FOR OFFICIAL USE ONLY

Project Number 772

NEW YORK UNIVERSITY

College of Engineering
RESEARCH DIVISION
University Heights, New York 53, N. Y.

Mathematics Department

Contract No. DA18-108-405-CML-788
DA18-108-CML-6617

SEVENTH QUARTERLY PROGRESS REPORT

PROGRESS REPORT 526.20

CW Mathematical Research

COVERING THE PERIOD

1 November 1961 - 31 January 1962

Prepared by

M.L. Norden
L.H. Herbach

DATE: 13 February 1962

COPY NO. 21 OF 32 COPIES

FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

ABSTRACT CARD

AD________________ Accession No.________ **UNCLASSIFIED**

Mathematics Department, Research Division College of Engineering, New York University, New York 53, New York, CW MATHEMATICAL RESEARCH - Monroe L. Norden, Leon H. Herbach

Seventh Quarterly Progress Report (Progress Report 526.20, 13 February 1962, Abstract card + 56 pp. Contract DA18-108-405-CML-788/DA18-108-CML-6617. Order No. CP-0-405-13614, Unclassified Report.

1) A method is given for improving the mathematical analysis of the effect of the lethality vs. dose function on the expected number of casualties, and a better understanding obtained of when certain approximations apply. It depends on the introduction of a random variable for lethality.

2) The effect of the lethality vs. dose function on the variance of casualties is obtained. This variance is now expressed in terms of the joint distribution of men, the lethality distribution, and joint distributions of the total dosages ingested.

3) The possibility is considered of using a second Monte Carlo procedure to compute the A(D) curve from each sample or trial obtained by the first or basic Monte Carlo procedure. A series of test calculations to be computed is given.

1. Mathematics
2. Probability
3. Casualty Production
4. Mathematical Models
5. Monte Carlo Techniques

I. Monroe L. Norden
II. Leon H. Herbach
III. Contract DA18-108-405-CML-788 DA18-108-CML-6617

NEW YORK UNIVERSITY
COLLEGE OF ENGINEERING
RESEARCH DIVISION

PROGRESS REPORT No. 526.20

Period: 1 November 1961 - 31 January 1962

CW MATHEMATICAL RESEARCH

Prepared by

Monroe L. Norden

Leon H. Herbach

Contract No.
DA18-108-405-Cml-788
DA18-108-Cml-6617
Order No. CP-0-405-13614
Weapon Research Division
Directorate of Research
Chemical Corps Research and Development Laboratories
Army Chemical Center, Maryland

I. Casualty Production

a) The Expected Number of Casualties $E(C)$.

In Appendix A we outline the mathematical and computing steps involved in the calculation of $G(x,y)$ without derivation or motivation. This is sort of a mathematical flow chart for the calculation. The various inputs and approximations needed are also described. From G, we obtain $E(C)$ by the Equation

$$(1) \qquad E(C) = \int_{x,y\epsilon A} \omega(x,y)G(x,y)dxdy$$

If the expected density of munitions $\eta(u,v)$ is constant η , then $G(x,y)$ is a constant G and we have

$$(2) \qquad E(C) = G \int\int_{x,y\epsilon A} \omega(x,y)dxdy$$

In Appendix B, Notes on the Calculation of G.I, we give a series of notes on the details involved in reducing the indicated functions, integrations and other operations to numerical quantities. E.g., the functional inputs needed, approximations, possible assumptions, alternative computing techniques, recommendations, etc. It is conceived that Appendix A gives the basic mathematic structure of the calculation, while Appendix B gives decisions and details that have to be made in realizing this structure. Appendix A should not require much change, Appendix B will require expansion as more details are added.

In Appendix D we describe a method for improving or simplifying the mathematical analysis of the effect of lethality on $E(C)$. This also leads to much better understanding of what occurs mathematically when the distribution of

(3) $\qquad D_i = D(x_i,y_i)$ = Total Dose Ingested by a man in Δ_i if he is there.

is not lognormal, and hence of when the assumption of lognormality leads to a good approximation of G.

b) The Variance of Casualties, σ_C^2

In Appendix D, the effect of lethality is now included in the mathematical expression for σ_C^2 . In previous work σ_C^2 was obtained in terms of the joint distribution of men, and the joint conditional distribution of casualties at a point if a man is there. Only mean and covariance functions are needed. The inclusion of the effect of lethality is obtained by expressing the necessary mean and covariance functions of conditional casualties at two points in terms of the lethality functions or distribution, and the joint distribution of the D_i.

If the bivariate distribution of D_i and D_j is bivariate lognormal, then the necessary calculations become greatly simplified. Results are given for both this assumption and the general case. In the general case, a method is given for obtaining a bivariate distribution of D_i and D_j with given marginal additive distributions, particularly the gamma or c.d. The results require the calculation, for each i and j of the five parameters $\mu_{D_i}, \mu_{D_j}, \sigma_{D_i}, \sigma_{D_j}, \rho_{D_i, D_j}$.

Of course, these five parameters describing the joint distribution of D_i and D_j , involve the rest of the problem, i.e., the joint distributions of breathing rate, the munition contaminant pattern, and the munition impact points.

The simplification that occurs if $\eta(u,v)$, the expected density of munitions, is constant is also given in Appendix D.

II. Monte Carlo Simulation of A(D) Curves

Monte Carlo simulation, or calculation, of the distribution of casualties or the distribution of the A(D) curve is being considered. In such a simulation, a field trial is simulated on a computer by selecting observed values, by the use of random numbers, from the appropriate probability distributions. For each such field trial , the problem arises of computing the obtained A(D) . This can be done by systematic mathematical procedures, but those tried to date require much calculation. The alternative possibility of using a second Monte Carlo procedure to compute each field trial A(D) curve is under consideration and the work is described in Appendix C.

Leon H. Herbach

Leon H. Herbach
Project Director

APPENDIX A

Memo 526.03 M. L. Norden December, 1961

An Outline of the Calculation of G

I Introduction

In the Fourth and Fifth Quarterly Progress Report (Progress Reports 526.17 and 526.18) the expected number of casualties, E(C), was mathematically derived in terms of various known functions and parameters. The motivations and derivations are given in these progress reports. We now wish to work toward the numerical calculation of E(C) . In this memo we shall outline the mathematical and computing steps involved, (without derivation or motivation) and describe the various inputs and approximations involved.

In general we wish to compute

$$(1) \qquad E(C) = \int_{x,y \in A}\int \omega(x,y)\,dxdy$$

where $\omega(x,y)$, the expected density of men, is given and $G(x,y)$, the probability that a man at (x,y) would become a casualty, depends in a complex way on the remainder of the problem.

In this memorandum, we shall assume that the munition density $\eta(u,v)$ is constant. This is, we shall assume

$$(2) \qquad \eta(u,v) = \eta$$

This assumption makes $G(x,y) = G$ = a constant. Hence

$$(3) \qquad E(C) = G \iint_{x,y \in A} \omega(x,y)\,dx\,dy$$

and we now wish to describe the calculation of G. At a latter time it will be desirable to not assume (2). Then $G(x,y)$ will have to be separately calculated at the different points (x,y), but the steps will be intrinsically the same as those given in this memorandum.

In this memorandum, we shall outline the mathematical and computing steps and details involved in the calculation of G, so that the choices involved in programming the numerical calculation of G can be examined, and the necessary flexibility of possible inputs to the calculation determined.

In Memo 526.04* we shall give Notes, which will contain more details, e.g. the functional inputs indicated, approximations, assumptions, alternative computing techniques, etc.

II The Calculation of G

G is defined as

$$(4) \qquad G = \int_0^\infty L(D)f(D)\,dD = \int_0^\infty [1-F(D)]\,\frac{dL(D)}{dD}\,dD$$

* Appended as Appendix B

where

(5) $$L(D) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\ell n D-\mu_L)/\sigma_L} e^{-y^2/2}\,dy$$ = Log Dosage vs Probit lethality

(6) $F(D)$ = Probability that the dosage at a point (x,y) is $\leq D$

$$f(D) = \text{Probability density} = \frac{dF(D)}{dD}$$

The probability distribution $F(D)$ will be mathematically specified in terms of its mean μ_D and variance σ_D^2. (Notes A, B,C,D,E).

We now assume the special breathing rate function model of the progress reports.

The quantities μ_D and σ_D^2 are then given by

(7) $$\mu_D = \mu_r \eta \int_{t=0}^{\infty} [1-H(t)]\, R(t)dt$$

(8) $$\mu_D^2 + \sigma_D^2 = 2(\sigma_r^2 + \mu_r^2) \int_{t=0}^{\infty} \int_{s=0}^{t} [1-H(t)]\,[\eta^2 R(t)R(s)+\upsilon R(t,s)]dsdt$$

where μ_L, σ_L, μ_r, σ_r, η and υ are known constants and

(9) $$\upsilon = \begin{cases} \eta & \text{if the munitions are placed at random.} \\ 0 & \text{if the munitions are at fixed points.} \end{cases}$$

(10) $H(t) = \text{Prob}\,[T \leq t]$.

$H(t)$ will be a known probability distribution, possibly a three parameter log normal. (Notes F,G,H) .

R(t) and R(t,s) can be defined by the following sequence of equations

$$(11) \qquad R(t) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \Psi(u,v,t)\,dv\,du$$

$$(12) \qquad R(t,s) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \Psi(u,v,t)\,\Psi(u,v,s)\,dv\,du$$

$$(13) \qquad \Psi(u,v,t) = \int_{\tau=0}^{t} \chi(u,v,\tau)\,dQ(t-\tau) = \int_{\tau=0}^{t} \chi(u,v,t-\tau)\,dQ(\tau)$$

$$(14) \qquad Q(t) = 1-ae^{-bt}$$

[Note that $Q(0) = 1-a$]

Hence

$$(15) \qquad \Psi(u,v,t) = (1-a)\chi(u,v,t) + ab \int_{\tau=0}^{t} \chi(u,v,t-\tau)e^{-b\tau}\,d\tau$$

$$= (1-a)\,\chi(u,v,t) + abe^{-bt} \int_{\tau=0}^{t} \chi(u,v,\tau)e^{+b\tau}\,d\tau$$

and finally

$$(16)\qquad \chi(x,y,t) = \frac{\lambda Q_0 e^{-\left[\frac{(x-\bar{u}t)^2}{(t^\beta + \alpha_x)k_x} + \frac{y^2}{(t^\beta + \alpha_y)k_y}\right]}}{\sqrt{\pi^3 k_x k_y k_z (t^\beta + \alpha_x)(t^\beta + \alpha_y)(t^\beta + \alpha_z)}}$$

where $\lambda, Q_0, \bar{u}, \beta, \alpha_x, \alpha_y, \alpha_z, k_x, k_y, k_z$ are given constants.

In this work, we have let t, s, x, y, u, v range symbolically to ∞ or $\pm\infty$. It may be necessary to truncate. Numerically noticable effects on G should physically not arise from large values of these variables.

The steps do not have to performed in this order. An alternative sequence is the following.

Combining the steps, we obtain for R(t)

$$(17)\qquad R(t) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \int_{\tau=0}^{t} \chi(u,v,\tau)\,dQ(t-\tau)\,dv\,du$$

Interchanging the order of integration, we obtain

$$(18) \qquad R(t) = \int_{\tau=0}^{t} \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \chi(u,v,\tau)\,dv\,du\; dQ(t-\tau)$$

$$= \int_{\tau=0}^{t} K(\tau)\,dQ(t-\tau) = \int_{\tau=0}^{t} K(t-\tau)\,dQ(\tau)$$

Substituting Equation (4), we obtain

$$(19) \qquad R(t) = (1-a)\,K(t) + ab \int_{\tau=0}^{t} K(t-\tau)e^{-b\tau}\,d\tau$$

$$= (1-a)\,K(t) + abe^{-bt} \int_{\tau=0}^{t} K(\tau)e^{+b\tau}\,d\tau$$

Where

$$(20) \qquad K(t) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \chi(u,v,t)\,du\,dv = \frac{\lambda Q_0}{\sqrt{\pi k_z(t^{\beta}+\alpha_z)}}$$

Hence

(21) $$R(t) = \frac{\lambda Q_0}{\sqrt{\pi K_z}} \left[\frac{1-a}{\sqrt{t^{\beta} + \alpha_z}} + ab \int_{\tau=0}^{t} \frac{e^{-b\tau}}{\sqrt{(t-\tau)^{\beta} + \alpha_z}} d\tau \right]$$

$$= \frac{\lambda Q_0}{\sqrt{\pi K_z}} \left[\frac{1-a}{\sqrt{t^{\beta} + \alpha_z}} + abe^{-bt} \int_{\tau=0}^{t} \frac{e^{+b\tau}}{\sqrt{\tau^{\beta} + \alpha_z}} d\tau \right]$$

Similarly for $R(t,s)$

(22) $$R(t,s) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \int_{\tau_1=0}^{t} \int_{\tau_2=0}^{s} \chi(u,v,\tau_1)\chi(u,v,\tau_2)dQ(s-\tau_2) \cdot$$

$$\cdot \, dQ(t-\tau_1) \, dvdu$$

$$= \int_{\tau_1=0}^{t} \int_{\tau_2=0}^{s} K(\tau_1,\tau_2)dQ(s-\tau_2)dQ(t-\tau_1)$$

where

Hence

$$(21) \qquad R(t) = \frac{\lambda Q_0}{\sqrt{\pi K_z}} \left[\frac{1-a}{\sqrt{t^\beta + \alpha_z}} + ab \int_{\tau=0}^{t} \frac{e^{-b\tau}}{\sqrt{(t-\tau)^\beta + \alpha_z}} \, d\tau \right]$$

$$= \frac{\lambda Q_0}{\sqrt{\pi K_z}} \left[\frac{1-a}{\sqrt{t^\beta + \alpha_z}} + abe^{-bt} \int_{\tau=0}^{t} \frac{e^{+b\tau}}{\sqrt{\tau^\beta + \alpha_z}} \, d\tau \right]$$

Similarly for $R(t,s)$

$$(22) \qquad R(t,s) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \int_{\tau_1=0}^{t} \int_{\tau_2=0}^{s} \chi(u,v,\tau_1)\chi(u,v,\tau_2) dQ(s-\tau_2) \cdot$$

$$\cdot \, dQ(t-\tau_1) \, dvdu$$

$$= \int_{\tau_1=0}^{t} \int_{\tau_2=0}^{s} K(\tau_1,\tau_2) dQ(s-\tau_2) dQ(t-\tau_1)$$

where

$$(23) \qquad K(t,s) = \int_{u=-\infty}^{+\infty} \int_{v=-\infty}^{+\infty} \chi(u,v,t)\, \chi(u,v,s)\, dv du$$

$$= \frac{\lambda^2 Q_0^2\, e^{-\frac{\bar{u}^2 (t-s)^2}{k_x (t^\beta + s^\beta + 2\alpha_x)}}}{\pi^2 \sqrt{k_x k_y k_z^2 (t^\beta + s^\beta + 2\alpha_x)(t^\beta + s^\beta + 2\alpha_y)(t^\beta + \alpha_z)(s^\beta + \alpha_z)}}$$

Again, because of the jump in $Q(t)$ at $t=0$, we have from Equation (22) and (4)

$$(24) \qquad R(t,s) = (1-a)^2 K(t,s) + (1-a)\, ab\, e^{-bt} \int_{\tau_1=0}^{t} K(\tau_1,s) e^{+b\tau_1}\, d\tau_1$$

$$+ (1-a)\, ab\, e^{-bs} \int_{\tau_2=0}^{s} K(t,\tau_2)\, e^{b\tau_2}\, d\tau_2$$

$$+ (ab)^2\, e^{-b(t+s)} \int_{\tau_1=0}^{t} \int_{\tau_2=0}^{s} K(\tau_1,\tau_2)\, e^{+b(\tau_1+\tau_2)}\, d\tau_2, d\tau_1$$

Memo 526.04 M.L. Norden December, 1961

Notes on the Calculation of G. I

All equations numbered in this memo have a letter and a number. Equations without a letter refer to Appendix A, above.

Note A : The Various Functional Forms for F(D)

Re : Equations (4) and (6) of Appendix A

F(D), fitted to μ_D and σ_D^2 will be.

(A-1) The Lognormal.(usual two parameter)

(A-2) The Scaled Gamma

(A-3) The Scaled C.D

(A-4) Etc.

These various alternatives shall be available from the final program.

Note B: The Mathematical Relationship between (μ_1, σ_1) and (μ_D, σ_D)

Re : Equations (4) and (6) and Note (A-1)

If F(D) is log normal, that is, if

$$F(D) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\ell nD-\mu_1)/\sigma_1} e^{-y^2/2} dy \tag{B-1}$$

then we have

$$\text{(B-2)} \quad \begin{aligned} \mu_1 &= E(\ell nD) = \ell n\mu_D - 1/2\, \ell n(1 + v_D^2) \\ \sigma_1^2 &= \sigma^2(\ell nD) = \ell n(1 + v_D^2) \end{aligned}$$

where

(B-3) $$v_D^2 = \sigma_D^2 / \mu_D^2$$

Note C : The Simplification if F(D) is Lognormal

Re : Equations (4) and (6) and Notes A and B.

If F(D) is log normal, then Equation (4) becomes

$$G = \frac{1}{\sqrt{2\pi}} \int_{h_1}^{\infty} e^{-x^2/2}\, dx$$

where

$$h_1 = \frac{\mu_L - \mu_1}{\sqrt{\sigma_L^2 + \sigma_1^2}}$$

and we have eliminated one troublesome integration.

Note D: The Relationship between (μ_L, σ_L) and More Usual Probit Lethality Notation

Re : Equation (5)

In the usual Probit terminology

$$\text{Probit} = a + b \ln D = 5 + \ln(D/D_{.50})$$

In this notation

(D-1) $$\mu_L = \ln(D_{.50}) = \frac{5 - a}{b}$$

$$\sigma_L^2 = \frac{1}{b^2} \quad ; \quad \sigma_L = 1/b$$

Note E: Integrating Equation (4)

Various suggestions for integrating Equation (4) when F(D) is not log normal are given in the Fourth Quarterly Progress Report [1] pages 19 to 23. In particular Technique 2 seems appealing.

Note F: Properties of H(T)

Re: Equations (7), (8) and (10)

Present data indicates that H(T) can be described as follows

$H(\theta) = \text{Prob}\ [T \leq \theta]$

(F-1) $\quad T = T_A + T_s$

That is, the random variable T is the sum of T_A the alarm time and T_s the time to stop intake of the contaminant. Then if we let

(F-2) $\quad T_r$ = reaction time = time to the stopping of breathing

(F-3) $\quad T_m$ = masking time

it was found that, particularly at high work rates, masking can occur before breathing stops. Hence

(F-4) $T_s = \min [T_r ; T_m]$

It was also found that the probability distribution of T_r can be approximated by a three parameter log normal distribution, the three parameters being a, μ_2 and σ_2. That is, the random variable $T_r + a$ is distributed in a log normal distribution with parameters.

$$\mu_2 = E [\ln(T_r + a)] \qquad \text{(F-5)}$$

$$\sigma_2^2 = \sigma^2 [\ln(T_r + a)]$$

Yes, this allows the random variable T_r to take on negative values. This is meant to correspond to the alarm being sounded while exhaling.

The probability distribution of T_r is also approximated by a log normal distribution. Here the choices seem to be

(a) A two parameter log normal distribution

(b) A three parameter log normal distribution

(c) A three parameter log normal distribution with the parameter "a" having the same value as for the distribution of T_r.

Note G : The Calculation of H(T)

Re. Equations (7), (8) and (10) and Note F

In Note F we defined

$$(G-1) \qquad T_s = \min [T_r \,;\, T_m]$$

The probability of T_s can be computed in two ways. The <u>first</u> or exact solution is as follows. Let $H_s(t)$, $H_r(t)$ and $H_m(t)$ be the probability distributions of T_s, T_r and T_m respectively. Then

$$(G-2) \qquad 1 - H_s(t) = [1 - H_r(t)]\,[1 - H_m(t)]$$

and so the correct answer can be obtained. Note that if $H_r(t)$ and <u>$H_m(t)$ are log normal, then $H_s(t)$ is not exactly log normal.</u>

However, we do feel that masking rarely occurs before breathing stops (the exact probabilities can be obtained from the data), and hence should have on a slight effect on $H_s(t)$. So perhaps some approximation can be given which is easier to compute. Hence the <u>second</u> solution would be to approximate the probability distribution of T_s presumably by a two or three parameter log normal distribution. That is we would try to approximate

$$(G-3) \qquad E(\ln T_s) \text{ and } \sigma^2(\ln T_s)$$

or

$$(G-4) \qquad E[\ln (T_s + a)] \text{ and } \sigma^2[\ln (T_s + a)]$$

from the parameters of T_s and T_m.

We could then use these values to approximate a log normal distribution. Such a technique has been derived.

Determination of whether the approximation or the exact solution are preferable for calculation needs to be made.

Note H : Re : Equations (7), (8), (9) and Notes F and G

A decision of what to do about the probability distribution of alarm time T_A still needs to be made. If it is assumed to be a constant, this value needs to be entered as a parameter.

Note I : Input Controls

Re : Equations (1) to (10)

We wish to be able to request these alternatives or switches from the program. (Position 1 is standard, i.e. our present best guess)

(a) The Distribution of F(D)

1. Log Normal
2. Scaled Gamma
3. Scaled C.D.
4. Etc.

(b) Breathing Rate

1. As Given
2. $\sigma_r = 0$ (constant breathing rate r)
3. $\sigma_{T_s} = 0$ (constant stopping time T_s)

4. $\sigma_r = 0$, $\sigma_{T_s} = 0$ (constant breathing function)

5. $H_m = 0$ (Neglect masking time)

(c) Drop of Munitions

1. $\upsilon = \eta$ Random Drop

2. $\upsilon = 0$ Fixed, known impact points

APPENDIX C

Memo 526.05 M.L.N./L.H. January 1962

Monte Carlo Estimation of the A(D) curve, I

This memo is a preliminary report on some of the thinking of NYU and CRDL personnel on the problem of determining the distribution of the A(D) curve by Monte Carlo techniques, where A(D) is the area covered by a given dose or dosage.

Conceptually we have a field trial where munitions impact, the contaminant spreads from the munitions, it is ingested (say by breathing) and a dose ingested occurs at each point of the target.* Hence, A(D), the area covered by a dose D, can be obtained for the trial. If another trial is run under "identical conditions" the munitions may impact at different points, etc., and another A(D) curve will be obtained. If very many trials are run, at least conceptually, we will obtain many trial A(D) curves, and these can be described by a probability distribution of A(D).

For given mathematical probability distributions or exact descriptions of munition impacts, contaminant dissemination patterns, breathing rates (and positions of men and lethality functions if we want casualties), we can proceed in two ways.

First a mathematical description of the probability distribution of A(D) can be obtained, and work is proceeding along these lines.

---

* If ingestion is neglected we have dosage, if the distribution of men and a lethality distribution are added as inputs, we will have casualties.

The second method is to model the trials, using Monte Carlo procedures, on a computer, hence obtaining the equivalent of a number of trials and hence of a number of trial values of $A(D)$, which can then be used to estimate the probability distribution of $A(D)$. This is done by selecting random numbers from the appropriate probability distribution. For example, if seven munitions are dropped on one aiming point with a known probability distribution of impact, then we can select seven random impact points from this distrubution and use these seven points in a computer trial. Similarly all other values for the trial are selected randomly, and we obtain a mathematically defined and computable $A(D)$ curve for the trial. This process is repeated until a sufficient number of trial $A(D)$ curves are obtained. Let us call this the first (or basic) Monte Carlo procedure.

At this point, the question arises of how to compute the $A(D)$ curve ~~for a given trial, and this is the topic of the remainder of this~~ memorandum.

This calculation has been tried by a systematic procedure. The possibility of using another Monte Carlo procedure to obtain the $A(D)$ for a given trial, is now being examined, a sort of second Monte Carlo procedure within a first Monte Carlo procedure.

The systematic procedure used breaks the target rectangle up into small areas of size $\Delta x \Delta y$ and computes $A(D)$ by counting the number of such areas with a central dose $\geq D$. Here the error is a "mathematical error" and depends on Δx and Δy. By "error" we mean the difference between the correct $A(D)$ curve and the $A(D)$ curve calculated. For given numerical values of D, Δx, Δy and the starting points of the interval, the error is a corresponding number, say $e(D)$. We call this number, $e(D)$ the "mathematical error". However, for this procedure, the number of calculations also depends inversely on Δx and Δy. So we find that if Δx and Δy are large, the number of calculations needed can be small, but the "mathematical error" will be large. Similarly, if Δx and Δy are made small, the mathematical error can be made small, but at the expense of a large number of machine calculations.

One alternative is to try to develop other way of systematic numerical calculation of $A(D)$. Here we can either, (a) try to improve the accuracy of the above procedure or, (b) attempt to develop essentially different systematic procedure. As an example of (a) , the above procedure computes $D(x_i, y_j)$ at a set of points $x_i = i\Delta x$, $y_j = j\Delta y$. If we look at the subareas of which these are the centers, it is clear that some improvement around D_{max} [which occurs at (0.0)] might be obtained by using the points $x_i = u + i\Delta_x$, $y_j = v + j\Delta_y$. In particular the values $u = \Delta_{x/2}$, $v = o$ or $u = \Delta_{x/2}$, $v = \Delta_{y/2}$ seem indicated. As an example of (b) , we might consider trying to compute or approximate the contour in the (x,y) plane corresponding to a given D, and then numerically integrate to obtain the area within it. This has been tried using simple interpolation to obtain the contour, and gave a larger error. Possibly the use of higher order interpolation, or the <u>fitting</u> of appropriate curves to the contours (say bivariate polynomicals) will give better results.

A second alternative is to use a second Monte Carlo procedure to compute $A(D)$. Here the number of calculations n and the size of Δx and Δy are unrelated. n is usually called the sample size, and for any Δx and Δy we can select any sample size n we wish. However, we now have a "mathematical error" depending on Δx and Δy and a statistical error depending on r . That is, the second Monte Carlo procedure results in an $A(D)$ curve computed from the n values of D obtained. However, unlike the systematic procedure, if the second Monte Carlo process or calculation is repeated on a computer, a different set of random numbers will be selected, hence a different set of n values of D, and hence a different A(D) curve will be obtained. That is, the A(D) obtained is now a random variable. For any D, A(D), while having an average or expected value E[A(D)] and a probability distribution about E[A(D)] which can be described in terms of σ^2[A(D)], the standard deviation of A(D). E[A(D)] is, for any choice of Δx and Δy, the value that would be calculated by the above systematic procedure based on the same Δx and Δy. Hence the difference between E[A(D)] and the correct value of A(D) is the mathematical error described above (called the "bias" in statistical terminology). Note that it does not depend on n. The "statistical" error of the A(D) calculated is described in terms of σ^2[A(D)] and depends on n, not on Δx and Δy.

In fact, since the random numbers used to select the points (x,y) are eight digit decimals, it turns out that it may be most convenient to use a Δx and Δy of 0.00000001, which is very small and results in what can only be called a very negligible mathematical error. The real error of the second Monte Carlo procedure is just the statistical one.

It should be pointed out that even for the same impact points, etc., the second Monte Carlo procedure (sampling of observation points) will introduce variability in the A(D) curve. However, for any set of chosen Δ's this can be reduced by sampling more points. The main problem is to determine just how many points to select.

It is deemed preferable to concentrate on the second part of the problem (estimating the A(D) curve for given drop data) before going to the first part (estimating average A(D) curve and its variability for a given tactical situation). In the language of stochastic processes, this second problem involves treating a single realization of the dosage field, rather than the ensemble of fields. The variability obtained in this phase will have to be added to the variability due to changing impact points, etc. (first phase of the problem).

To explore the possibilities of using such a second type of Monte Carlo procedure, the following machine calculations were agreed upon.

1. <u>The single munition pattern</u>

The area covered by $D/Q = \theta$ is to be obtained, where

$$\theta = \frac{D}{Q} = \frac{2}{\pi C_y C_z \bar{u}\, x^{2-n}}\, e^{-\frac{y^2}{C_y^2 x^{2-n}}} \tag{1}$$

And

$C_y = 0.211$ meters $^{n/2}$

$C_z = 0.135$ meters $^{n/2}$

$\bar{u} = 2$ meters/second

$n = \frac{1}{4} = 0.25$

D is in units of gram seconds/(meter)^3

Q is in units of grams.

Hence

θ is in units of seconds/(meter)^3

A total area going from $x = 0$ to 250 meters in the x coordinate and from -18 to +18 meters in the y coordinate is considered [9,000 sq. meters].

<u>Note</u>: The exact A(D) curve for this pattern can be obtained mathematically and will be used for comparison with the numerical calculations to be described.

2. Cutoffs and class intervals

2.1 Upper and lower dosage cutoffs* will be used. The details are being examined.

2.2 The values of θ are then tabled into 100 class intervals giving $A(\theta)$ at 100 values of θ. The selection of these class intervals is being examined in more detail. One point is that the lowest interval seems to define a second kind of zero**.

---

* Upper and lower dosage cutoffs must be used. Equation (1) implies that dosage is positive for all x and y and can easily give values greater than the machine can handle. Dosages above a lethal value can be reduced to the lethal value without any loss. It is, of course, also necessary to cut down the size of the dosage field so that it can fit in the computer. This has the effect of truncating at the low dosage values.

** The first kind of zero occurs when the dosage field is "boxed" in by a finite "target". This second kind occurs when values are left in because of the shape of the contours. See Figure 1, where we suppose that θ = .005 is the minimum contour. The points outside the rectangle represent the first kind of zero, and x represents the second kind, which consists of the points between the bounding rectangle and the minimum contour.

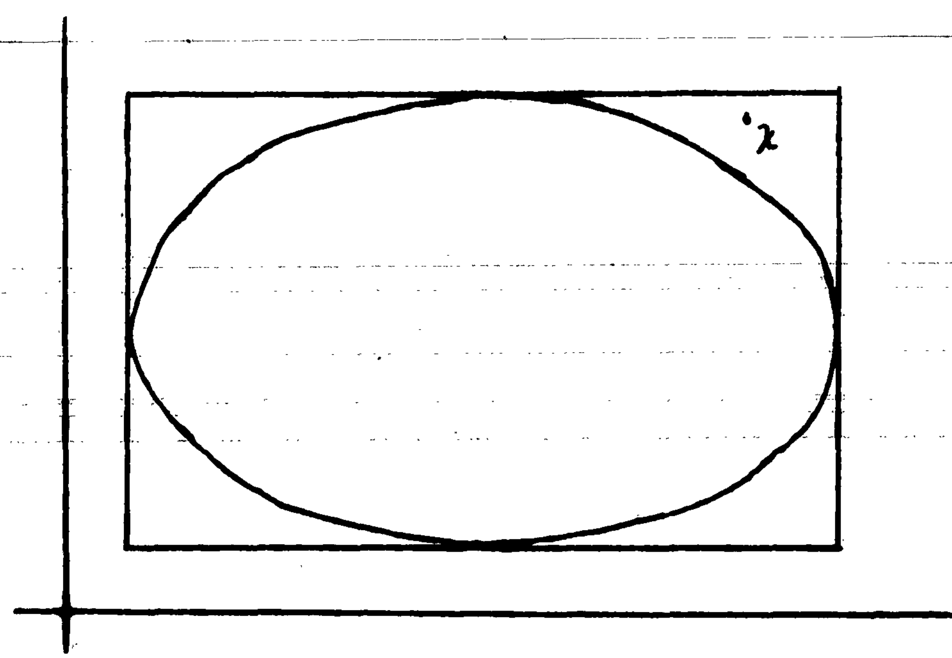

Figure 1. θ = .0005 contour. "Minimum" contour.

A second point is that logarithmic intervals in θ would be preferable. For example, 100 logarithmic intervals from 5 to 500 would make each division about 1.047 times previous division. If we make the ratio 1.05 (that is 5% increments) then 100 logarithmic intervals gives a ratio of about 132 from the first to the 100th interval*.

3. Calculations of $A(\theta)$ by use of a systematic grid have been performed. It was decided to

3.1 Repeat the systematic calculations for the single munition for the two subareas

a) $\Delta x = 50$, $\Delta y = 2$ and

b) $\Delta x = 1$, $\Delta y = .5$.

3.2 Perform Monte Carlo calculations for $A(\theta)$ for a single munition as follows:

3.2.1 Eight digit random numbers were to be used making $\Delta x = \Delta y = 0.00000001$.

3.2.2 6400 ~~hundred~~ points were to be computed in the following manner. 16 runs of 400 points each were to be made, and the resulting $\tilde{A}(\theta)$ tabulated in the 100 class intervals for each run. The successive 4 runs were to be tabulated, giving 4 runs of 1600 points each. The total, 1 run of 6400 points, is also to be tabulated.

3.2.3 Each of these estimated $\tilde{A}(\theta)$ (16 from 400 point, 4 from 1600 points, 1 from 6400 points) is to be compared, presumably graphically, with the exact $A(\theta)$ which is known in the single munition situation.

* Since $\theta = D/Q$, corresponding changes in log θ are the same as corresponding changes in log D. However, a change in D is Q times the corresponding change in θ.

3.2.4 The curves $A(\theta) \pm \sigma[\tilde{A}(\theta)]$ and $A(\theta) \pm 2\sigma[\tilde{A}(\theta)]$ are also to be computed and presented for comparison.

3.3 Perform Monte Carlo calculations similar to those made in Section 3.2 for the <u>two</u> fixed munitions situation.

APPENDIX D

The Concept of a Random Variable for Lethality, its Effect on $E(C)$, and the Inclusion of the Effect of Lethality on $\sigma^2(C)$.

TABLE OF CONTENTS of Memo 526.06

Memo 526.06 M.L.N. February 1962

The Concept of a Random Variable for Lethality, its Effect on $E(C)$, and the Inclusion of the Effect of Lethality on $\sigma^2(C)$.

I. Introduction

In Progress Reports 526.17, 526.18, 526.19 [1] [2] [3], a mathematical model of the structure of casualty production was given, the expected number of casualties $E(C)$ derived, and work toward σ_C^2 the variance of casualties described. In this memo we shall describe further improvements in the mathematical analysis of the effect of lethality on $E(C)$ and continue the work of [3] to include the effect of lethality on σ_C^2.

In Equation (25) of Progress Report 526.17, the probit versus log dose lethality function was defined as

$$(1) \qquad L(D) = \frac{1}{\sqrt{2\pi}\,\sigma_L} \int_{-\infty}^{\ln D} \exp\{-(x-\mu_L)^2/(2\sigma_L^2)\}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\ln D-\mu_L)/\sigma_L} e^{-y^2/2}\, dy$$

where

(2) $L(D)$ = probability that a man who has ingested a total dose D will be a casualty.

and

(3) μ_L = logarithm of the 50% lethal dose.

(4) σ_L = the reciprocal of the slope of the $L(D)$ versus D curve when plotted on log normal or log probit probability paper.

Then, in Equation (3) of Progress Report 526.19, we defined the random variable*

$$(5)\quad k_1 = \begin{cases} 0 \text{ if a man, assuming he in } \Delta_1\text{, is \underline{not} a casualty} \\ 1 \text{ if a man, assuming he is in } \Delta_1\text{, is a casualty} \end{cases}$$

Then we defined

$$(6)\quad G_1 = G(x_1, y_1) = \text{Prob}\ (k_1 = 1) = E(k_1)$$

and

$$(7)\quad \begin{cases} F_1(\theta) = \text{probability that the random variable } D_1 \text{ is} \leq \text{the number } \theta. \\ f_1(\theta) = \dfrac{dF_1(\theta)}{d\theta} \end{cases}$$

* The k_1 defined in equation (13) and used on page 11 of Progress Report 526.17 is not the same but rather corresponds to the conditioned random variable $k_1(D_1)$ given that the dose ingested in Δ_1 is D_1. The k_1 used on page 12 of 526.17 is the unconditional random variable, as we now define k_1.

Then on pages 11 and 12 of 526.17, we showed that

$$(8) \quad G_i = \int_0^\infty L(D) f_i(D) dD = \int_0^\infty L(D) dF_i(D)$$

$$= \int_0^\infty [1-F_i(D)] \, dL(D) = \int_0^\infty [1-F_i(D)] \frac{dL(D)}{dD} dD$$

and pages 15 to 24 are devoted to a discussion of methods of computing G_i.

II. The Introduction of a Random Variable, h_i, for Lethality

The introduction of a random variable for the lethality considerably simplifies the proof of some of the results, and leads to the next step after P.R. 526.19, the inclusion of lethality into the calculation of σ_C^2. The idea is as follows. Let*

(9) h_i = a random variable called lethality.

That is, h_i is a random variable with a cumulative probability distribution function $L(D)$ or $L(\theta)$. That is, were

(10) $\text{Prob}\,[h_i \leq \theta] = L(\theta)$

---

* Again a slight notational inconsistency arises. The symbol h_i was used in P.R. 526.17 for the constant a_i to be defined in Equation (18) of this report.

and the h_i are statistically independant of each other and of all the other random variables in the casualty model.

Then, if we define k_i as

$$(11) \qquad k_i = \begin{cases} 0 & \text{if } h_i > D_i \\ 1 & \text{if } h_i \leq D_i \end{cases}$$

the model is just as before. We have just introduced a symbolism for the lethality random variable rather than just its probability distribution $L(\theta)$.

III. <u>The Use of This New Random Variable for the Calculation of G_i</u>

The conditional expectation of k_i given D_i is

$$(12) \qquad E(k_i|D_i) = \text{Prob } (k_i = 1|D_i) = \text{Prob } (h_i \leq D_i|D_i) = L(D_i)$$

and the unconditional expectation is

$$(13) \qquad G_i = E(k_i) = E[L(D_i)] = \int_0^\infty L(D)dF_i(D)$$

as given by Equation (8).

However, this formulation also gives us other results. We can write

$$(14) \quad G_i = E(k_i) = \text{Prob}\ (h_i \leq D_i) = \iint\limits_{h \leq D} F_i(D)\ \frac{dL(h)}{dh}\ dDdh$$

Now the distribution $L(\theta)$ of h_i was assumed to be probit vs. log dose or what is commonly called, log normal. That is, the distribution of $\ln h_i$ will be a normal distribution. In particular, from Equation (1)

$$(15) \quad E(\ln h_i) = \mu_L$$

$$\sigma^2(\ln h_i) = \sigma_L^2$$

This also implies that h_i is non negative, that is, it lies in the range 0 to $+\infty$. D_i, by its very definition, is also never negative. Hence we can also write

$$(16) \quad G_i = E(k_i) = \text{Prob}\ (h_i \leq D_i) = \text{Prob}\ (\ln h_i \leq \ln D_i)$$

$$= \text{Prob}\ [0 \leq \ln D_i - \ln h_i] = \text{Prob}\ [0 \leq z_i]$$

where we have defined

$$(17) \quad z_i = \ln D_i - \ln h_i$$

Now let

(18) $E(\ln D_i) = \mu_i$

$\sigma^2(\ln D_i) = \sigma_i^2$

Then

(19) $E(z_i) = \mu_i - \mu_L$

$\sigma^2(z_i) = \sigma_i^2 - \sigma_L^2$

IV. <u>The Assumption that the Probability Distribution of D_i is Lognormal</u>

Now if D_i also has a lognormal distribution, then, since the difference between two normally distributed random variables is also normally distributed, $\ln D_i - \ln h_i$ is distributed in a normal distribution. Hence

(20) $G_i = \text{Prob}\,[0 \leq z_i] = \frac{1}{2\pi} \int_{a_i}^{\infty} e^{-x^2/2} dx$

$a_i = (\mu_L - \mu_i) / \quad \sigma_L^2 + \sigma_i^2$

a result that took quite a bit of mathematical manipulation to obtain in Progress Report 526.17.

This result greatly simplifies the calculation of G_i if D_i is lognormally distributed. However, Equation (16) also allows us to have a better understanding of what is happening if D_i is not so distributed and even to understand when Equation (20) is a good approximation.

V. Calculation of G_i when D_i is not Log Normal, and the Effect of Approximating by a Log Normal Distribution

If the probability distribution of D_i is not lognormal, we can obtain the distribution of z_i as the distribution of the difference between $\ln D_i$ and $\ln h_i$, and then integrate to obtain G_i by Equation (16). The result can be placed in the form given by Equation (45) of P.R. 526.17[1].

Another approach is described in Technique 5 on page 23 of P.R. 526.17. That is to expand the distribution of $\ln D_i$ in a Gram Charlier Series. Here there are two points. The first, as pointed out in 526.17, is that we feel that the distribution of $\ln D_i$ is "approximately" normal. Surely it is much "closer" to a normal distribution than is the distribution of D_i.

However, we now see a second point. Since z_i is the difference between two random variables, one of which is normally distributed, the distribution of z_i will be much "closer" to a normal distribution than is the distribution of $\ln D_i$. We can see this by looking at the r'th cumulants of the random variables.

For a normally distributed random variable, all the cumulants above the second are zero. That is

(21) $K_1 = \mu$

$K_2 = \sigma^2$

$K_r = 0 \quad \text{if} \quad r \geq 3$

This defines a normal distribution.

Standardized (or unitless) cumulants α_r are defined as

(22) $$\alpha_r = \frac{K_r}{\sigma^r} = \frac{K_r}{(K_2)^{r/2}}$$

The closeness of α_r to zero (for $r \geq 3$) is the appropriate criteria for closeness to normality.

In this notation we have

(23) $K_1(\ln h_i) = \mu_L$

$K_2(\ln h_i) = \sigma_L^2$

$K_r(\ln h_i) = 0 \quad \text{if } r \geq 3$

(24) $K_1(\ln D_i) = \mu_i$

$K_2(\ln D_i) = \sigma_i^2$

$K_r(\ln D_i)$ as given by the distribution

Hence,

$$(25) \qquad \alpha_r(\ln D_i) = \frac{K_r(\ln D_i)}{\sigma_i^r} \qquad \text{if } r \geq 2$$

Now, by the additivity of cumulants of the sum (or difference) between independent random variables, we have

$$(26) \qquad K_r(\ln z_i) = K_r(\ln D_i - \ln h_i) = K_r(\ln D_i) + K_r(-\ln h_i)$$

Hence

$$(27) \qquad K_1(\ln z_i) = \mu_i - \mu_L$$

$$K_2(\ln z_i) = \sigma_i^2 + \sigma_L^2$$

$$K_r(\ln z_i) = K_r(\ln D_i) = \sigma_i^r \alpha_r(\ln D_i) \qquad \text{if } r \geq 3$$

And so, if $r \geq 3$

$$(28) \qquad \alpha_r(\ln z_i) = \frac{K_r(\ln z_i)}{[K_2(\ln z_i)]^{r/2}} = \left[\frac{\sigma_i^2}{\sigma_i^2 + \sigma_L^2}\right]^{r/2} \alpha_r(\ln D_i)$$

$$= \frac{\alpha_r(\ln D_i)}{(1 + \sigma_L^2/\sigma_i^2)^{r/2}}$$

Clearly $\alpha_r(\ln z_i)$ is less than $\alpha_r(\ln D_i)$.

For example, if $\sigma_L = \sigma_1$, then

(29) $$\alpha_3(\ln z_1) = \frac{\alpha_3(\ln D_1)}{2\sqrt{2} = 2.82}$$

$$\alpha_4(\ln z_1) = \frac{\alpha_4(\ln D_1)}{4}$$

If $\sigma_L = 2\sigma_1$, then

(30) $$\alpha_3(\ln z_1) = \frac{\alpha_3(\ln D_1)}{5\sqrt{5} = 11.2} \quad ; \quad \alpha_4(\ln z_1) = \frac{\alpha_4(\ln D_1)}{125}$$

We are tempted to guess that if $\sigma_L \geq \sigma_1$ and $\alpha_3(\ln D_1)$ is not too large, that G_1 is quite satisfactorily approximated by Equation (20), that is, by assuming the z is normally distributed.

Of course, a more detailed examination of the various approximations to G_1 when the distribution of D_1 is Gamma or C.D., would be desirable.

VI. The Inclusion of Lethality in the Calculation of $\sigma^2(c)$.

The Basic Equation

In the last progress report, Progress Report 526.19, the variance of casualties, σ_c^2 was obtained in terms of the joint distribution of men and the joint (conditional) distribution of casualties at a point (x,y) if a man is there. Essentially, only mean and covariance functions of the two joint distributions are needed.

If the target is broken into small subareas Δ_i; the joint distribution of men was described by the quantities

$$(31)\quad \begin{cases} E(m_i) = \overline{m} = \omega_i \Delta_i \\ \sigma(m_i : m_j) = E(m_i m_j) - \overline{m}_i \overline{m}_j \\ P_m(x_i, y_i;\ x_j, y_j) = \text{Prob}\ [m_i = 1;\ m_j = 1] = E(m_i m_j) \end{cases}$$

and the joint distribution of the k_j was described by the quantities

$$(32)\quad \begin{cases} E(k_i) = \overline{k}_i = G_i \\ \sigma(k_i : k_j) = E(k_i k_j) - G_i G_j \\ P_k(x_i, y_i;\ x_j, y_j) = \text{Prob}\ [k_i = 1;\ k_j = 1] = E(k_i k_j) \end{cases}$$

A continuous representation was also obtained, and for this the functions $\omega(x_1 y)$, $G(x_1 y)$, $K_k(x_1, y_1; x_2, y_2)$, etc. introduced, where

$$(33)\quad \begin{cases} \omega(x_i, y_i) = \omega_i \\ G(x_i, y_i) = G_i \\ K_k(x_i, y_i; x_j, y_j) = \sigma(k_i : k_j) \\ R_1(x_i, y_i; x_j, y_j)\sqrt{\Delta_i \Delta_j} + W(x_i, y_i; x_j, y_j)\,\Delta_i \Delta_j = \sigma(m_i : m_j) \end{cases}$$

For both the discrete and continuous representations σ_c^2 was obtained in terms of the above quantities. A general solution was obtained and a number of special cases given. At various times in the argument, the derivations in terms of the bivariate probability functions P_m and P_k seem surer, and these were used to obtain $E(C^2)$. From which, of course, we could obtain σ_c^2 by the equation

$$\sigma_C^2 = E(C^2) - [E(C)]^2 \qquad (34)$$

Table 1 lists these results.

TABLE 1

Equation Numbers of Results in Progress Report 526.19 Giving σ_c^2 or $E(C)^2$ in terms of the Probability Distributions of Men and Conditional Casualties Under Various Assumptions

	σ_C^2		$E(C)^2$	
	Discrete	Continuous	Discrete	Continuous
General	(17)(18)	(69)	(26)	(53)
m_i (men) are statistically independent	(30)(34)	(35)(36)(37)	(59)	(60)
Men placed with no error	(41)(45)	(42)(45)	(47)	(47)
m_i and k_i are statistically independent		(39)		

In this work the random variables k_i depend on the lethality, the breathing rate, the munition contaminant pattern and the munition impact points, and hence summarize everything but the placement of men.

The joint distribution of the k_i enters into σ_C^2 through the mean G_i or $G(x_i y_i)$ and the covariances $\sigma(k_i:k_j)$ or $K_k(x_1,y_1;x_2,y_2)$ or equivalent functions. The calculation of G has been described in the previous progress reports. Let us now obtain the covariances in terms of the distribution of lethality h_i and the dose ingested D_i.

Now

$$(35) \qquad K_k(x_i,y_i;x_j,y_j) = \sigma(k_i:k_j) = E(k_ik_j) - G_iG_j$$

$$= P_k(x_i,y_i;x_j,y_j) - G_iG_j$$

and we see that it will be sufficient to obtain

$$(36) \qquad E(k_ik_j) = \text{Prob }(k_i=1;k_j=1) = P_k(x_i,y_i;x_j,y_j)$$

Now, from Equation (11), we have

$$(37) \qquad E(k_ik_j) = \text{Prob }(k_i=1;k_j=1) = \text{Prob}(D_i \geq h_i;D_j \geq h_j)$$

By the nonnegativity of the h_i and D_i, and the definition of z_i, Equation (16), this becomes

$$(38) \qquad E(k_i k_j) = \text{Prob}\,[D_i - h_i \geq 0;\ D_j - h_j \geq 0]$$

$$= \text{Prob}\,[z_i \geq 0;\ z_j \geq 0]$$

$$= \text{Prob}\,[z(x_i, y_i) \geq 0;\ z(x_j, y_j) \geq 0]$$

These are the basic equations from which $E(k_i k_j)$ is to be obtained.

VII. $E(k_i k_j)$ If the Bivariate Distributions of D_i and D_j are Bivariate Log Normal.

Now if the individual probability distributions of the D_i are log normal, then the individual distributions of the z_i are normal. If we make the slight additional assumption that $\ln D_i$ and $\ln D_j$ have a bivariate normal distribution, then z_i and z_j will also be so distributed. (Remember that the h_i were assumed to be statistically independent of each other and the other random variables, but the D_i were correlated with each other. Substituting Equation (38) into the standard bivariate normal probability density function, we obtain

(39)
$$E(k_i k_j) = \frac{\displaystyle\int_{x=0}^{\infty}\int_{y=0}^{\infty} e^{-\frac{1}{2(1-\rho^2_{z_i z_j})}\left[\left(\frac{x-\mu_{z_i}}{\sigma_{z_i}}\right)^2 - \frac{2\rho_{z_i z_j}(x-\mu_{z_i})(x-\mu_{z_j})}{\sigma_{z_i}\sigma_{z_j}} + \left(\frac{y-\mu_{z_j}}{\sigma_{z_j}}\right)^2\right]} dxdy}{2\pi\, \sigma_{z_i}\sigma_{z_j}\sqrt{1-\rho^2_{z_i z_j}}}$$

$$= \frac{\displaystyle\int_{t_1 = \frac{-\mu_{z_i}}{\sigma_{z_i}}}^{\infty}\int_{t_2 = \frac{-\mu_{z_j}}{\sigma_{z_j}}}^{\infty} e^{-(t_1^2 - 2\rho_{z_i z_j} t_1 t_2 + t_2^2)/2(1-{}^2_{z_i z_j})} dt_1 dt_2}{2\pi\sqrt{1-\rho^2_{z_i z_j}}}$$

$$= M(-\mu_{z_i}/\sigma_{z_i},\ -\mu_{z_j}/\sigma_{z_j}, \rho_{z_i z_j})$$

Where $M(h,k;r)$ defined as in [5], [6] and [7] is

(40) $$M(h,k;r) = \frac{1}{2\pi\sqrt{1-r^2}} \int_h^{\infty}\int_k^{\infty} \exp\left[-(x^2-2rxy+y^2)/2(1-r^2)\right]dxdy$$

and

(41) $$\mu_{z_i} = E(z_i) = E(\ln D_i - \ln h_i) = \mu_i - \mu_L$$

(42) $$\sigma^2_{z_i} = \sigma^2_i + \sigma^2_L$$

Now to obtain

(43) $$\rho_{z_i,z_j} = \frac{\sigma_{z_i,z_j}}{\sigma_{z_i}\sigma_{z_j}}$$

Note that

(44) $$\sigma_{z_i,z_j} = \sigma(\ln D_i - \ln h_i;\ \ln D_j - \ln h_j)$$

If we define

(45) $$\sigma_{i,j} = \sigma(\ln D_i, \ln D_j)$$

Then

(46) $$\sigma_{z_i,z_j} = \begin{cases} \sigma(\ln D_i,\ \ln D_j) = \sigma_{i,j} & \text{if } i \neq j \\ \sigma^2(\ln D_i - \ln h_i) = \sigma^2_i + \sigma^2_L & \text{if } i = j \end{cases}$$

Hence

(47) $$\rho_{z_i,z_j} = \begin{cases} \dfrac{\sigma_{i,j}}{\sqrt{(\sigma^2_i+\sigma^2_L)(\sigma^2_j+\sigma^2_L)}} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Substituting these in Equation (39) we obtain

$$(48)\qquad E(k_i k_j) = \begin{cases} M\left[(\mu_L-\mu_i)/\sqrt{\sigma_i^2+\sigma_L^2};\ (\mu_L-\mu_j)/\sqrt{\sigma_j^2+\sigma_L^2};\ \sigma_{i,j}/\sqrt{(\sigma_i^2+\sigma_L^2)(\sigma_j^2+\sigma_L^2)}\right] & \text{if } i \neq j \\ G_i & \text{if } i = j \end{cases}$$

If we also define

$$(49)\qquad \rho_{i,j} = (\ln D_i;\ln D_j) = \sigma_{i,j}/\sigma_i\,\sigma_j$$

Then

$$(50)\qquad \rho_{z_i z_j} = \begin{cases} \dfrac{\sigma_i\sigma_j\rho_{i,j}}{\sqrt{(\sigma_i^2+\sigma_L^2)(\sigma_j^2+\sigma_L^2)}} = \dfrac{\rho_{i,j}}{\sqrt{\left(1+\dfrac{\sigma_L^2}{\sigma_i^2}\right)\left(1+\dfrac{\sigma_L^2}{\sigma_j^2}\right)}} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

This result is very useful and convenient because the function $M(h,k;r)$ has been extensively tabled and it is believed that adequate computer programs and approximations are available for its calculation.

One of the first tables of $M(h,k,\ )$ is given by Karl Pearsen on pages 78 - 137 of [4]. The most extensive table is given in [8] where $M(h,k,r)$

is called $L(h,k;r)$. Extensive bibliographies are given in [6] and [8]. A discussion of computing techniques is given in [4][5][6][7] and [8].

VIII. A Simplification if $\eta(u,v)$, the Expected Density of Munitions, is Constant.

If η_i or $\eta(u,v)$ is constant then the calculation of σ_C^2 or $E(k_i k_j)$ is greatly simplified in two ways; in the calculation of σ_C^2 and in the calculation of M. We can see this as follows.

The distribution of D_i or $D(x,y)$ will now be the same at every point on the target. The correlation coefficients ρ_{D_i,D_j} or the co-variances $\sigma(D_i;D_j)$ will now only depend on the difference between the two points (x_i,y_i), (x_j,y_j), that is on $x_i - x_j$ and $y_i - y_j$. Hence the $\sigma(k_i:k_j)$ will have the same property. That is, the functions

$$K_k(x_i,y_i;x_j,y_j) = \sigma(k_i:k_j)$$

and

$$P_k(x_i,y_i;x_j,y_j) = E(k_i k_j)$$

will be functions of the two quantities x_i-x_j, y_i-y_j not the four quantities x_i,x_j,y_i,y_j. This simplifies the calculation of σ_C^2 or $E(k_i k_j)$ given in Table 1.

In addition, <u>if we assume</u> that the distributions of the z_i are normal, the M calculation that now enters into the calculation of each $E(k_i k_j)$ is also simplified. That is, if $\eta(u,v)$ is constant, we have for $\ln D_i$

$$\mu_i = \mu_j = \mu_1 \qquad \text{for all } i \text{ and } j \tag{51}$$

$$\sigma_i = \sigma_j = \sigma_1 \qquad \text{for all } i \text{ and } j$$

and similarly for z_i

$$\mu_{z_i} = \mu_z = \mu_1 - \mu_L \qquad \text{for all } i \tag{52}$$

$$\sigma^2_{z_i} = \sigma^2_z = \sigma^2_L + \sigma^2_1 \qquad \text{for all } i$$

Then, if the distribution of z_i and z_i are bivariate normal, we have

$$E(k_i k_j) = M(b,b,r) \tag{53}$$

where

$$b = \frac{-\mu_z}{\sigma_z} = \frac{\mu_L - \mu_1}{\sqrt{\sigma^2_L + \sigma^2_1}} \tag{54}$$

and

$$(55) \qquad r = \frac{\sigma_{i,j}}{\sigma_z^2} = \frac{\rho_{i,j}}{1 + \frac{\sigma_L^2}{\sigma_1^2}}$$

Of course, $\sigma_{i,j} = \sigma(\ln D_i : \ln D_j)$ and $\rho_{i,j}$ will be functions of the two quantities $x_i - x_j$ and $y_i - y_j$, not the four x_i, x_j, y_i, y_j.

Now the calculation of $M(b,b;r)$ or $M(h,h;r)$ is even simpler than the calculation of $M(h,h;\rho)$ and considerable simplification occurs.

For example, from Equation 5 of [5], we obtain

$$(56) \qquad M(h,h,r) = 2M(h,0; -\sqrt{(1-r/2)})$$

From the Equations on page 7-8 of [6], we obtain, for $h \geq 0$

$$(57) \qquad M(h,h,r) = 2T(h,\infty) - 2T\left(h, \sqrt{\frac{1-r}{1+r}}\right)$$

Where

$$(58) \qquad T(h,a) = \frac{1}{2\pi} \int_0^a \frac{e^{-\frac{h^2}{2}(1+x^2)}}{1+x^2}\, dx$$

For negative h, we use the result

$$(59) \qquad M(-h,-h,r) = M(h,h,r) + \frac{1}{2\pi} \int_{-h}^{+h} e^{-t^2/2}\, dt$$

These are typical of the kinds of simplifications available. However the M function can be expressed in other ways, and other formulations of $M(h,h,r)$ may be preferable for calculations.

Because of this great simplification available, we strongly feel that calculations of σ_c^2 should, at least at first, be based on the above results, assuming a bivariate lognormal distribution of D_i and D_j.

In the next section we shall give results for general bivariate distributions of D_i and D_j. This can then be used to check the degree of approximation to σ_c^2 obtained by the use of the lognormal assumption.

IX. $E(k_i k_j)$ in the General Case, in Terms of the Bivariate Distribution of D_i and D_j.

If the distributions of the D_i are not multivariate Log Normal, then $E(k_i k_j)$ is much more difficult to obtain. We shall give a general result, but, of course, will be thinking of numerical calculations based on the gamma, c.d. or similar distributions. Note also that $E(k_i k_j)$ only depends on the bivariate distribution of D_i and D_j. Let us introduce the following notation for this bivariate distribution.

(60) $$F_{i,j}(\theta_1;\theta_2) = \text{Prob}[D_i \le \theta_1;\ D_j \le \theta_2]$$

(61) $$f_{i,j}(\theta_1,\theta_2) = \frac{d^2F_{i,j}(\theta_1;\theta_2)}{d\theta_1 d\theta_2}$$

(62) $$U_{i,j}(\theta_1,\theta_2) = \text{Prob}\ [D_i > \theta_1;\ D_j > \theta_2]$$

$$= 1-F_i(\theta_1) - F_j(\theta_2) + F_{i,j}(\theta_1;\theta_2)$$

Then we can obtain $E(k_ik_j)$ in two ways. The first result is obtained by looking at the conditional expectation of k_ik_j, holding D_i and D_j fixed. By equation (37) and (10), we obtain

(63) $$E(k_i;k_j \mid D_i,D_j) = L(D_i)L(D_j)$$

Hence the unconditional expectation is

(64) $$E(k_i,k_j) = \int_{D_i}\int_{D_j} L(D_i)L(D_j)f_{i,j}(D_i,D_j)dD_idD_j$$

$$= \int_{\theta_1}\int_{\theta_2} L(\theta_1)L(\theta_2)f_{i,j}(\theta_1,\theta_2)d\theta_1 d\theta_2$$

The second result is obtained by first holding h_i and h_j fixed. Then the conditional expectation of $k_i k_j$ is, by Equation (37) and (62)

$$(65) \qquad E(k_i k_j | h_i, h_j) = U_{i,j}(h_i, h_j) = 1 - F_i(h_i) - F_j(h_j) + F_{i,j}(h_i, h_j)$$

the last step by Equation (62).

Then the unconditional expectation is

$$(66) \qquad E(k_i k_j) = \int_{h_i} \int_{h_j} U_{i,j}(h_i, h_j) dL(h_i) dL(h_j)$$

$$= \int_{\theta_1} \int_{\theta_2} U_{i,j}(\theta_1, \theta_2) dL(\theta_1) dL(\theta_2)$$

These two results depend on expressing the bivariate distribution of D_i and D_j in terms of the functions $F_{i,j}(\theta_1, \theta_2)$, $f_{i,j}(\theta_1, \theta_2)$ or $K_{i,j}(\theta_1, \theta_2)$. These must now be obtained if these results are to be useful.

X. A Method for Obtaining Bivariate Distributions of D_i and D_j That Have Given Additive Marginal Distributions.

Now if the distribution of D_i is assumed to be gamma, or c.d., there is no unique bivariate distribution with a gamma (or c.d.) distribution for the marginal distributions. However, the bivariate gamma and bivariate c.d. that have been suggested have been produced by the same technique. This technique also produces the bivariate normal if the marginal are normal, and will work if the distribution used is additive (e.g., the sum of the two gamma (or c.d.) random variables is gamma (or c.d.)).

The technique is as follows. Let

$$
\begin{aligned}
D_i &= S_1(x_1 + x_3) \\
D_j &= S_2(x_2 + x_3)
\end{aligned}
\qquad (67)
$$

where the random variables x_1, x_2 and x_3 are statistically independent and have the appropriate underlying distribution, i.e. gamma, c.d., normal, etc. If the distributions of x_1, x_2 and x_3 are gamma (c.d.) with parameters α_1, α_2 and α_3 (a_1, a_2, a_3) respectively, then, by the additive properties of the gamma (c.d.) distribution, the distribution of D_i will be scaled gamma (c.d.) with scale factor S_1 and parameter $\alpha_1 + \alpha_3$ $(a_1 + a_3)$. Similarly, the distribution of D_j will be scaled gamma (c.d.) with scale factor S_2 and parameter $\alpha_2 + \alpha_3$ $(a_2 + a_3)$. And, of course, D_i and D_j are correlated because of the common random variable x_3. There are five parameters involved, $S_1, S_2, \alpha_1, \alpha_2$ and α_3 $(S_1, S_2, a_1, a_2, a_3)$ and these can be estimated from μ_{D_i}, μ_{D_j}, σ_{D_i}, σ_{D_j}, and ρ_{D_i,D_j}. Such a procedure is described in [9] for the c.d. distribution and is the basis for the bivariate gamma used in [10].

Now let Ψ_i be the underlying distribution of x_i , that is let

(68) $\Psi_i(t) = \text{Prob}[x_i \leq t]$

where Ψ is the cumulative gamma, or c.d., or whatever we are using, probability distribution, and let

(69) $\phi_i(t) = \dfrac{d\Psi_i(t)}{dt}$

Now by Equations (60) and (67)

(70) $$F_{i,j}(\theta_1,\theta_2) = \text{Prob}(D_i \le \theta_1; D_j \le \theta_2)$$

$$= \text{Prob}[S_1(x_1 + x_3) \le \theta_1;\ S_2(x_2 + x_3) \le \theta_2]$$

Now the condition probability holding x_3 fixed is

(71) $$\text{Prob}\ [S_1(x_1 + x_3) \le \theta_1;\ S_2(x_2 + x_3) \le \theta_2 | x_3]$$

$$= \text{Prob}\ [x_1 \le \frac{\theta_1}{S_1} - x_3;\ x_2 \le \frac{\theta_2}{S_2} - x_3 | x_3]$$

$$= \Psi_1(\frac{\theta_1}{S_1} - x_3)\ \Psi_2(\frac{\theta_2}{S_2} - x_3)$$

the last step by Equation (68) and the independence of x_1, x_2, and x_3.

Hence the unconditional probability is

(72) $$F_{i,j}(\theta_1,\theta_2) = \int_{x_3} \Psi_1(\frac{\theta_1}{S_1} - x_3)\ \Psi_2(\frac{\theta_2}{S_2} - x_3)\ d\Psi_3(x_3)$$

By a similar argument we obtain

(73) $$U_{i,j}(\theta_1,\theta_2) = \text{Prob}(D_i > \theta_1;\ D_j > \theta_2)$$

$$= \int_t [1 - \Psi_1(\frac{\theta_1}{S_1} - t)][1 - \Psi_2(\frac{\theta_2}{S_2} - t)]\ d\Psi_3(t)$$

Differentiating $F_{i,j}$ as given by Equation (72) we obtain

$$(74) \qquad f_{i,j}(\theta_1,\theta_2) = \frac{1}{S_1S_2} \int_t \phi_1\left(\frac{\theta_1}{S_1} - t\right) \phi_2\left(\frac{\theta_2}{S_2} - t\right) \phi_3(t)\, dt$$

We can now obtain the final expressions for $E(k_ik_j)$. Substituting Equation (74) into (64) we obtain

$$(75) \qquad E(k_ik_j) = \frac{1}{S_1S_2} \int_{\theta_1} \int_{\theta_2} \int_t L(\theta_1)L(\theta_2)\phi_1\left(\frac{\theta_1}{S_1} - t\right)\phi_2\left(\frac{\theta_2}{S_2} - t\right)\phi_3(t)\,dt\,d\theta_1\,d\theta_2$$

Substituting Equation (73) into (66), we obtain

$$(76) \qquad E(k_ik_j) = \int_{\theta_1} \int_{\theta_2} \int_t \left[1-\Psi_1\left(\frac{\theta_1}{S_1} - t\right)\right]\left[1-\Psi_2\left(\frac{\theta_2}{S_2} - t\right)\right] d\Psi_3(t)\,dL(\theta_1)\,dL(\theta_2)$$

References

[1] 4th Quar. Prog. Report 526.17, 24 May 1961.

[2] 5th Quar. Prog. Report 526.18, 14 August 1961.

[3] 6th Quar. Prog. Report 526.19, 15 November 1961.

[4] Pearson, Karl (editor) Tables for Statisticians and Biometricians, Part II, Cambridge University Press, England, 1931 (Tables VIII and IX)

[5] Owen, D. B. "Tables for Computing Bivariate Normal Probabilities", Annals of Mathematical Statistics, Volume 27, 1956, pages 1075-1090

[6] Owen, D. B. The Bivariate Normal Probability Distribution, Research Report SC-3831(TR) Sandia Corp., Albuquerque, N.M., March 1957.

[7] Owen, D. B. and J. M. Wiesen "A Method of Computing Bivariate Normal Probabilities" The Bell System Technical Journal, Volume 38, 1959 pages 1-20.

[8] Tables of the Bivariate Normal Distribution Function and Related Functions, National Bureau of Standards, Washington, D.C., June 1959.

[9] Max A. Woodbury and Monroe L. Norden. Technical Report 526.02 CW MATHEMATICAL RESEARCH Tasks 1 and 3 Multiple Source Coverage Problems, Constant Density Distribution and Monte Carlo Techniques. 20 March 1960. Contract DA18-108-Cml-6404/DA18-108-405-Cml-171. Research Division, New York University, New York 53, N.Y. (Confidential)

[10] Knut A. Krieger, etc. Final Report, Casualty Rate Assessment, Project Cerami. 31 August 1959. Contract DA18-108-Cml-5956/DA18-108-405-Cml-1-50. The Institute for Cooperative Research, University of Pennsylvania, Philadelphia, Penn. (Confidential)