Mathematical Geology, Vol. 12, No. 1, 1980

Progressively Censored Sampling of Rock Joint Traces[1]

Gregory B. Baecher[2]

A number of sampling problems in geology and engineering geology involve geometric variables, and must deal with the almost pervasive biases that accompany geometric sampling. Among these biases is the fact that not all elements of the sampled population are fully observable. Some members, usually the largest, are censored. Inferences cannot ignore the censored members of the sample, because the censoring is often related to the variable being inferred—for example, the case of sampling for feature size. Inferences from samples are conceptually straightforward, and for the simple case of exponential parent distributions, mathematically tractible. Maximum likelihood and Bayesian results are given for the exponential case, and examples are drawn from joint surveys in rock mechanics.

KEY WORDS: censored sampling, joint traces, maximum likelihood.

INTRODUCTION

The question of censored observations in sampling from known distributions has been recognized for several decades, having arisen in life testing (Epstein, 1954), insurance statistics (Fisher, 1931; Hald, 1949) and other fields. In geology, the problem has been notable in joint surveys for rock mechanics, in resource estimation, and other sampling problems involving geometric variables. For simplicity, the present discussion draws examples from joint surveys, but the methods are sufficiently general to apply to a range of sampling problems.

The standard definition of joints holds them to be fractures or cracks (in rock) along which there has been little movement. Empirically, joints tends to form subparallel groups, to be randomly rather than systematically spaced, and to have a finite measurable extent. The importance of jointing to the engineering behavior of excavations, openings, and foundations is almost an underlying principle of rock mechanics. Although there is considerable divergence of opinion on

[1]Manuscript received 1 March 1979; revised 19 June 1979.

[2]Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

0020-5958/80/0200-0033$03.00/0

the gensis of jointing (e.g., Price, 1966), for the purposes of engineering this controversy is somewhat unimportant.

The geometric properties of jointing are inferred primarily from observations in outcrops and openings. While advances in statistical techniques for inferring fracture patterns from drill cores are being made, from a practical point of view these are yet to find application. The observations made in outcrop are of joint traces, that is, of the intersections of joint planes with the outcrop. Several geometric properties are measured, but of present concern is the distribution of trace lengths, from which joint size and persistence are ultimately inferred.

A number of biases exist in sampling trace lengths and imferring joint size. These have been discussed in Baecher, Lanney, and Einstein (1977a, b) among other places and will not be repeated here. The question of censoring involves joint traces that are not completely observable. The most common reason a joint trace is not completely observable is that it runs off the outcrop, or into a wall (Fig. 1). Thus, for that particular observation one knows only that the actual trace length is longer than observed. Because longer traces have a greater probability of being censored than do shorter ones, these incomplete observations cannot be ignored.

The problem of inferring the distribution of trace lengths from a censored sample is, in principle, easy. Given a set of observations partitioned into three groups:

$$x = \{x_1, \ldots, x_n\}: \text{ trace lengths with both ends observable,}$$

$$y = \{y_1, \ldots, y_m\}: \text{ trace lengths with one end observable,}$$

$$z = \{z_1, \ldots, z_k\}: \text{ trace lengths with no ends observable} \quad (1)$$

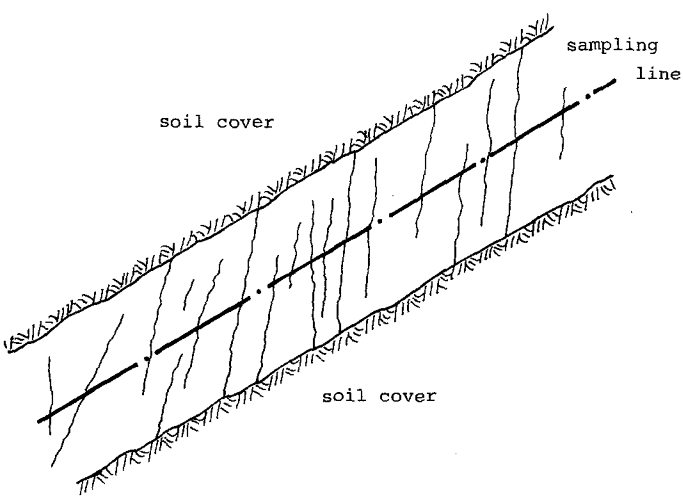

Fig. 1. Joint traces sampled in construction trench. Note, not all traces have fully observable length.

and a density function for trace length $f(l\,|\,\theta)$, with parameter θ, the likelihood of the observations is simply

$$L(x,y,z\,|\,\theta) = \prod_{i=1}^{n} f(x_i|\theta) \times \prod_{i=1}^{m} \int_{y_i}^{\infty} f(y_i|\theta)\,dy \times \prod_{i=1}^{k} \int_{z_i}^{\infty} f(z_i|\theta)\,dz \tag{2}$$

While equation (2) is easily written, its maximum is less easily found. Numerical solutions by enumeration are possible for any family of density functions, but closed form solutions are conveniently calculated only for those families having analytical cumulative distributions. This has made the exponential family a favored choice, even though empirical evidence sometimes suggests distributions more of the lognormal or gamma shape (e.g., Baecher, Lanney, and Einstein, 1977a; Epstein, 1954).

For exponential trace length

$$L(x,y,z\,|\,\theta) = (\theta^n e^{-\theta\Sigma x_i}) \times (e^{-\theta\Sigma y_i}) \times (e^{-\theta\Sigma z_i}) \tag{3}$$

Equating the derivative of (3) to zero yields the maximum likelihood estimate originally due to Epstein (see also Cruden, 1977)

$$\begin{aligned}\hat{\theta} &= n/(\Sigma x_i + \Sigma y_i + \Sigma z_i)\\ &= n/\mathcal{L}\end{aligned} \tag{4}$$

where $\mathcal{L} = (\Sigma x_i + \Sigma y_i + \Sigma z_i)$.

The simplicity of this estimate is attractive. The difficulty is that the sampling variance of $\hat{\theta}$ becomes large as n, the number of two-ended observations, becomes small.

The variance of $\hat{\theta}$ is found in the normal way

$$V[\hat{\theta}] \cong -E^{-1}\,\{[\partial^2 \log L(x,y,z\,|\,\theta)]/\partial\theta^2\} \tag{5}$$

$$= \theta^2/n \tag{6}$$

showing the sensitivity to n. Since θ is not known, it is replaced by $\hat{\theta}$ to obtain an estimate from the sample. In this case $\hat{\theta}$ is easily shown to be biased. Since first and second derivatives of equation (3) exist for all θ, the estimator $\hat{\theta}$ is asympptotically normal (Kendall and Stuart, 1973).

From a Bayesian point of view the natural conjugate to censored Exponential sampling remains the gamma distribution,

$$f^0(\theta\,|\,\alpha,\beta) \propto \theta^{\alpha-1} e^{-\beta\theta} \tag{7}$$

From which the posterior parameters are found by

$$\begin{aligned}\alpha' &= \alpha^0 + n\\ \beta' &= \beta^0 + \mathcal{L}\end{aligned} \tag{8}$$

where (α^0, β^0) are the prior parameters and (α', β') the posterior. For the noninformative prior

$$\alpha^0 \to 0$$
$$\beta^0 \to 0 \tag{9}$$

The predictive density on trace length starting from a noninformative prior is

$$f_l(l) = \int_0^\infty f_l(l \mid \theta) f'(\theta \mid x, y, z)\, d\theta \tag{10}$$

$$= \int_0^\infty (\theta e^{-\theta l}) \{ [\mathcal{L}^n / \Gamma(n)]\, \theta^{n-1} e^{-\theta \mathcal{L}} \}\, d\theta \tag{11}$$

$$= \frac{n\mathcal{L}^n}{(\mathcal{L} + l)^{n+1}} \tag{12}$$

This can vary substantially from the pdf conditioned on $\hat{\theta}$, the maximum likelihood estimate (Fig. 2). However, since the predictive pdf incorporates both nat-

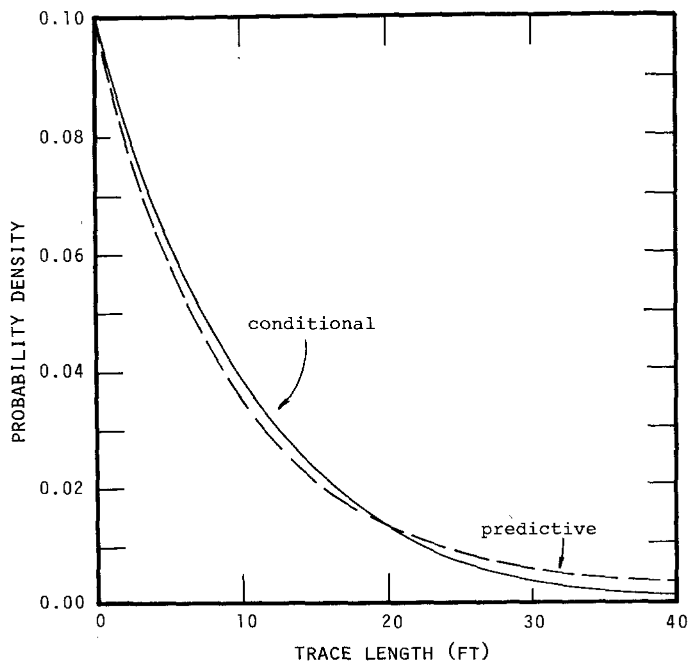

Fig. 2. Predictive density over trace length and density conditioned on maximum likelihood θ ($\mathcal{L}$ = 100 ft, n = 10).

ural and statistical uncertainty it is the appropriate density function for statistical prediction and reliability analysis.

EXAMPLE

Figure 3 shows joint trace length data collected in outcrops by measuring every trace intersecting a sampling circle of fixed diameter.[3] The data is divided according to the number of observable ends, and can be summarized as

$$\sum_{i=1}^{n} x_i = 147.2 \qquad n = 128$$

$$\sum_{i=1}^{n} y_i = 131.2 \qquad m = 64$$

$$\sum_{i=1}^{n} z_i = \underline{25.8} \qquad k = \underline{12}$$

$$304.2 \qquad 204$$

Thus, from (3) and (6)

$$\hat{\theta} = 0.421$$

$$V(\hat{\theta}) = 0.0014$$

$$\mathrm{CV}(\hat{\theta}) = 9\%$$

and the Bayesian posterior pdf is shown in Figure 4. This estimate, of course, is much smaller than that obtained merely from the complete observations

$$\hat{\theta}_x = n/\Sigma x_i = 0.87$$

and is also smaller than that obtained from the uncorrected observations

$$\hat{\theta}_{x+y+z} = (n + m + k)/\Sigma x_i + \Sigma y_i + \Sigma z_i) = 0.67$$

The reason is clear: large traces have a greater probability of being censored than do small ones.

The predictive density function for the data is,

$$f(l \mid z, y, z) = \frac{128\,(304.2)^{128}}{(304.2 + l)^{129}}$$

which is compared to the pdf conditioned on $\hat{\theta}$ in Figure 5. As $n \to \infty$ the condition pdf approaches the predictive because uncertainty in θ goes to zero.

[3]Length bias is ignored here, although it can be easily accommodated in principle. In practice length bias leads to censored gamma sampling with accompanying analytical difficulties.

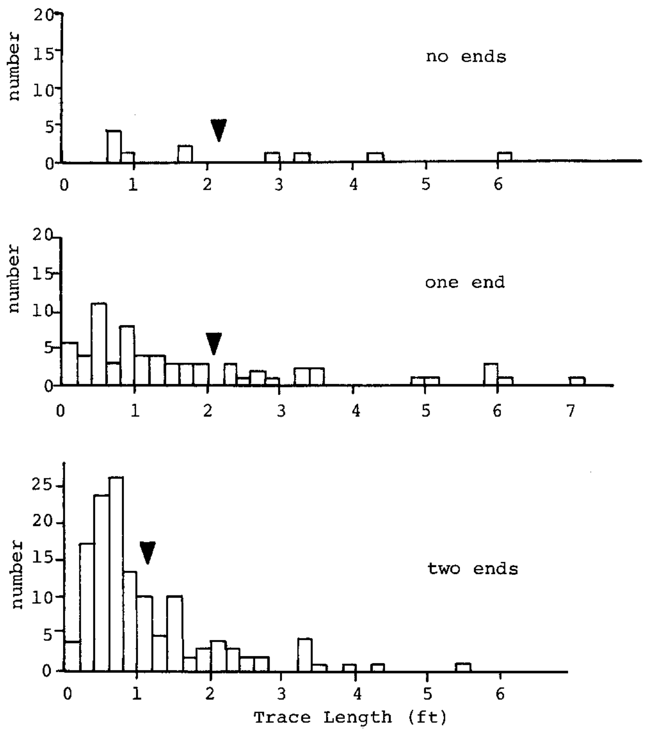

Fig. 3. Joint trace length data grouped by the number of observable end points. Triangles mark sample means.

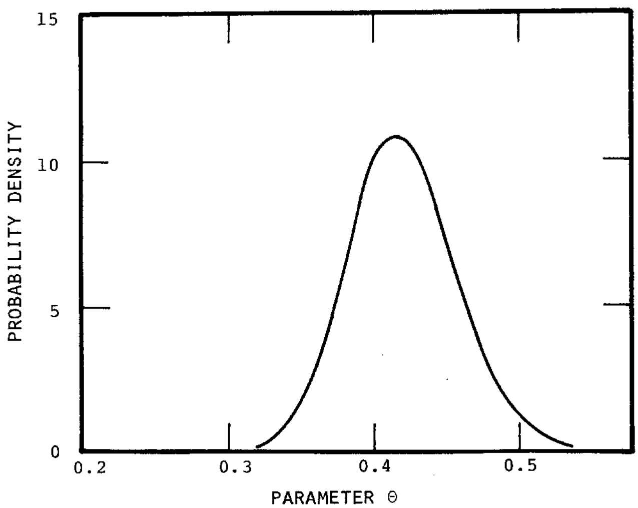

Fig. 4. Bayesian posterior distribution on θ.

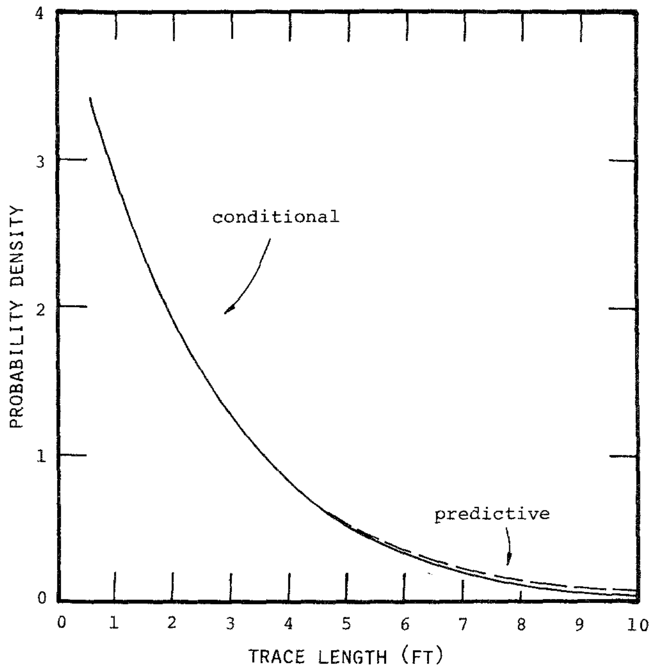

Fig. 5. Predictive and conditional density of trace length ($\mathcal{L}$ = 304.2, n = 128).

CONCLUSIONS

While the problem of censored sampling from exponential distributions has a convenient analytical solution, certain problems in sampling geometric data remain. Perhaps the most troubling is the fact that many geometric properties display lognormal or gamma distributions, and the cumulative functions of these distributions do not have nice analytical forms. For a fixed point of censoring, tabulated solutions are given by Hald, but otherwise the problem is unsolved. It is certainly true that inferences from progressively censored lognormal and gamma samples can be made using electronic computers, but these are not always available. Another problem is the relation between trace length and joint geometry. That is, the trace is an intersection between some two- or three-dimensional features and an outcrop. Inferences of the size of the feature itself generally require shape assumptions and often lead to mathematical complexity (e.g., Baecher, Lanney, and Einstein, 1977 b; Epstein, 1954). Nevertheless, because of its simplicity, the exponential model is in wide use in engineering geology, and tractable solutions are available for a broad group of sampling problems associated with it.

ACKNOWLEDGMENTS

This work was funded by the U.S. Bureau of Mines under Contract J0275015, and the National Science Foundation under Grant ENG 76-24592. The author would like to acknowledge the comments and ideas of Paul J. Visca and John M. Marek of Pincock, Allen and Holt, Inc.; Herbert H. Einstein of the Massachusetts Institute of Technology; Nicholas Lanney of the Stone and Webster Corporation; and Benjamin Epstein of the Technion. The author also acknowledges the helpful comments of a reviewer.

REFERENCES

Aitchison, J. and Dunsmore, I. R., 1978, Statistical prediction analysis: Cambridge University Press.

Baecher, G. B., Lanney, N. A., and Einstein, H. H., 1977a, Trace length bias in joint surveys, 20th U.S. national symposium on rock mechanics: Lake Tahoe, Nevada.

Baecher, G. B., Lanney, N. A., and Einstein, H. H., 1977b, Statistical descriptions of rock properties and sampling, 19th U.S. national symposium on rock mechanics: Keystone, Colorado.

Bartholomew, D. J., 1957, A problem in life testing: Jour. Am. Stat. Assn., v. 52, pp. 350–55.

Cruden, D. M., 1977, Describing the size of discontinuities: Rock Mech. Min. Sci., v. 14, pp. 133–137.

Epstein, B., 1954, Truncated life tests in the exponential case: Ann. Math. Statist., v. 25, pp. 555ff.

Fisher, R. A., 1931, The truncated normal distribution: British Assn. Adv. Sci. Math., table I, pp. XXXIII-XXXIV.

Hald, A., 1949, Maximum likelihood estimators of the parameters of a normal distribution truncated at a known point: Skand. Aktur. Tidsk., v. 32, pp. 119ff.

Kendal, M. G. and Stuart, A., 1973, The advanced theory of statistics: Hafner, New York, v. 2, 3rd ed.

Price, N. J., 1966 Fault and joint development in brittle and semi-brittle rock: Pergamon, London.