Covariance Analysis Technique Based on Bivariate Log-Normal Distribution with Weather Modification Applications

PAUL W. MIELKE, JR., AND JAMES S. WILLIAMS

Department of Statistics, Colorado State University, Fort Collins 80523

SING-CHOU WU

Department of Computer Science and Statistics, California Polytechnic State University, San Luis Obispo 93407

(Manuscript received 29 July 1976, in revised form 3 January 1977)

ABSTRACT

A statistical technique based on the bivariate log-normal distribution is presented for analyses of treatment (cloud seeding) effects on measurements (precipitation amounts) when appropriate covariates (correlated control area measurements) are available. An example is given which evaluates effects of seeding on specific 500 mb temperature partitions of 24 h precipitation amount data from the 1964–70 Wolf Creek Pass wintertime orographic cloud seeding experiment. In addition, an appendix includes an analogous analytic technique based on the bivariate log-normal distribution for cross-over designs.

1. Introduction

The purpose of this article is to present a useful covariance analysis technique based on the bivariate log-normal distribution. A specific application of this technique involves analyses of possible effects by cloud seeding on target area precipitation data when appropriate (viz., non-contaminated, well correlated, etc.) control area precipitation data are available.

The primary considerations affecting the choice of a bivariate distribution involving asymmetric marginals for analyses of paired measurements such as daily precipitation amounts at two locations are noted. The first consideration is the ability of the distribution to describe the measurements in question. In particular, the log-normal distribution is a reasonable candidate for fitting both precipitation and related measurements in comparison to alternative candidates such as the gamma distribution and closed form beta distributions of the second kind (Mielke and Johnson, 1974). Other arguments in support of the log-normal distribution for describing precipitation have also been proposed (Biondini, 1976). The second consideration involves the tractability of applying bivariate distributions for this purpose. In this respect, the bivariate log-normal distribution is relatively simple to employ. In contrast, a bivariate gamma distribution with similar properties requires cumbersome computations to implement such a covariance analysis technique (Moran, 1969). Also, other contenders such as the bivariate beta distribution of the second kind are similarly plagued with severe implementation problems.

Section 2 includes a description and properties of the bivariate log-normal distribution. The analysis technique based on this distribution is presented in Section 3. A detailed numerical example is given in Section 4 which involves an evaluation of cloud seeding effects on specific 500 mb temperature partitions of 24 h precipitation amount data from the 1964–70 Wolf Creek Pass wintertime orographic cloud seeding experiment sponsored by the State of Colorado (Grant and Elliott, 1974). Finally, an analytic technique based on the bivariate log-normal distribution for a cross-over design is contained in the Appendix. Although the cross-over design is often limited in application because of contamination problems, it has on occasion been effectively utilized in weather modification studies including the outstanding Israeli artificial rainfall stimulation experiment (Gabriel, 1967; Mielke, 1974).

2. Bivariate log-normal distribution

Let X and Y denote bivariate log-normal random variables with marginal distribution shape parameters α and β, and scale parameters A and B, respectively. Then the bivariate log-normal density function of X and Y is

\[
fx,y(x,y) = \left[2\pi xy\alpha(1-\rho^2)^{\frac{3}{2}}\right]^{-1} \times \exp\left( -\frac{1}{2(1-\rho^2)} \left[ \frac{\ln(x/A)}{\alpha} \right]^2 \right.
\]

\[
-2\rho \left[ \frac{\ln(x/A)}{\alpha} \right] \left[ \frac{\ln(y/B)}{\beta} \right] + \left[ \frac{\ln(y/B)}{\beta} \right]^2 \right) \],
\]
where $0<\min(x,y,\alpha,\beta,A,B)$ and $\rho$, the dependence parameter, satisfies $|\rho|<1$. If $\rho=0$, then $X$ and $Y$ are independent random variables.

The moments of $X$ and $Y$ about zero are given by

$$E(X^t Y^r) = A_t B_r \exp[(\sigma_2^2 + \rho \beta^2 + 2 \rho \sigma_2 \beta)/2].$$

In particular, the mean and variance of $X$ are

$$E(X) = A e^{\sigma_2^2} \text{ and } \text{Var}(X) = A^2 e^{\sigma_2^2} (e^{\rho^2} - 1),$$

the mean and variance of $Y$ are

$$E(Y) = B e^{\beta^2} \text{ and } \text{Var}(Y) = B^2 e^{\beta^2} (e^{\rho^2} - 1),$$

and the covariance and correlation of $X$ and $Y$ are

$$\text{Cov}(X,Y) = AB (e^{\rho^2 - 1}) e^{(\sigma_2^2 \beta^2)/2},$$

$$\text{Cor}(X,Y) = (e^{\rho^2 - 1}) / \sqrt{(e^{\sigma_2^2} (e^{\rho^2} - 1) (e^{\beta^2} - 1))}.$$

3. Covariance analysis technique

Let $(x_i,y_i)$, $i=1, \ldots, m+n$, denote $m+n$ independent pairs of precipitation amount measurements from target and control areas, respectively, in either a weather modification experiment or project. In this instance $x$ and $y$ denote the variates designated for treatment and covariate, respectively. For convenience, we assume the first $m$ and last $n$ pairs of measurements designate the non-seeded and seeded events in the target area, respectively. (Actually the arrangement of pairs is inconsequential for this model with time-independent responses. Any pattern, predetermined or resulting from randomization, is analyzed in the same manner.) This design is a continued-covariate design (also known as a target-control design). If seeding has an effect, this effect is assumed to appear only in scale parameter differences of the log-normal responses associated with non-seeded and seeded target area measurements. Thus we let $A_N$ and $A_S$ denote scale parameters associated with non-seeded target area measurements, respectively, in the bivariate log-normal distribution. It is assumed that $\alpha$, $\beta$, $B$ and $\rho$ are unaffected by seeding. The null hypothesis that seeding does not change target area measurements in any way is therefore quantified by $A_N=A_S$.

The covariance analysis technique presented here is directly related to an optimal small-sample analysis technique (which is also asymptotically optimal) for the continued-covariate design described by Wu et al. (1972). The continued-covariate design is commonly a necessity in weather modification experiments involving target and control areas because of anticipated downwind seeding contamination in the target area if the control area is seeded (and not vice versa). The example in Section 4 is typical of this situation. Thus, while the optimal design for this example described by Wu et al. (1972) is not feasible due to contamination (i.e., the prescribed optimal design would have been a balanced cross-over design had contamination not existed), the analysis technique associated with the continued-covariate design used is appropriate.

The following covariance analysis technique based on the bivariate log-normal distribution for testing the null hypothesis ($H_0$: $A_N=A_S$) versus the alternative hypothesis ($H_1$: $A_N\neq A_S$) is the same analytic technique, following a transformation to a bivariate normal distribution, which was described by Wu et al. (1972) for continued-covariate designs. They showed that these simple design and analysis variations of the classical covariance technique have optimal properties for testing the hypotheses in question.

The description of this covariance analysis technique is simplified by using the following set of sufficient statistics:

$$u_{1N} = m^{-1} \sum_{i=1}^{m} \ln x_i,$$

$$u_{1S} = n^{-1} \sum_{i=m+1}^{m+n} \ln x_i,$$

$$u_{2N} = m^{-1} \sum_{i=1}^{m} (\ln x_i)^2,$$

$$u_{2S} = n^{-1} \sum_{i=m+1}^{m+n} (\ln x_i)^2,$$

$$v_{1N} = m^{-1} \sum_{i=1}^{m} \ln y_i,$$

$$v_{1S} = n^{-1} \sum_{i=m+1}^{m+n} \ln y_i,$$

$$v_{2N} = m^{-1} \sum_{i=1}^{m} (\ln y_i)^2,$$

$$v_{2S} = n^{-1} \sum_{i=m+1}^{m+n} (\ln y_i)^2,$$

$$(uv)_{N} = m^{-1} \sum_{i=1}^{m} (\ln x_i)(\ln y_i),$$

$$(uv)_{S} = n^{-1} \sum_{i=m+1}^{m+n} (\ln x_i)(\ln y_i).$$

Under $H_0$: $A_N=A_S$, the test statistic given by

$$T = \left[ \frac{m+n}{mn} (1+H) s^2 \right]^{-1}$$

is distributed as Student's $t$ with $m+n-3$ degrees of freedom where

$$r = u_{1S} - u_{1N} - (v_{1S} - v_{1N}),$$

$$G = \frac{m [(uv)_{N} - u_{1N} v_{1N}] + n [(uv)_{S} - u_{1S} v_{1S}]}{m (v_{2N} - v_{1N}^2) + n (v_{2S} - v_{1S}^2)},$$

$$H = \frac{mn}{(v_{1S} - v_{1N})^2}$$

$$s^2 = \frac{m(u_{2N} - u_{1N}^2) + n(u_{2S} - u_{1S}^2) - G^2 [m (v_{2N} - v_{1N}^2) + n (v_{2S} - v_{1S}^2)]}{m+n-3}.$$
The point estimate of $A_S/A_N$ is given by
\[ \hat{A}_S/\hat{A}_N = e^r \]
and the $1-\gamma$ confidence limits for $A_S/A_N$ are given by
\[ \exp\left\{ r \pm \left[ \frac{m+n}{m}\left(1+H^2\right)s^2 \right] \frac{t_{m+n-2}(\gamma/2)}{m+n-2} \right\}, \]
where $t_{m+n-2}(\gamma/2)$ is that point of the t distribution with $m+n-3$ degrees of freedom which is exceeded by chance alone $\gamma/2$ proportion of the time. In particular, $100\left[(A_S/A_N) - 1\right]$ may be interpreted as the percentage increase due to seeding (a negative value denotes a decrease).

Specific point estimates of scale parameters $A_N$, $A_S$, and $B$ are given by
\[ \hat{A}_N = \exp(u_{1N}) \]
\[ \hat{A}_S = \hat{A}_{NE}^r \]
\[ \hat{B} = \exp\left[ (nv_{1N} + nv_{1S})/(m+n) \right] \]
Point estimates of the shape and dependence parameters $\alpha$, $\beta$, and $\rho$ are also given by
\[ \hat{\alpha} = \left( \frac{m(u_{2N} - u_{1N}) + n(u_{3S} - u_{1S})}{(m+n)} \right)^{1/4}, \]
\[ \hat{\beta} = \left( \frac{m(v_{2N} - v_{1N}) + n(v_{3S} - v_{1S})}{(m+n)} \right)^{1/4}, \]
\[ \hat{\rho} = \frac{m(uv_{2N} - uv_{1N}) + n(uv_{3S} - uv_{1S})}{(m(u_{2N} - u_{1N}) + n(u_{3S} - u_{1S}))^{1/2}}. \]

The above point estimates of $\alpha$, $\beta$ and $\rho$ are also appropriate as given above (i.e., with no modification of the sufficient statistics) for the same parameters of the bivariate log-normal distribution associated with the analytic technique given in the Appendix for the cross-over design.

4. Numerical example

The data used to illustrate this covariance analysis technique were obtained during the 1964-70 Wolf Creek Pass (WCP) wintertime orographic cloud seeding experiment. This experiment was sponsored by the State of Colorado and was both designed and operated by Colorado State University under the direction of Lewis O. Grant. A description of this experiment has recently been given by Grant and Elliott (1974).

In particular, the target area of this experiment was the WCP vicinity. Three precipitation sensors spaced at roughly 5 mi intervals were used during each of the six winter seasons of this experiment. These three precipitation sensors were designated WCP West (2899 m), WCP Summit (3249 m) and WCP East (2819 m). The target area was seeded during three winter seasons of this experiment (1964-65, 1966-67 and 1968-69). The target area was non-seeded during the other three winter seasons of this experiment (1965-66, 1967-68 and 1969-70). Contrary to a statement by Grant and Elliott (1974), the WCP experiment was initiated during the 1964-65 winter season.

The winter season was the experimental unit of the WCP experiment since a primary goal of this effort was to compare non-seeded and seeded winter season snowpack runoff amounts. The concurrent Climax wintertime orographic cloud seeding experiment (cf. Mielke et al., 1971) was designed to compare non-seeded and seeded daily precipitation amounts.

As a consequence, a 24 h period was the experimental unit of the Climax experiment. Since non-seeded and seeded 24 h periods of the WCP experiment were intentionally not randomized, comparisons of non-seeded and seeded 24 h precipitation amounts from the three WCP precipitation sensors per se are highly questionable because of climatological differences between winter seasons. However, the use of a covariate based on appropriate control data provides a vehicle to adjust comparisons of non-seeded and seeded 24 h period precipitation amounts for the climatological differences between winter seasons. The control area (an area 50-100 mi west of WCP) data consisted of six National Weather Service (NWS) precipitation sensors located at Durango (NWS #0-2432, 1996 m), Mesa Verde (NWS #0-5531, 2155 m), Ouray (NWS #0-6203, 2359 m), Pleasant View (NWS #0-6591, 2091 m), Silverton (NWS #0-7656, 2841 m) and Telluride (NWS #0-8204, 2669 m). Since each of these control area precipitation sensors is situated 50-100 mi west of WCP, contamination of any control area precipitation sensor by seeding of the WCP experiment is exceedingly remote.

In this example $x$ is the mean of precipitation amount measurements (inches of water) from the three WCP precipitation sensors for a 24 h period providing (i) no more than one of the three measurements is missing and (ii) the mean precipitation amount is at least 0.015. Also $y$ is 0.01 plus the mean of precipitation measurements (inches of water) from the six control area NWS precipitation sensors for a 24 h period providing at least three of the six measurements are non-missing (the purpose of adding 0.01 to each value is simply to insure that all covariates
are positive prior to taking logarithms). Among all the 357 pairs of target and control area measurements \((x_i, y_i)\) obtained during the WCP experiment, 185 (172) pairs involved a non-seeded (seeded) target area. A further point worth noting is that lag correlation estimates associated with these data were negligible (near zero).

Using the previously described data of the WCP experiment, we will illustrate the covariate analysis technique by evaluating two disjoint and exhaustive 500 mb temperature partitions of this data \((\text{vis.}, \quad < -23^\circ C \quad \text{and} \quad \geq -23^\circ C)\). As in previous investigations, the 500 mb temperature is merely an estimate of cloud top temperature.

\[
\begin{align*}
500 \text{ mb temperature} & \quad < -23^\circ C \quad (m = 83, \quad n = 72) \\
\bar{u}_1 & = -2.0073 \quad \bar{v}_1 & = -1.9620 \quad G = 0.55332 \\
\bar{v}_1 & = 5.2744 \quad \varepsilon_2 & = 4.8009 \quad H = 0.000001 \\
\bar{v}_1 & = -2.8964 \quad \bar{v}_1 & = -2.8987 \quad s_2 = 0.76566 \\
\varepsilon_2 & = 9.6943 \quad \varepsilon_2 & = 9.4144 \quad \tau = 0.04657 \\
(uw)_1 & = 6.5825 \quad (uw)_2 & = 6.1937 \quad t_{12}(0.025) = 1.975 \\
T & = 0.33 \quad [2\text{-sided} \quad P\text{-value} = 0.74] \\
\bar{A}_s/\bar{A}_N & = 1.05 \quad [\text{estimated} \ 5\% \ \text{increase} \ \text{(non-significant)} \ \text{due to seeding}] \\
(0.79, 1.38) & \text{are } 0.95 (95\%) \ \text{confidence limits for } \bar{A}_s/\bar{A}_N \\
\bar{A}_N & = 0.134 \quad \bar{A}_s & = 0.140 \quad \bar{B} = 0.055 \\
\alpha & = 1.053 \quad \beta & = 1.081 \quad \beta & = 0.568 \\
500 \text{ mb temperature} & \quad \geq -23^\circ C \quad (m = 102, \quad n = 100) \\
\bar{u}_i & = -2.0481 \quad \bar{v}_1 & = -1.6170 \quad G = 0.68861 \\
\bar{v}_i & = 5.8963 \quad \varepsilon_2 & = 3.9933 \quad H = 0.000034 \\
\bar{v}_1 & = -2.6571 \quad \bar{v}_1 & = -2.6420 \quad s_2 = 0.73743 \\
\varepsilon_2 & = 8.6894 \quad \varepsilon_2 & = 8.7331 \quad \tau = 0.42070 \\
(uw)_i & = 6.6600 \quad (uw)_2 & = 5.3812 \quad t_{12}(0.025) = 1.972 \\
T & = 3.48 \quad [2\text{-sided} \quad P\text{-value} = 0.0006] \\
\bar{A}_s/\bar{A}_N & = 1.52 \quad [\text{estimated} \ 52\% \ \text{increase due to seeding}] \\
(1.20, 1.93) & \text{are } 0.95 (95\%) \ \text{confidence limits for } \bar{A}_s/\bar{A}_N \\
\bar{A}_N & = 0.129 \quad \bar{A}_s & = 0.196 \quad \bar{B} = 0.071 \\
\alpha & = 1.236 \quad \beta & = 1.300 \quad \beta & = 0.724 \\
\end{align*}
\]

Not only do the above results illustrate this covariance analysis technique based on the bivariate log-normal distribution, they also yield additional support for previous findings pertaining to the cloud top temperature association with wintertime orographic cloud seeding (Grant and Elliott, 1974).

Acknowledgments. This research was performed as part of the National Hail Research Experiment (Sub-contract SS013), managed by the National Center for Atmospheric Research and sponsored by the Weather Modification Program, Research Application Directorate, National Science Foundation, and also was jointly supported by NSF, RANN, under Grant ERT-71-01885-A03, and a cooperative agreement with the Rocky Mountain Forest and Range Experiment Station and the Northern Forest Fire Laboratory, USDA Forest Service. The authors appreciated comments by a referee.

APPENDIX

Analytic Technique Based on Bivariate Log-Normal Distribution for Cross-Over Designs

Suppose the \(m\) non-seeded events of the continued-covariate design described in Section 3 are replaced with modified events such that the target area is non-seeded and the control area is seeded. If now the target area is termed target area I and the control area is termed target area II and the two target areas are alternately seeded, then this modified design is a cross-over design.

For a cross-over design to be effective, precipitation amount measurements from the two target areas must be well correlated and neither target area can be seriously contaminated by seeding of the other target area. As in Section 3, let \(A_N\) and \(A_S\) denote scale parameters associated with non-seeded and seeded target area I measurements, respectively. However, now let \(B_N\) and \(B_S\) denote scale parameters associated with seeded and non-seeded target area II measurements, respectively. Thus either scale parameters \(A_N\) and \(B_N\) or scale parameters \(A_S\) and \(B_S\) are associated with a specific event. It is assumed that \(A_s/A_N = B_N/B_S\) (i.e., the effect of seeding on either target is the same). Thus \(A_N = A_N\) implies \(B_S = B_N\) and visa versa.

Suppose \(x_i\) and \(y_i\) are replaced with \(w_i = x_i/y_i\) and \(z_i = x_i y_i\), respectively, in each sufficient statistic listed in Section 3. Then test statistic \(T\) in Section 3 is again distributed as Student's \(t\) with \(m + n - 3\) degrees of freedom under \(H_S: A_N = A_S\) and where each of the statistics \(\tau, G, H\) and \(s_2\) now involve the modified sufficient statistics described in the previous sentence. However, the point estimate of either \(A_s/A_N\) or \(B_N/B_S\) is now given by

\[
\bar{A}_s/\bar{A}_N = \bar{B}_N/\bar{B}_S = e^{\tau_n} \\
\text{and the } 1 - \gamma \ \text{confidence limits for either } A_s/A_N \text{ or } B_N/B_S \text{ are given by} \\
\exp \left\{ \pm \left( \frac{m+n}{24mn} (1+H) s_2^2 \right)^{\frac{1}{m+n-3}} \right\}.
\]

Also specific point estimates of scale parameters \(A_N, A_S, B_N\) and \(B_S\) are given by

\[
\bar{A}_N = \exp(u_{1N}), \quad \bar{B}_N = \exp(v_{1S}), \\
\bar{A}_S = \bar{A}_N e^{\tau_n}, \quad \bar{B}_S = \bar{B}_N e^{\tau_n}.
\]

The point estimates of the shape and dependence parameters \(\alpha, \beta\) and \(\rho\) are identical (i.e., do not in-
volve modification of sufficient statistics given above) to those point estimates of $\alpha$, $\beta$ and $\rho$ given in Section 3.

If the choice of a design is optional (i.e., no problem such as contamination exists), the optimal design both can and should be calculated with the algorithm given by Wu et al. (1972). In particular, the balanced cross-over design would be a most informative design for the example in Section 4 (as a consequence of $\alpha \approx \beta$).

REFERENCES


Mielke, P. W., 1974: Squared rank test appropriate to weather modification cross-over design. Technometrics, 16, 13–16.


