

EXACT CREDIBILITY FOR WEIGHTED OBSERVATIONS

BY ROB KAAS, DENNIS DANNENBURG and MARC GOOVAERTS¹

University of Amsterdam

ABSTRACT

This note generalizes Jewell's theorem on exact credibility from the classical Bühlmann model to the (weighted) Bühlmann-Straub model.

1. INTRODUCTION

A well-known theorem of Jewell (1974) states that exact credibility, which is the concurring of the Bayesian estimator (posterior mean) with the credibility estimator of a contract mean, is found for a class of examples which includes many common situations. In a nutshell, exact credibility obtains when the observations are drawn from distributions in the exponential family, with natural conjugate prior distributions for the risk parameter. Surprisingly, Jewell's theorem pertains only to the classical Bühlmann model, and does not hold in case different variances of the observations are allowed, as in the Bühlmann-Straub model. In this contribution we prove exact credibility to hold for the (weighted) Bühlmann-Straub model as well, thus allowing the observations to be averages of varying numbers of observations, also in case of Poisson and Binomial distributions. The parametrization used coincides with the one used in the theory of Generalized Linear Models. In the original form of Jewell's theorem, and in ours as well, rather cumbersome reparametrizations are required to prove that ordinary distributions like Poisson and Gamma are special cases of this theorem. This is remedied in Gerber (1995) by choosing a more convenient parametrization.

Our extension of Jewell's theorem still does not incorporate Jewell's hierarchical model. Exact credibility for this model, and also for even more complicated ones like Hachemeister's regression credibility model, can, however, easily be proven for the normal-normal model. This is because conditional expectations of multinormal random variables are linear in the conditions.

Consider a portfolio consisting of J contracts, for which we have data of the past claims. These observations are assumed to have been generated by a Bayesian chance mechanism: first a contract-specific risk parameter Θ_j , $j = 1, \dots, J$, is drawn from a structure distribution with known parameters, called hyperparameters. In this contribution, we will concentrate on one particular contract j . Since the observations on other, independent, contracts do not appear in the estimators of $\mu(\Theta_j)$ used, we will simply write the observations as X_1, \dots, X_T , and not incorporate the contract number j in our

¹ The authors acknowledge the contributions of Ria Kuzee.

notation. The unobservable random variable Θ represents the risk characteristics of the contract in question. These characteristics may be unobservable, or if they are, unusable for reasons such as social acceptability. The observations, conditionally given Θ , are independently drawn from some distribution of which the mean is a function of the parameter Θ . The quantity of interest is not Θ itself, but the risk premium for contract j , traditionally denoted by $\mu(\Theta) = E[X|\Theta]$. The risk variable Θ acts as a parameter of the distribution of the risks X_1, X_2, \dots, X_T ; conditionally on Θ , the risks will be independent with mean $\mu(\Theta)$. They are not necessarily identically distributed, since the conditional variance of X_i is taken inversely proportional to some known weight w_i , just as in the Bühlmann-Straub model.

The best estimator of $\mu(\Theta)$, in the least squares sense, in the class consisting of all random variables of type $g(X_1, X_2, \dots, X_T)$ where $g(\cdot)$ is any function, is the one with minimal mean squared error $E\{[g(X_1, X_2, \dots, X_T) - \mu(\Theta)]^2\}$. It is obtained by taking $g = g^*$ with

$$g^*(X_1, X_2, \dots, X_T) = E[\mu(\Theta)|X_1, X_2, \dots, X_T] \quad (1)$$

Thus, we see that the best predictor of $\mu(\Theta)$ is just the conditional mean of $\mu(\Theta)$, given the observations, or in the idiom of Bayesian estimation, the posterior mean. These posterior means may have a rather unpleasant form, which is the reason why in credibility theory the restriction to linear functions of the observations is imposed. The estimator thus obtained is not only the best approximation to $\mu(\Theta)$, but it is also closest to the posterior mean (1). It can be shown that if the simultaneous distribution of Θ and X_1, \dots, X_T is of particular type, $g^*(\cdot)$ happens to be a linear function of the data already, and thus is the credibility estimator. In the second section of this note, we investigate conditions for which this holds.

2. EXACT CREDIBILITY

When the optimal Bayes estimator (1) is linear, it is obviously equal to the credibility estimator, since they both minimize the mean squared error. In this case we say that the credibility estimator is exact Bayesian, or equivalently, that exact credibility holds. Jewell (1974, 1975) showed that exact credibility is found when the observations X_1, \dots, X_T , given the value of the structure parameter Θ , are an iid random sample from the exponential family of distributions; moreover, the prior distribution of Θ must be the so-called natural conjugate prior, which ensures that the posterior distribution of Θ , given X_1, \dots, X_T , is of the same type as the prior distribution. In Jewell's original theorem, the X_i are iid, given Θ , as is the case in the original Bühlmann model. To be able to apply the theorem to the more general Bühlmann-Straub model, we have to account for the observations having different variances (weights). For definitions and assumptions of these credibility models, consult e.g. Goovaerts et al. (1990).

The well-known exponential family of distributions contains many frequently used distributions. Prominent members are the Normal, Poisson, Binomial, Gamma and Inverse Gaussian distributions. The densities in it can be written as :

$$f_X(x; \theta) = \exp \left[\frac{x\theta - b(\theta)}{\phi/w} + c(x, \phi/w) \right], x \in A_w. \quad (2)$$

The parameter θ of the distribution of X will be regarded as a realization of a structure random variable Θ . The other parameter ϕ is a dispersion parameter, like σ^2 in normal distributions. It may be assumed known or unknown. For one-parameter distributions, for example the Poisson, ϕ is taken to be 1. The weight $w > 0$ is known. Since (2) involves only the ratio ϕ/w , we might also say that only the relative weights of the contracts are known. Just as in the Bühlmann-Straub model, the variance of X , given Θ , is proportional to $1/w$. This is the case when X is an average of w elementary claims (natural weight), as we will prove later on, but w is not necessarily an integer. The set A_w consists of possible values of the claims. If the elementary risks are for instance Poisson, then $A_w = \{0, 1/w, 2/w, \dots\}$. In the sequel, we assume X to be continuous. In the discrete case, integrals over $x \in A_w$ below should be replaced by summations.

The above parametrization of the exponential family is sometimes called 'natural', in view of the fact that the part of it depending on both x and θ has the form $e^{x\theta}$. As we will see later on, it proves that the natural parametrization is not always the customary one, which is generally chosen because it is the most convenient. The one Gerber (1995) uses makes the reparametrizations much easier, but gives problems when incorporating weights. The parametrization we use closely resembles the one standard in the theory of Generalized Linear Models, see, e.g., McCullagh and Nelder (1989) or Nelder and Verrall (1995). Here ϕ is, without much gain of generality, replaced by $a(\phi)$.

We can evaluate the moment generating function with density (2) as follows:

$$\begin{aligned} m_X(r) &= \int e^{rx} \exp \left[\frac{x\theta - b(\theta)}{\phi/w} + c(x, \phi/w) \right] dx \\ &= \int \exp \left[\frac{x\{\theta + r\phi/w\} - b(\theta + r\phi/w)}{\phi/w} + c(x, \phi/w) \right] dx \times \\ &\quad \times \exp \left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w} \right] \\ &= \exp \left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w} \right]. \end{aligned} \quad (3)$$

Note that the second integral in (3) equals one, because it is the integral over a density of type (2), with θ replaced by $\theta + r\phi/w$. Mean and variance of (2) follow easily from the cumulant generating function $\kappa_X(r) = \log m_X(r)$:

$$\mu(\theta) = E[X|\Theta = \theta] = \kappa'_X(0) = b'(\theta); \quad (4)$$

$$\text{Var}[X; \theta] = \kappa''_X(0) = b''(\theta)\phi/w.$$

The function $b(\theta)$ is sometimes referred to as the cumulant function.

Having found the moment generating function, we can show that density (2) truly represents the density of an average $\bar{X} = \frac{1}{w} \sum_i X_i$ of w iid random variables X_1, \dots, X_w , with the same density (2), but with weight 1. Indeed we have

$$m_{\bar{X}}(r) = m_{\sum_i X_i} \left(\frac{r}{w} \right) = \{m_{X_i} \left(\frac{r}{w} \right)\}^w = \exp \left[\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w} \right]. \quad (5)$$

Assume that Θ has a prior density which is the so-called *natural conjugate prior*, i.e., of which the θ -dependent part is the same as in (2), and $x_0, \phi/w_0$ are parameters:

$$f_{\Theta}(\theta) = \exp \left[\frac{\theta x_0 - b(\theta)}{\phi/w_0} + d(x_0, \phi/w_0) \right]. \quad (6)$$

The normalizing function $d(x_0, \phi/w_0)$ is chosen in such a way that the density, which ranges over some θ -interval, integrates to one. Assume further that, conditionally given $\Theta = \theta$, the random variables X_t are independent drawings from density (2) with parameters θ, ϕ and weight $w_t, t = 1, \dots, T$. Then the posterior density of Θ , given $X_1 = x_1, \dots, X_T = x_T$, is found to be, apart from division by a normalizing constant equal to the integral over θ of the resulting expressions:

$$\begin{aligned} f_{\Theta|X_1, \dots, X_T}(\theta|x_1, \dots, x_T) &\propto f_{\Theta}(\theta) \prod_{i=1}^T \exp \left[\frac{x_i \theta - b(\theta)}{\phi/w_i} + c(x_i, \phi/w_i) \right] \\ &\propto \prod_{i=0}^T \exp \left[\frac{x_i \theta - b(\theta)}{\phi/w_i} \right] = \exp \sum_{i=0}^T \left[\frac{w_i x_i \theta - w_i b(\theta)}{\phi} \right] = \exp \left[\frac{\theta x_{\bullet} - b(\theta)}{\phi/w_{\bullet}} \right], \end{aligned} \quad (7)$$

$$\text{where } w_{\bullet} = \sum_{i=0}^T w_i \text{ and } x_{\bullet} = \sum_{i=0}^T \frac{w_i}{w_{\bullet}} x_i. \quad (8)$$

Thus, posterior and prior distribution are of the same type, but with parameter x_0 replaced by x_{\bullet} and w_0 by w_{\bullet} .

As a corollary to the above discussion we formulate the main theorem of exact credibility:

Theorem 2.1 (Posterior mean equals credibility estimator of exponential family with natural conjugate prior)

Suppose that, conditionally on $\Theta = \theta$, X_1, \dots, X_T are independent random variables with density (2) for fixed ϕ and weights $w_t, t = 1, \dots, T$. Further, let Θ have a prior distribution (6) with parameters x_0 and w_0 . Then the posterior mean $E[\mu(\Theta)|X_1, \dots, X_T]$ is an inhomogeneous linear form in X_1, \dots, X_T provided the prior density (6) vanishes at the endpoints of the θ -interval.

Proof. We must prove that the following expression is linear in x_1, \dots, x_T :

$$\int \mu(\theta) f_{\Theta|x_1, \dots, x_T}(\theta|x_1, \dots, x_T) d\theta. \tag{9}$$

Since $\mu(\theta) = b'(\theta)$ by (4) and the posterior density is proportional to (7), we must compute

$$\begin{aligned} \int \mu(\theta) f_{\Theta}(\theta; x_{\bullet}, w_{\bullet}) d\theta &= \frac{\int b(\theta) \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\} d\theta}{\int \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\} d\theta} \\ &= x_{\bullet} - \frac{\int \{x_{\bullet} - b'(\theta)\} \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\} d\theta}{\int \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\} d\theta} \\ &= x_{\bullet} - \frac{\int \frac{\phi}{w_{\bullet}} d \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\}}{\int \exp \left\{ \frac{w_{\bullet}}{\phi} [\theta x_{\bullet} - b(\theta)] \right\} d\theta} = x_{\bullet}, \end{aligned} \tag{10}$$

where the numerator vanishes because by assumption, $\theta x_{\bullet} - b(\theta) = -\infty$ at both endpoints of the integration interval. By (8), this expression is indeed inhomogeneous linear in x_1, \dots, x_T . ■

Remark 2.2 (Credibility factor and virtual experience)

By (8), we may write the estimator X_{\bullet} resulting from (10) as follows:

$$X_{\bullet} = \frac{w_0 x_0 + \sum_{l=1}^T w_l X_l}{w_0 + \sum_{l=1}^T w_l} = z X_w + (1 - z)m, \tag{11}$$

where $z = \frac{w_{\Sigma}}{w_0 + w_{\Sigma}}$ for $w_{\Sigma} = \sum_{l=1}^T w_l$ is the credibility factor,

$$X_w = \sum_{l=1}^T \frac{w_l}{w_{\Sigma}} X_l \tag{12}$$

and $m = E[\mu(\Theta)] = x_0$ (see (10)).

So the premium is the ratio of total claims and exposure, where a ‘virtual experience’ of m on average in w_0 exposure units is added to the actual experience of X_w on average, with a total weight (exposure) of w_{Σ} . ■

Remark 2.3 (Credibility estimator equals posterior mode)

Under the same conditions of the previous theorem, the maximum of the posterior density is found when θ is such that $\mu(\theta) = x_{\bullet}$, as well, since the derivative of (7) is zero when $x_{\bullet} = b'(\theta)$, which equals $\mu(\theta)$ by (4). ■

In the examples that follow, two special cases are given where the credibility estimators of the Bühlmann-Straub model are exact Bayesian. The third example shows that not all cases of exact credibility are covered by Theorem 2.1.

Example 2.4 (Poisson observations with Gamma prior)

Suppose that the risks X_i represent average numbers of claims in homogeneous cells with w_i policies in it, which, given $\Lambda = \lambda$, are Poisson(λ) distributed, for some positive structure random variable Λ . In automobile insurance, this risk parameter represents the 'accident-proneness' of the drivers in the cell considered. In general, the Gamma distribution proves to describe the spread of Λ rather well. The conditional density is

$$\begin{aligned} f_{X|\Lambda}(x|\lambda) &= \frac{e^{-\lambda w} (\lambda w)^{wx}}{(wx)!}, \quad x \in A = \{0, 1/w, 2/w, \dots\} \\ &= \exp\{w\{x \log \lambda - \lambda\} + wx \log w - \log((wx)!\}\}. \end{aligned}$$

From the last expression we see that this density belongs to the exponential family (2), with

$$\theta = \log \lambda, \quad \phi = 1, \quad b(\theta) = e^\theta, \quad c(x, \phi/w) = wx \log w - \log((wx)!). \quad (14)$$

By (6), the natural prior of $\Theta = \log \Lambda$ is, apart from the normalization constant $d(x_0 1/w_0)$:

$$f_\Theta(\theta) \propto \exp\left[\frac{\theta x_0 - e^\theta}{1/w_0}\right], \quad -\infty < \theta < \infty, \quad (15)$$

for some parameters $x_0 > 0$ and $w_0 > 0$. The corresponding density for Λ is then

$$f_\Lambda(\lambda) = f_\Theta(\log(\lambda)) \left| \frac{d\theta}{d\lambda} \right| \propto \lambda^{x_0 w_0 - 1} e^{-\lambda w_0}, \quad \lambda > 0, \quad (16)$$

in which we immediately recognize the Gamma(α, β) distribution with $\alpha = x_0 w_0$, $\beta = w_0$.

It is easy to verify that the extra condition of Theorem 2.1 is met, since $\theta x_0 - b(\theta)$ tends to $-\infty$ both for $\theta \rightarrow -\infty$ and $\theta \rightarrow \infty$. Therefore we know that the original Bühlmann inhomogeneous credibility estimator of $\mu(\Theta) = E[X|\Theta] (= \exp(\Theta))$ is exact Bayesian. As a consequence, the conditional mean of $\mu(\Theta)$, given X_1, \dots, X_T , is linear in X_1, \dots, X_T . Because Λ is a one-to-one function of Θ , we have also

$$\mu(\Theta) = E[X|\Theta] = E[X|\log(\Theta)] = E[X|\Lambda]. \quad (17)$$

So we conclude that the conditional expected value of $\mu(\Theta) = E[X|\Lambda]$, given X_1, \dots, X_T , is linear as well. This means that if the claims are averages of Poisson distributions with the usual parametrization of the first expression in (13), and the prior distribution of Λ is Gamma, the inhomogeneous credibility estimator for the Bühlmann-Straub model is exact Bayesian. ■

Example 2.5 (Normal distribution with Normal prior in the Bühlmann-Straub model)

Another example of exact credibility arises if the risks X_i are independent and $N(\theta, s^2/w_i)$ distributed, conditionally given $\Theta = \theta$, where Θ is an $N(m, a)$ distributed random variable. This model arises when $X_i = m + \Xi + \varepsilon_i$ for independent normal ε -components, with $\varepsilon \sim N(0, a)$ and $\varepsilon_i \sim N(0, s^2/w_i)$. Then $\Theta = m + \varepsilon$. To determine the credibility estimator, only the first and second order moments matter, and they are just those of the Bühlmann-Straub model. Recall that contracts of other cells appear neither in the posterior mean, nor in the inhomogeneous credibility estimator, by the independence between the cells.

The conditional density of the X_i can be written as

$$\begin{aligned}
 f_{X_i|\Theta}(x_i|\theta) &= \frac{1}{\sqrt{2\pi s^2/w_i}} \exp \frac{-(x_i - \theta)^2}{2s^2/w_i} \\
 &= \exp \left[\frac{x_i\theta - 1/2\theta^2}{s^2/w_i} - 1/2 \left\{ \frac{x_i^2}{s^2/w_i} + \log(2\pi s^2/w_i) \right\} \right], \tag{18}
 \end{aligned}$$

which is (2) when

$$b(\theta) = 1/2\theta^2, \quad \phi = s^2, \quad c(x_i, s^2/w_i) = -1/2 \left\{ \frac{x_i^2}{s^2/w_i} + \log(2\pi s^2/w_i) \right\}. \tag{19}$$

The natural conjugate prior density is again normal, see (6) and (18), so

$$f_{\Theta}(\theta) = \frac{1}{\sqrt{2\pi a}} \exp \frac{-(\theta - m)^2}{2a} \propto \exp \frac{\theta m - 1/2\theta^2}{s^2/(s^2/a)}, \tag{20}$$

which, apart from the normalization constant, equals expression (6) when

$$x_0 = m, \quad \phi = s^2, \quad w_0 = s^2/a. \tag{21}$$

Because $\theta x_w - 1/2\theta^2$ again tends to $-\infty$ both for $\theta \rightarrow -\infty$ and $\theta \rightarrow \infty$, we find from Theorem 2.1 that the posterior mean equals the Bühlmann-Straub estimator

$$E[\Theta|X_1, \dots, X_T] = \frac{w_0 x_0 + \sum_{i=1}^T w_i X_i}{w_0 + \sum_{i=1}^T w_i} = z \frac{\sum_{i=1}^T w_i X_i}{\sum_{i=1}^T w_i} + (1 - z)m,$$

with $z = \frac{a \sum_{i=1}^T w_i}{a \sum_{i=1}^T w_i + s^2}$. (22) \blacksquare

Other situations, apart from Examples 2.4 and 2.5, in which exact credibility holds are (Negative) Binomial data with Beta prior, and Inverse Gaussian data with the corresponding natural prior.

Example 2.6 (Exact credibility in normal-normal models)

Theorem 2.1, which extends Jewell's original theorem to the weighted case, cannot be applied to Jewell's hierarchical credibility model, see, e.g., Goovaerts et al. (1990). Written in the same additive components form of the previous example, the statistic for sector p , cell j , and time period t is

$$X_{pj,t} = m + \Xi_p + \Xi_{pj} + \Xi_{pj,t}, \quad (23)$$

where the Ξ -components of the risks are independent with mean zero and variances b , a and $s^2/w_{pj,t}$. First we try to consider only one sector p . Then, as is required in Theorem 2.1, the observations of other sectors are independent of the ones considered. Conditionally on Θ , which in this case is Ξ_p , the observations in sector p have the same mean. They are, however, not independent, since the observations in cell j of this sector have a common risk component Ξ_{pj} . If on the other hand we only look at a specific cell j in some sector p , taking $\Theta = \Xi_p + \Xi_{pj}$ we do have that conditionally given Θ , the observations are independent and have equal mean. But in this case the other observations cannot be disregarded when estimating the risk premium of this cell, since observations in cell $i \neq j$ of sector p are dependent on those of cell j through the common component Ξ_p .

Still, when we assume in addition that the Ξ -components are normally distributed, the credibility estimators for the Jewell model can easily be shown to be exact Bayesian. This is because for each choice $M = m + \Xi_p$, $M = m + \Xi_p + \Xi_{pj}$ and $M = X_{pj, T+1}$, M has a multivariate normal joint distribution with the vector of observations \vec{X} . This, as is well-known and can be found in any statistics text of a reasonable level, implies that $E[M|\vec{X}]$ is linear in \vec{X} .

Also under normality assumptions, the estimators in Hachemeister's regression credibility model can be shown to be exact Bayesian. ■

Remark 2.7 (Variance components outlook on credibility theory)

In the authors' opinion, credibility is currently taught in an unnecessarily complicated way. For didactic reasons, models should not be formulated using a hard-to-explain risk variable Θ , a function $\mu(\Theta)$ of which is the variable of interest. Setting credibility in a Bayesian framework also isn't exactly helpful for the acceptance of credibility techniques by practitioners, especially in Europe. Since in most countries actuaries generally are not fully qualified mathematicians, formulating credibility estimation as a projection in a Hilbert space, however elegant mathematically, is also aiming too high. Rather, one should formulate the credibility models as additive independent variance components models such as (23). As argued in Dannenburg, Kaas and Goovaerts (1996), this presents no loss of generality, since only the first and second moments of the data and (weighted) averages thereof are needed for the calculation of credibility estimators. To calculate covariances and correlations is almost trivial in this framework, but a much more laborious process via the conditional expectations, given Θ , needed in the more usual model. ■

REFERENCES

- DANNENBURG, D.R., R. KAAS, M.J. GOOVAERTS (1996); Practical actuarial credibility models, *Institute of Actuarial Science and Econometrics, University of Amsterdam*, Amsterdam.
- GERBER, H.U. (1995); A teacher's remark on exact credibility, *ASTIN Bulletin*, **25**, 189-192.
- GOOVAERTS, M.J., R. KAAS, A.E. VAN HEERWAARDEN & T. BAUWELINCKX (1990); Effective Actuarial Methods; *North-Holland*, Amsterdam.
- JEWELL, W.S. (1974); Credible means are exact Bayesian for exponential families; *ASTIN Bulletin*, **8**, 77-90.
- JEWELL, W.S. (1975); Regularity conditions for exact credibility; *ASTIN Bulletin*, **8**, 336-341.
- MCCULLAGH, P. and NELDER, J.A. (1989), Generalized Linear Models, 2nd edition. *Chapman & Hall*, London.
- NELDER, J.A. and R.J. VERRALL (1997). Credibility Theory and Generalized Linear Models, *ASTIN Bulletin*, **27.1**, 71-82.

ROB Kaas, DENNIS DANNENBURG and MARC GOOVAERTS
Universiteit van Amsterdam
Instituut voor Actuarie en Econometrie
Roetersstraat 11
1018 WB Amsterdam, The Netherlands

