce qui concerne les distributions à une dimension de la famille des exponentielles et les distributions a priori conjuguées qui s'y rapportent. La présente étude traite du cas à plusieurs dimensions et des conditions dans lesquelles la formule habituelle de crédibilité est égale à l'exacte espérance mathématique a posteriori. En admettant comme fonctions de structure des distributions exponentielles à plusieurs dimensions et leur plus simple conjuguée, l'espérance mathématique a posteriori correspondante est linéaire, mais triviale, car la fonction d'estimation ne comprend que des observations sur la composante à estimer. Si, toutefois, dans un cas particulier, important dans la pratique, on tient également compte de certaines distributions a priori plus générales, on obtient de véritables formules de crédibilité à plusieurs dimensions. La même démarche conduit aussi à de nouvelles distributions généralisées, conjuguées à la distribution normale dont la valeur moyenne et la déviation standard sont inconnues.

Riassunto

In un lavoro anteriore l'autore ha dimostrato che le formule di credibility sono – nel caso di una dimensione – esatte per la famiglia esponenziale del tipo Koopman-Darmois e le loro leggi conjugate. Questo lavoro tratta il caso multidimensionale e le condizioni sotto quali le formule di credibility sono identiche colla media a posteriori. Prendendo la famiglia di distribuzioni esponenziali multidimensionali e le loro leggi conjugate le più semplici, si trova una media a posteriori lineare ma triviale perchè tiene conto soltanto delle osservazioni della singola componente. In un caso speciale importante per la pratica l'autore dimostra come la famiglia delle distribuzioni a priori può essere estesa per ottenere formule di credibility non triviali. Lo stesso procedere genera anche nuove leggi conjugate generalizzate per la distribuzione multidimensionale normale con media e scarto sconosciuto.

Abstract

In a previous paper, it was shown that the linear exponential family of likelihoods, together with their natural conjugate priors, gave *exact credibility* in the one-dimensional case; that is, a Bayesian forecast of the mean observation which is linear in the data. This paper considers the multidimensional case, and the conditions under which exact credibility again holds. Using the multidimensional version of the exponential family likelihoods, and a certain simple natural conjugate prior, it is shown that the prediction is linear, but self-dimensional, that is, only data from the predicted component is used. However, in an important special case of this family (called the linearly dependent exponential family), it is possible to provide additional hyperparameters to enrich the prior, thus giving full-multidimensional credibility formulae. This approach also gives a new, enriched prior for the multinormal, with unknown mean and precision.