ARMY RESEARCH LABORATORY

Fitting Expectation to Identify Plausible Explanations

by Andrew A. Thompson

ARL-TN-320 July 2008

NOTICES

Disclaimers

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

ARL-TN-320 **July 2008**

Fitting Expectation to Identify Plausible Explanations

Andrew A. Thompson
Weapons and Materials Research Directorate, ARL

REPORT DOCUMENTATION PAGE						*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* July 2008	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)** October 2007–March 2008

4. TITLE AND SUBTITLE Fitting Expectation to Identify Plausible Explanations	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Andrew A. Thompson	**5d. PROJECT NUMBER** AH80
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory ATTN: AMSRD-ARL-WM-BF Aberdeen Proving Ground, MD 21005-5066	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TN-320
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The investigation of a model of combustion led to a novel estimation procedure. The a priori knowledge was the equation: "impulse is equal to average force multiplied by time duration," or $I = \bar{f}t$. The data provided was the mean and standard deviation of the impulse and time. The information requested was the mean and standard deviation of a factor representing average force. The procedure is based on the formula for the variance of a product of random variables. Starting with a simplified model, terms are added until a feasible solution exists. Some of the terms required by the model are not available, and a simulation-based expectation fitting algorithm is used to estimate these unknown values. Through simulation and an assumption about the type of distribution, it is possible to investigate questions that do not have a closed form solution.

15. SUBJECT TERMS

expectation, products of random variables, plausible, correlation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Andrew A. Thompson
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED	UL	16	**19b. TELEPHONE NUMBER** (*Include area code*) 410-278-6805

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

Acknowledgments

The author would like to thank David Webb for his insightful review and comments that improved the quality of this report.

1. Introduction

The investigation of a model of combustion led to a novel estimation procedure. The a priori knowledge was the equation: "impulse is equal to average force multiplied by time duration," or $I = \bar{f}t$. The data provided was the mean and standard deviation of the impulse and time. The information requested was the mean and standard deviation of a factor representing average force. The procedure is based on the formula for the variance of a product of random variables. Starting with a simplified model, terms are added until a feasible solution exists. Some of the terms required by the model are not available, and a simulation-based expectation fitting algorithm is used to estimate these unknown values. Through simulation and an assumption about the type of distribution, it is possible to investigate questions that do not have a closed form solution.

2. Problem Statement

Combustion data was available on several fabrications of explosive jets. Averages and standard deviations were available for the impulse of an explosive jet and its duration. From this data, it was requested that the standard deviation and mean of the average force be determined. The impulse is the product of the duration and the average force. This seemingly simple request does not have a precise answer. It will be shown that there are several plausible explanations.

3. Data Description

The original data was the mean and standard deviation of the jet strength and its duration. Only the means and variances were available; information on the individual trials was not given. The data from combustor 8 has an impulse average of 3.7 mN-s, with a standard deviation of 1.2 and an average duration of 1.4 ms, with a standard deviation of 0.9. Data is available on the product and one of the factors. The information requested is the mean and standard deviation of the other factor. The provided data is treated as if it were the population mean and population standard deviation for the impulse and duration random variables. The number of samples was unknown, so no attempt was made to investigate the effects of variations in these values.

4. Product of Random Variables

First consider the possibility that the impulse is the result of a constant term multiplied by the duration. If this were true, then the variance of the impulse would be the square of the constant multiplied by the variance of the duration. Using the variance to find the constant yields a value of 4/3; if this were the case, the mean value of the impulse would be 1.8667. The observed mean impulse is 3.7, which discounts the plausibility of this functional form. Although it is possible that the impulse can be modeled as some function of duration, there is not a good physical reason to do so. Duration can be considered a random variable as it is partially determined by reaction rate and the initiation of the reaction. The force can be considered a random variable because it is determined by the amount of combustible material and the jet formation process. These variables contain common elements; correlation is anticipated.

The product of two random variables has been discussed by Bohrnstedt and Goldberger.[1] Their paper gives the following two equations for the expectation and variance of the product of two random variables:

$$E(xy) = E(x)E(y) + C(x, y), \tag{1}$$

and

$$\begin{aligned} Var(xy) = {} & E^2(x)Var(y) + E^2(y)Var(x) + 2E(x)E(y)C(x, y) - C^2(x, y) \\ & + E(dx^2dy^2) + 2E(x)E(dxdy^2) + 2E(y)E(dx^2dy), \end{aligned} \tag{2}$$

where $C(x, y)$ symbolizes the covariance between x and y and dx is the difference between x and the mean of x. The first two terms of the variance expression are positive. The covariance can be positive or negative. The last two terms will be zero if the marginal distributions are symmetric; these terms can be negative for asymmetric random variables. A simplified formula results from assuming uncorrelated random variables. In this case, the covariance terms are zero and the variance expression becomes

$$Var(xy) = E^2(x)Var(y) + E^2(y)Var(x) + Var(x)Var(y). \tag{3}$$

Each term in this equation is positive. If the means are large compared to the variance, then it may be possible to ignore the term that is the product of two variances.

[1] Bohrnstedt, G. W.; Goldberger A. S. On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association* **1969**, *64*.

5. Benign Case

The impulse data has a mean of 3.7 and a variance of 1.44, and the duration has a mean of 1.4 and a variance of 0.81. In the previous equation, $I = xy$ and the duration is represented by *y*. Thus, *x* represents the random variable for the average force $\overline{f}$. Using the assumption of no correlation and normal variables gives the following equations:

$$3.7 = 1.4\overline{f}\ , \tag{4}$$

and

$$1.44 = \overline{f}^{2} 0.81 + 1.4^{2} Var(\overline{f}) + 0.81 Var(\overline{f})\ . \tag{5}$$

The average force is 2.64. Substituted into the variance equation, this value leads to negative variance. This contradiction demands that the assumptions be changed.

The correlation terms will be added to the previous model. For a given quantity of combustible material, the average force and time duration should have a correlation of –1. This correlation value is a boundary in the sense that the correlation cannot be made more extreme.

The assumption of zero correlation will be replaced with an assumption that there is correlation. The following equations result from the addition of correlation:

$$E(xy) = E(x)E(y) + C(x,y)\ , \tag{6}$$

and

$$\begin{aligned} Var(xy) = {} & E^{2}(x)Var(y) + E^{2}(y)Var(x) + 2E(x)E(y)C(x,y) - C^{2}(x,y) \\ & + E(dx^{2}dy^{2})\ . \end{aligned} \tag{7}$$

First note that the presence of correlation changes the expected value of the product. If the correlation is positive, then the covariance term will cause the expectation of the product to increase since errors in each random variable will tend to have the same sign. Negative correlation will diminish the expectation of the product. The final term for bivariate normally distributed variables has been shown by Anderson[2] as follows:

$$E(dx^{2}dy^{2}) = Var(x)Var(y) + 2C(x,y)\ . \tag{8}$$

Incorporating this result in equation 8 results in the following impulse variance equation:

$$\begin{aligned} Var(xy) = {} & E^{2}(x)Var(y) + E^{2}(y)Var(x) + 2E(x)E(y)C(x,y) - C^{2}(x,y) \\ & + Var(x)Var(y) + 2Cov(x,y)\ . \end{aligned} \tag{9}$$

[2] Anderson T. W. *An Introduction to Multivariate Statistics*; John Wiley and Sons: New York, NY, 1958.

Notice the assumption that the force and duration are bivariate normal variables has been added. Using this result and substituting in numbers yields the following equations for a correlation of –1; in this case, some of the terms cancel. (Note that $C(x,y) = -\sigma_x \sigma_y$.)

$$3.7 = 1.4\bar{f} + C(x,y) = 1.4\bar{f} + (-1)(0.9)\sigma_f \,. \tag{10}$$

$$1.44 = \bar{f}^2 0.81 + 1.4^2 \sigma_f^2 - 2\bar{f}1.4(0.9)\sigma_f - .81\sigma_f^2 + .81\sigma_f^2 - 2(0.9)\sigma_f \,. \tag{11}$$

By solving equation 10 for the average force and substituting this value into equation 11, the quadratic formula can be used to find the standard deviation of the force. These two steps can be used as an iteration to approximate a solution. For the numeric values in equations 10 and 11, the process converges to an average force of 3.3261, with a standard deviation of 1.0628. A plausible solution has been achieved; a range of plausible solutions should be considered. This process was repeated for different amounts of correlation (ρ). The results are presented in table 1.

Table 1. Correlation results.

Correlation ρ	**Average Force $\bar{f}$**	**Standard Deviation σ_f**
–1	3.1687	0.8181
–0.9	2.3519	0.5029
–0.8	2.3448	0.5794
–0.7	2.3329	0.6888
–0.6	2.3088	0.8662
–0.5	Imaginary	Imaginary

The range of plausible solutions stops between a correlation of –0.6 and –0.5. It can be seen in table 1 that the lowest standard deviation is around a –0.9 correlation. More calculations could be done to find the correlation associated with the lowest standard deviation, and a minimum variance estimate could be proposed. The question remains as to what can be done if there are reasons to believe the distribution is not symmetric. Adding asymmetric terms requires information which is not available; this information will be generated through simulation.

6. Terms That Capture Nonsymmetrical Distribution

At this point, it has been determined that the final two terms of the original variance expression be included. These terms will be zero when the distribution is symmetric. To capture these terms, there is insufficient data; thus, assumptions are necessary. First, a distribution will be chosen for the random variable duration; then, a distribution will be chosen for the force. If these distributions are acceptable, the impulse mean and variance should result from the product of a simulation of the random variable's duration and force. The time duration data can be modeled with a gamma, or Weibull distribution. (Note that an exponential distribution can be ruled out since the variance should be the square of the mean.) The duration mean of 1.4 and variance of 0.81 make an exponential distribution seem unlikely. Each variable was assumed to result from a gamma distribution, and sets of simulated data were used to generate the values needed to find the variance of the product. This process was repeated, as described next, until the simulated data resulted in values matching the known means and variances.

The gamma distribution is not symmetric distribution. Two parameters are used to describe this distribution—the shape parameter is α and the scale parameter is β. The gamma is often used to simulate the time taken to complete a task. For a gamma distribution, the product of the parameters is the population mean, $\mu = \alpha\beta$; the variance is equal to the product of the mean and beta, or $\sigma^2 = \alpha\beta^2$. Assuming that the gamma distribution is the proper distribution, the duration yields $\alpha = 2.42$ and $\beta = 0.58$ as estimates of the parameters. These parameters were used to generate a set of gamma random variables. At this point, it was not possible to generate gamma variables with a correlation of –1. Instead, variables were generated that corresponded with each other in probability. Thus, if the cumulative distribution function of one distribution was p(x), the corresponding value from the other distribution was 1 – p(x). This allowed the generation of related pairs of data; further, this relation could be weakened. Using the cumulative distribution function, the probability associated with each value was found. This value was subtracted from 1 to get the probability of a uniform random variable with a correspondence of –1. Using these values, the parameters of the force distribution were approximated as if it were a similar gamma distribution. In this case, similar means that the ratio of the parameters is the same for each distribution. This requirement results in $\alpha = 3.85$ and $\beta = 0.91$ for the force distribution. Next, using the time duration distribution and the force distribution, the mean and variance of the product were calculated and compared to the obtained impulse mean and variance. With these values, the mean of the impulse was matched; however, the variance was too low. A method was devised to decrease the probability correspondence. As the correspondence moved from –1 towards 0, the variance increased. At a correspondence of –0.78, the simulated mean and variance agreed with the observed mean and variance for the impulse data. The following formula was used to vary the probability correspondence:

$$\text{icp=w*igdp+(1-w)*rand(size(igdp))}, \tag{12}$$

where w is the weight to adjust probability correspondence and igdp is the vector containing the cumulative gamma values with a probability correspondence of –1.

The argument presented shows only that duration and force distribution could have correlated gamma distributions. While specific cases have been ruled out, there can be no claim of a unique solution; only plausibility has been established. The force could be distributed as a gamma distribution with parameters $\alpha = 3.85$ and $\beta = 0.91$ and a probability correspondence of –0.78 with the duration data. The correlation coefficient of the simulated variables was –0.86.

Thus, the concept of a probability correspondence is different than correlation. Note that the minimum variance estimator for the previously discussed symmetric case had a correlation of about –0.9.

7. Discussion of Implications of Principles of Explosive Behavior

In many situations, an investigation of the physical laws describing the phenomenon will provide insight into the nature of the random variables. In some situations, it is worthwhile to perform a sensitivity analysis of the pertinent variables to gain insight. The following material is taken from AMCP 706-180.[3] Chapter 10 of this engineering design handbook discusses thermal explosion. Thermal explosion occurs when explosive systems undergo internal heating. This internal heating can be initiated by external sources; when the chemical reaction releases enough heat, this reaction turns into an explosion. The reaction rate of an explosive varies exponentially with temperature. Expressed as the Arrhenius Law of reaction rate, the relationship is as follows:

$$k_r = Ze^{\frac{-E}{RT}}, \tag{13}$$

where Z is the preexponential factor, E is the activation energy for the reaction, R is the gas constant, and T is absolute temperature. The decomposition of the explosive is exothermic; opposing this is heat loss to the surrounding environment. If the heat generation from the chemical reaction dominates, an explosion will occur. If environmental cooling dominates, the material will react relatively slowly until it is used up. The heat gain equation from the chemical reaction is as follows:

$$\dot{q}_1 = mk_r T_1 Q, \tag{14}$$

where T_1 is the temperature of the reactant and Q is the heat of decomposition. Note the reaction rate term is also a function of temperature. Environmental cooling is represented by the following equation:

$$\dot{q}_2 = hA(T_0 - T_1), \tag{15}$$

where T_0 represents the temperature of the surroundings, A is the surface area exposed to the surroundings, and h is the heat transfer constant. The temperature of the explosive is then expressed by the following equation:

[3] AMCP 706-180. *Principles of Explosive Behavior* **1972**.

$$\frac{dT_1}{dt} = \frac{(\dot{q}_1 + \dot{q}_2)}{mc}, \tag{16}$$

where c represents the specific heat. As time increases, the material decomposes and releases heat. If this heat is conducted away fast enough, the explosion may be delayed and the amount of explosive material will be reduced. Data from a different combustor was available. Observing this data indicates that in the material used to form the jets, the cooling rate and heating rate are almost in balance. This perspective could explain some of the variation in the data from a phenomenological perspective.

For some of the data, the preignition time was rather long. This could be caused by environmental cooling slowing down the reaction rate. In other situations, the energy released in the impulse appears to be a low outlier. This could be explained by only a portion of the energetic material reacting. If only the material on one side of the heating coil ignites or if some material detaches and blows out of the camber before ignition, the impulse strength would be reduced. It is also possible that the orifice could become partially clogged with material during the event.

8. Discussion

For control purposes, the start time, strength, and duration of a jet need to be known precisely. In a spinning round, timing errors cause the force to be applied at suboptimal angles. Procedures reducing the timing errors of the time delay until ignition and the pulse duration should be investigated. While impulse strength does the work, it is only valuable if applied properly.

The variables duration and average force cannot be considered separately because they are highly correlated. The gamma distribution was used to model both of these variables; however, this distribution was chosen only because it was not symmetric and its parameters were easy to estimate from the mean and variance. Weibull and Pearson distributions are also plausible choices. The constraint that both distributions need to be the same could be relaxed.

The calculation of average force is complicated since average force and duration are random variables. Working with an assumed distribution for duration, the parameters of the candidate distribution were adjusted until the mean and variance of the products matched the impulse statistics. Thus, there is only a procedure and not a formula that allows the mean and variance of average force to be found. The simulation step overcomes the data limitation imposed by having access only to the mean and variance of the data. The distributions chosen by the investigator represent the constraint that allows the problem to be solved. Thus, the investigator should provide a justification for the distribution selected. The procedure discussed is general and does not rely on the properties of a specific distribution. A plausible solution is possible through this simulation-based expectation fitting algorithm.

NO. OF COPIES	ORGANIZATION
1 (PDF ONLY)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA 8725 JOHN J KINGMAN RD STE 0944 FORT BELVOIR VA 22060-6218
1	US ARMY RSRCH DEV & ENGRG CMD SYSTEMS OF SYSTEMS INTEGRATION AMSRD SS T 6000 6TH ST STE 100 FORT BELVOIR VA 22060-5608
1	DIRECTOR US ARMY RESEARCH LAB IMNE ALC IMS 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL RD ADELPHI MD 20783-1197
1	DIRECTOR US ARMY RESEARCH LAB AMSRD ARL CI OK T 2800 POWDER MILL RD ADELPHI MD 20783-1197
	ABERDEEN PROVING GROUND
1	DIR USARL AMSRD ARL CI OK TP (BLDG 4600)

NO. OF COPIES	ORGANIZATION
1	DREXEL UNIV DEPT OF MECHL ENGRG B CHANG 3141 CHESTNUT ST PHILADELPHIA PA 19104
1	DIRECTOR US ARMY RSRCH LAB AMSRD ARL CI NT R PRESSLEY 2800 POWDER MILL RD ADELPHI MD 20783-1197

ABERDEEN PROVING GROUND

NO. OF COPIES	ORGANIZATION
2	COMMANDER USAATC TEDT AT ADR B GILLICH S CLARK 400 COLLERAN RD TRAILER T1 APG MD 21005-5059
1	COMMANDER USAATC 400 COLLERAN RD TEDT AT ADF S NOVAK APG MD 21005-5059
20	DIR USARL AMSRD AAR AEF T M ANDRIOLO AMSRD ARL CI IC B BODT AMSRD ARL CI IB AMSRD ARL SL BD J COLLINS L MOSS AMSRD ARL WM BA R MCGEE T BROWN T HARKINS M ILG AMSRD ARL WM BF J WALD J WALL D WEBB M ARTHUR A THOMPSON (4 CPS) B FLANDERS R PEARSON B OBERLE AMSRD ARL WM MB R KASTE

INTENTIONALLY LEFT BLANK.