ON COMPUTER GENERATION OF LOGNORMAL RANDOM VARIABLES

by

P. A. Firey and C. L. Wood

Biometrics Unit, Cornell University, Ithaca, New York 14853

BU-624-M* August, 1977

Abstract

The necessary equations for generation of lognormal random variables of given mean and variance are given, as well as FORTRAN and PL/I subroutines for the calculations. Also discussed is the generation of lognormals with random location parameters, with suggested formulae and methods, and appropriate subroutines.

1. Generating Lognormal Variables

In simulations it is often necessary to generate lognormal random variables. This is most easily done by exponentiating normal random variables, obtained from any standard package of subroutines. (Recommended by the authors is GGUSN, of the IMSL subroutine package.) One appropriate question is what mean and variance should the normal variables, the generating variables, have to produce a given mean and variance for the lognormal, the target variables?

The appropriate equations for the mean, μ_t and variance σ_t^2 of the target lognormal are

---

*Paper No. BU-624-M in the Biometrics Unit Mimeograph Series, Cornell University, Ithaca, New York 14853.

(1) $$\mu_t = \exp(\mu_g + 1/2 \cdot \sigma_g^2) ,$$

and

(2) $$\sigma_t^2 = \exp(2 \cdot \mu_g) \cdot [\exp(2 \cdot \sigma_g^2) - \exp(\sigma_g^2)] ,$$

where μ_g and σ_g^2 are the mean and variance, respectively, of the generating normal.

The solution of this pair of equations in terms of the desired target parameters is

(3) $$\mu_g = \ln(\mu_t) - 1/2\left\{\ln\left[\left(\sigma_t^2/\mu_t^2\right) + 1\right]\right\} ,$$

and

(4) $$\sigma_g^2 = \ln\left[\left(\sigma_t^2/\mu_t^2\right) + 1\right] .$$

In the Appendix are FORTRAN and PL/I subroutines that can be easily added to a program to do the necessary calculations.

One common error in simulations is to think in terms of the underlying or generating normal parameters, as often the theory can be more easily expressed that way. Care must be taken, however, as reasonable generating parameters can produce unreasonable actual, or target lognormal distributions. Indeed, in many simulations, it is perhaps better to think in terms of the parameters of the lognormal, especially in cases where the simulation is mimicking an actual experiment. There what would be reportable, i.e., calculable, to the experimenter is indeed the target lognormal parameters, not the generating parameters. It is hoped that these subroutines would help in this regard, as one can use them as "black boxes" for generating lognormals with a given mean and variance.

2. Generation of Lognormals with Random Location Parameter

Sometimes in simulations it is also desirable to generate random variables of the form

$$(5) \qquad Y_{ij} = \exp(\mu_i + \epsilon_{ij}) \qquad i = 1, \cdots, a; \quad j = 1, \cdots, n_{ij}$$

where $\mu_i \sim NID(\bar{\mu}_g, \sigma^2_{g\alpha})$ and independent of $\epsilon_{ij} \sim NID(0, \sigma^2_{g\epsilon})$.

In other words, lognormals with a random treatment effect are desired. Care must be taken here that one generates exactly what one wants. There are four basic equations which define target parameters in terms of generating parameters. They are

$$(6) \qquad \bar{\mu}_t = \exp\left[\bar{\mu}_g + 1/2\left(\sigma^2_{g\alpha} + \sigma^2_{g\epsilon}\right)\right] ,$$

$$(7) \qquad \sigma^2_{t\alpha} = \exp(2\cdot\bar{\mu}_g)\left[\exp\left(2\cdot\sigma^2_{g\alpha}\right) - \exp\left(\sigma^2_{g\alpha}\right)\right] ,$$

$$(8) \qquad \sigma^2_{t\epsilon} = \exp(2\cdot\bar{\mu}_g)\left[\exp\left(2\cdot\sigma^2_{g\epsilon}\right) - \exp\left(\sigma^2_{g\epsilon}\right)\right] ,$$

and

$$(9) \qquad \sigma^2_t = \exp(2\cdot\bar{\mu}_g)\left\{\exp\left[2\left(\sigma^2_{g\alpha} + \sigma^2_{g\epsilon}\right)\right] - \exp\left(\sigma^2_{g\alpha} + \sigma^2_{g\epsilon}\right)\right\} ,$$

where

$\bar{\mu}_t$ is the mean of the target means; i.e., in a sampling framework, if one were to repeat the experiment (or simulation) many times the average value of the means of the treatments would be $\bar{\mu}_t$,

σ^2_t is the overall variance of the target random variables,

$\sigma^2_{t\alpha}$ is the variance for the target means,

$\sigma^2_{t\epsilon}$ is the within-treatment variance of the target lognormals, and

$\bar{\mu}_g$, $\sigma^2_{g\epsilon}$ and $\sigma^2_{g\alpha}$ are the mean of the generating means, the within-generating variance, and the variance of generating means, respectively, as defined in equation (5).

Any three equations will define the system completely; which three are chosen depends on what is desired in the final sample. The authors feel that $\bar{\mu}_t$ and $\sigma^2_{t\alpha}$ are specified in most situations. Then three cases are possible:

Case 1: Specify an overall target variance.

Case 2a: Specify a variance for within each treatment for the target random variables.

Case 2b: Specify an average within-variance, allowing the within-variance to vary such that the coefficient of variance remains constant.

Case 1. Here the appropriate equations are (6), (7) and (9) and the solutions are

(10) $$\bar{\mu}_g = \ln\left(\bar{\mu}_t \Big/ \sqrt{1 + \sigma^2_t/\bar{\mu}^2_t}\right) ,$$

(11) $$\sigma^2_{g\alpha} = \ln\left[\left(1 + \sqrt{1 + 4\left(\sigma^2_{t\alpha}/\bar{\mu}^2_t + \sigma^2_{t\alpha}\sigma^2_t/\bar{\mu}^4_t\right)}\right)\Big/ 2\right] ,$$

and

(12) $$\sigma^2_{g\epsilon} = \ln\left\{\left[2\left(1 + \sigma^2_t/\bar{\mu}^2_t\right)\right]\Big/\left[1 + \sqrt{1 + \left(\sigma^2_{t\alpha}/\bar{\mu}^2_t + \sigma^2_{t\alpha}\sigma^2_t/\bar{\mu}^4_t\right)}\right]\right\} .$$

Appropriate subroutines for calculating (10) – (12) are provided in the Appendix.

If one uses the same generating within-variance for each μ_i, then the C.V. will remain constant. In order for the solution to exist, the following constraint on choice of target parameters must hold:

(13) $$1 + \sigma^2_t/\bar{\mu}^2_t > \left[1 + \sqrt{1 + 4\left(\sigma^2_{t\alpha}/\bar{\mu}^2_t + \sigma^2_{t\alpha}\sigma^2_t/\bar{\mu}^4_t\right)}\right]\Big/ 2 .$$

The supplied subroutines check for this. If one does the calculations by hand, it will become obvious if one of the generating variances is negative.

Case 2a and Case 2b. The appropriate equations here are (6), (7) and (8) and the solutions are

(14) $$\bar{\mu}_g = \ln \bar{\mu}_t + 1/2\left\{\ln\left[\left(1 + \sigma^2_{t\alpha}/\bar{\mu}^2_t\right)\Big/\left(1 + \sigma^2_{t\epsilon}/\bar{\mu}^2_g\right) - \sigma^2_{t\alpha}/\bar{\mu}^2_t\right]\right.$$

$$+ \ln\left[\left(1 + \sigma^2_{t\epsilon}/\bar{\mu}^2_t\right)/\left(1 + \sigma^2_{t\alpha}/\bar{\mu}^2_t\right) - \sigma^2_{t\epsilon}/\bar{\mu}^2_t\right]\Big\} \ ,$$

(15) $$\sigma^2_{g\alpha} = -\ln\left[\left(1 + \sigma^2_{t\epsilon}/\bar{\mu}^2_t\right)/\left(1 + \sigma^2_{t\alpha}/\bar{\mu}^2_t\right) - \sigma^2_{t\epsilon}/\bar{\mu}^2_t\right] \ ,$$

and

(16) $$\sigma^2_{g\epsilon} = -\ln\left[\left(1 + \sigma^2_{t\alpha}/\bar{\mu}^2_t\right)/\left(1 + \sigma^2_{t\epsilon}/\bar{\mu}^2_t\right) - \sigma^2_{t\alpha}/\bar{\mu}^2_t\right] \ .$$

For Case 2a, solve the equations to obtain $\sigma^2_{g\alpha}$ and $\bar{\mu}_g$. Use these to generate the μ_i, and for each μ_i solve the below equation:

(17) $$\sigma^2_{g\epsilon_i} = \ln\left[\left(1 + \sqrt{1 + 4\exp(-2\mu_i)\sigma^2_{t\epsilon}}\right)/2\right] \ .$$

For Case 2b, one only needs to solve the equations once, and use the answers obtained throughout. This will produce the desired effect, namely a constant C.V. equal to

(18) $$\left[\sigma^2_{t\epsilon}/\mu^2_t\left(1 - \sigma^2_{t\alpha}/\mu^2_t\right)\right]/\left(1 - \sigma^2_{t\epsilon}\sigma^2_{t\alpha}/\bar{\mu}^4_t\right) \ .$$

The constraint on choice of the target parameters is

(19) $$\sigma^2_{t\epsilon}\sigma^2_{t\alpha}/\bar{\mu}^4_t < 1 \ .$$

Care must be taken in both cases that the overall variance is not too large. Substituting the obtained values into equation (9) will check for this. The variances are almost multiplicative, <u>not</u> additive.

3. Example

A case where the above might be used is where one wished to model and simulate the behavior of new varietal types of barley or other grains. The new varietals have means which are random, and there is reason to believe that these means would have a lognormal distribution. The errors are proportional to the means, so either Case 1 or Case 2b could be used, depending on what bounds ones wished to put on the simulation.

Appendix

```
LOGTR1:PROC(TARGET_MEAN,TARGET_VARIANCE,GENERATING_MEAN,GENERATING_VARIANCE):
       DCL(TARGET_MEAN,           /*SUPPLIED*/
          TARGET_VARIANCE,        /*SUPPLIED*/
          GENERATING_MEAN,        /*PASSED BACK*/
          GENERATING_VARIANCE     /*PASSED BACK*/)FLOAT DEC(16);
       GENERATING_VARIANCE=LOG(((TARGET_VARIANCE/(TARGET_MEAN*
          TARGET_MEAN))+1.0E00));
       GENERATING_MEAN=LOG(TARGET_MEAN)-.5E0*GENERATING_VARIANCE;
       END LOGTR1;

LOGTR2:PROC(TARGET_MEAN,TARGET_ALPHA_VAR,TARGET_OVERALL_VAR,GENERATING_MEAN,
       GENERATING_ALPHA_VAR,GENERATING_EPS_VAR);
       DCL(TARGET_MEAN            /*SUPPLIED*/,
          TARGET_ALPHA_VAR        /*SUPPLIED*/,
          TARGET_OVERALL_VAR      /*SUPPLIED*/,
          GENERATING_MEAN         /*PASSED BACK*/,
          GENERATING_ALPHA_VAR    /*PASSED BACK*/,
          GENERATING_EPS_VAR      /*PASSED BACK*/) FLOAT DEC(16);
       DCL(CVA,CVO,TEST)FLOAT DEC(16);
       CVA=TARGET_ALPHA_VAR/(TARGET_MEAN*TARGET_MEAN):
       CVO=TARGET_OVERALL_VAR/(TARGET_MEAN*TARGET_MEAN);
       GENERATING_ALPHA_VAR=LOG((.5E0*(1.E0+SQRT((1.E0+4.E0*(CVA+CVA*CVO))))));
       GENERATING_EPS_VAR=LOG((1.E0+CVO))-GENERATING_ALPHA_VAR;
       GENERATING_MEAN=LOG((TARGET_MEAN/SQRT((1.E0+CVO))));
       IF(GENERATING_EPS_VAR<0.E0) THEN PUT SKIP LIST(
       'GENERATING_EPS_VAR IS NEGATIVE-ATTEMPT TO CRASH PROGRAM');
       TEST=SQRT(GENERATING_EPS_VAR);
       END LOGTR2;

LOGTR3:PROC(TARGET_MEAN,TARGET_ALPHA_VAR,TARGET_EPS_VAR,GENERATING_MEAN,
       GENERATING_ALPHA_VAR,GENERATING_EPS_VAR);
       DCL(TARGET_MEAN            /*SUPPLIED*/,
          TARGET_ALPHA_VAR        /*SUPPLIED*/,
          TARGET_EPS_VAR          /*SUPPLIED*/,
          GENERATING_MEAN         /*PASSED BACK*/,
          GENERATING_ALPHA_VAR    /*PASSED BACK*/,
          GENERATING_EPS_VAR      /*PASSED BACK*/) FLOAT DEC(16);
       DCL(CVA,CVE,TEST,RATIO)FLOAT DEC(16);
       CVA=TARGET_ALPHA_VAR/(TARGET_MEAN*TARGET_MEAN);
       CVE=TARGET_EPS_VAR/(TARGET_MEAN*TARGET_MEAN);
       RATIO=(1.E0+CVE)/(1.0E0+CVA);
```

```
        GENERATING_ALPHA_VAR=-LOG((RATIO-CVE));
        GENERATING_EPS_VAR=-LOG(((1.EO/RATIO)-CVA));
        GENERATING_MEAN=LOG(TARGET_MEAN)-.5EO*(GENERATING_ALPHA_VAR+
           GENERATING_EPS_VAR);
        IF((GENERATING_ALPHA_VAR<0.EO)|(GENERATING_EPS_VAR<0.EO))  THEN
        PUT SKIP LIST(
           'ONE OR BOTH VARIANCES NEGATIVE-ATTEMPT CRASH');
        TEST=SQRT(GENERATING_ALPHA_VAR)+SQRT(GENERATING_EPS_VAR);
        END LOGTR3;

LOGTR4:PROC(GENERATING_MEAN,TARGET_EPS_VAR,GENERATING_EPS_VAR);
        DCL(GENERATING_MEAN          /*SUPPLIED*/,
           TARGET_EPS_VAR            /*SUPPLIED*/,
           GENERATING_EPS_VAR        /*PASSED BACK*/) FLOAT DEC(16);
        GENERATING_EPS_VAR=LOG((.5EO*(1.EO+SQRT((1.EO+4.EO*EXP((-2.EO*
           GENERATING_MEAN))*TARGET_EPS_VAR)))));
        END LOGTR4;

        SUBROUTINE LOGTR1(TRGTMU,TRGTVR,GENRMU,GENRVR)
C       TRGTMU-TARGET MEAN-SUPPLIED
C       TRGTVR-TARGET VARIANCE-SUPPLIED
C       GENRMU-GENERATING MEAN-PASSED BACK
C       GENRVR-GENERATING VARIANCE-PASSED BACK
        IMPLICIT REAL*8(A-H,)-Z)
        GENRVR=DLOG(((TRGTVR/(TRGTMU*TRGTMU))+1.ODO))
        GENRMU=DLOG(TRGTMU)-.5DO*GENRVR
        RETURN
        END

        SUBROUTINE LOGTR2(TRGTMU,TRGTVA,TRGTVO,GENRMU,GENRVA,GENRVE)
C       TRGTMU-TARGET MEAN-SUPPLIED
C       TRGTVA-TARGET EFFECTS VARIANCE-SUPPLIED
C       TRGTVO-OVERALL TARGET VARIANCE-SUPPLIED
C       GENRMU-GENERATING MEAN-PASSED BACK
C       GENRVA-GENERATING EFFECTS VARIANCE-PASSED BACK
C       GENRVE-GENERATING WITHIN VARIANCE-PASSED BACK
        IMPLICIT REAL*8(A-H,O-Z)
        CVA=TRGTVA/(TRGTMU*TRGTMU)
        CVO=TRGTVO/(TRGTMU*TRGTMU)
        GENRVA=DLOG((.5DO*(1.DO+DSQRT((1.DO+4.DO*(CVA+CVA*CVO))))))
        GENRVE=DLOG((1.ODO+CVO))-GENRVA
        GENRMU=DLOG((TRGTMU/DSQRT((1.ODO+CVO))))
        IF(GENRVE.LT.O.DO) WRITE(6,10)
     10 FORMAT(1H,51H***GENRVE IS NEGATIVE-SUBROUTINE WILL ATTEMPT CRASH)
        TEST=DSQRT(GENRVE)
        RETURN
        END

        SUBROUTINE LOGTR3(TRGTMU,TRGTVA,TRGTVE,GENRMU,GENRVA,GENRVE)
C       TRGTMU-TARGET MEAN-SUPPLIED
C       TRGTVA-TARGET EFFECTS VARIANCE-SUPPLIED
C       TRGTVE-TARGET WITHIN VARIANCE-SUPPLIED
C       GENRMU-GENERATING MEAN-PASSED BACK
```

```
C       GENRVA-GENERATING EFFECTS VARIANCE-PASSED BACK
C       GENRVE-GENERATING WITHIN VARIANCE-PASSED BACK
        IMPLICIT REAL*8(A-H,O-Z)
        CVA=TRGTVA/(TRGTMU*TRGTMU)
        CVE=TRGTVE/(TRGTMU*TRGTMU)
        RATIO=(1.0D0+CVE)/(1.0D0+CVA)
        GENRVA=-DLOG((RATIO-CVE))
        GENRVE=-DLOG(((1.0D0/RATIO)-CVA))
        GENRMU=DLOG(TRGTMU)-.5D0*(GENRVA+GENRVE)
        IF((GENRVE.LT.0.D0).OR.(GENRVE.LT.0.D0)) WRITE(6,10)
     10 FORMAT(1H,50H***WARNING-ONE OR BOTH VARIANCES NEGATIVE-ATTEMPT,
       *21HWILL BE MADE TO CRASH)
        TEST=DSQRT(GENRVE)+DSQRT(GENRMU)
        RETURN
        END

        SUBROUTINE LOGTR4(GENRMU,TRGTVE,GENRVE)
C       GENRMU-GENERATING MEAN-SUPPLIED
C       TRGTVE-TARGET WITHIN VARIANCE-SUPPLIED
C       GENRVE-GENERATING WITHIN VARIANCE-PASSED BACK
        IMPLICIT REAL*8(A-H,O-Z)
        GENRVE=DLOG((.5D0*(1+DSQRT((1.D0+4.D0*DEXP((-2.D0*GENRMU))*
       *TRGTVE)))))
        RETURN
        END
```