Miszellen

D.A.S. Fraser/Philip McDunnough

Some Remarks on Conditional and Unconditional Inference for Location-Scale Models

Key Words and Phrases: location-scale analysis, conditional analyses, ancillary statistic, necessary reductions, maximum-likelihood estimator, length, power.

Abstract

Conditional and unconditional confidence intervals have been compared by Grice, Bain, and Engelhardt (Commun. Statist. B7 (1978), 515-524) in terms of the location-scale model with double-exponential distribution form. Preference was found for the conditional intervals based on mean length and coverage probability for untrue parameters values. These two criteria for a location-scale system are shown to be inappropriate criteria for assessing the conditional versus unconditional approaches to inference. The usual ancillarity concept is also noted to be inappropriate. Support for many conditional analyses, however, is found in a more careful formulation of the statistical model.

1. Introduction

Recently, Grice, Bain, and Engelhardt (1968) have examined certain problems of inference for the location-scale model with density functions

$$f(y:\mu,\sigma) = \sigma^{-1} g(\sigma^{-1}(y-\mu)) \qquad (1.1)$$

($\mu \varepsilon R$, $\sigma \varepsilon R^+$) in the particular case

$$g(z) = \frac{1}{2}\exp\{-|z|\}$$

Their analysis compared (I) the confidence intervals for μ derived from the marginal distribution of

$$t = (\hat{\mu} - \mu)/\hat{\sigma} \qquad (1.2)$$

for a sample of n with (II) the confidence intervals for μ derived from the conditional distribution of t given the configuration $(d_1,\dots,d_n)$, $d_i = (y_i - \hat{\mu})/\hat{\sigma}$.

We recall that the conditional analysis Π was proposed by Fisher (1934) as a means to "recover information lost" through use of the not-sufficient maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}$. Recently Hinkley (1978) (see also Hinkley and Efron (1978)) has proposed some large-sample approximations for the conditional distribution of t; these approximations include finding the MLE and typically require the use of a computer. It should, however, be pointed out that the exact conditional analysis for the location-scale model is readily available (Fraser, 1976) using the computer program (Fick, 1976).

Grice, Bain, and Engelhardt examine the relative merits of the conditional and the unconditional confidence intervals in terms of three criteria. The first of these is the ancillarity property of the configuration $\underset{\sim}{d}$. Ancillarity is often proferred as grounds for the conditional analysis. However, while supported by a few appealing examples (e.g. Cox and Hinkley, 1974), ancillarity as a general principle of inference has severe shortcomings. Indeed the existence of conflicting ancillaries exhibits the principle as self-contradictory (cf. Fraser, 1973, 1979). Support for the conditional analysis must therefore be found elsewhere.

Grice, Bain, and Engelhardt also consider the criteria of minimizing mean length and maximizing power of confidence intervals. For the double exponential, they find that intervals derived from the conditional analysis tend to have shorter mean length and higher power than those from the unconditional analysis. This support for the conditional analysis appears tempered, however, by the Welch (1939) example in which intervals obtained from the conditional analysis were in fact longer for the case with g(z) given by the uniform(-1/2, 1/2) distribution.

The purpose of this note is to show that short mean length and high power are not suitable criteria for assessing the conditional versus the unconditional approaches to inference. This is not to deny, however, certain merit to the two criteria when other things are equal.

Also we note that, though ancillarity, shortest mean length, and highest power do not provide adequate support for the conditional analysis, such support does exist; section 4 discusses briefly this alternative support.

In addition we emphasize particularly that inferences coming from the conditional analysis do not depend on choosing $\hat{\mu}$ and $\hat{\sigma}$ to be the maximum likelihood estimates. In fact any location-scale statistics $\hat{\mu}$, $\hat{\sigma}$ will lead to the same conditional inferences; note that $\hat{\mu}$, $\hat{\sigma}$ are location, scale statistics if $\hat{\mu}(a\underset{\sim}{1} + c\underset{\sim}{y}) = a + c\hat{\mu}(\underset{\sim}{y})$, $\hat{\sigma}(a\underset{\sim}{1} + c\underset{\sim}{y}) = c\hat{\sigma}(\underset{\sim}{y})$ for all a in R and c in R^+. Thus the choice of estimator vanishes in the conditional context. Indeed if there are certain ultimate estimators as are sought by the robustness analyst then such are automatically being used in the conditional context. For the remainder of this article $\hat{\mu}$ and $\hat{\sigma}$ will refer to arbitrary location, scale statistics.

2. The mean length criterion

We consider a sample of size n from the distribution (1.1). A β_* level confidence interval for μ is of the form

$$(\hat{\mu} - t''_{\underset{\sim}{d}}\,\hat{\sigma},\ \hat{\mu} - t'_{\underset{\sim}{d}}\,\hat{\sigma}) \qquad (2.1)$$

where $(t'_{\underset{\sim}{d}}, t''_{\underset{\sim}{d}})$ is an interval containing β_* probability for the conditional distribution of t given the configuration $\underset{\sim}{d}$. Note the special form of the interval as required by the typical nonsymmetry of the conditional distribution. We see that (2.1) is an overall β level confidence interval for μ where $\beta = E(\beta_*)$ is calculated using the marginal distribution for $\underset{\sim}{d}$.

Now let $\underset{\sim}{d}_1$ and $\underset{\sim}{d}_2$ be two possible values of $\underset{\sim}{d}$ with respective probabilities p_1 and p_2. We present our arguments in a discrete context but note that they carry over to the continuous case by the use of limiting arguments involving neighbourhoods. Also, to simplify formulae we assume the scale statistic $\hat{\sigma}$ has been specially selected so that $E(\hat{\sigma}|\underset{\sim}{d}) = k\sigma$, is constant with respect to $\underset{\sim}{d}$.

Consider a β-lèvel confidence interval with conditional confidence β_1 at $\underset{\sim}{d}_1$ and β_2 at $\underset{\sim}{d}_2$. We examine the effect on the mean length of increasing β_1 to $\beta_1 + \varepsilon_1$ by changing outward one of the boundaries $t'_{\underset{\sim}{d}_1}$, $t''_{\underset{\sim}{d}_1}$, say t_1 by an amount τ_1 and decreasing β_2 to $\beta_2 - \varepsilon_2$ by changing inward one of the boundaries $t'_{\underset{\sim}{d}_2}$, $t''_{\underset{\sim}{d}_2}$, say t_2, by an amount τ_2. We then have the local relationship $\tau_i h_i = \varepsilon_i$ where h_i is the conditional density for t given $\underset{\sim}{d}_i$ evaluated at t_i.

Under the preceding changes the overall confidence level β will remain fixed if

$$p_1\tau_1 h_1 = p_2\tau_2 h_2.$$

The mean length, however, will typically change. This change is given by

$$(p_1\tau_1 - p_2\tau_2)\ k\sigma$$

which by (2.2) reduces to

$$k\sigma p_1\tau_1(1 - h_1/h_2)$$

and is negative if $h_1 > h_2$.

Thus to minimize mean length we should increase conditional confidence on intervals with high boundary t density and decrease conditional confidence on intervals with low boundary t density. Or more intuitively we would place higher conditional confidence on intervals formed from narrow precise conditional distributions for t, and lower conditional confidence on intervals formed from wide imprecise conditional distributions. Moreover the conditional β interval with $\beta_\star = \beta$ will only be shortest if the conditional distribution is essentially the same (at the critical value) for all $\underset{\sim}{d}$, in which case the conditional and unconditional analyses are in fact identical.

Thus minimizing mean length runs counter to the very essense of conditional analysis: you can always do better in the mean-length sense by using the unconditional analysis - unless the conditional and unconditional analyses are the same, as with the normal. The operating principle from the minimum length viewpoint is get as much confidence coverage as possible where it is cheap - where the conditional distribution is precise.

The implications of this are overwhelmingly apparent in the simple location example considered by Welch (1939). Let y_1, y_2 be uniform ($\theta \pm 1/2$). The minimum mean length β-level confidence interval depends on the range $r = 1 - y_{(2)} - y_{(1)}$ of possible θ values; it has the form

$$\begin{array}{lll} \bar{y} \pm r/2 & \text{if} & r < \surd\beta \\ \phi & & r \geq \surd\beta, \end{array}$$

in other words the full range of possible θ values when that range is small and the empty set (!) otherwise.

3. The power criterion

Somewhat similar results are obtained with the power criterion. For illustration we take the somewhat simpler location model with known scale, say $\sigma = 1$. For this model the configuration is given by $\underset{\sim}{d} = \underset{\sim}{y} - \hat{\mu}\underset{\sim}{1}$ where $\hat{\mu}$ is a location statistic, $\hat{\mu}(a\underset{\sim}{1} + \underset{\sim}{y}) = \tilde{a} + \tilde{\hat{\mu}}(\underset{\sim}{y})$ and the appropriate t statistic is given by $t = \hat{\mu} - \mu$.

A $\beta_\star$-level conditional confidence interval has the form

$$(\hat{\mu} - t''_{\underset{\sim}{d}},\ \hat{\mu} - t'_{\underset{\sim}{d}}) \tag{3.1}$$

where $(t'_{\underset{\sim}{d}}, t''_{\underset{\sim}{d}})$ is a $\beta_\star$ interval for the conditional distribution of t. We have that (3.1) is an overall β confidence interval where $\beta = E(\beta_\star)$ with respect to the distribution of $\underset{\sim}{d}$. Also we see easily that a shift in $\hat{\mu}$ results in a compensating shift in $(t'_{\underset{\sim}{d}}, t''_{\underset{\sim}{d}})$ so that the interval is independent of the choice of location statistic.

The power of a confidence interval is defined as the probability of missing a false parameter value say $\mu + \Delta$, $\Delta \neq 0$. The power of (2.3) is given by

$$1 - P(\hat{\mu} - t'_{\tilde{d}} < \Delta + \mu < \hat{\mu} - t'_{\tilde{d}})$$

$$= 1 - P(t'_{\tilde{d}} + \Delta < t < t''_{\tilde{d}} + \Delta).$$

Without loss of generality we examine the case $\Delta > 0$.

As in Section 2 we let $\tilde{d}_1$ and $\tilde{d}_2$ be two possible values of $\tilde{d}$ with probabilities p_1 and p_2. We consider a β-level interval with conditional confidence β_1 at $\tilde{d}_1$ and β_2 at $\tilde{d}_2$. The upper boundary of the interval is critical for the $\Delta > 0$ case and thus we examine the effect of increasing β_1 to $\beta_1 + \varepsilon_1$ by changing outward the first upper boundary by an amount τ_1 and decreasing β_2 to $\beta_2 - \varepsilon_2$ by changing inward the second upper boundary by an amount τ_2; then $\tau_i h_i = \tau_i$ where h_i is the conditional density for $\tilde{d}$ given $\tilde{d}_i$ evaluated at the lower boundary point t_i.

The overall confidence remains unchanged if $p_1\tau_1 h_1 = p_2\tau_2 h_2$. The change in power is given by

$$- p_1\tau_1 h_1^{\Delta} + p_2\tau_2 h_2^{\Delta}$$

where h_i^{Δ} is the conditional (on $\tilde{d}$) density evaluated at $t_i + \Delta$ or is the $-\Delta$ displaced conditional density evaluated at t_i. This reduces to

$$\tau_1 p_1 h_1 \left(\frac{h_2^{\Delta}}{h_2} - \frac{h_1^{\Delta}}{h_1}\right)$$

which is positive if the likelihood ratio h_1^{Δ}/h_1 is small with respect to h_2^{Δ}/h_2. Indeed this is what we would expect from hypothesis testing theory: put points into the acceptance region for small values of the likelihood ratio.

How does this effect the conditional confidence levels? For a reference case consider a central β-level conditional confidence interval and in particular the $(1-\beta)/2$ boundary point in the left tail of the conditional distribution. If the conditional distribution is very concentrated then the likelihood ratio will be very small, and if the conditional distribution becomes diffuse then the likelihood ratio becomes larger. The implication is that the interval should be increased in the concentrated case and decreased in the diffuse case.

Thus we see that the criterion of maximum power leads to high conditional confidence for the precise distributions and low conditional confidence for the diffuse conditional distributions. The second criterion also runs counter to the conditional analysis.

In conclusion we note that both minimum length and maximum power lead to high confidence levels for the precise distributions and low levels for the imprecise distributions. From a conditional point of view this situation is untenable and amounts to a statistician playing off one client against another in order to produce, on the average, short powerful intervals. In brief get your coverage where its cheap.

4. Formulation of the model

Support for many conditional analyses is found in a more careful formulation of the model used. Consider, for instance, a distribution $g(\underset{\sim}{z})$ with a corresponding variable $\tilde{z}$. Then it should be more or less obvious that the model (I') $\tilde{y} \overset{d}{=} \mu\underset{\sim}{1}+\sigma\underset{\sim}{z}$, where $\overset{d}{=}$ denotes equal-in-distribution, is in some way different from the model $(\Pi')\underset{\sim}{y} = \mu\underset{\sim}{1} + \sigma\underset{\sim}{z}$.

The first model (I') simply provides a description of the possible distributions of the response $\underset{\sim}{y}$. It is identical with the model that asserts that the response $\underset{\sim}{y}$ has possible distributions $\{\sigma^{-n}g(\sigma^{-1}(\underset{\sim}{y} - \mu\underset{\sim}{1})) : -\infty < \mu < \infty, 0 < \sigma < \infty\}$. An observed $\underset{\sim}{y}$ is then assessed in accordance with these possibilities. On the other hand, model (Π') asserts that an observed $\underset{\sim}{y}$ is in fact some (unknown) non-random transformation or presentation of an (unknown) realized $\underset{\sim}{z}_0$. Knowing $\underset{\sim}{z}_0$ as well as $\underset{\sim}{y}_0$, is tantamount to knowing the transformation (and thus μ and σ). This, of course, is not the case with model (I'). A statistical model plus data has been termed an <u>inference base</u>. For model (Π'), the variation-based model, an observed $\underset{\sim}{y}$ leads to a reduction in the possibilities for $\underset{\sim}{z}_0$ (all but two dimensions of $\underset{\sim}{z}_0$ are determined). The possible values for this realized $\underset{\sim}{z}_0$ thus lead to the alternatives with which to assess the observed $\underset{\sim}{y}$. This leads (Fraser, 1976, 1979) <u>necessarily</u> to the conditional analysis: the initial inference base has been replaced by a simpler inference base.

Model I', the classical model, is a response-based model. The conditional inference (on $\underset{\sim}{d}$) involves a reduction of the initial inference base (classical model plus data); it replaces an assessment of the observed $\underset{\sim}{y}$ against all possible $\underset{\sim}{y}$, with a comparison of the observed $\underset{\sim}{y}$ against those $\underset{\sim}{y}$'s having a specified $\underset{\sim}{d}$ (in light of course of the possible distributions). Such a reduction of the response based model can only be accomplished in conjunction with some principle. As we have noted ancillarity, "shortest length" and "largest power" are not appropriate.

Thus the question of whether to use a conditional or an unconditional analysis becomes rephrased in terms of whether to use model (I') or (II'). Both entities can be viewed as potential models for a location-scale problem. Which is more "appropriate"? An answer to this question requires some minimal basic <u>axioms</u> pertaining to the ingredients of a statistical model. We suggest that one minimal requirement is for the model to be <u>descriptive</u> in the sense that components of the model correspond to real objective components of the system being modelled and vice versa. It has been shown (Fraser, 1979 and Brenner and Fraser, 1979) that distribution form (and thus $\underset{\sim}{z}$) of the location scale system is an objective component of the system. Accordingly it should be included in the statistical model. Thus for location scale problems model (II') with the consequent conditional analysis should be employed.

Acknowledgements

This research was supported in part by National Science and Engineering Research Council grants A3011, A3161.

References

Brenner, D. and Fraser, D.A.S. (1979): The identification of distribution form. Department of Statistics, University of Toronto.

Cox, D.R. and Hinkley, D. (1974): Theoretical Statistics, London: Chapman and Hall.

Fick, G. (1976): LSAI: Location-scale analysis implemented, Department of Mathematics and Statistics, University of Calgary.

Fisher, R.A. (1934): Two new properties of maximum likelihood. Proc. R. Soc. A144, 285-307.

Fraser, D.A.S. (1973): The elusive ancillary, Multivariate Statistical Inference, Ed D.G. Kabe and R.P. Gupta, Amsterdam: North Holland Publishing Co.

Fraser, D.A.S. (1976): Necessary analysis and adaptive inference. Jour. Amer. Statist. Assoc. 71, 99-113.

Fraser, D.A.S. (1979): Inference and Linear Models. London: McGraw-Hill.

Grice, J.V., Bain, L.J., and Engelhardt, M. (1978): Comparison of conditional and unconditional confidence intervals for the double exponential distribution. Commun. Statist. B7, 515-524.

Hinkley, D.V. (1978): Likelihood inference about location and scale parameters. Biometrika, 65, 253-261.

Hinkley, D.V. and Efron, B. (1978): Assessing the accuracy of the maximum likelihood estimator: Observed versus expected information. Biometrika, 65, 457-487.

Welch, B.L. (1939): On confidence limits and sufficiency, with particular reference to parameters of location. Ann. Math. Statist. 10, 58-69.

Gunter Lorenzen

Spezifikationsfehler beim Messen der Abhängigkeit in Kontingenztabellen

1. Einleitung

Der Begriff des Spezifikationsfehlers ist in der Ökonometrie vertraut. Im weitesten Sinne werden unter diesem Begriff die Konsequenzen eines fehlerhaft oder unvollständig spezifizierten Regressionsmodells verstanden. Beispielsweise kann gezeigt werden, daß die OLS-Schätzfunktionen der Regressionskoeffizienten nicht erwartungstreu sind, wenn eine 'relevante' Variable in der Regressionsbeziehung nicht berücksichtigt wurde (z.B. Schneeweiß (1971), S. 150). Da es bei der Regressionsanalyse um das Messen von Abhängigkeiten zwischen Variablen geht, heißt dies, daß bestehende Abhängigkeiten zwischen Variablen systematisch fehlerhaft ermittelt werden, wenn nicht alle 'relevanten' Variablen im Regressionsmodell berücksichtigt werden.

Die Regressionsanalyse wird in der Regel nur für quantitative Variable benutzt, allenfalls werden unter den unabhängigen Variablen (Regressoren) des Regressionsmodells qualitative Variable in Form von (0,1)-Variablen zugelassen. Das übliche Regressionsmodell ist dagegen ungeeignet, wenn auch die abhängige Variable eine (0,1)-Variable ist, da dann die üblichen Annahmen über die stochastische Residualvariable nicht länger erfüllt sein können. Um Aussagen über Zusammenhänge zwischen (0,1)-Variablen zu machen, stehen andere Methoden zur Verfügung, beispielsweise der χ^2-Unabhängigkeitstest.

Im folgenden wollen wir uns mit speziellen Spezifikationsfehlern beim χ^2-Unabhängigkeitstest befassen, d.h. mit dem Problem, daß der χ^2-Unabhängigkeitstest zu inkorrekten Schlußfolgerungen hinsichtlich des Zusammenhangs zwischen zwei Variablen führt, wenn 'relevante' Variable aus der Untersuchung ausgeklammert werden.

Das Problem ist seit langem bekannt, aber in der Regel nur qualitativ abgehandelt worden, siehe z.B. Simpson (1951), Birch (1963), Bishop (1971). Ansätze zu einer Quantifizierung des Spezifikationsfehlers finden sich bei Kendall and Stuart, Vol. II (1961) bzw. Kendall (1975). Allerdings wird der Spezifikationsfehler dort mit Hilfe eines unüblichen Kontingenzmaßes quantifiziert, daß keinen unmittelbaren Schluß auf das vor allem in der schließenden Statistik wichtige χ^2-Maß zuläßt. Demgegenüber wird in vorliegender Arbeit der Spezifikationsfehler im χ^2-Maß ausgedrückt.

Im nächsten Abschnitt werden wir das im folgenden benutzte χ^2-Maß kurz vorstellen. Im dritten Abschnitt soll das Spezifikationsproblem anhand eines Beispiels erörtert werden. Im