Technische Universität Dresden
Institut für Mathematische Stochastik

Dresdner Schriften zur
Versicherungsmathematik
3/2001

Klaus Th. Hess

Estimation of the Mean Under Vague Prior Information

Herausgeber: Die Professoren des Instituts für Mathematische Stochastik
ISSN 0946–4727

Estimation of the Mean Under Vague Prior Information

Klaus Th. Hess

Institut für Mathematische Stochastik

Technische Universität Dresden

Abstract

We consider an array of random variables in which all random variables have the same expectation, random variables of different rows are uncorrelated, and random variables in the same row have the same variance and the same covariance. If the number of observations varies from row to row, a uniformly best linear unbiased estimator does not exist. For this case we calculate a minimax estimator under vague prior information on the variance and the covariance in the rows.

1 Introduction

We consider an array of random variables

$$\begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n_1} \\ X_{21} & X_{22} & \dots & X_{2n_2} \\ \vdots & \vdots & & \\ X_{m1} & X_{m2} & \dots & X_{mn_m} \end{pmatrix}$$

in which all random variables have the same expectation, random variables of different rows are uncorrelated, and random variables in the same row have the same variance and the same covariance. The problem is to estimate the common expectation μ.

This model can be interpreted as follows: There is a certain characteristic, which we shall observe for some objects $1, \dots, m$. For object i, we have n_i observations. We assume that there is no interference of observations between different objects, but interference between the observations of the same object is permitted. Here, the interference is independent of the order of the observations.

Let us consider two extreme cases:

– If all random variables X_{ij} are i.i.d., it is well–known, that

$$\frac{1}{\sum_{k=1}^m n_k} \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij} \quad = \quad \sum_{i=1}^m \frac{n_i}{\sum_{k=1}^m n_k} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

is the best linear unbiased estimator of μ.

– If all random of the same row are identical, it is obvious that

$$\frac{1}{m} \sum_{i=1}^m X_{i1} = \sum_{i=1}^m \frac{1}{m} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

is the best linear unbiased estimator of μ.

In both cases, the best linear unbiased estimator can be written as a weighted mean of the sample means of the rows; however, the weights depend on the assumption on the underlying distribution. The present paper studies the optimal choice of the weights when vague prior information on the variance structure is available.

2 The model

Throughout this paper let $(\Omega, \mathcal{F})$ be a measurable space. We consider an array of random variables $\Omega \to \mathbf{R}$

$$\begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n_1} \\ X_{21} & X_{22} & \dots & X_{2n_2} \\ \vdots & \vdots & & \\ X_{m1} & X_{m2} & \dots & X_{mn_m} \end{pmatrix}$$

with $m, n_1, \dots, n_m \in \mathbf{N}$.

We are interested in a linear estimator of the common expectation of theses random variables. Let Π denote the set of all probability measures $\mathbf{P} : \mathcal{F} \to [0, 1]$ such that

(i) all random variables X_{ij} are square–integrable,

(ii) there exists some $\mu \in \mathbf{R}$, such that

$$\mathbf{E}[X_{ij}] \quad = \quad \mu$$

holds for all $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n_i\}$,

(iii) any two random variables of different rows are uncorrelated, and

(iv) there exist $\{\lambda_i\}_{i \in \{1,\dots,m\}} \subseteq \mathbf{R}$ and $\{\varphi_i\}_{i \in \{1,\dots,m\}} \subseteq \mathbf{R}_+$ with $\lambda_i + \varphi_i/n_i > 0$ such that

$$\mathbf{cov}[X_{ij}, X_{il}] \quad = \quad \begin{cases} \lambda_i & \text{if } j \neq l \\ \lambda_i + \varphi_i & \text{if } j = l \end{cases}$$

holds for all $i \in \{1, \dots, m\}$ and $j, l \in \{1, \dots, n_i\}$.

2.1 Remarks.

– The parameters μ, φ_i, and λ_i depend on the choice of $\mathbf{P} \in \Pi$.

– Condition (iv) is fulfilled if the random variables of each row are exchangeable or even conditionally i.i.d. with respect to some sub–σ–algebra of $\mathcal{F}$. In the last case, λ_i is the variance of the conditional expectation and φ_i is the expectation of the conditional variance.

– Condition (iv) implies $\mathbf{var}\,[X_{ij}] > 0$.

– Let $\lambda, \varphi \in \mathbf{R}$ and $n \in \mathbf{N}$. Since

$$\begin{aligned} &\lambda \left(\sum_{j=1}^{n} \beta_j \right)^2 + \varphi \sum_{j=1}^{n} \beta_j^2 \\ &= n\varphi \left(\frac{1}{n} \sum_{j=1}^{n} \beta_j^2 - \left(\frac{1}{n} \sum_{j=1}^{n} \beta_j \right)^2 \right) + \left(n^2 \lambda + n\varphi \right) \left(\frac{1}{n} \sum_{j=1}^{n} \beta_j \right)^2 \end{aligned}$$

holds for all $\{\beta_j\}_{j \in \{1,\ldots,n_i\}} \subseteq \mathbf{R}$, the (n,n)–matrix

$$\begin{pmatrix} \lambda + \varphi & & \lambda \\ & \ddots & \\ \lambda & & \lambda + \varphi \end{pmatrix}$$

is

- positive semi definite, if and only if $\varphi \geq 0$ and $\lambda + \varphi/n \geq 0$.
- positive definite, if and only if $\varphi > 0$ and $\lambda + \varphi/n > 0$.

This explains the restrictions on λ_i and φ_i in condition (iv). Corollary 2.3 will show that the case $\lambda_i + \varphi_i/n_i = 0$ does not lead to a statistical problem since the estimator is almost surely identical with μ.

Now we study some properties of the probability measures $\mathbf{P} \in \Pi$.

2.2 Lemma. *For each $\mathbf{P} \in \Pi$, the identity*

$$\mathbf{cov}\left[\sum_{j=1}^{n_i} \beta_j X_{ij}, \sum_{j=1}^{n_i} \gamma_j X_{ij} \right] = \lambda_i \left(\sum_{j=1}^{n_i} \beta_j \right) \left(\sum_{j=1}^{n_i} \gamma_j \right) + \varphi_i \sum_{j=1}^{n_i} \beta_j \gamma_j$$

holds for all $i \in \{1, \ldots, m\}$ and $\{\beta_j\}_{j \in \{1,\ldots,n_i\}}, \{\gamma_j\}_{j \in \{1,\ldots,n_i\}} \subseteq \mathbf{R}$.

Proof. Using the covariance structure of the rows, we obtain

$$\begin{aligned} \mathbf{cov}\left[\sum_{j=1}^{n_i} \beta_j X_{ij}, \sum_{j=1}^{n_i} \gamma_j X_{ij} \right] &= \sum_{j=1}^{n_i} \sum_{l=1}^{n_i} \beta_j \gamma_l \, \mathbf{cov}[X_{ij}, X_{il}] \\ &= \sum_{j=1}^{n_i} \left(\sum_{l=1}^{n_i} \beta_j \gamma_l \lambda_i + \beta_j \gamma_j \varphi_i \right) \\ &= \lambda_i \left(\sum_{j=1}^{n_i} \beta_j \right) \left(\sum_{j=1}^{n_i} \gamma_j \right) + \varphi_i \sum_{j=1}^{n_i} \beta_j \gamma_j \end{aligned}$$

for all $i \in \{1, \ldots, m\}$. $\square$

2.3 Corollary. *For each* $\mathbf{P} \in \Pi$*, the identity*

$$\mathbf{var}\left[\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\right] \;=\; \lambda_i + \varphi_i/n_i$$

holds for all $i \in \{1,\ldots,m\}$*.*

2.4 Corollary. *For each* $\mathbf{P} \in \Pi$*, the identity*

$$\mathbf{var}\left[\sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha_i X_{ij}\right] \;=\; \sum_{i=1}^{m} (\lambda_i + \varphi_i/n_i)\, n_i^2 \alpha_i^2$$

holds for all $\{\alpha_i\}_{i\in\{1,\ldots,m\}} \subseteq \mathbf{R}$*.*

We will need Lemma 2.2 and its corollaries in the following sections.

Now we define the set Δ of all admissible estimators for μ by letting

$$\Delta := \left\{ \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha_{ij} X_{ij} \;\middle|\; \{\alpha_{ij}\}_{\substack{i\in\{1,\ldots,m\}\\ j\in\{1,\ldots,n_i\}}} \subseteq \mathbf{R}, \sum_{i=1}^{m}\sum_{j=1}^{n_i} \alpha_{ij} = 1 \right\}$$

and we define a payoff function $L : \Pi \times \Delta \to \mathbf{R}$ by letting

$$L(\mathbf{P},\delta) := \mathbf{E}\left[(\delta-\mu)^2\right]$$

Then (Π, Δ, L) is an abstract game, where player 1 (nature) chooses a strategy $\mathbf{P} \in \Pi$ and player 2 (statistician) chooses a strategy $\delta \in \Delta$.

Each estimator $\delta \in \Delta$ is unbiased with respect to each $\mathbf{P} \in \Pi$ since the sum of all α_{ij} is equal to one. Hence

$$L(\mathbf{P},\delta) \;=\; \mathbf{var}\,[\delta]$$

In the sequel we need the following definitions: A probability measure $\mathbf{P}^* \in \Pi$ is *least favourable* if it satisfies

$$\inf_\Delta L(\mathbf{P}^*,\delta) \;=\; \sup_\Pi \inf_\Delta L(\mathbf{P},\delta)$$

An estimator $\delta^* \in \Delta$ is *minimax* if it satisfies

$$\sup_\Pi L(\mathbf{P},\delta^*) \;=\; \inf_\Delta \sup_\Pi L(\mathbf{P},\delta)$$

An estimator $\delta^* \in \Delta$ is *Bayes* with respect to $\mathbf{P}^* \in \Pi$ if it satisfies

$$L(\mathbf{P}^*,\delta^*) \;=\; \inf_\Delta L(\mathbf{P}^*,\delta)$$

A pair of strategies $(\mathbf{P}^*,\delta^*) \in \Pi \times \Delta$ is a *saddlepoint* if it satisfies

$$\sup_\Pi L(\mathbf{P},\delta^*) \;=\; \inf_\Delta L(\mathbf{P}^*,\delta)$$

The connections between the definitions is given by the Sattlepoint Theorem:

Proposition 2.5 (Saddlepoint Theorem). *For* $\mathbf{P}^* \in \Pi$ *and* $\delta^* \in \Delta$, *the following are equivalent*:
(i) $(\mathbf{P}^*, \delta^*)$ *is a saddlepoint.*
(ii) $\mathbf{P}^*$ *is least favourable and* δ^* *is minimax and Bayes with respect to* $\mathbf{P}^*$.

The notation above agrees with Schmidt (2000); for applications in statistics, see e. g. Lehmann and Casella (1998) and Witting (1985).

3 The Bayes estimator

Throughout this section let $\mathbf{P} \in \Pi$ be a fixed probability measure. In the present section, we show that

$$\delta^* \quad := \quad \sum_{i=1}^{m} \sum_{j=1}^{n_i} \alpha_i^* X_{ij}$$

with

$$\alpha_i^* \quad := \quad \frac{1}{n_i} \frac{(\lambda_i + \varphi_i/n_i)^{-1}}{\sum_{k=1}^{m} (\lambda_k + \varphi_k/n_k)^{-1}}$$

is Bayes with respect to $\mathbf{P}$. For an interpretation of δ^* it is convenient to use the representation

$$\delta^* \quad = \quad \sum_{i=1}^{m} \frac{(\lambda_i + \varphi_i/n_i)^{-1}}{\sum_{k=1}^{m} (\lambda_k + \varphi_k/n_k)^{-1}} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

This shows that the estimator is a weighted mean of the sample mean of the rows, with weights being inversely proportional to the variance of the respective sample mean; see Corollary 2.3.

For the proof of the main result of this section we need the following lemma:

3.1 Lemma. *The identity*

$$\mathbf{cov}[\delta, \delta^*] \quad = \quad \frac{1}{\sum_{k=1}^{m} (\lambda_k + \varphi_k/n_k)^{-1}}$$

holds for all $\delta \in \Delta$.

Proof. Since condition (iii) on $\mathbf{P}$ and Lemma 2.2, we obtain

$$\begin{aligned}
\mathbf{cov}[\delta, \delta^*] &= \mathbf{cov}\left[\sum_{i=1}^{m} \sum_{j=1}^{n_i} \alpha_{ij} X_{ij}, \sum_{k=1}^{m} \sum_{l=1}^{n_k} \alpha_k^* X_{kl} \right] \\
&= \sum_{i=1}^{m} \alpha_i^* \left(n_i \lambda_i + \varphi_i \right) \sum_{j=1}^{n_i} \alpha_{ij} \\
&= \frac{1}{\sum_{k=1}^{m} (\lambda_k + \varphi_k/n_k)^{-1}} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \alpha_{ij} \\
&= \frac{1}{\sum_{k=1}^{m} (\lambda_k + \varphi_k/n_k)^{-1}}
\end{aligned}$$

as was to be shown. □

3.2 Theorem. *The estimator δ^* is Bayes with respect to $\mathbf{P}$ and every estimator $\delta \in \Delta$ which is Bayes with respect to $\mathbf{P}$ is $\mathbf{P}$–almost surely equal to δ^*.*

Proof. For all $\delta \in \Delta$, Lemma 3.1 yields

$$\mathbf{var}[\delta] \;=\; \mathbf{var}[\delta - \delta^*] + \mathbf{var}[\delta^*]$$

and hence

$$\mathbf{var}[\delta^*] \;\leq\; \mathbf{var}[\delta]$$

If $\delta \in \Delta$ is Bayes with respect to $\mathbf{P}$, then $\mathbf{var}[\delta] = \mathbf{var}[\delta^*]$. Hence, we get

$$\mathbf{var}[\delta - \delta^*] \;=\; 0$$

which is fulfilled if and only if δ is $\mathbf{P}$–almost surely equal to δ^*. □

3.3 Remarks.

– If $\lambda_i = 0$, then

$$\delta^* \;=\; \frac{1}{\sum_{k=1}^m n_k} \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij}$$

In this case, all random variables have the same weight. The condition $\lambda_i = 0$ is especially fulfilled, if all random variables are i.i.d.

– If $\varphi_i = 0$, then

$$\delta^* \;=\; \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

In this case, all rows have the same weight.

– If $\lambda_i = \lambda \neq 0$ and $\varphi_i = \varphi$ for admissible $\lambda, \varphi \in \mathbf{R}$, then

$$\delta^* \;=\; \sum_{i=1}^m \frac{(1+\kappa/n_i)^{-1}}{\sum_{k=1}^m (1+\kappa/n_k)^{-1}} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

with $\kappa := \varphi/\lambda = (\varphi+\lambda)/\lambda - 1$. In this case, the weights depend on the ratio of the variance and the covariance.

– If $\lambda_i = \lambda \neq 0$ and $\varphi_i = \varphi$ with admissible $\lambda, \varphi \in \mathbf{R}$ and if $n_i = n$ for some $n \in \mathbf{N}$, then

$$\delta^* \;=\; \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij}$$

which is independent of λ and φ. Moreover, the case $n = 1$ gives the classical result for m uncorrelated observations having the same variance.

Since δ^* depends on λ_i and φ_i, δ^* is a *pseudoestimator*. Its structure, as described above, is what a statistician expects.

4 Minimax estimators

The Bayes estimator of Theorem 3.2 depends on the parameters λ_i and φ_i. We assume that λ_i and φ_i are unknown, but that we have vague prior information on these parameters. For this case, we consider the minimax estimator depending on the priori information.

For fixed $\tau, \sigma \in \mathbf{R}$ with $\sigma > 0$ and $(\max\{n_1, \ldots, n_m\} - 1)\tau + \sigma^2 > 0$ denote by Π' the set of all $\mathbf{P} \in \Pi$ with

$$\begin{aligned} \mathbf{cov}[X_{ij}, X_{il}] &\leq \tau \\ \mathbf{var}\,[X_{ij}] &\leq \sigma^2 \end{aligned}$$

for all $i \in \{1, \ldots, m\}$ and $j, l \in \{1, \ldots, n_i\}$ with $j \neq l$ and define

$$\begin{aligned} \lambda &:= \min\{\tau, \sigma^2\} \\ \varphi &:= \max\{0, \sigma^2 - \tau\} \end{aligned}$$

The inequality on τ and σ implies $\lambda + \varphi/n_i > 0$ for all $i \in \{1, \ldots, m\}$.

4.1 Theorem. *Consider the game (Π', Δ, L). If $\mathbf{P}^* \in \Pi'$ satisfies*

$$\begin{aligned} \mathbf{cov}^*[X_{ij}, X_{il}] &= \min\{\tau, \sigma^2\} \\ \mathbf{var}^*[X_{ij}] &= \sigma^2 \end{aligned}$$

for all $i \in \{1, \ldots, m\}$ and $j, l \in \{1, \ldots, n_i\}$ with $j \neq l$ then $\mathbf{P}^*$ is least favourable, the estimator $\delta^* \in \Delta$ with

$$\delta^* = \sum_{i=1}^{m} \frac{(\lambda + \varphi/n_i)^{-1}}{\sum_{k=1}^{m} (\lambda + \varphi/n_k)^{-1}} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

is minimax and Bayes with respect to $\mathbf{P}^*$ and satisfies

$$\mathbf{var}^*[\delta^*] = \frac{1}{\displaystyle\sum_{k=1}^{m} (\lambda + \varphi/n_k)^{-1}}$$

Proof. First note that $\mathbf{P}^*$ is defined such that the identities

$$\begin{aligned} \lambda_i^* &= \lambda \\ \varphi_i^* &= \varphi \end{aligned}$$

hold for all $i \in \{1, \ldots, m\}$.

Let $\mathbf{P} \in \Pi'$. In the case $\tau < \sigma^2$, we have

$$\begin{aligned} \lambda_i + \varphi_i/n_i &= \lambda_i(1 - 1/n_i) + (\lambda_i + \varphi_i)/n_i \\ &\leq \tau(1 - 1/n_i) + \sigma^2/n_i \\ &= \tau + (\sigma^2 - \tau)/n_i \end{aligned}$$

In the case $\tau \geq \sigma^2$, we have $\lambda_i \leq \lambda_i + \varphi_i \leq \sigma^2$ and hence

$$\begin{aligned}\lambda_i + \varphi_i/n_i &= \lambda_i(1 - 1/n_i) + (\lambda_i + \varphi_i)/n_i \\ &\leq \sigma^2(1 - 1/n_i) + \sigma^2/n_i \\ &= \sigma^2\end{aligned}$$

Both cases can be summarized as

$$\begin{aligned}\lambda_i + \varphi_i/n_i &\leq \min\{\tau, \sigma^2\} + \max\{0, \sigma^2 - \tau\}/n_i \\ &= \lambda + \varphi/n_i\end{aligned}$$

Theorem 3.2 implies that δ^* is Bayes with respect to $\mathbf{P}^*$. Hence, for arbitrary $\mathbf{P} \in \Pi'$ and $\delta \in \Delta$, we get from Corollary 2.4

$$\begin{aligned}\mathbf{var}\,[\delta^*] &= \sum_{i=1}^{m} (\lambda_i + \varphi_i/n_i) \left(\frac{(\lambda + \varphi/n_i)^{-1}}{\sum_{k=1}^{m} (\lambda + \varphi/n_k)^{-1}} \right)^2 \\ &\leq \sum_{i=1}^{m} (\lambda + \varphi/n_i) \left(\frac{(\lambda + \varphi/n_i)^{-1}}{\sum_{k=1}^{m} (\lambda + \varphi/n_k)^{-1}} \right)^2 \\ &= \mathbf{var}^*[\delta^*] \\ &\leq \mathbf{var}^*[\delta]\end{aligned}$$

This implies

$$\begin{aligned}\sup_{\Pi'} \mathbf{var}\,[\delta^*] &= \mathbf{var}^*[\delta^*] \\ &= \inf_{\Delta} \mathbf{var}^*[\delta]\end{aligned}$$

and the assertion follows from the Sattlepoint Theorem. □

4.2 Remarks.

– In the case $\tau < \sigma^2$, we have

$$\delta^* = \sum_{i=1}^{m} \frac{n_i/[(n_i - 1)\tau + \sigma^2]}{\sum_{k=1}^{m} n_k/[(n_k - 1)\tau + \sigma^2]} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

– In the case $\sigma^2 \leq \tau$, we have

$$\delta^* = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \right)$$

In this case, we have no information on the covariance. Theorem 4.1 shows, that in this situation the minimax estimator is the arithmetic mean of the sample mean of the rows, whatever the number of observations in the rows may be.

References

Lehmann, E. L. and Casella, G. (1998): *Theory of Point Estimation.* New York: Springer.

Schmidt, K. D. (2000): *Statistical decision problems and linear prediction under vague prior information.* Statistics & Decisions **18**, 429–442.

Witting, H. (1985): *Mathematische Statistik I.* Stuttgart: Teubner.

Klaus Th. Hess
Institut für Mathematische Stochastik
Technische Universität Dresden
D–01062 Dresden
Germany

e–mail: `hess@math.tu-dresden.de`
internet: `www.math.tu-dresden.de\~hess`

December 10, 2001