KYBERNETIKA — VOLUME *30* (1994), NUMBER 4, PAGES 433 – 443

# BAYESIAN RANK ESTIMATION WITH APPLICATION TO FACTOR ANALYSIS

Miroslav Kárný and Martin Šámal

Rank estimation is a common sub-problem met in various fields exploiting matrix algebra. Estimation of the number of factors in factor analysis is of this type. Due to large noise contents in the analyzed matrix, standard procedures, like deterministic inspection of singular values, fail.

In the paper, a novel procedure is proposed. It is gained by straightforward application of Bayesian statistics to a carefully selected model which fits to the target area, namely, factor analysis of dynamic scintigraphic studies. The formal solution consists of an exactly feasible part and a maximum-likelihood type one. The latter is justified by large dimensions of the data matrices containing analyzed images.

Properties of the procedure are illustrated on simulated and real data.

## 1. INTRODUCTION

Rank estimation is a common sub-problem of the applied matrix algebra. Singular value decomposition [2] followed by a selection of significant singular values is the most successful procedure used. Often, the distinction of significant and insignificant singular values is "visible" by naked eyes and the rank estimation is easy. The situation becomes different when the inspected matrix is gained from noisy data and no sharp bound can be seen between both types of singular values. We have met this problem when estimating the number of factors in factor analysis of dynamic scintigraphic studies [3]. This method provides a computer assistance to medical doctors analyzing image data. For a routine use, a reliable algorithmic support in choosing number of factors is of importance. We have tested several procedures, for instance [4], with various degrees of success. At the end, no one was found sufficiently satisfactory for the wide range of data tested. For this reason, we have tried to use the methodology of Bayesian structure estimation [5] in order to overcome the observed drawbacks. The choice of the methodology was stimulated by successes in other applications.

This paper describes the procedure and provides illustrative results both for simulated and real image data. The derivation presented should clarify the assumptions and restrictions under which the procedure is expected to work properly.

## 2. BAYESIAN STRUCTURE ESTIMATION

Bayesian structure estimation [5] can be interpreted as a standard solution of a special estimation problem with a multivariate parameter. One of parameter entries is a pointer to structures $\mathcal{S}(k)$, $k = 1, \ldots, m < \infty$, and the rest consists of the related finite dimensional unknown parameter $\Theta(k) = \Theta(\mathcal{S}(k))$.

The pointer distinguishes competitive descriptions of the same observed data $Y$, namely, probability density functions (p.d.f.s) conditioned on the structure and related parameter

$$p(Y|\Theta(k), \mathcal{S}(k)). \tag{1}$$

Assigning prior distribution to the unknown entities

$$p(\Theta(k), \mathcal{S}(k)) = p(\Theta(k)|\mathcal{S}(k))p(\mathcal{S}(k)) \tag{2}$$

(with $p(\mathcal{S}(k))$ being a probability function (p.f.)), the Bayes' formula [5] provides the posterior probabilities of the structures

$$p(\mathcal{S}(k)|Y) = \frac{p(Y|\mathcal{S}(k))p(\mathcal{S}(k))}{\sum_{i=1}^{m} p(Y|\mathcal{S}(i))p(\mathcal{S}(i))}. \tag{3}$$

The needed data description – within $k$th structure and *without "nuisance" parameter* $\Theta(k)$ – is given by the standard relation between the joint and marginal p.d.f.s

$$p(Y|\mathcal{S}(k)) = \int p(Y|\Theta(k), \mathcal{S}(k))p(\Theta(k)|\mathcal{S}(k))\,\mathrm{d}\Theta(k). \tag{4}$$

**Remark.** In order to keep presentation as simple as possible, the following sidesteps from a mathematical correctness are made:

- the existence of used p.d.f.s with respect to a dominating measure (Lebesque measure in the continuous case and counting one the in discrete case) is assumed;
- $\int \cdot\, \mathrm{d}\Theta$ denotes generally multivariate integral over whole support of the integrand with respect to Lebesque measure supposing this operation to be meaningful;
- random variables, their realizations and p.d.f. arguments are not distinguished in notation;
- various p.d.f.s are labelled by $p$ and distinguished through their arguments only; even probabilities for discrete random variable are denoted in the same way.

## 3. PROBLEM FORMULATION

This paper contains essentially an application of the Bayes formula (3) to the appropriately chosen constituents (1), (2) which are believed to model a wide class of situations under which rank has to be estimated. Thus, the problem formulation is completed by specifying the elements in the formula (3).

### 3.1. Data normality and covariance structure

The raw data $Y$ (the matrix of $(m,n)$-type with $m \leq n$) are assumed to be normally distributed with the expected value $\mu$ which explains all relations leaving zero correlations among $Y - \mu$ entries. A common non-zero dispersion $\lambda$ of all entries $Y_{ij}$, $i = 1, \ldots, m$; $j = 1, \ldots, n$ is supposed. Thus, the p.d.f. of $Y$ conditioned on the unknown $(\mu, \lambda)$ is

$$p(Y|\mu,\lambda) \propto \lambda^{-0.5mn} \exp[0.5\lambda^{-1}\mathrm{tr}(Y-\mu)'(Y-\mu)] \equiv \mathcal{N}_Y(\mu, \lambda I_m \otimes I_n), \quad (5)$$

where $\propto$ means equality up to a factor which is uniquely defined by the p.d.f. normalization, tr denotes the matrix trace, $'$ transposition, $I_n$ the unit matrix of the $(n,n)$-type and $\otimes$ the Kronecker product.

### 3.2. Rank dependent parametrization

The following simple algebraic proposition expresses dependences among entries of $\mu$ if $\mu$ has rank $k\,(\leq m \leq n)$.

**Proposition 1.** [*Decomposition of* $\mu$] Let $\mu$ be a real non-zero $(m,n)$-matrix of the rank $k$, $k \leq m \leq n$. Then, it can be *uniquely factorized* into the product

$$\mu = FG, \quad (6)$$

where the matrix $F$ belongs to the set $F^*(k)$ of full rank $(m,k)$-matrices with orthogonal columns

$$F'F = I_k \quad (7)$$

and with the leading $(k,k)$-submatrix being lower triangular one with positive diagonal. The matrix $G$ is a full rank matrix of the $(k,n)$-type.

P r o o f . Singular value decomposition and the rank assumption imply existence of the unique decomposition $\mu = SDV$ with $S'S = VV' = I_k$ and the diagonal $(k,k)$-matrix $D$ having non-increasing values of the non-negative diagonal entries. Introducing $\tilde{G} = DV$ we get the decomposition $\mu = S\tilde{G}$ which is unique up to an "inner" regular transformation $T$, i. e. $\mu = STT^{-1}\tilde{G}$. The required orthonormality of $F = ST$ enforces the orthogonality of $T$ and the "triangularity" requirement (together with the assumed full rank of $S$) implies the uniqeness of $T$. □

### 3.3. Parametrized model

For any of the possible structures: $\mathcal{S}(k) \equiv$ *rank of the data expectation* $\mu$ *is* $k$ (k=0,1...,m), we adopt as the data model the Gaussian one, uniquely parametrized by $\Theta(k) = (F(k), G(k), \lambda(k))$

$$p(Y|\Theta(k),\mathcal{S}(k)) = p(Y|F(k),G(k),\lambda(k),\mathcal{S}(k)) = \mathcal{N}_Y(F(k)G(k), \lambda(k)I_m \otimes I_n) \quad (8)$$

with $F(k) \in F^*(k)$.

### 3.4. Priors

Uniform prior p.f. is assigned to the competitive structures using the argument of insufficient reasons. This can be easily modified if such reasons exist.

The definition of $\Theta(k)$ items and the chain rule imply

$$p(\Theta(k)|\mathcal{S}(k)) = p(G(k)|F(k),\lambda(k),\mathcal{S}(k))p(F(k)|\lambda(k),\mathcal{S}(k))p(\lambda(k)|\mathcal{S}(k)) \tag{9}$$

thus, the prior-p.d.f. $p(\Theta(k)|\mathcal{S}(k))$ specification can be made in terms of the particular factors.

We shall assume that

$$p(G(k)|F(k),\lambda(k),\mathcal{S}(k)) = p(G(k)|\lambda(k),\mathcal{S}(k)) = \mathcal{N}_{G(k)}\left(0, \frac{\lambda(k)}{\varepsilon(k)} I_k \otimes I_n\right), \tag{10}$$

where $\varepsilon(k) > 0$ can be interpreted as noise-to-signal-uncertainty ratio,

$$p(F(k)|\lambda(k),\mathcal{S}(k)) = p(F(k)) = \text{ uniform on } F^*(k) \tag{11}$$

and

$$p(\lambda(k)|\mathcal{S}(k)) \propto \lambda^{-\alpha(k)-1}(k)\exp[-\beta(k)\lambda^{-1}(k)], \tag{12}$$

where the "shaping" parameters $\alpha(k), \beta(k)$ have to be positive in order to get a proper prior p.d.f.

The self-reproducing forms are selected in order to keep algebraic level of evaluations. A specific choice of the shaping parameters $\alpha(k), \beta(k), \varepsilon(k)$ is discussed below.

The assumed zero prior expectation of $G$ is the only serious restriction. It implies that unconditional expectation of data is assumed to be zero. This property can be, at least approximately, met by subtracting sample means of matrix rows.

## 4. SOLUTION

The solution of the formulated problem is decomposed in two parts: the first one follows exactly the above theory, the second one contains an approximation of an analytically unfeasible integration.

### 4.1. Evaluation of $p(Y|F(k),\mathcal{S}(k))$

**Proposition 2.** [*Distribution* $p(Y|F(k),\lambda(k),\mathcal{S}(k))$] For the parametrized model (8) and the prior p.d.f. (10), it holds

$$p(Y|F(k),\lambda(k),\mathcal{S}(k)) = \mathcal{N}_Y(0,\lambda(k)\,\mathcal{C}(k)\otimes I_n) =$$
$$\propto [\lambda(k)]^{-0.5mn}\left(\frac{\varepsilon(k)}{1+\varepsilon(k)}\right)^{0.5kn}\exp[-0.5\,\mathrm{tr}(\lambda^{-1}(k)\,\mathcal{C}^{-1}(k)R)] \tag{13}$$

with

$$\mathcal{C}^{-1}(k) = I_m - \frac{F(k)F'(k)}{1+\varepsilon(k)} \tag{14}$$

$$R \equiv YY'. \tag{15}$$

P r o o f . It is based on a standard integration of multivariate Gaussian distributions, see e. g. [1]. The orthogonality of $F$-columns implies the simple form of

$$\text{determinant}[\mathcal{C}^{-1}(k)] = \left(\frac{\varepsilon(k)}{1+\varepsilon(k)}\right)^{k}, \tag{16}$$

which is needed when integrating. □

**Proposition 3.** [*Distribution* $p(Y|F(k),\mathcal{S}(k))$] For the parametrized model (8) and prior p.d.f.s (10), (12) it holds

$$p(Y|F(k),\mathcal{S}(k))) \propto \left(\frac{\varepsilon(k)}{1+\varepsilon(k)}\right)^{0.5nk} [2\beta(k) + \text{tr}(\mathcal{C}^{-1}(k)R)]^{-0.5(2\alpha(k)+mn)}. \tag{17}$$

The normalizing factor is independent of $F(k)$ and this formula is valid even for $(\alpha(k),\beta(k))$ approaching zero.

P r o o f . It is implied by chain rule [5] and by the simple substitution $x := \lambda(k)/(2\beta(k) + \text{tr}(\mathcal{C}^{-1}(k)R))$. The independency of the normalizing factor of $F(k)$ is obvious from (16). □

### 4.2. Approximate evaluation of $p(Y|\mathcal{S}(k))$ and $p(\mathcal{S}(k)|Y)$

The p.d.f. $p(Y|\mathcal{S}(k))$, needed for completing the evaluations, is given by the formula

$$p(Y|\mathcal{S}(k)) = \text{vol}^{-1}(F^*(k)) \int_{F^*(k)} p(Y|F(k),\mathcal{S}(k))\, \mathrm{d}F(k), \tag{18}$$

where $\text{vol}(F^*(k)) = \int_{F^*(k)} \mathrm{d}F(k)$ is clearly finite due to definition of $F^*(k)$. The complex forms of the integration range and of the integrated function leave no chance for exact analytical evaluation and the dimensionality curse prevents the use of numerical integration. For this reason, a simple upper bound is found and used further on in evaluations.

**Proposition 4.** [*Upper bound on* $p(Y|\mathcal{S}(k))$] For the parametrized model (8) and prior p.d.f.s (10), (11), (12), it holds

$$p(Y|\mathcal{S}(k)) \le c \left(\frac{\varepsilon(k)}{1+\varepsilon(k)}\right)^{0.5nk} \left[1 + \gamma(k) - \frac{\Lambda(k)}{1+\varepsilon(k)}\right]^{-0.5(2\alpha(k)+mn)}, \tag{19}$$

where

$$\gamma(k) = 2\beta(k)/\text{tr}[R], \tag{20}$$

$c$ is a universal, structure independent factor and

$$\Lambda(k) = \sum_{i=1}^{k} e_i, \tag{21}$$

where $e_i$ are eigenvalues of

$$\tilde{R} = \frac{R}{\mathrm{tr}[R]} \tag{22}$$

indexed in the non-increasing order.

P r o o f. Clearly, $p(Y|\mathcal{S}(k)) \leq \max_{F(k)\in F^*(k)} p(Y|F(k),\mathcal{S}(k))$. The maximizing matrix $F(k)$ has to maximize $\mathrm{tr}[F'(k)\tilde{R}F(k)]$, see (14), (17). We factorize the positive semidefinite $\tilde{R} = T\mathcal{D}T'$ where $T$ is an orthogonal matrix and $\mathcal{D}$ a non-negative diagonal one with non-increasing entries. Factorizing $T = [T(k), T^c(k)]$ we see that $T(k)Q(k) \in F^*(k)$ for the $(k,k)$-orthogonal matrix $Q(k)$ introducing lower triangular structure into leading $k$ rows of $T(k)$. For $\bar{F}(k) = T(k)Q(k)$ we have

$$\mathrm{tr}[\bar{F}'(k)\tilde{R}\bar{F}(k)] = \sum_{i=1}^{k} \mathcal{D}_i.$$

Taking any $F(k) \in F^*(k)$, we have $\mathrm{tr}[F'(k)\tilde{R}F(k)] = \sum_{i=1}^{k} f_i'(k)\tilde{R}f_i(k)$ where $f_i(k)$ are the orthonormal columns of $F(k)$. Elementary properties of eigenvalues forming the diagonal entries of $\mathcal{D}$ imply that $\sum_{i=1}^{k} f_i'(k)\tilde{R}f_i(k) \leq \sum_{i=1}^{k} \mathcal{D}_i$, thus $\bar{F}(k)$ is the maximizing point in $F^*(k)$. Inserting $\bar{F}(k)$ into $p(Y|F(k),\mathcal{S}(k))$ (17) we get the formula (19) after simple manipulations. □

If the p.d.f. $p(Y|F(k),\mathcal{S}(k))$ is well peaked, which happens for large $n$, then the approximation (19) is sharp enough and the probability function

$$\hat{p}(\mathcal{S}(k)|Y) \propto \left(\frac{\varepsilon(k)}{1+\varepsilon(k)}\right)^{0.5nk} \left[1 + \gamma(k) - \frac{\Lambda(k)}{1+\varepsilon(k)}\right]^{-0.5(2\alpha(k)+mn)} p(\mathcal{S}(k)) \tag{23}$$

is a reasonable approximation of $p(\mathcal{S}(k)|Y)$.

### 4.3. Selection of shaping parameters

The estimation results are undoubtedly influenced by the shaping parameters $\alpha(k)$, $\beta(k)$, $\varepsilon(k)$ which should be supplied by the user. The end-user will exploit the proposed procedure if requirements put on him by their choice will be weaker than the direct choice of the rank. Thus, our option has to be as universal as possible.

The parameters $\alpha(k)$, $\beta(k)$ determine the prior p.d.f. of $\lambda(k)$. Both will be formally set to zero as admitted by Proposition 3.

It is motivated by the following reasoning. These parameters determine prior expected ($\mathcal{E}$) value of $\lambda(k)$ and its dispersion (disp) as follows

$$\mathcal{E}[\lambda(k)|\mathcal{S}(k)] = \frac{\beta}{\alpha-1} \text{ for } \alpha > 1, \quad \mathrm{disp}[\lambda(k)|\mathcal{S}(k)] = \frac{\mathcal{E}^2[\lambda(k)|\mathcal{S}(k)]}{\alpha-2} \text{ for } \alpha > 2.$$

The dispersion is influenced by $\alpha$ in a decisive manner and it should be relatively large in order to get universal flat priors. Thus, we should take $\alpha$ not very far from 2. This value is, however, added to the much larger product $mn/2$ (cf. (23)) and thus it can be formally set to zero.

The parameter $\beta(k)$ determines guess of the noise level. It enters, however, the final formula just through the value $\gamma(k) = 1+2\beta(k)/\mathrm{tr}[R]$ in which the second term should be less than $2/(mn)$. Thus, we can put $\gamma(k) \approx 1$ which is equivalent to the option $\beta(k) = 0$.

It is intuitively clear and it was experimentally verified that the choice of the parameter $\varepsilon(k)$ is critical. Its interpretation (noise-to-signal ratio) indicates that it should be chosen independently of $\mathcal{S}(k)$. Moreover, the information about its appropriate value should be contained in data. It would be potentially possible to take it as hyperparameter of the $p(Y|\mathcal{S}(k))$ and estimate it, too. It would complicate the evaluations. For this reason, the following simple guess is constructed: Let $\tilde{m}$ be conservative guess of the number of non-zero eigenvalues of $R$. Then $(m - \tilde{m})(1 - \Lambda(\tilde{m}))\mathrm{tr}[R]$ is proportional to the uncertainty attributed to $m - \tilde{m}$ "zero" eigenvalues of $R$ and $\tilde{m}\Lambda(\tilde{m})\mathrm{tr}[R]$ to that of significant eigenvalues. This together with the definition of $\varepsilon$ motivates the $\varepsilon$ estimate

$$\hat{\varepsilon} = \frac{1 - \Lambda(\tilde{m})}{\Lambda(\tilde{m})} \frac{m - \tilde{m}}{\tilde{m}}. \tag{24}$$

### 4.4. Rank estimate – algorithm and expected properties

Assuming well peaked p.d.f. $p(Y|F(k), \mathcal{S}(k))$ and the discussed options of shaping parameters we get the final approximate p.f.

$$\hat{p}(\mathcal{S}(k)|Y) \propto \left(\frac{\hat{\varepsilon}(k)}{1 + \hat{\varepsilon}(k)}\right)^{0.5nk} \left[1 - \frac{\Lambda(k)}{1 + \hat{\varepsilon}(k)}\right]^{-0.5mn} p(\mathcal{S}(k)) \tag{25}$$

which is expected to be close to $p(\mathcal{S}(k)|Y)$.

The following algorithm summarizes the necessary evaluations and provides, as an example, (approximate) maximum posterior probability (MAP) estimate of the rank.

**MAP ESTIMATE OF RANK**

**Pre-processing**

1. Measure raw data $(m, n)$-matrix $Z$.
2. Cut off outlying entries using physically justified ranges of data and their changes.
3. Subtract sample means of $Z$-rows (images) from them.
4. If need be, scale the resulting matrix in order to reach approximately common level of uncertainty for all entries of the resulting data $Y$.
5. Assign prior probabilities $p(\mathcal{S}(k))$ to particular ranks $\mathcal{S}(k)$.

**Processing**

1. Evaluate $\tilde{R} = YY'/\mathrm{tr}[YY']$.
2. Determine $\tilde{m} \leq m$ eigenvalues $e_i \geq e_{i+1}$ of $\tilde{R}$ which are a priori non-zero and larger than the numerical-noise level.
3. Compute $\Lambda(k) = \sum_{i=1}^{k} e_i,\ k = 1, \ldots, \tilde{m}$.
4. Set $\hat{\varepsilon} = (1 - \Lambda(\tilde{m}))(m - \tilde{m})/(\Lambda(\tilde{m})\tilde{m})$.
5. Evaluate

$$\tilde{\mathcal{L}}(k) = \frac{k}{m}\log(1+\hat{\varepsilon}^{-1}) + \log(1 - \frac{\Lambda(k)}{1+\hat{\varepsilon}}) - 0.5(nm)^{-1}\log(p(\mathcal{S}(k))),\ k = 1, \ldots, \tilde{m}. \tag{26}$$

6. Take $\hat{k}$ *minimizing* $\tilde{\mathcal{L}}(k)$ as approximate MAP rank estimate.

**Commentaries**

1. Outliers are known to make the most severe deviation from normality, their removal should make the deviation small.
2. Removal of sample mean makes the assumption of zero expectation of $G$ realistic.
3. Stopping of evaluations at numerical noise level spares computation time: it removes those singular values which are zero "by naked eyes", moreover it helps to provide sensible estimate of $\hat{\varepsilon}$.
4. Uniform prior probabilities are chosen as a rule. This user's input makes possible to fix a range of physically sensible ranks. For instance, if the rank corresponds to the number of possible compartments displayed on the images treated such boundaries are realistically known.
5. Determination of eigenvalues (squares of singular values) is needed for the further processing in factor analysis, thus the extra effort for the rank estimation is negligible.
6. The minimized quantity $\tilde{\mathcal{L}}(k)$ is equal to $-\frac{2}{nm}\log$(of $k$-dependent part of $\hat{p}(\mathcal{S}(k)|Y)$). It
   (a) is independent of $n$ for the usual uniform prior on $\mathcal{S}(k)$,
   (b) is sum of
   – increasing term reflecting uncertainty of the parameter $G(k)$ (this part is the key consequence of the Bayesian treatment and has no counterpart in fully maximum-likelihood based approach)
   – decreasing term expressing misfit of data when using their low-rank approximation ($\log(1 - \Lambda(k)/(1+\hat{\varepsilon}))$)
   – corrective term given by prior p.f.

## 5. EXPERIMENTAL VERIFICATION

The following tables contain eigenvalues of $\tilde{R}$ gained for some simulated and real data. The number of tested values is always $\tilde{m} = 5$. Approximate MAP estimate $\hat{k}$ is provided together with approximate expected value $\mathcal{E}[k]$ gained from approximate p.f. gained by normalizing the upper bound (19) to p.f. It gives a handy measure of posterior p.d.f. concentration. For information, $\hat{\varepsilon}$ is provided, too.

The selection of a more extensive set of experiments should illustrate properties of the proposed procedure especially its robustness. For this reason, variations in dimensions and noise level are made and various forms (changes in sampling, smoothing, number of exploited images etc.) of the same real data analyzed.

**Estimation results for simulated studies**

| No | $m$<br>$n$ | data<br>type | eigenvalues $*$ 100 | true<br>$k$ | $\hat{k}$ | $\hat{\mathcal{E}}[k]$ | $\hat{\varepsilon}$ |
|---|---|---|---|---|---|---|---|
| 1 | 12 | case $A$ | | | | | |
| | 256 | no noise | 87.89 12.09 0 0 0 | 2 | 2 | 2 | 0.00028 |
| 2 | 12 | case $A$ | | | | | |
| | 256 | low noise | 87.317 11.942 0.176 0.139 0.11 | 2 | 2 | 2 | 0.00444 |
| 3 | 12 | case $A$ | | | | | |
| | 256 | medium noise | 74.225 11.802 2.851 2.523 1.857 | 2 | 2.00009 | 2 | 0.10121 |
| 4 | 12 | case $A$ | | | | | |
| | 256 | strong noise | 14.601 12.621 10.633 9.882 9.176 | 2 | 4.17298 | 5 | 1.05989 |
| 5 | 12 | case $B$ | | | | | |
| | 256 | medium noise | 73.082 8.555 3.112 2.865 2.14 | 2 | 1.99947 | 2 | 0.15982 |
| 6 | 54 | case $C$ | | | | | |
| | 256 | low noise | 88.566 4.85 0.76 0.717 0.652 | 2 | 2.55125 | 2 | 0.45695 |
| 7 | 14 | case $C$ | | | | | |
| | 256 | low noise | 87.533 5.852 2.639 1.714 0.978 | 2 | 4.01605 | 4 | 0.02341 |
| 8 | 54 | case $D$ | | | | | |
| | 256 | low noise | 83.03 10.05 1.814 0.448 0.427 | 3 | 3.07952 | 3 | 0.43296 |
| 9 | 14 | case $D$ | | | | | |
| | 256 | low noise | 81.033 12.006 2.325 1.485 1.058 | 3 | 2.70891 | 3 | 0.03848 |
| 10 | 54 | case $E$ | | | | | |
| | 256 | low noise | 85.429 10.361 1.754 0.246 0.206 | 3 | 3.05586 | 3 | 0.20041 |
| 11 | 14 | case $E$ | | | | | |
| | 256 | low noise | 83.248 12.158 2.433 0.655 0.474 | 3 | 3 | 3 | 0.01877 |
| 12 | 54 | case $F$ | | | | | |
| | 256 | low noise | 95.003 1.599 0.533 0.365 0.309 | 2 | 2.23165 | 2 | 0.21953 |
| 13 | 14 | case $F$ | | | | | |
| | 256 | low noise | 94.639 2.326 1.262 0.662 0.334 | 2 | 2.97495 | 3 | 0.01410 |

**Estimation results for real-data studies**

| No | $m$<br>$n$ | data<br>type | eigenvalues $*$ 100 | "true"<br>$k$ | $\hat{k}$ | $\hat{\mathcal{E}}[k]$ | $\hat{\varepsilon}$ |
|---|---|---|---|---|---|---|---|
| 1 | 50 | case $G$ | | | | | |
| | 1533 | | 61.673 8.822 2.953 1.259 1.046 | 3 | 3.1579 | 3 | 2.88072 |
| 2 | 40 | case $G$ | | | | | |
| | 1533 | | 65.313 7.285 3.605 1.504 1.205 | 3 | 3.0563 | 3 | 1.87064 |
| 3 | 30 | case $G$ | | | | | |
| | 1533 | | 70.346 6.178 3.866 1.795 1.344 | 3 | 3.00096 | 3 | 0.98594 |
| 4 | 25 | case $G$ | | | | | |
| | 1533 | | 72.569 6.679 3.67 1.684 1.438 | 3 | 2.99316 | 3 | 0.64900 |
| 5 | 20 | case $G$ | | | | | |
| | 1533 | | 75.022 7.834 3.432 1.777 1.357 | 3 | 2.58469 | 3 | 0.35488 |
| 6 | 15 | case $G$ | | | | | |
| | 1533 | | 75.972 9.925 3.612 1.888 1.445 | 3 | 2.05573 | 2 | 0.15420 |
| 7 | 10 | case $G$ | | | | | |
| | 1533 | | 76.847 12.511 3.4 2.435 1.443 | 3 | 2 | 2 | 0.03481 |
| 8 | 24 | case $H$ | | | | | |
| | 2289 | | 87.722 3.165 0.657 0.493 0.474 | 2 | 2 | 2 | 0.30762 |
| 9 | 24 | case $H$ | | | | | |
| | 2289 | | 85.646 3.056 0.861 0.623 0.603 | 2 | 1.99917 | 2 | 0.38553 |
| 10 | 56 | case $I$ | | | | | |
| | 1035 | | 86.654 6.503 1.927 0.406 0.295 | 3 | 3.00122 | 3 | 0.44885 |
| 11 | 28 | case $I$ | | | | | |
| | 1035 | | 86.028 6.941 2.225 0.547 0.448 | 3 | 3 | 3 | 0.18225 |
| 12 | 14 | case $I$ | | | | | |
| | 1035 | | 84.827 7.997 2.903 0.789 0.668 | 3 | 3 | 3 | 0.05216 |
| 13 | 56 | case $J$ | | | | | |
| | 526 | | 83.829 7.769 2.271 0.717 0.463 | 3 | 3.15339 | 3 | 0.53131 |
| 14 | 28 | case $J$ | | | | | |
| | 526 | | 82.962 8.21 2.93 1.111 0.629 | 3 | 3.00498 | 3 | 0.19957 |
| 15 | 14 | case $J$ | | | | | |
| | 526 | | 81.293 9.516 4.303 1.239 0.774 | 3 | 3 | 3 | 0.05328 |
| 16 | 56 | case $J$ | | | | | |
| | 526 | | 87.635 8.047 2.175 0.523 0.333 | 3 | 3.27655 | 3 | 0.13299 |
| 17 | 56 | case $J$ | | | | | |
| | 1035 | | 90.051 6.509 1.777 0.309 0.206 | 3 | 3.00015 | 3 | 0.11846 |
| 18 | 14 | case $J$ | | | | | |
| | 526 | | 84.691 9.528 3.795 0.712 0.446 | 3 | 3 | 3 | 0.01503 |
| 19 | 56 | case $J$ | | | | | |
| | 526 | | 89.289 7.525 1.992 0.39 0.264 | 3 | 3.92924 | 4 | 0.05538 |
| 20 | 56 | case $J$ | | | | | |
| | 1035 | | 91.297 6.204 1.57 0.248 0.168 | 3 | 3.00146 | 3 | 0.05260 |

## 6. CONCLUSIONS

We have derived a simple, reasonably justified estimate of the matrix rank. It outperforms the available tests on data we have tried which are – according to

our best knowledge – representative in our application field, i. e. factor analysis of dynamic scintigraphic studies. We believe that the found simple formula suits well in a wide range of other applications, too, as a useful statistical complement of the popular singular value decomposition. This belief, however, has to be supported both experimentally and theoretically in the future.

(Received March 2, 1993.)

REFERENCES

[1] M. H. DeGroot: Optimal Statistical Decisions. McGraw-Hill Company New York – St. Louis – San Francisco – London – Sydney – Toronto 1970.
[2] G. H. Golub and C. F. van Loan: Matrix Computations. John Hopkins University Press, Baltimore – London 1989.
[3] M. Šámal, M. Kárný and D. Zahálka: Confirmatory aspects in factor analysis of image sequences. In: Information Processing in Medical Imaging, (A. C. F. Colchester and D. J. Hawkes, eds.), Springer Verlag, Berlin – Heidelberg 1991, pp. 397–407.
[4] P. Hannequin, J. C. Liehm and J. Valeyre: The determination of the number of statistically significant factors in factor analysis of dynamic structures. Phys. Med. Biol. *34* (1989), 2, 1213–1227.
[5] V. Peterka: Bayesian system identification. In: Trends and Progress in System Identification, (P. Eykhoff, ed.), Pergamon Press, Oxford 1981, pp. 239–304.

*Ing. Miroslav Kárný, DrSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of The Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.*

*MUDr. Martin Šámal, DrSc., Ústav nukleární mediciny, 1. Lékařská fakulta Karlovy university (Institute of Nuclear Medicine, 1st Medical Faculty, Charles University), Salmovská 3, Praha 2. Czech Republic.*