Biometrika (1979), **66**, 3, *pp.* 655–8
Printed in Great Britain

On some methods of analysis for weather experiments

BY ANDREY FEUERVERGER
Department of Statistics, University of Toronto

SUMMARY

Using generalized densities, maximum likelihood and ancillarity, methods are developed for the analysis of randomized weather modification experiments that incorporate predictor variables. The linear logistic form is used to model the distributional singularity at the origin, while a gamma model with predictors is developed for the component conditional on the presence of precipitation. Connexions between the component submodels are explored and illustrated using an Australian data set.

Some key words: Ancillarity; Gamma distribution model with predictors; Generalized density; Linear logistic model; Maximum likelihood; Single cloud weather modification experiment.

1. INTRODUCTION

This paper summarizes some efforts to develop methods for the analysis of data collected in certain randomized experiments in climate modification. Methods related to likelihood are used to examine the problem of zero precipitation cases and the use of predictor variables in conjunction with the gamma distribution. The procedures proposed were implemented for testing on data given by Bethwaite *et al.* (1966).

2. SINGULARITY AND THE LIKELIHOOD

Owing to the singularity at the origin we use generalized densities (Zacks, 1971, p. 41) for rainfall. These take values 0, $1-p$, $pf(y)$ according as $y<0$, $y=0$, $y>0$ with p the rainfall probability and $f(y)$ the density conditional on occurrence. It should be noted (Zacks, 1971, Chapter 5) that under regularity the usual asymptotics for maximum likelihood still hold.

Now suppose the experiment to consist of n independent cases, each giving a precipitation measure Y_i, and l fixed predictors $X_i = (X_{1i}, \ldots, X_{li})$ $(i = 1, \ldots, n)$. Here X_{1i} may be a seeding indicator with the remaining predictors unaffected by seeding; the cases may be clouds that satisfy preconditions; and the measuring may require instrumented aircraft. The AgI seeding is conducted in blind randomized trials. The joint density for $Y_1, \ldots, Y_n$ will be

$$\Pi_D\{1-p(X_i,\mu)\}\,\Pi_R\, p(X_i,\mu)\, f(y_i, X_i, \nu),$$

where the products are over the dry and rain cases. In this expression the probability, p, and the conditional density, f, are assumed related to X_i by means of models with parameterizations μ and ν respectively. This then leads to a likelihood function $L(\mu,\nu)$.

It follows that to specify a distributional family we need to specify $p(X,\mu)$ and $f(y,X,\nu)$. We also see that questions of real interest concerning the state of nature are raised by enquiring about relations among the parameter sets μ and ν. Initially, specific information will be lacking, so that we regard μ and ν to be unrelated. The likelihood $L(\mu,\nu) = L_1(\mu)\,L_2(\nu)$ will then factorize into a probability component which concerns only the aspect of occurrence and a conditional component which concerns only the situation given rainfall. By virtue of the ancillarity which ensues (Cox & Hinkley, 1974, p. 31–5) the factorization means that estimation of the submodel parameters proceeds as independent problems and that the estimates resulting for μ and ν may be regarded as independent. On this basis one may then enquire into relations, e.g. proportionalities, among the parameters of μ and ν. If none is tenable, inference proceeds on a basis of independent components. An assumed relationship, on the other hand, increases power but couples the estimation procedure. Here one could consider to maximize the whole likelihood $L(\mu,\nu)$, but in § 5 we propose a simpler approach which preserves the asymptotic efficiency.

3. The probability component

To model rainfall probability we use (Cox, 1970) the linear logistic model

$$\lambda_i = \log\{p_i/(1-p_i)\} = \mu_0+\mu_1 X_{1i}+\ldots+\mu_l X_{li}.$$

We note here that stability of Newton–Raphson maximum likelihood procedures is assured in view of the known unimodality of the likelihood surface.

4. The conditional component

The conditional model obtains upon dropping zeros. Neyman, Scott & Vasilevskia (1960) and Neyman & Scott (1960) considered Gaussian regression in conjunction with transformed precipitation measures and correction for bias due to transformation. For a conditional gamma model with predictors, the distribution form $\{\Gamma(\alpha)\}^{-1}(\pi/\alpha)^{-\alpha}y^{\alpha-1}\exp(-\alpha y/\pi)$ $(0\leqslant y<\infty)$ is convenient. Modelling shape as constant, and scale as

$$\log\pi_i = -A_i = \theta_0+\theta_1 X_{1i}+\ldots+\theta_l X_{li} \quad (i\in R)$$

we obtain the log likelihood

$$-n\log\Gamma(\alpha)+\alpha\Sigma A_i+n\alpha\log\alpha+(\alpha-1)\Sigma\log Y_i-\alpha\Sigma Y_i e^{A_i},$$

where now, as throughout this section, n is the number of rainfall cases, and sums are over R.

The structure of this surface is less complex than first appears, for upon centring predictors at their R-means so that $\Sigma A_i = -n\theta_0$, writing A_i' for $A_i+\theta_0$, and rearranging we obtain

$$[-n\log\Gamma(\alpha)+n\alpha\log\alpha+\alpha\{\Sigma\log Y_i+n\theta_0-e^{\theta_0}(\Sigma Y_i e^{A_i'})\}]$$

and this may be maximized in stages. First, the surface in innermost parentheses has (j,k)th second partial derivative $\Sigma Y_i X_{ji} X_{ki} e^{A_i'}$ $(j,k = 1,\ldots,l)$ and unimodality follows from positive-definiteness. Hence we may find the minimum M by a Newton–Raphson procedure and then set $\theta_0 = \log(n/M)$ as required by the intermediate expression. Finally, maximization of the outermost parentheses follows the familiar gamma procedure for scale.

It is of interest that the simplified likelihood equations with R-centred predictors are

$$\log\alpha-\psi(\alpha) = \theta_0-n^{-1}\Sigma\log Y_i, \quad \Sigma Y_i e^{A_i} = n, \quad \Sigma Y_i X_{ji} e^{A_i'} = 0 \quad (j = 1,\ldots,l),$$

where ψ is the digamma function. Hence the $\hat\theta$'s are chosen so $Y_i/\hat\pi_i$ averages to unity and is orthogonal to X_{ji}, for all j. The equations have a unique solution, and the estimate $\hat\alpha$ is independent of the $\hat\theta$'s, by Basu's theorem. The asymptotic covariance for $\hat\alpha,\hat\theta_0,\ldots,\hat\theta_l$ is

$$\begin{bmatrix} n^{-1}\{\psi'(\alpha)-\alpha^{-1}\}^{-1} & 0 \\ 0 & \alpha^{-1}E^{-1} \end{bmatrix},$$

where ψ' is the trigamma function, and E is $(l+1)\times(l+1)$ with entries $\Sigma Y_i X_{ji} X_{ki} e^{A_i}$ $(j,k = 0,1,\ldots,l)$. Note α^{-1} plays the role of a variance, and the effectiveness of the predictors, which is related to the quantity $\log(n^{-1}\Sigma\pi_i) - n^{-1}\Sigma\log\pi_i$, depends on the increase in α attributable to the predictors.

5. On synthesis and evaluation

The natural approach to problems of inference concerning relations across the submodel parameters involves cumbersome maximization of the coupled likelihood $L(\mu,\nu)$. To avoid this we exploit first the ancillarity, thus regarding the submodel estimates $\hat{\mu}, \hat{\nu}$ as independent, and secondly the asymptotic form of the log likelihood:

$$-\tfrac{1}{2}(\hat{\mu}-\mu)'\Omega_\mu^{-1}(\hat{\mu}-\mu) - \tfrac{1}{2}(\hat{\nu}-\nu)'\Omega_\nu^{-1}(\hat{\nu}-\nu).$$

Here the covariances are estimated but considered fixed, and the unrestricted maximum is zero. Now consider a relation such as $\mu = \kappa\nu$ among some corresponding submodel parameters. A likelihood ratio approach based on the asymptotic form gives

$$\nu = (\kappa^2\Omega_\mu^{-1} + \Omega_\nu^{-1})^{-1}(\kappa\Omega_\mu\hat{\mu} + \Omega_\nu^{-1}\hat{\nu}), \quad \kappa = (\nu'\Omega_\mu^{-1}\hat{\mu})/(\nu'\Omega_\mu^{-1}\nu),$$

where uncapped symbols are now the estimates sought. These are solved by recursion. As one would expect that factors governing rainfall occurrence are nearly those which govern quantity given rain, one may reflect that evidence contrary to proportionality would be indeed remarkable. We note the covariance estimate for κ, ν under proportionality:

$$\begin{bmatrix} \nu'\Omega_\mu^{-1}\nu & (2\kappa\nu-\hat{\mu})'\Omega_\mu^{-1} \\ \Omega_\mu^{-1}(2\kappa\nu-\hat{\mu}) & \kappa^2\Omega_\mu^{-1}+\Omega_\nu^{-1} \end{bmatrix}^{-1}.$$

The model and inference possibilities that now arise are numerous: for the submodels may be treated as related or not, and if related, the proportionality may include the seeding parameter or not. Further, there is the question of whether seeding alters the meteorological dynamics thus invalidating the use of 'parallel' models for seeded and unseeded cases. Also there is the matter of producing unbiased estimates, asymptotically at least, of percentage seeding effect in units of untransformed precipitation. We note that gamma expectations behave simply under power transform: if Y has parameters α, π then

$$E(Y^c) = (\pi/\alpha)^c\Gamma(\alpha+c)/\Gamma(\alpha).$$

Finally, concerning robustness, we note that the Gaussian and gamma models of § 4 differ also in structural form, one being additive in seeding, the other multiplicative, and that the range of these models is further extended by power transformation (Box & Cox, 1964): the implication of robustness is enhanced if conclusions stay relatively invariant over different models and transformation intervals judged plausible, say, by means of quantile plots. For some further details see an unpublished University of Toronto report by the author. For other discussions relevant to this work see Neyman & Scott (1967), Dawkins et al. (1977) and Brillinger, Jones & Tukey (1978).

6. Numerical study with Australian data

Data from a single cloud experiment given by Bethwaite *et al.* (1966) were used to guide this work. Difficulties noted by the experiment authors were ignored and we used all 69 cases reported, 33 rain and 36 dry, and three predictors: X_1, cloud top temperature in negative degrees Celsius; X_2, cloud height to sea level in km; X_3, cloud depth in km. The seeding variable was omitted. Fortran programs were developed on the University of Toronto IBM 360/165 to carry out the linear logistic procedure, to fit a gamma model with predictors, and to produce quantile and other plots of $Y_i/\hat{\pi}_i$. Convergence in general was extremely rapid.

The linear logistic relation fitted was

$$\lambda = -0{\cdot}14 + 0{\cdot}16X_1 - 0{\cdot}69X_2 + 1{\cdot}74X_3,$$

the t values for the three regression coefficients being respectively 1·83, −1·60 and 2·85. The signs are as expected with the last coefficient highly significant, and the other two not quite. An interesting kind of ancillary nonparametric verification of fit is possible: the actual precipitation values for rain cases may be compared to the probabilities determined by the relation above. Using the randomness test of Lehmann (1975, § 6·2) we obtained, in the natural direction, an observed level of significance of 0·043. This we are content to leave one sided, but caution that to do so means to impute a certain lack of perversity on the part of nature!

Quantile plots revealed very good fit of the conditional gamma model combined with power transforms of precipitation with exponent between 0·55 and 0·65. We quote the relation for 0·60 whose fit was remarkably good:

$$\log \pi_i = 2{\cdot}77 + 0{\cdot}064X_1 - 0{\cdot}54X_2 + 0{\cdot}78X_3.$$

The t values of 1·37, −2·32 and 2·55 were very stable in the exponent. Further details and a Gaussian analysis may be found in the unpublished report mentioned above. The computer programs described are available from the author.

This research was supported by the National Research Council, Canada and the Atmospheric Environment Service, Canada. It is a pleasure to acknowledge very helpful conversations with Dr A. J. Chisholm.

References

Bethwaite, F. D., Smith, E. J., Warburton, J. A. & Hefferman, K. J. (1966). Effects of seeding isolated cumulus clouds with silver iodide. *J. Appl. Meteorol.* **5**, 513–20.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations (with discussion). *J. R. Statist. Soc.* B **26**, 211–43.

Brillinger, D. R., Jones, L. V. & Tukey, J. W. (1978). The role of statistics in weather resources management. *The Management of Weather Resources*, Vol. 2, Report of the Statistical Task Force to the Weather Modification Advisory Board. Washington, D.C.: Department of Commerce.

Cox, D. R. (1970). *The Analysis of Binary Data*. London: Methuen.

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

Dawkins, S. M., Neyman, J., Scott, E. L. & Wells, M. A. (1977). Preliminary statistical analysis of precipitation amounts. Chapters 13, 14. In *Pyramid Project*, The Final Report on the Pyramid Lake Pilot Project 1970 to 1975, Vol. 1. Ed. P. Squires. Reno, Nevada: Desert Research Institute.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden Day.

Neyman, J. & Scott, E. L. (1960). Correction for bias introduced by a transformation of variables. *Ann. Math. Statist.* **31**, 643–55.

Neyman, J. & Scott, E. L. (1967). Note on techniques of evaluation of single rain stimulation experiments. *Proc. 5th Berkeley Symp.* **5**, 371–84.

Neyman, J., Scott, E. L. & Vasilevskia, M. (1960). Statistical evaluation of the Santa Barbara randomized cloud-seeding experiment. *Bull. Am. Meterol. Soc.* **41**, 531–47.

Zacks, S. (1971). *The Theory of Statistical Inference*. New York: Wiley.

[*Received February* 1978. *Revised March* 1979]