

CHAPTER 1

INTRODUCTION

Since the discovery of so-called Stein effect, the multiparameter estimation which occurs rather naturally in statistical decision-theoretic problems has been studied in many literatures. More recently the estimation problem in which procedures revolve mainly around the eigenstructures of random and parameter matrices has been received a lot of attention. There is a growing literature relating to the eigenvalues estimation: A useful review paper by Muirhead[41] gives many references. This thesis is concerned with possible decision theoretic approach to the following problems:

1. Let X be an $m \times p$ matrix normally distributed with matrix of mean B and a covariance matrix $I_m \otimes \Sigma$ and let S be a $p \times p$ Wishart matrix with n degrees of freedom and mean $n\Sigma$. We wish to estimate the mean matrix B .
2. Let U be a $p \times p$ random matrix, having the multivariate F-distribution with a scale matrix Δ and degrees of freedom k_1 and k_2 . We wish to estimate the eigenvalues of Δ .

The usual or classical estimator for these problems, i.e., the maximum likelihood estimator or the unbiased estimator, is fully equivariant. It means that, under any groups of transformations, it transforms in the same way as its estimand. Equivariance is desirable but not compelling property as shown in Stein[48] that, in estimating a p -variate normal mean vector with identity covariance matrix, the sample mean is inadmissible in terms of sum of the squared-error loss if $p \geq 3$. Moreover, such a classical estimate ignores information about the ordered eigenvalues of the random and parameter matrices in the multivariate situation. All these arguments suggest that a systematic treatment of above problems can probably give superior alternatives to the commonly used estimator.

1.1. DEFINITION

The notation $A : p \times q$ is a matrix of p rows and q columns. The transpose of A is denoted by A' . When $p = q$, then $\det A$ or $|A|$ is the determinant of the square matrix A . The trace of A is denoted by $\text{tr } A$, that is, the sum of diagonal elements of square matrix A . $D = \text{diag}(d_1, d_2, \dots, d_p)$ is a diagonal matrix with diagonal elements d_1, d_2, \dots, d_p . The expected value of a random variable X is denoted by $\mathbf{E} X$.

Next let us recall some of terminology and definitions involved in decision-theoretic estimation.

Let X denote a random variable whose distribution depends on an unknown parameter θ . Here X can be a vector or matrix, as can θ . Let $\hat{\theta}(X)$ denote an estimate of θ . We shall denote its estimate of parameter by "hat" over parameter. A *loss function* $L(\theta, \hat{\theta}(X))$ is a non-negative function of θ and $\hat{\theta}(X)$. The *risk function* $\mathbf{R}(\theta, \hat{\theta}(X))$ is the expectation of this loss function with respect to the distribution of X when θ represents the true value of the parameter. An estimate $\theta_1(X)$ is said to be *better than* or *beat* estimate $\hat{\theta}_2(X)$ if

$$\mathbf{R}(\theta, \hat{\theta}_1(X)) \leq \mathbf{R}(\theta, \hat{\theta}_2(X)) \quad \text{for } \forall \theta \in \Omega,$$

where Ω is the parameter space of the distribution of X , and

$$\mathbf{R}(\theta, \hat{\theta}_1(X)) < \mathbf{R}(\theta, \hat{\theta}_2(X)) \quad \text{for at least one } \theta.$$

An estimate is said to be *admissible* if there exists no estimate which beats it. If there is an estimate which beats it, it is called *inadmissible*. An estimate $\theta_0(X)$ is said to be *minimax* if there is no other estimate whose risk function has smaller supremum, i.e.,

$$\sup_{\theta \in \Omega} \mathbf{R}(\theta, \hat{\theta}_0(X)) = \inf_{\hat{\theta}(X) \in \mathbf{C}} \sup_{\theta \in \Omega} \mathbf{R}(\theta, \hat{\theta}(X)),$$

where \mathbf{C} denotes the class of estimators of θ .

To introduce invariance into decision problem, let G be a topological group which acts on \mathbf{X} , Ω , and \mathbf{A} where \mathbf{X} is a sample space and \mathbf{A} is an action space. Let $\mathcal{P} = \{P_\theta | \theta \in \Omega\}$ be a family of distribution of X . The family \mathcal{P} is said to be *invariant* under G if

$$gP_\theta = P_{g\theta} \quad \text{for } g \in G, \theta \in \Omega.$$

The loss function is said to be *invariant* under G if

$$\mathbf{L}(g\theta, g\hat{\theta}) = \mathbf{L}(\theta, \hat{\theta}) \quad \text{for } g \in G, \theta \in \Omega, \hat{\theta} \in \mathbf{A}.$$

An estimation problem is said to be *invariant* under G if the family of distributions and the loss function are invariant.

An estimator $\hat{\theta}_0(X)$ is said to be *equivariant* under G if

$$\hat{\theta}_0(gX) = g\hat{\theta}_0(X) \quad \text{for } g \in G, X \in \mathbf{X}.$$

Eaton[13] provides a good introduction to invariant decision problem and some techniques for finding "good" invariant decision rules.

1.2. THE GENERAL METHOD OF IMPROVING UPON USUAL ESTIMATOR

As seen in Stein[48], a fully equivariant estimator such as the maximum likelihood estimator or the unbiased estimator can be improved by simply requiring less invariance. However, once we look outside the class of fully equivariant estimators, the expression of its risk function becomes intractable in the multivariate situation. To overcome this, we shall employ the following method, which was first introduced by Stein[48]. The following description of this method comes from Loh[36].

1. Narrow the class of the estimators, for example, using equivariance in the problem and work out the form of the estimators.
2. Compute the unbiased estimate of the risk of the estimators under consideration using the integration by parts formula.
3. Determine promising alternative estimators from the unbiased estimate of risk.

By reducing the size of contenders, the problem reduces the difficulty of the search for a superior alternative to the commonly used estimator. Furthermore, equivariance is imposed on pooling across rows and columns, which implies that the resulting unbiased risk estimate depends only on the maximum invariant statistics if the estimation problem considered is invariant. The final step is rather difficult to deal with, since it has not been established to obtain the widely applicable way of deriving promising estimators (which has frequentist risk uniformly smaller than the classical estimator) from unbiased risk estimate. Since the introduction of this approach, numerous researchers have applied this technique to the problems in statistical decision theory. The literature includes Berger[2], Dey and Srinivasan[11], Efron and Morris[15], Haff[19, 22, 23], Loh[35, 36, 37], Muirhead and Verathworn[42], and Perron[46]. See Berger[3] for an extensive reference.

1.3. SUMMARY OF RESULTS

In Chapter 2, the problem of estimating matrix of normal mean is considered relative to the invariant loss function where the covariance matrix is unknown. Certain form of invariant estimators is introduced. Then the unbiased estimate of the risk is obtained, which depends on the eigenvalues of the usual F-matrix $X'XS^{-1}$. It facilitates an extensive search for superior alternatives to the commonly used estimator X . New classes of invariant minimax estimators are proposed, which are multivariate extensions of the estimators of James and Stein[24], Baranchik[1], and Lin and Tsai[35]. These results also extend the estimators of mean matrix of Stein[48] and Zheng[54] for the case in which the covariance matrix is known. Furthermore, following an approach by Haff[23], an alternative estimator as the solution of the Euler-Lagrange system of partial differential equations is derived.

Chapter 3 is devoted to the estimation of the eigenvalues in the multivariate F-distribution. First, using the F-identity we derive the second order moments of this distribution, which is useful to the statistical inference. For the Stein's loss function, we obtain the improved estimators (Haff-type and Perron-type) which beat the unbiased estimator. Furthermore, we derive the orthogonally invariant minimax estimators when the scale matrix is 2×2 . For the squared loss function, we give the Haff-type improved estimators which beat the best scalar multiple estimator.