551.509.617 : 551.574.11

Bayesian statistical analysis as applied to cloud seeding experiments

P. N. SHARMA and R. K. KAPOOR

Indian Institute of Tropical Meteorology, Pune

(Received 12 December 1979)

ABSTRACT. Bayesian statistical analysis technique adapted to the evaluation of cloud seeding experiments offers great advantages by enabling to determine not only the direction of the seeding effect but also its magnitude. Since the seeding experiments are carried out sequentially, the technique immensely helps in the decision analysis regarding the period, economics, environmental and social impacts of the experimentation. The technique is elucidated. Results of Indian cloud seeding experiments are computed with this technique and compared with those obtained from powerful classical statistical tests.

1. Introduction

Despite nearly a quarter century of experimentation in precipitation enhancement by cloud seeding, the production of increased rainfall as a result of seeding is still regarded as uncertain though promising. However these experiments have made it amply clear that, just raingauge data analysed by classical statistical methods alone would not be sufficient to achieve meaningful results in a reasonable period of experimentation. Results of many an experiment remained inconclusive because the experiment was not conducted for a long enough period. Efforts are therefore made to supplement classical statistics with Bayesian statistics, numerical modelling and detailed and accurate measurements of the clouds and rainfall. Statistical problems here are deeply intermixed with problems of meteorology and cloud physics.

Simpson and her group were the first to adapt the techniques of Bayesian analysis to the cloud modification experiments (Simpson *et al.* 1973). Because of the time and cost factor involved in conducting the cloud seeding experiments, it is very important to utilize all the information that one can have. The Bayesian approach is a part of the development towards more effective utilization of all relevant data in statistical analysis. It gives most weight to the existing data and enables to determine not only the direction of the seeding effect but also its magnitude.

The advantage of the Bayesian statistical technique is that, it can be applied sequentially; the posterior probability from the first stage of experimental work serving as the prior probability of next stage. The method is specially useful in the evaluation of cloud seeding experiments, where data is being collected in stages. As evidence is being gathered, one can stop and see if the current posterior opinions determined by applying Bayes' theorem are sufficient to justify terminating the experiment. This is invaluable in the decision analysis, when we have to calculate early in an experiment, or even before it, the number of years of experimentation required to establish a postulated range of seeding effects to a specified significance level.

2. Bayesian statistics

It is accepted that relative frequency of an event approaches its probability, which quantifies our opinion or belief about some hypothesis or uncertain event. Prior opinions are changed by the data to yield posterior opinions or probabilities through the operation of Bayes' theorem. This revision of opinion in the light of new information is the basis of the Bayesian statistics. Bayes' theorem prescribes the amount of revision of opinion that should occur in the light of new information or data. It is, therefore, a non-controversial consequence of the theorem of inverse probability.

Bayes' theorem can be stated in simple form as follows :

Posterior probability $=\dfrac{\text{Prior probability} \times \text{Likelihood}}{\text{Datum probability}}$

or

$$p(H/D) = \frac{p(H) \times p(D/H)}{p(D)} \qquad (1)$$

where,

$D=$ Data or events that bear on our uncertainty about the hypothesis

$H=$ Hypothesis.

It may be noted that $p(H)$ is the unconditional probability for the hypothesis to be true. It is called prior probability because it represents opinion before any data are obatined. Some degree of knowledge regarding the natural distribution is required. The prior probability assignment may either incorporate prior knowledge, if it exists, or may be diffuse or unprejudiced. Prior opinion in the form of a probability distribution can usually be closely approximated by one of just a few common distributions that are easy to use in Bayes' theorem. $p(H/D)$ is the posterior probability so called because it indicates opinion that has been revised in the light of data, i. e., opinion after observing D. $p(D/H)$ is called likelihood; it is the probability associated with a particular data, given that event H has occurred.

The main objection to Bayesian approach is in the assigning of prior probabilities to one or more variables. It is contended that the subjective choice of prior probabilities introduces bias. But the rationale of Bayesian technique is that, inferences are made by combining the available information independent of data, provided by prior probabilities, with that given by the sample data obtained from subsequent experiments. Bayes' theorem automatically weights the relative contributions of prior and sample information to the posterior. The weights are determined by the relative precisions of the prior and the population. We rather intuitively expect that the weighted estimate of the current situation would be somewhere between the most likely prior value and the sample estimate of data. Further advantage is derived here by weighting the complete prior distribution by the information in the data to obtain an augmented probability (posterior) distribution as an expression of the combined inference. We shall briefly elucidate and follow the Bayesian approach applied to cloud seeding experiments, outlined by Simpson et al. (1973). The technique is applied to the evaluation of Indian cloud seeding experiments performed till now.

3. Bayesian analysis applied to cloud seeding

In a randomized seeded/control cloud seeding experiment two independent rainfall data are replicated. One set giving the rainfall over the seeded area and the other when the area is left unseeded. The aim of the statistical analysis is to find if there is an effect on the rainfall of the area caused by seeding and to determine its magnitude.

An essential requirement for applying Bayesian statistics to cloud seeding experiments however is that, the distribution of natural rainfall over the area should be known. This means that we should know the distribution function and its statistics, viz., mean and standard deviation of the natural rainfall, which are assumed to remain stationary in time.

It has been observed that convective rainfall is highly variable in both space and time, with most clouds raining little and a few very much. The rainfall distribution is thus highly skewed. As reported in meteorological literature, the gamma distribution function has been found to fit a large class of rainfall data, for Indian rainfall [Mooley and Crutcher (1968)]. In the present study it is seen that the seeded and control area rainfall data in respect of 19 seasons (1957-1966), of Delhi, Agra and Jaipur regions fitted into gamma distributions, after square root transformation, with nearly invariant shape parameter. Gamma distribution was fitted to the data by the method of maximum likelihood. A simplified procedure for estimating the shape and scale parameters based on this method is provided by Mooley (1974).

Considering x to be the daily area rainfall for the experimental period, which is well fitting a gamma distribution, whose probability density function is given by

$$p(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \qquad \alpha>0, \ \beta>0, \ x>0, \qquad (2)$$

where α and β are the shape and scale parameters respectively. Γ is the gamma function. Its mean μ and variance σ^2 are

$$\mu = <x> = \alpha/\beta; \qquad \sigma^2 = \alpha/\beta^2 \qquad (3)$$

Suppose that seeded and control data can be separately fitted into gamma distributions with nearly the same shape parameter and have scale parameters β_s and β_c respectively. Also, let $<x>_s$ and $<x>_c$ be the expected values of rainfall for seeded and control distributions. Let the seeding factor F be defined by

$$F = \frac{<x>_s}{<x>_c} = \frac{\beta_c}{\beta_s} \qquad (4)$$

where the last ratio is obtained from (3). A smaller value of β_s, therefore, implies more rainfall in seeded cases compared to non-seeded cases. Further it is assumed that the effect of seeding is multiplicative, *i. e.*, the seeding multiplies the natural rainfall amount per seeded experimental day by some factor F rather than merely producing a change in number of zero rainfall days or providing some additional rainfall per exprimental day. A seeding factor, *e. g.*, $F=1.30$ represents 30 per cent increase and $F=0.90$ represents 10% decrease. Alternatively, the effect may be assumed to be additive in which case the result is expressed as, *e.g.*, a 4mm increase in rainfall per seeded day. The multiplicative effect assumption is compatible with the test statistic for the crossover design, *viz.*, root double ratio (RDR) defined as,

$$\mathrm{RDR} = \sqrt{\frac{T_s}{C_{NS}} \cdot \frac{C_s}{T_{NS}}} \qquad (5)$$

which is utilised for evaluating the experiments by the classical statistical tests. Here T_s and C_s refer to area average rainfall over target and control areas on respective seeded days and C_{NS} and T_{NS} to rainfall on corresponding non-seeded days. Moreover, since a gamma distribution fits well into the transformed data for both target and control area rainfalls with the shape parameter nearly constant, it indicates that the seeding effect is multiplicative.

By virtue of the relation (4) we can make use of Bayes' equation for the scale parameter β to determine actual seeding factors.

Bayes' equation for β is

$$p(\beta/S) = \frac{p(\beta) \cdot p(S/\beta)}{p(S)} \qquad (6)$$

where,

$p(\beta/S)=$ Posterior probability density distribution of the scale parameter β in reference to the seeded data,

$p(\beta) =$ Prior probability assigned to the scale parameter,

$p(S/\beta)=$ Likelihood of the seeded data for the scale parameter,

$p(S) =$ Probability of the seeded data; may here be regarded as just a normalizing constant.

Thus, if the prior distribution of β and the likelihood function are known, posterior distribution of β can be computed. Considering different discrete values of prior F and the expected value of control rainfall $<x>_c$, expected value of prior $<\beta>_s$ were computed from the relation :

$$<\beta>_s = \frac{a}{<x>_s} = \frac{a}{F \cdot <x>_c} \qquad (7)$$

which is derived from relations (3) and (4).

Now Bayes' Eqn. (6) can be solved by assuming prior probability distribution of β either a gamma function or a uniform distribution. In the following analysis a prior gamma distribution has been assumed for β to simplify the analysis and because of the ability of the distribution to incorporate a wide range of prior information. This in no way limits the scope of the analysis though sensitivity is somewhat affected by assuming peaked or flat gamma, distributions. In the beginning of the of experiment a spread out prior distribution of β is preferable though of course a uniform assignment of β, within a wide range of seeding effect, would be the best thing to avoid any prejudice on the posterior β.

Let the prior distribution of β be,

$$p(\beta) = \frac{K_2 \cdot \beta^{K_1 - 1} \cdot e^{-K_2 \beta}}{\Gamma(K_1)} \qquad (8)$$

where K_1 and K_2 are respectively the shape and the scale parameters.

$$\text{Prior } <\beta>_s = \frac{K_1}{K_2} \qquad (9)$$

K_2 is calculated from this relation for an assumed value of K_1.

Since the seeding trials are independent, the likelihood function $p(S/\beta)$ can be written as

$$p(S/\beta) = \prod_{i=1}^{n} \frac{\beta^a x_i^{a-1}}{\Gamma(a)} \cdot e^{-\beta x_i} \qquad (10)$$

where n is the number of seeded cases.

Substituting (10) and (8) into (6), the gamma distribution for posterior β_s is given by

$$p(\beta/S) = \frac{\left(\sum_{i=1}^{n} x_i + K_2\right)^{na + K_1}}{\Gamma(na + K_1)} \cdot \beta^{na + K_1 - 1} e^{-\beta\left(\sum_{i=1}^{n} x_i + K_2\right)} \qquad (11)$$

Its shape parameter is $(na + K_1)$ and

scale parameter is $\left(\sum_{i=1}^{n} x_i + K_2\right)$

The expected value of posterior β_s is, therefore,

$$<\beta/S> = \frac{na + K_1}{\sum\limits_{i=1}^{n} x_1 + K_2} = \frac{a + K_1/n}{\bar{x}_i + K_2/n} \qquad (12)$$

With this posterior $<\beta/S>$ the actual seeding factor is computed using relation (7). When shape parameter K_1 of prior β distribution exceeds about ten, the posterior gamma distribution tends to be Gaussian (Johnson & Kotz 1970).

4. Application of Bayesian analysis to Indian experiments

Randomized warm cloud seeding experiments were first conducted in India in the Delhi, Agra and Jaipur regions during the period (1957-1966). A single target-control design was utilized in these experiments. A fresh series of experiments using aerial seeding commenced from 1973 in the drought prone area in the Pune region. These experiments utilize an area crossover design with a central buffer zone. In this design paired target areas are set up and either area is seeded at random in each test event, the unseeded one serving as the control for that event. Standard randomization procedure has been followed to avoid bias in the selection of the area for seeding on seedable days. The location of two areas is chosen such that they have a high correlation (>0.75) of daily rainfall and are oriented to avoid any contamination of one area due to seeding in the other area. Experimetal days in each pair, which are suitable for conducting cloud seeding experiments over the areas, are chosen so that similar meteorological conditions prevail on both the days. For this purpose if the second seedable day does not fall within five days of the first day of a pair, the first day is also rejected and a fresh pair is started. Also the days, on which there is continuous rainfall over the area during the time of experiment, are considered as unseedable days and the experiments are not conducted on such naturally rainy days.

The experimental unit for the area experiment is all clouds passing over the area during a fixed period on the seeding day. This does not imply that all clouds have to be seeded even though all rainfall in the area during the period is included in the evaluation. The concept of randomization is not compromised by the exercise of judgement concerning which clouds are to be treated or the amount of seeding material to be delivered to an individual cloud. In any one series of experiments the same seeding material and seeding technique is continued to be employed from the beginning till the termination of the experiments. The seeding material utilized in the experiments reported here was a micro-pulverized mixture of common salt and soapstone in the

TABLE 1

Bayesian analysis

Results : Delhi, Agra and Jaipur (Season-wise)
Area experiment (Period : 1957-1966)

Prior β	$K_1=10$ Gamma function		Posterior β Gaussian	
Prior $<\beta>s$	F	Post $<\beta>_s$	F	F for 95% probability
2.3445	0.5	1.4044	1.3934	1.1337–1.7533
1.6578	1.0	1.3967	1.4152	1.1414–1.7816
1.3535	1.5	1.3857	1.4311	1.1590–1.8116

ratio of ten to one respectively with particle diameter size of the order of ten microns. The experiments in the Delhi, Agra and Jaipur regions were ground based. The seeding particles were pumped up by high pressure air compressors operated from the ground, to form a rising plume during hours of maximum insolation, on seedable days. The experiment in the Pune area are being conducted from an aircraft flying through suitable clouds and delivering the seeding material into them through a gadget fitted inside the seeder aircraft. Details of the above mentioned experiments along with the corresponding data have been reported earlier (Ramana Murty and Biswas 1968; Krishna et. al. 1974, 1976).

Rainfall data pertaining to 19 monsoon seasons (July to September) which comprise 9 seasons at Delhi, 6 seasons at Agra and 4 seasons at Jaipur is considered as a single data set since these three regions are climatologically nearly similar. For Pune (crossover) area experiment results of 17 pairs, period 1973-74, are analysed. The details of these experiments and the data are given in the publications referred to above and are therefore not being repeated. Seasonal rainfall for seeded area x_s and control area x_c in respect of 19 seasons mentioned above has been found to be fitting well into gamma distributions after square root transformation. The shape parameter was found to be nearly invariant with value about 16 and the scale parameter varied between 1.3 and 1.8. Bayesian analysis applied to the scale parameter. By assuming different prior seeding factors (F) posterior seeding factors were evaluated. The results are presented in Table 1.

It is noted that even though prior seeding factors were assumed 50 per cent decrease, no effect and 50 per cent increase in rainfall, the posterior seeding factors ranged between 1.393 and 1.431 showing thereby rainfall increases between 39.3 and 43.1 per cent. The 95 per cent confidence limits of the posterior seeding factors are also listed in the last two columns of the same table. Thus, the positive effects of seeding were nearly constant and they were about 40 per cent. The results of Bayesian analysis applied to the Pune experiments are given in Table 2.

TABLE 3

Levels of significance (One sided)

Results : Pune (Crossover) Area Experiment
(Period : 1973-1974), No. of pairs : 17

Mann-Whitney	Wilcoxon Matched pairs	K.S. two samples	Likelihood ratio
0.1894	0.1056	0.2020	0.3272

TABLE 2

Bayesian analysis

Results : Pune (Crossover) Area Experiment
(Period : 1973-1974) No. of Pairs : 17

$K_1=10$ Prior β Gamma function			Posterior β Gaussian	
Prior $<(\beta)>s$	F	Post $<\beta>s$	F	F for 95% probability
0.3607	0.5	0.2066	0.8730	0.6130-1-5159
0.2254	0.8	0.1789	1.0079	0.7078-1-7 504
0.1803	1.0	0.1643	1.0978	0.7709-1-9062
0.1503	1.2	0.1518	1.1878	0.8341-2-0623
0.1202	1.5	0.1363	1.3226	0.9288-2-2966

The results are also not statistically significant by any of these tests. Preliminary statistical analysis shows that more experiments are needed at Pune to achieve any significant results.

5. Conclusion

Bayesian statistical method can be applied to the evaluation of area cloud seeding experiments with advantage when the natural rainfall distribution over the area is known. The technique not only helps in finding whether cloud seeding causes an increase or decrease in rainfall but also the magnitude of the seeding effect. The results of earlier Indian experiments have been independently confirmed with this method. For the on-going experiments in the Pune region, analysis indicate that further experiments are necessary to obtain significant results.

It is seen that the results are not consistent, ranging from a 13 per cent decrease in the first case to a 32 per cent increase in rainfall in the last case when the prior seeding factors are assumed to vary from 0.5 (50 per cent decrease) to 1.5 (50 per cent increase). It is felt that more data is required to obtain conclusive results.

Three different powerful classical tests, *viz.*, Mann-Whitney, Wilcoxon matched pairs signed ranks and K-S two sample test have been applied to Pune experimental data. The results are given in Table 3.

References

Johnson, Norman I. and Kotz, Samuel, 1970, Distributions in statistics—Continuous univariate distributions: I, p. 170; Houghton Miflin Company, Boston, USA.

Krishna, K., Kapoor, R.K., Ramachandra Murty, A.S., Selvam, A.M., Kanuga, K.K., Khemani. L.T., Mukherjee, B.K., Paul, S.K., Vijay Kumar, R. and Ramana, Murty, Bh. V., 1974, A randomized cloud seeding experiment in the Deccan Plateau, India—Preliminary Results, J. Weath. Modification, 7, 1, 266-277.

Krishna, K., Kapoor, R.K., Ramachandra Murty, A.S., Rajamani, S., Selvam, A.M., Khemani, L.T., Kanuga, K.K., Mukherjee, B.K., Paul, S.K., Reddy, R.S., Manohar, G.K., Brij Mohan, Vijay Kumar, R. and

Ramana Murty, Bh.V., 1976, Results of two years of randomized cloud seeding experiment in the Deccan Plateau, India, *J. Weath. Modification*, **8**, 1, 134-144.

Mooley, D.A. and Crutcher, H. L., 1968, An application of gamma distribution function to Indian rainfall, ESSA Tech. Report, U.S. Department of Commerce, Silver Spring, M.D.

Mooley, D.A., 1974, Suitable tables for application of gamma probability model to rainfall, Research Report RR-018, Indian Institute of Tropical Meteorology, Pune.

Ramana Murty, Bh. V. and Biswas, K.R., 1968, Weather Modification in India, *J. met. Soc. Japan*, **46**, 160-165.

Simpson, J., Woodley William, L., Olsen Anthony and Eden, Jane C., 1973, Bayesian Statistics applied to dynamic modification experiments on Florida Cumulus Clouds, *J. atmos. Sci.*, **30**, 1178-1190.