ON A PARADOX CONCERNING INFERENCE ABOUT A COVARIANCE MATRIX[1]

By A. P. Dempster

Harvard University

1. Introduction and summary. Suppose a $p \times p$ dispersion matrix $\mathbf{T}$ is considered to have the Wishart distribution $W(\mathbf{\Sigma}, n)$, c. f. Anderson (1958) p. 158, where $\mathbf{\Sigma}$ is an arbitrary full rank covariance matrix and $n \geqq p$. Suppose $\mathbf{T}$ is observable but $\mathbf{\Sigma}$ is unknown, and suppose a posterior distribution is to be assigned to $\mathbf{\Sigma}$ given $\mathbf{T}$, where the term posterior is meant in a wide sense to allow the use of a Bayesian or fiducial or any other form of reasoning in arriving at the posterior distribution. A lemma is proved in Section 3 giving a property of all such posterior distributions which possess a natural linear invariance property. The paradoxical nature of this property is discussed in Section 4.

2. Notation and preliminaries. Two basic properties of the Wishart distribution are:

(I) *For any* $p \times 1$ *vector* $\mathbf{a}$, *the ratio*

$$P(\mathbf{a}) = \mathbf{a}'\mathbf{T}\mathbf{a}/\mathbf{a}'\mathbf{\Sigma}\mathbf{a} \tag{2.1}$$

has the χ_n^2 *distribution, and*

(II) *For any* $p \times 1$ *vector* $\mathbf{b}$, *the ratio*

$$Q(\mathbf{b}) = \mathbf{b}'\mathbf{\Sigma}^{-1}\mathbf{b}/\mathbf{b}'\mathbf{T}^{-1}\mathbf{b} \tag{2.2}$$

has the χ_{n-p+1}^2 *distribution.*

Particular examples of (I) and (II) are very familiar. For example, suppose t_{11} and σ_{11} denote the first diagonal elements of $\mathbf{T}$ and $\mathbf{\Sigma}$. Then the χ_n^2 distribution of t_{11}/σ_{11} is an example of (I). Again, regarding $\mathbf{T}$ as a dispersion matrix of a set of p variates, suppose $t_{11\cdot 23\cdots p}$ denotes the residual dispersion of the first variate after fitting the best linear function of the remaining $p - 1$ variates. More precisely, $t_{11\cdot 23\cdots p} = 1/t^{11}$ where t^{11} is the first diagonal element of $\mathbf{T}^{-1}$. Similarly, define $\sigma_{11\cdot 23\cdots p} = 1/\sigma^{11}$ where σ^{11} is the first diagonal element of $\mathbf{\Sigma}^{-1}$. Then the familiar χ_{n-p+1}^2 distribution of $t_{11\cdot 23\cdots p}/\sigma_{11\cdot 23\cdots p} = \sigma^{11}/t^{11}$ is an example of (II). The general forms of (I) and (II) follow from the above particular examples together with the linear invariance property of the normal $N(\boldsymbol{\mu}, \mathbf{\Sigma})$ distribution related to the $W(\mathbf{\Sigma}, n)$ distribution.

It is convenient to have a representation of this situation in geometrical terms

---

Received January 10, 1963.

[1] This research has been supported in part by the United States Navy through the Office of Naval Research, under contract Nonr 1866(37).

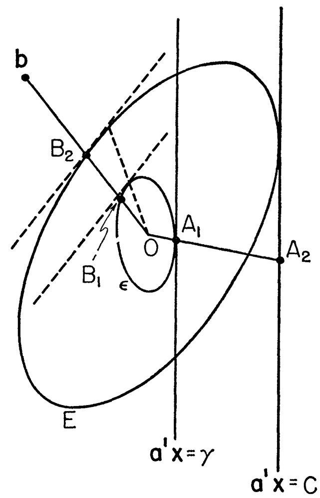

FIG. 1. The geometrical entities required for a geometrical description of $\mathbf{T}$, $\mathbf{\Sigma}$, $P(\mathbf{a})$ and $Q(\mathbf{b})$.

as in Figure 1. Here E and ϵ represent the concentration ellipsoids with equations $\mathbf{x}'\mathbf{T}^{-1}\mathbf{x} = 1$ and $\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x} = 1$, respectively. O denotes the origin, and B_1 and B_2 are points where the line joining O and $\mathbf{b}$ meets ϵ and E, respectively. The hyperplanes $\mathbf{a}'\mathbf{x} = \gamma$ and $\mathbf{a}'\mathbf{x} = C$ are tangent to ϵ and E, respectively; A_1 is the point of contact of the former and A_2 is the intersection of the line OA_1 with the latter.

It is easily checked that $P(\mathbf{a})$ and $Q(\mathbf{b})$ in (2.1) and (2.2) are related to line segment ratios as in Figure 1. Specifically,

$$P(\mathbf{a}) = (OA_2/OA_1)^2, \quad \text{and} \tag{2.3}$$

$$Q(\mathbf{b}) = (OB_2/OB_1)^2. \tag{2.4}$$

Note that, given $\mathbf{\Sigma}$, $\mathbf{a}$ and $\mathbf{b}$ may be chosen so that A_1 and B_1 coincide. It is then clear that $OA_2 \geqq OB_2$ so that $P(\mathbf{a}) \geqq Q(\mathbf{b})$ for this $\mathbf{a}$ and $\mathbf{b}$. This inequality is reflected in the χ_n^2 and χ_{n-p+1}^2 distributions assigned to $P(\mathbf{a})$ and $Q(\mathbf{b})$, the former being larger in the sense that the c. d. f. of χ_n^2 is uniformly less than the c. d. f. of χ_{n-p+1}^2.

3. A lemma. Suppose $H(\mathbf{T})$ denotes a posterior distribution assigned to $\mathbf{\Sigma}$ when $\mathbf{T}$ is observed. For any non-singular $\mathbf{C}$ the original coordinates $\mathbf{x}$ for the p underlying variates may be replaced by $\mathbf{y} = \mathbf{Cx}$ and, in the new coordinates, the dispersion and covariance matrices become $\mathbf{CTC}'$ and $\mathbf{C\Sigma C}'$, respectively. For statistical methods based on the multivariate normal distribution it is natural to require that an inference be free of the choice of the coordinates $\mathbf{y}$. Accordingly, $H(\mathbf{T})$ will be said to be *linearly invariant* if, when $\mathbf{\Sigma}$ has the $H(\mathbf{T})$ distribution, then $\mathbf{C\Sigma C}'$ has the $H(\mathbf{CTC}')$ distribution. Linear invariance of $H(\mathbf{T})$ implies that $P(\mathbf{a})$ and $Q(\mathbf{b})$ in (2.1) and (2.2), regarded now as functions of random $\mathbf{\Sigma}$ for given $\mathbf{T}$, again have distributions which are free of the choice of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{T}$.

LEMMA. *For any linearly invariant posterior distribtuion of* $\mathbf{\Sigma}$ *given* $\mathbf{T}$, *the distribution of* $P(\mathbf{a})$ *must be smaller than or equal to the distribution of* $Q(\mathbf{b})$, *in the sense that the* c. d. f. *of the former must be uniformly greater than or equal to the* c. d. f. *of the latter.*

The proof is easily given in terms of line segment ratios. When E is regarded as fixed and ϵ as random, $\mathbf{a}$ and $\mathbf{b}$ may be chosen so that $\mathbf{a}'\mathbf{x} = C$ is tangent to E at B_2. (See the dotted lines in Figure 1.) In this case it is clear from the geometry that

$$OA_2/OA_1 \leqq OB_2/OB_1 \tag{3.1}$$

so that $P(\mathbf{a}) \leqq Q(\mathbf{b})$ and the result follows immediately.

4. Discussion. A consequence of the above lemma is that it is impossible, when $\mathbf{T}$ is regarded as fixed and $\mathbf{\Sigma}$ is regarded as random, to continue to regard $P(\mathbf{a})$ and $Q(\mathbf{b})$ as having the χ_n^2 and χ_{n-p+1}^2 distributions as specified by properties (I) and (II) in Section 2. Indeed, to continue to regard $P(\mathbf{a})$ as having the χ_n^2 distribution is to imply that $Q(\mathbf{b})$ has a distribution at least as large, where the term large is used as above. Or, to continue to regard $Q(\mathbf{b})$ as having the χ_{n-p+1}^2 distribution is to imply that $P(\mathbf{a})$ has a distribution at least as small.

On the other hand, I believe that $P(\mathbf{a})$ and $Q(\mathbf{b})$ are often regarded as pivotal quantities on which to base confidence or fiducial statements about $\mathbf{a}'\mathbf{\Sigma}\mathbf{a}$ or $\mathbf{b}'\mathbf{\Sigma}^{-1}\mathbf{b}$, and that the essential feature of such pivotal quantities is that one shall continue to regard their stated distributions as valid for inferences when the observable $\mathbf{T}$ is fixed. Consider a specific situation where $p = n = 100$. From Property (I) of Section 2, t_{11}/σ_{11} is regarded as a χ_{100}^2 random variable given $\mathbf{\Sigma}$. From this, one is led to make confidence or fiducial statements consistent with the assertion that t_{11}/σ_{11} is a χ_{100}^2 random variable where t_{11} is fixed at its observed value and σ_{11} is regarded as an unknown variable. Thus one is led to the loose assertion:

(A) The unknown σ_{11} is roughly of the order of $t_{11}/100$ where t_{11} is known.

Similarly, from Property (II) of Section 2, one may base confidence or fiducial statements on the χ_1^2 pivotal quantity $t_{11\cdot 23\cdots p}/\sigma_{11\cdot 23\cdots p}$ and be led to the loose assertion:

(B) The unknown $\sigma_{11\cdot 23\cdots p}$ is roughly of the same order as $t_{11\cdot 23\cdots p}$ where $t_{11\cdot 23\cdots p}$ is known.

Within the tradition of the past thirty years, the temptation is very strong to regard assertions such as (A) and (B) as roughly acceptable inferences about the unknown quantities σ_{11} and $\sigma_{11\cdot 23\cdots p}$. Yet, if one desires inferences consistent with an overall linearly invariant posterior distribution, the foregoing inferences are contradictory, and one or other of them must be altered by a factor of roughly 100. In general, the paradox appears in extreme form only when p is close to n

and both are large, but, as a matter of principle, the contradiction between the two types of confidence or fiducial statements appears for any p and n.

Lindley (1958) has also demonstrated the incompatibility of certain confidence or fiducial statements with any posterior statement, but with a restriction to Bayes posterior statements. Again, the present example has features in common with that of Stein (1959) who gave an example of wide discrepancy between a fiducial statement and a confidence statement for the same unknown parameter. In the present situation there is a pair of confidence statements for different parameters, at least one of which must be widely discrepant with any fiducial statement deduced from a linearly invariant joint fiducial distribution of all the parameters. One might suspect that this phenomenon is due to my requirement of linear invariance, but I incline rather to the view that it is simply an illustration of the general failure of confidence statements to agree with posterior distributions, and I suspect that similar examples could be found where an invariance restriction is not involved. The present example reinforces my general view that the confidence argument cannot be trusted when posterior probability statements are required.

It is of interest to see what becomes of Assertions (A) and (B) when certain obvious linearly invariant posterior distributions are assigned to $\mathbf{\Sigma}$, namely when $\mathbf{\Sigma}^{-1}$ is assigned one of the family of Wishart distributions $W(\mathbf{T}^{-1}, k)$. The particular choice $k = n$ appears plausible because the original assumption of a $W(\mathbf{\Sigma}, n)$ distribution for $\mathbf{T}$ is simply inverted into a $W(\mathbf{T}^{-1}, n)$ distribution for $\mathbf{\Sigma}^{-1}$. For general k, the $W(\mathbf{T}^{-1}, k)$ distribution results from a formal Bayesian argument with prior density of $\mathbf{\Sigma}^{-1}$ proportional to

$$|\mathbf{\Sigma}|^{\frac{1}{2}(-n+p+1-k)},$$

c.f., Geisser and Cornfield (1963).

From Properties (I) and (II) of Section 2 with $\mathbf{T}$, $\mathbf{\Sigma}$ and n replaced by $\mathbf{\Sigma}^{-1}$, $\mathbf{T}^{-1}$ and k, respectively, the quantities $P(\mathbf{a})$ and $Q(\mathbf{b})$ are assigned χ^2_{k-p+1} and χ^2_k distributions, respectively, as posterior distributions depending on random $\mathbf{\Sigma}$ given $\mathbf{T}$. This is to be contrasted with the χ^2_n and χ^2_{n-p+1} distributions originally assigned to the same pair of quantities. Note that the posterior distribution interchanges the sizes of the two distributions in accordance with the lemma of Section 3. Consider the specific case $p = n = 100$ and suppose that $k = n$. Then, under the posterior law, Assertion (A) must be replaced by:

(A*) The unknown σ_{11} is roughly of the same order as t_{11} where t_{11} is known.

Similarly Assertion (B) is replaced by

(B*) The unknown $\sigma_{11\cdot 23\cdots p}$ is roughly of the order of $t_{11\cdot 23\cdots p}/100$ where $\sigma_{11\cdot 23\cdots p}$.

The change from (A) and (B) to (A*) and (B*) is quite startling. Nor are matters improved by altering the choice of k, for this merely has the effect of making (A) more like (A*) at the cost of making (B) less like (B*), or vice versa. The question of what constitutes a reasonable posterior inference about $\mathbf{\Sigma}$ appears to me to be wide open.

REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

GEISSER, SEYMOUR and CORNFIELD, JEROME (1963). Posterior distributions for multivariate normal parameters. To appear in *J. Roy. Statist. Soc. Ser. B*.

LINDLEY, D. V. (1958). Fiducial distributions and Bayes' theorem. *J. Roy. Statist. Soc. Ser. B* **20** 102–107.

STEIN, CHARLES (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.* **30** 877–879.