Monte Carlo Technique to Simulate Aflatoxin Testing Programs for Peanuts

T.B. WHITAKER, J.W. DICKENS, ARS, USDA, and **E.H. WISER,** [2]
Biological and Agricultural Engineering Department, North Carolina University,
Raleigh, North Carolina 27607

ABSTRACT

A computer model that accounts for sampling, subsampling, and analytical variability was developed to simulate aflatoxin testing programs. Monte Carlo solution techniques were employed to account for conditional probabilities that arise from multiple samples, subsamples, and/or analyses being used in testing programs. The aflatoxin testing program to be used on the 1974 peanut crop was evaluated by use of the described model.

INTRODUCTION

The total variability associated with aflatoxin test results for shelled peanuts is composed of at least three components: sampling, subsampling, and analysis (1). An observed aflatoxin test result $\bar{x}$ may be represented as follows:

$$\bar{x} = \mu + \alpha + \beta + \gamma \qquad \text{(I)}$$

where μ is the true aflatoxin concentration in the lot being tested, α is the random error due to sampling with expected value zero and variance $\sigma^2_{\bar{x}(s)}$, β is the random error due to subsampling with expected value zero and variance $\sigma^2_{\bar{x}(ss)}$, and γ is the random error due to analysis with expected value zero and variance $\sigma^2_{\bar{x}(a)}$. Eq. I suggests that three distributions must be considered in describing the variability of $\bar{x}$ about the lot concentration μ: (a) distribution of sample concentrations $\bar{x}_s$ about the lot concentration μ (b) distribution of subsample concentrations $\bar{x}_{ss}$ about the sample concentration μ_s, and (c) distribution of analytical determinations $\bar{x}_a$ about the subsample concentration μ_{ss}.

Aflatoxin testing programs may specify acceptance of lots with low test results and further testing of other lots. The additional tests may require that additional samples be drawn, that additional subsamples be analyzed, or that additional analyses be made on the same subsample. The conditional probabilities associated with this type of testing program cannot be handled by analytical means. This paper describes the development of a model coupled with Monte Carlo solution techniques to simulate aflatoxin testing programs, including those which involve conditional probabilities.

[1] Paper 4869 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, NC 27607.

[2] ARS, USDA, and Biological and Agricultural Engineering Department, North Carolina State University, Raleigh, NC 27607.

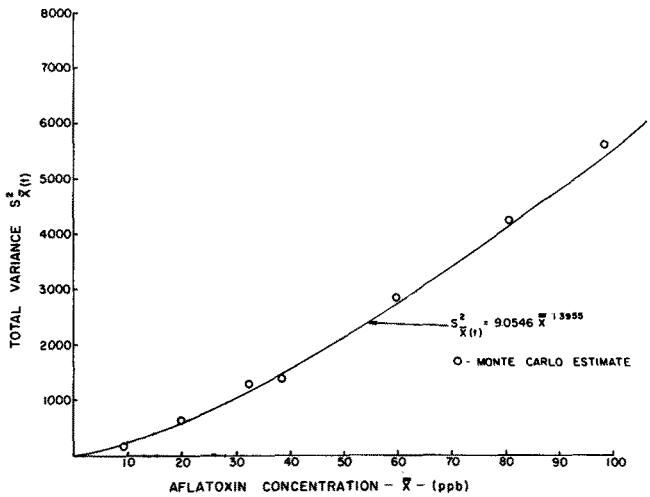

FIG. 1. Monte Carlo estimates of the total variance $s^2_{\bar{x}(t)}$.

DISTRIBUTION EQUATIONS

Previous studies have indicated that the negative binomial distribution function adequately describes the distribution of peanut sample aflatoxin concentrations $\bar{x}_s$ about the lot aflatoxin concentration μ (2).

$$F_s(N_s\bar{x}_s) = \sum_{r=0}^{N_s\bar{x}_s} \left[(\Gamma(r+N_sK_s)/(r!\ \Gamma(N_sK_s))) \right.$$
$$\left. (K_s/(K_s+\mu))^{N_sK_s} (\mu/(K_s+\mu))^r \right] \qquad \text{(II)}$$

where Γ is the gamma function, N_s is the sample size in number of kernels, and K_s is the "shape parameter" determined by the aflatoxin concentration in the lot. Since the distribution of aflatoxin-contaminated particles in the comminuted sample is probably similar to the distribution of contaminated kernels in the sample before comminution, the negative binomial distribution function may be used to describe the distribution of subsample concentrations $\bar{x}_{ss}$ about the sample concentration μ_s.

$$F_{ss}(N_{ss}\bar{x}_{ss}) = \sum_{r=0}^{N_{ss}\bar{x}_{ss}} \left[(\Gamma(r+N_{ss}K_{ss}) / (r!\ \Gamma(N_{ss}K_{ss}))) \right.$$
$$\left. (K_{ss}/(K_{ss}+\mu_s))^{N_{ss}K_{ss}} (\mu_s/(K_{ss}+\mu_s))^r \right] \qquad \text{(III)}$$

where Γ is the gamma function, N_{ss} is the number of communited particles in the subsample, and K_{ss} is the "shape parameter" determined by the aflatoxin concentration in

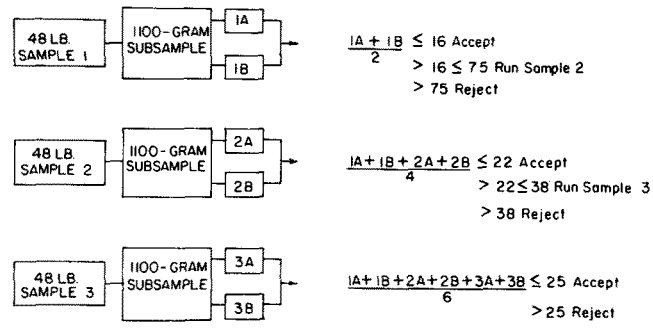

FIG. 2. Schematic of the aflatoxin testing program used on the 1975 peanut crop.

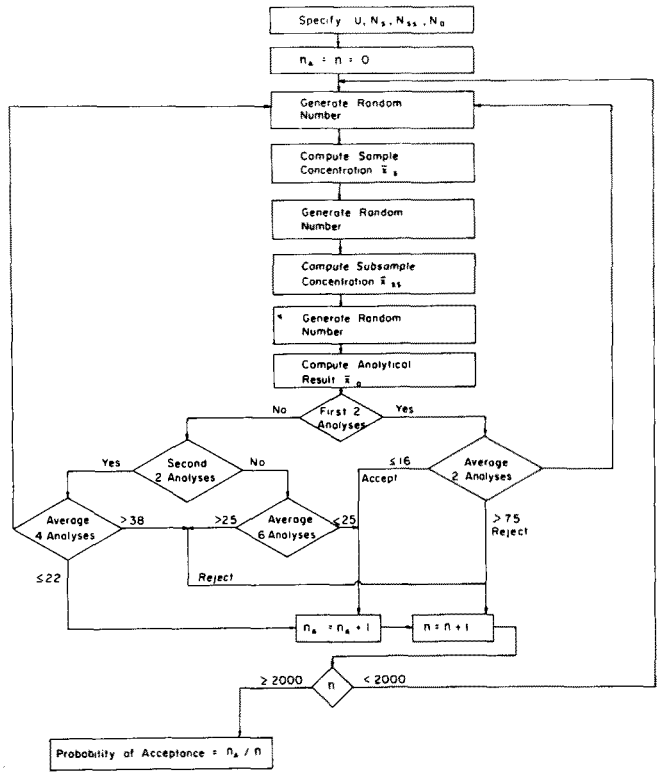

FIG. 3. Flow chart describing the computer model and Monte Carlo solution techniques.

the sample. A study of subsampling variability (1) indicated that the subsampling variance σ^2_{ss} is large compared to μ_s, which is characteristic of skewed distributions such as the negative binomial function.

Previous studies indicated that analytical results $\bar{x}_s$ may be normally distributed about the subsample concentration μ_{ss}. The normal and the log-normal distributions were compared to experimental observations, and the normal distribution gave a better fit than did the log-normal distribution (3).

$$F(\bar{x}_a) = \int_0^{\bar{x}_a} (1/\sqrt{2\pi}\,\sigma_{\bar{x}(a)})\, \mathrm{EXP}\,(-(\bar{x}_a-\mu_{ss})^2 / (2\sigma^2_{\bar{x}(a)}))\, d\bar{x}_a \quad \text{(IV)}$$

Sample size N_s, subsample size N_{ss}, and the number of analyses N_a are defined by the test procedure. The parameters K_s, K_{ss}, and $\sigma^2_{\bar{x}(a)}$ for the sampling, subsampling, and analytical distribution equations, respectively, must be computed. The shape parameter K for the negative binomial distribution is defined as:

$$K = \mu^2/(\sigma^2 - \mu) \quad \text{(V)}$$

where σ^2 is the variance among individual members of a population and μ is the population mean. In a previous study, empirical relationships were derived for sampling variance σ^2_s, subsampling variance σ^2_{ss}, and analytical variance σ^2_a (1).

$$\sigma^2_s = 10{,}634(9.0546\mu^{1.3955} - 0.3494\mu^{1.7867}) \quad \text{(VI)} \quad \text{(VI)}$$

$$\sigma^2_{ss} = 1{,}456{,}000(0.3494\mu_s^{1.7867} - 0.0637\mu_s^{1.9339}) \quad \text{(VII)}$$

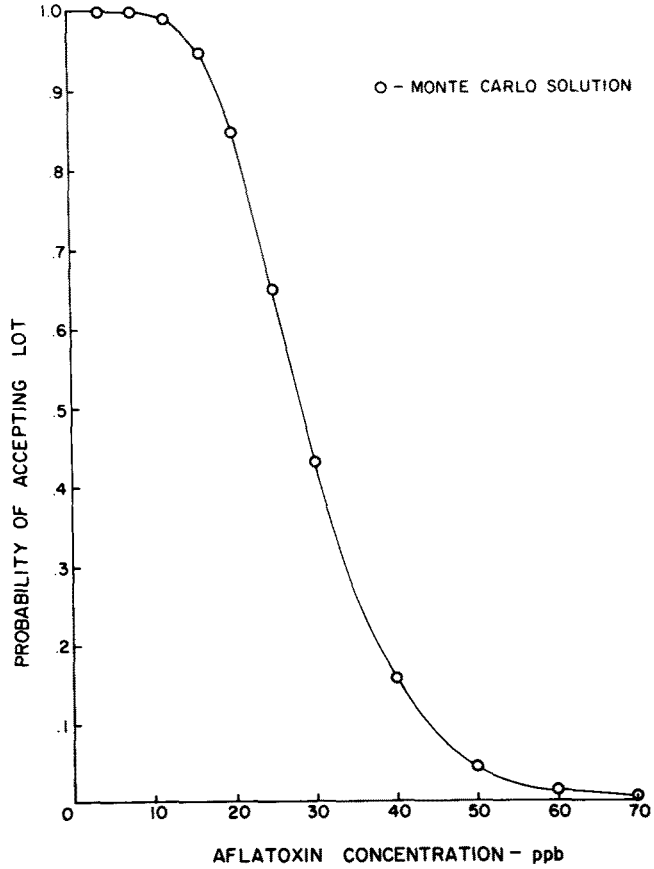

FIG. 4. Operating characteristic curve of the aflatoxin testing program used on the 1975 peanut crop.

$$\sigma^2_a = 0.0637\mu_{ss}^{1.9339} \quad \text{(VIII)}$$

Expressions for σ^2_s, σ^2_{ss}, and σ^2_a in Eqs. VI, VII, and VIII were evaluated for μ values up to ca. 200 parts per billion (ppb). Expressions for K_s and K_{ss} may be obtained by substitution of Eqs. VI and VII, respectively, for σ^2 in Eq. V. From Eq. VIII, the analytical variance $\sigma^2_{\bar{x}(a)}$ can be determined for a given subsample concentration μ_{ss}.

MONTE CARLO TECHNIQUE

With the Monte Carlo method, a random number generator (4-7) is used to simulate the random selection of a sample, subsample, or analysis. To simulate taking a sample from a contaminated lot with aflatoxin concentration μ, a random number, uniformly distributed between 0 and 1, is generated. This number is taken as the value of $F_s(N_s\bar{x}_s)$ in Eq. II for which the corresponding value of $N_s\bar{x}_s$ is determined. Then sample size N_s is specified, and the sample concentration $\bar{x}_s$ is computed.

To simulate drawing a subsample from the above sample with aflatoxin concentration $\bar{x}_s$, another random number, uniformly distributed between 0 and 1, is generated. This number is taken as the value of $F_{ss}(N_{ss}\bar{x}_{ss})$ in Eq. III for which the corresponding value of $N_{ss}\bar{x}_{ss}$ is determined ($\mu_s = \bar{x}_s$). Then the subsample size N_{ss} is specified, and the subsample concentration $\bar{x}_{ss}$ is computed.

To simulate a chemical analysis on the above subsample with aflatoxin concentration $\bar{x}_{ss}$, another random number, uniformly distributed between 0 and 1, is generated. This number is taken as the value of $F_a(\bar{x}_a)$ in Eq. IV for which the corresponding value of $\bar{x}_a$ is determined ($\mu_{ss} = \bar{x}_{ss}$).

ACCURACY OF MONTE CARLO

With the above procedure, the end result of one pass through the three distributions by use of Monte Carlo is one $\bar{x}_a$ value. Multiple $\bar{x}_a$ values are generated in order to estimate the distribution of $\bar{x}_a$ values given μ, N_a, N_{ss}, and N_s. The error between the Monte Carlo solution and the exact solution of the model is a function of the number of Monte Carlo trials N_G. Buslenko et al. (7) showed that the error is inversely proportional to the square root of N_G. To get a 10-fold increase in accuracy, a 100-fold increase in the number of trials is required. Generally a compromise is made between solution cost and solution error. Usually the number of trials N_G used in Monte Carlo simulation is between 10^3 and 10^5 (7).

The accuracy of the Monte Carlo solution also depends on the correct description of the distributions and parameters associated with each step of the inspection procedure.

The total variance of aflatoxin test results computed by the Monte Carlo technique was compared with the variance of replicated aflatoxin tests for lots of shelled peanuts when one analysis, one 280 g subsample, and one 12 pound sample were used for each test (1,2). The variance for the aflatoxin tests was estimated by regression analysis to be

$$s^2_{\bar{x}(t)} = 9.0546 \, \bar{\bar{x}}^{1.3955} \qquad \text{(IX)}$$

where $\bar{\bar{x}}$ is the estimated lot concentration.

The curve for Eq. IX and the data points for the Monte Carlo estimates are plotted in Figure 1. The variance estimates are based upon 2000 values. Agreement is reasonable between the data points and the regression curve, with differences attributed to both types of errors discussed above.

SIMULATION OF THE 1975 TESTING PROGRAM

The model described above was used to simulate the aflatoxin testing program adopted for the 1975 crop of peanuts (8). A diagram of the testing program is shown in Figure 2. A 48 pound sample of kernels is comminuted in a subsampling mill, and a 1100 g subsample is extracted for aflatoxin analysis. Two analyses, identified as 1A and 1B, are made on the extract. If the average of 1A and 1B is 16 ppb aflatoxin or less, the lot is accepted; if the average is $>$ 75 ppb, the lot is rejected. Otherwise, a second 48 pound sample is processed and two more analyses are made (2A and 2B). If the average of 1A, 1B, 2A, and 2B is 22 ppb or less, the lot is accepted; if the average is $>$38 ppb, the lot is rejected. Otherwise, a third 48 pound sample is processed and two more analyses are made (3A and 3B). If the average of 1A, 1B, 2A, and 2B, 3A, and 3B is 25 ppb or less, the lot is accepted; if the average is greater than 25 ppb, the lot is rejected.

A computer program was written to determine the probability of accepting a lot with a specified aflatoxin concentration μ under the aflatoxin testing program described above. The following values were specified: N_s = 48 pounds, N_{ss} = 1100 g, and N_a = 2. A flow chart describing the computer program is shown in Figure 3. Analytical results simulating the testing of 2000 lots with the same μ value were generated. The acceptance probability was then computed by dividing the number of lots accepted by 2000. The above procedure was repeated for various values of μ.

The probabilities of accepting a lot for μ values of 4, 8, 12, 16, 20, 25, 30, 40, 50, 60, and 70 ppb for the 1975 testing program are shown in Figure 4. A smooth curve, called an "operating characteristics (OC) curve," is drawn through the points to provide estimates of the acceptance probability for lot concentrations other than those mentioned. The area above the curve is related to the number of lots rejected, while the area below the curve is related to the number of lots accepted by the testing program (9). Assuming 25 ppb is the decision between "good" and "bad" lots, the area above the OC curve and below 25 ppb is related to the number of "good" lots rejected (processors' risk), and the area below the curve and above 25 ppb is related to the number of "bad" lots accepted (consumers' risk).

Other distributions may be used without changing the basic procedure for solution. The model coupled with the Monte Carlo solution technique would allow evaluation of complex aflatoxin testing programs in granular material which involve conditional probabilities generated by sequential sampling, subsampling, and analyses.

REFERENCES

1. Whitaker, T.B., J.W. Dickens, and R.J. Monroe, JAOCS 51:214 (1974).
2. Whitaker, T.B., J.W. Dickens, R.J. Monroe, and E.H. Wiser, Ibid. 49:590 (1972).
3. Quesenberry, C.P., T.B. Whitaker, and J.W. Dickens, Biometrics (In press).
4. Naylor, T.H., J.L. Balintfy, D.S. Curdeck, and Kong Chu, "Computer Simulation Techniques," John Wiley and Sons, New York, NY, 1967, pp. 43-67.
5. Hammersley, J.M., and D.C. Handscomb, "Monte Carlo Methods," Associated Book Publishers Ltd., London, England, 1967, pp. 25-75.
6. Tocher, K.D., "The Art of Simulation," England, The English University Press, London, England, 1963, pp. 6-71.
7. Buslenko, N.P., K.I. Golenko, Yu. A. Shrieder, I.M. Sobol, and V.G. Sragovich, "The Monte Carlo Method," Pergamon Press, London, England, 1967, pp. 1-90.
8. Peanut Administrative Committee, Marketing Agreement, 1975.
9. Whitaker, T.B., J.W. Dickens, and E.H. Wiser, JAOCS 47:501 (1970).

[Received January 8, 1976]