NASA TECHNICAL NOTE

NASA TN D-6639

c. 1

NASA TN D-6639

OBJECTIVE ESTIMATES BASED ON EXPERIMENTAL DATA AND INITIAL AND FINAL KNOWLEDGE

by Burt M. Rosenbaum

Lewis Research Center
Cleveland, Ohio 44135

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JANUARY 1972

1. Report No. NASA TN D-6639	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle OBJECTIVE ESTIMATES BASED ON EXPERIMENTAL DATA AND INITIAL AND FINAL KNOWLEDGE		5. Report Date January 1972
		6. Performing Organization Code
7. Author(s) Burt M. Rosenbaum		8. Performing Organization Report No. E-6382
		10. Work Unit No. 132-15
9. Performing Organization Name and Address Lewis Research Center National Aeronautics and Space Administration Cleveland, Ohio 44135		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Note
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D. C. 20546		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

An extension of the method of Jaynes, whereby "least-biased" probability estimates are obtained, permits such estimates to be made which objectively take into account any experimental data on hand as well as prior (before data acquisition) and posterior (after data acquisition) knowledge. These estimates can be made for both discrete and continuous sample spaces. The method allows a simple interpretation of Laplace's two rules: (1) the principle of insufficient reason, and (2) the rule of succession. Several examples are analyzed by way of illustration.

17. Key Words (Suggested by Author(s)) Bayes rule; Decision theory; Entropy; Estimation; Information theory; Probability; Distribution; Statistics; Mathematical model	18. Distribution Statement Unclassified - unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	42	$3.00

* For sale by the National Technical Information Service, Springfield, Virginia 22151

CONTENTS

OBJECTIVE ESTIMATES BASED ON EXPERIMENTAL DATA AND INITIAL AND FINAL KNOWLEDGE

by Burt M. Rosenbaum

Lewis Research Center

SUMMARY

A reformulation of the expression for "entropy" in the presence of experimental data enables the determination of "least-biased" probability estimates in the a posteriori case for both discrete and continuous sample spaces. The method is an extension of that used by Jaynes in 1963 by means of which he estimated least-biased probability values for the a priori situation for a discrete sample space where his probability estimates objectively took into consideration any prior knowledge that an experimenter might have at his disposal. The modification employed herein allows for the adjustment of the influence of prior knowledge on the posterior distribution in accordance with the subjective degree of belief in the accuracy of the prior knowledge. Also, as part of the modified method, any posterior knowledge is treated in the same way as the prior knowledge. Thus, the "best" posterior estimates depend on initial or prior information, expermental data, and final or posterior information. The postulated method agrees with the rule of Bayes and affords an insight into the rules of Laplace, namely, the principle of insufficient reason and the rule of succession. Several simple examples are treated in detail to illustrate the procedure followed.

INTRODUCTION

Scientific inference techniques based on Bayesian methods are internally consistent (refs. 1 and 2). As soon as a prior distribution has been set up by the statistician, the rule of Bayes states how the experimental data modifies the prior distribution to yield a unique posterior distribution.

The difficulty associated with the application of Bayesian concepts to a particular problem can be attributed to the arbitrariness of the assumed prior distribution (refs. 3 and 4). Two statisticians with the same initial knowledge will, in general, not assume the same priors and, hence, their posteriors or final distributions will also differ. It has been argued that the two posteriors obtained would not be significantly different and, in any event, if a prior is smooth, the experimental evidence will eventually overwhelm any bias that might exist in the prior; therefore, any prior that is consistent with the initial information is satisfactory (ref. 5). However, while this may be the case when "sufficient" data is on hand, the question as to whether a particular prior is or is not biased still remained unanswered.

The establishment of unbiased priors was the problem Jaynes considered (ref. 6). Jaynes used the maximum entropy principle employed in information theory (ref. 7) and statistical mechanics (refs. 8 and 9) to obtain estimates of prior probabilities. Using these estimates, Jaynes calculated expectation values of a loss function associated with each allowable decision. The solutions that he obtained showed that the optimum decisions - those yielding the smallest expected loss values - were the same as one would normally choose based on a common sense approach. When the initial knowledge or constraints governing a particular decision problem changed so that it was obvious that the optimum decision should change, it was found that the maximum entropy concept also dictated a like change in the optimum decision. When the problem under consideration became so complicated that common sense or intuition could not definitely single out the optimum decision, the mathematical approach based on maximum entropy still was able to indicate unambiguously a particular decision as optimum. The major point to be noted is that Jaynes had constructed a quantitative method for establishing values of prior probabilities that could be said to be objective and "optimally" reasonable.

Although Jaynes had devised a method for calculating objective optimum values for prior probabilities from initial knowledge, these values could not be used to calculate unique values of posterior probabilities. In order to use the Bayes rule in going from initial to final probabilities, what is needed is the prior probability density function for the probabilities themselves, and Jaynes' method does not generate this function.

This report extends the method of Jaynes so that probability estimates based on both initial information and experimental data may be made. The extension affords a new interpretation of Laplace's principle of insufficient reason and introduces a weight parameter into the calculation dependent on the statistician's degree of belief in his original "guesstimates".

SYMBOLS

a_i number of times outcome i would be expected to occur based on initial information

$\vec{a}$ vector $\left\{a_1, a_2, \ldots, a_n\right\}$

$\mathscr{C}$ adjustable parameter

$E\left[p_i/D\right]$ expectation value of p_i based on the posterior probability density function

$E[X]$ expectation of X

$f(p_i)$ probability density function for p_i

$f(\vec{p})$ multivariate probability density function for variables $\left\{p_1, p_2, . . ., p_n\right\}$

$f(t)$ probability density function for t

$f(\sigma)$ probability density function for σ

$f(x)$ probability density function for x

$g_k(i)$ function of index i

$\langle g_k \rangle$ expectation of g_k

H modification of entropy function defined in eq. (36)

$h_q(i)$ function of index i

$\langle h_q \rangle$ expectation of h_q

K number of prior constraining equations

M number of actual experimental measurements

M_0 number of initial hypothetical measurements

M_r number of measurements of runout time

m_i number of times outcome i occurs

$\vec{m}$ vector $\left\{m_1, m_2, . . ., m_n\right\}$

n number of possible outcomes

n_i multiplicity associated with outcome i

$P\left(D/\vec{p}\right)$ probability of data given that $\vec{p}$ takes on values $\left\{p_1, p_2, . . ., p_n\right\}$

p probability

$\vec{p}$ vector $\left\{p_1, p_2, . . ., p_n\right\}$

Q number of posterior constraining equations

S entropy defined in eq. (4)

S_D "entropy" function defined in eq. (35)

t time-to-failure, hr

t_a, t_b times-to-failure, $t_a < t_b$

W	number of randomly selected white balls
X	bounded continuous random variable
x	value of X
B(q,r)	Beta function of q and r
$\Gamma(q)$	Gamma function of q
$\delta(x)$	Dirac delta function with argument x
λ_k	k^{th} Lagrangian multiplier, $k \neq 0$
μ_t	mean of t
μ_X	mean of X
ν_q	q^{th} Lagrangian multiplier, $q \neq 0$
σ	alternating fatigue stress
σ_t^2	variance of t
σ_X^2	variance of X

Subscripts:

A	altered (after incorporation of some of the data)
D	final or posterior (after acquisition of data)
i	outcome i
k	k^{th} prior constraint
q	q^{th} posterior constraint
α_i	α_i^{th} outcome of n_i possible outcomes
0	initial or prior (before acquisition of data)

Superscripts:

$\hat{}$	least-biased estimate
$\tilde{}$	objective (not necessarily least-biased) estimate
$\bar{}$	average value

METHOD OF JAYNES

We consider an experimental measurement which can take on any one of n mutually exclusive distinct results where the possible results are labeled by the numbers

$1, 2, \ldots, n$. Let p_i, $i = 1, 2, \ldots, n$, denote the probability that the system is in the state that yields the i^{th} result where it is assumed that the nature of the experiment is such that p_i is independent of time. We have

$$\sum_{i=1}^{n} p_i = 1 \tag{1}$$

Suppose, before any experimental measurements are made, we know or can make a guess of the values of the expectations of certain functions of i and that this prior knowledge or information may be expressed by K independent relations of the form

$$\sum_{i=1}^{n} p_i g_k(i) = \langle g_k \rangle , \; k = 1, 2, \ldots, K \tag{2}$$

where

$$K \leq n - 1 \tag{3}$$

According to Jaynes, the least-biased prior probability estimates $\left(\hat{p}_0\right)_i$, $i = 1, 2, \ldots, n$, are those that maximize the entropy

$$S = - \sum_{i=1}^{n} p_i \ln p_i \tag{4}$$

subject to the constraints given by equations (1) and (2).

Notice that if there is no initial information about the experiment other than the possible results, then there are no constraints except that of equation (1), which always applies, and the entropy S is greatest when all p_i are the same; that is,

$$\left(\hat{p}_0\right)_i = \frac{1}{n} \tag{5}$$

Hence, in the absence of any information, the least-biased prior probability estimates are the same for every possible result. These estimates agree with Laplace's principle of insufficient reason, sometimes attributed to Bernoulli, which assumes that each result is as likely to be true as any other result unless there exists some reason for assuming otherwise. Jaynes has noted that the maximum entropy criterion leads to prior probability estimates that are as uniform as possible subject to the fact that the constraints must hold.

In reference 6, Jaynes applies his method to the situation wherein a plant manager must decide which single color to paint the day's output of 200 "widgets". This article by Jaynes represents a stimulating and plausible argument for the method in general.

ROWLINSON'S CRITICISM OF JAYNES' METHOD

Although Jaynes' arguments appeared convincing to some (e.g., refs. 10 and 11), it was still opposed by the classical statistician for whom the probability of an outcome can only be interpreted as the frequency with which it occurs in a given experiment or a given idealization of an experiment. From this viewpoint, parameters in a probability distribution are constants which cannot be said to possess probability distributions. On the other hand, the Bayesian statistician treats any physical parameter about which he has less than full knowledge as a random variable where the probability that the parameter under consideration may take on a particular value or lie in a given interval represents a person's degree of belief in that happenstance based on what the person knows or feels at the time. This latter viewpoint is anathema to the classical statistician.

Rowlinson (ref. 12) discussed a game of chance in which the score on each turn could be any integer from 1 to 6. If it is known that the average score is 4.5, then, employing Jaynes' analysis, we have

$$\sum_{i=1}^{6} i p_i = 4.5 \tag{6}$$

and we can maximize

$$S = -\sum_{i=1}^{6} p_i \ln p_i \tag{7}$$

subject to the constraints given by equations (1) and (6) to yield Jaynes' least-biased estimates $\left(\hat{p}_0\right)_i$ for the probability of each integer on any turn. Carrying out this process gives

$$\left(\hat{p}_0\right)_i = \frac{(1.45)^i}{\left[\sum_{j=1}^{6} (1.45)^j\right]} \tag{8}$$

Rowlinson, however, states that there is really no reason for trusting that the probabilities as generated by the Jaynes' method are actually correct. For instance, he says that each turn of the previously mentioned game of chance might consist of noting the number W of white balls in a sample of five balls randomly selected (with replacement) from an urn containing seven white and three black balls, the score i for each turn being (W + 1). For this game,

$$p_i = \binom{5}{i-1} 0.7^{i-1} 0.3^{6-i}, \quad i = 1, 2, \ldots, 6 \tag{9}$$

The p_i's for equations (8) and (9) are as follows:

Method	Score, i					
	1	2	3	4	5	6
	p_i					
Equation (8)	0.055	0.080	0.120	0.165	0.235	0.345
Equation (9)	.000243	.02835	.1323	.3087	.36015	.16807

It is seen that the two probability distributions are markedly different. On this basis, Rowlinson rejects the principle of maximizing the entropy as a "useful way of attacking the problem" in the first place.

The rationale behind Rowlinson's argument is that taken by the classical statistician. Rowlinson throws out the maximum entropy estimates of p_i because there are situations that exist when these estimates are not correct. In fact, Rowlinson would object to any prior estimate of p_i because we can always manufacture games where the prior estimate would be far from being correct. Indeed, Rowlinson maintains that on the basis of the information given, namely, that the score can be any of the integers from 1 to 6 and that the average score is 4.5, there is <u>no</u> means of estimating the p_i's.

The position of Jaynes is that, although the p_i's as given by equation (8) are the least biased, any set of p_i's satisfying the constraints might actually be correct. Because the average score per turn in the example is 4.5 and this value is larger than it would be if all p_i were equal, one would reasonably expect the p_i's associated with the larger values of i to be larger than those associated with the smaller i-values. This is what the maximum entropy concept objectively accomplishes. Knowing only that the average score is 4.5 does not make the assumption that p_5 is more than twice as large as p_6 (as is given by eq. (9)) reasonable. The maximum entropy criterion can be said to give equal p_i's unless there is a good reason for not doing so and, in the event that there is such a reason, the criterion still endeavors to keep the p_i's as nearly equal as possible while satisfying the constraints imposed by prior knowledge. The more prior information that is available, the more one would expect the least-biased probability estimates to be closer to reality. However, in any but the complete information case, the chance that these probability estimates are exactly correct is for all practical purposes zero because the number of possible distributions is infinite.

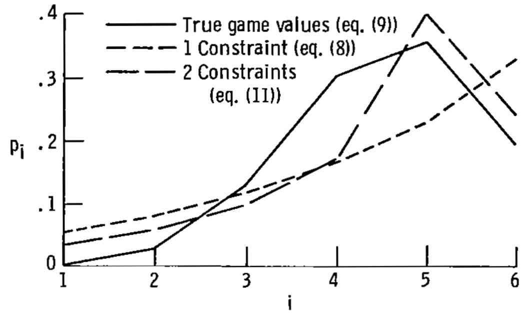
To illustrate that least-biased probability estimates become better approximations to correct values as prior knowledge increases, let it be suspected that p_5 is twice p_6 so that the additional constraint that

$$p_5 = 2p_6 \tag{10}$$

is imposed. The least-biased probability estimates become

$$\left.\begin{aligned}
\left(\hat{p}_0\right)_i &= \frac{(1.653)^i}{1.5(2)^{1/3}(1.653)^{16/3} + \sum_{j=1}^{4}(1.653)^j} \qquad i = 1, 2, 3, 4 \\
\left(\hat{p}_0\right)_5 &= \frac{2^{1/3}(1.653)^{16/3}}{1.5(2)^{1/3}(1.653)^{16/3} + \sum_{j=1}^{4}(1.653)^j} \\
\left(\hat{p}_0\right)_6 &= \frac{1}{2}\left(\hat{p}_0\right)_5
\end{aligned}\right\} \tag{11}$$

The probability distributions given by equations (8), (9), and (11) are plotted in the following sketch:



And we see that, for every value of i except for $i = 3$, the p_i given by equation (9) is closer to the p_i of equation (11) than to that of equation (8).

MODIFICATION OF THE FORM OF THE ENTROPY FUNCTION TO INCORPORATE MULTIPLICITY CONSIDERATIONS

We consider a slightly altered model from the one heretofore considered. Suppose each possible experimental measurement labeled i, where $i = 1, 2, \ldots, n$, arises from any of a group of n_i mutually exclusive distinct outcomes. Let $p_{i\alpha_i}$ represent the probability that an experimental result is the α_i^{th} outcome in the set of n_i outcomes associated with the i^{th} possible experimental measurement. We have

$$\sum_{\alpha_i=1}^{n_i} p_{i\alpha_i} = p_i \qquad i = 1, 2, \ldots, n \tag{12}$$

where equation (1) still applies so that

$$\sum_{i=1}^{n} \sum_{\alpha_i=1}^{n_i} p_{i\alpha_i} = 1 \tag{13}$$

Again, suppose that the prior knowledge is given by equations (2). Then, we can proceed with Jaynes' method where the entropy now is given by

$$S = -\sum_{i=1}^{n}\sum_{\alpha_i=1}^{n_i} p_{i\alpha_i} \ln p_{i\alpha_i} \tag{14}$$

Maximizing S subject to the constraints given by equations (13) and (2) yields the least-biased probability estimates

$$\left(\hat{p}_0\right)_{i\alpha_i} = \exp\left[-\lambda_0 - \sum_{k=1}^{K} \lambda_k g_k(i)\right] \tag{15}$$

where $\lambda_0 - 1, \lambda_1, \lambda_2, \ldots, \lambda_k$ are Lagrangian multipliers whose values are determined by substituting equation (15) into the constraining equations (13) and (2).

From equation (12) and the fact that $\left(\hat{p}_0\right)_{i\alpha_i}$ is independent of α_i

$$\left(\hat{p}_0\right)_i = \sum_{\alpha_i=1}^{n_i} \left(\hat{p}_0\right)_{i\alpha_i} = n_i \exp\left[-\lambda_0 - \sum_{k=1}^{K} \lambda_k g_k(i)\right] \tag{16}$$

Hence, the maximum value of S subject to the constraints occurs when each of the outcomes corresponding to a given value of i has the same probability of happening. Note that the value of $\left(\hat{p}_0\right)_i$ as given in equation (16) is modified by the factor n_i where n_i is the multiplicity associated with the i^{th} experimental measurement.

When each of the n_i outcomes, $\alpha_i = 1, 2, \ldots, n_i$, has the same probability, we have

$$p_{i\alpha_i} = \frac{p_i}{n_i}$$

and equation (14) takes on the form

$$S = -\sum_{i=1}^{n}\sum_{\alpha_i=1}^{n_i} \frac{p_i}{n_i} \ln \frac{p_i}{n_i} = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{n_i} \tag{17}$$

Thus, the least-biased estimates as given by equation (16) result from maximizing equation (17) for S subject to the appropriate constraints. Because the multiplicity associated with a given experimental measurement can never be neglected, the expression for S that applies to every problem should always be that given in equation (17). However, in the absence of any information about the multiplicity of the various experimental results, n_i is taken as unity for each i and the expression for S reverts to that given in equation (4).

As an example of the application of equation (16), let us turn once again to the game of chance suggested by Rowlinson. For this game, n_i is the number of ordered ways of selecting (i - 1) white balls and (6 - i) black balls so that

$$n_i = \binom{5}{i-1}, \quad i = 1, 2, \ldots, 6 \tag{18}$$

The equation of constraint is equation (6) where $g(i) = i$. By equation (16)

$$\left(\hat{p}_0\right)_i = \binom{5}{i-1} e^{-\lambda_0 - \lambda_1 i} \tag{19}$$

Employing equation (1) to eliminate $e^{-\lambda_0}$ gives

$$\left(\hat{p}_0\right)_i = \frac{\binom{5}{i-1} e^{-\lambda_1 i}}{\sum_{j=1}^{6} \binom{5}{j-1} e^{-\lambda_1 j}} \tag{20}$$

Substituting equation (20) into constraining equation (6) results in

$$\sum_{i=1}^{6} \binom{5}{i-1} i\, e^{-\lambda_1 i} = 4.5 \sum_{i=1}^{6} \binom{5}{i-1} e^{-\lambda_1 i}$$

or, using the binomial theorem,

$$\xi \frac{d}{d\xi}\left[\xi(1+\xi)^5\right] = 4.5\, \xi(1+\xi)^5 \tag{21}$$

where

$$\xi \equiv e^{-\lambda_1} \tag{22}$$

We get from equation (21)

$$e^{-\lambda_1} = \frac{7}{3}$$

and equation (20) becomes

$$\left(\hat{p}_0\right)_i = \binom{5}{i-1} (0.7)^{i-1}(0.3)^{6-i}, \quad i = 1, 2, \ldots, 6 \tag{23}$$

Hence, the least-biased estimate $\left(\hat{p}_0\right)_i$ for p_i is identical to the correct values as given in equation (9).

The fact that the least-biased probability estimates are exactly those given by a binomial distribution (when the binomial multiplicity or degeneracy of an experimental measurement is included in the entropy formulation and the constraining condition fixes the true average) was pointed out by Jaynes (ref. 13). Indeed, Jaynes went so far as to state "if the experiment fails to confirm the maximum-entropy prediction, and this disagreement persists on indefinite repetition of the experiment, then we will conclude that the physical mechanism of the experiment must contain additional constraints which were not taken into account in the maximum-entropy calculation. The observed deviations then provide a clue as to the nature of the new constraints".

CONSEQUENCE OF THE RULE OF BAYES WHEN NO PRIOR INFORMATION IS AVAILABLE

We now turn our attention to a situation that has been well documented in the literature. The problem that will be reviewed in this section will serve as a limiting case for the generalized method to be proposed later in this report and, as such, is important in providing necessary insight.

Again the index i denotes the i^{th} possible result of n mutually exclusive possible results and p_i denotes the probability of getting the i^{th} result. We suppose that no prior information is available and that, in a total of M repetitive measurements, the first result has turned up m_1 times, the second result m_2 times, and so forth. We have

$$\sum_{i=1}^{n} m_i = M \tag{24}$$

What we wish to calculate on the basis of the data $\left\{m_1, m_2, \ldots, m_n\right\}$ is a set of objective (not necessarily least-biased) posterior estimates $\left(\tilde{p}_D\right)_i$ of the probabilities p_i. The procedure we shall follow is to first set up the prior distribution functior for the probabilities, then to employ the rule of Bayes to obtain the posterior distribution function for the probabilities, and finally to calculate the desired objective probability estimates by finding the expectation values of the probabilities based on the posterior distribution function.

Because at the outset nothing is known about the occurrence of any of the outcomes except for the conditions that $p_i \geq 0$, $i = 1, 2, \ldots, n$, and equation (1) holds, all possible vectors $\vec{p} = \left\{p_1, p_2, \ldots, p_n\right\}$ are equally likely. In other words, every point in the n-dimensional hypercube

$$0 \leq p_1 \leq 1$$

$$0 \leq p_2 \leq 1$$

$$\vdots$$

$$0 \leq p_n \leq 1$$

is as likely to be true as any other point provided both points satisfy equation (1). The prior probability density function consistent with this viewpoint is

$$f_0(\vec{p}) = f_0(p_1, p_2, \ldots, p_n) = c \tag{25}$$

for all points on the surface S where S is given by the equation

$$\sum_{i=1}^{n} p_i = 1$$

The value of the normalization constant c must be such that integration of the probability density function over the (n-1)-dimensional surface S gives unity; that is,

$$\iint \cdots \int_S f_0(\vec{p})\,dS = \underset{\substack{0 \le p_1 + p_2 + \ldots + p_{n-1} \le 1 \\ p_n = 1 - (p_1 + p_2 + \ldots + p_{n-1})}}{\iint \cdots \int} f_0(\vec{p}) \sqrt{1 + \left(\frac{\partial p_n}{\partial p_1}\right)^2 + \left(\frac{\partial p_n}{\partial p_2}\right)^2 + \ldots + \left(\frac{\partial p_n}{\partial p_{n-1}}\right)^2}\; dp_{n-1} \ldots dp_2\, dp_1$$

$$= c\sqrt{n} \int_0^1 \int_0^{1-p_1} \cdots \int_0^{1-p_1-p_2-\ldots-p_{n-2}} dp_{n-1} \ldots dp_2\, dp_1 = 1 \tag{26}$$

The multiple integral in equation (26) may be evaluated by using the Beta function identity (ref. 14):

$$B(q,r) = \int_0^1 t^q (1 - t)^r\, dt = \frac{\Gamma(q + 1)\Gamma(r + 1)}{\Gamma(q + r + 2)}; \quad q,r > -1 \tag{27}$$

Here $\Gamma(q + 1)$ denotes the Gamma function with agrument $(q + 1)$. Changing the variable of integration in equation (27) from t to $x = at$, we get

$$\int_0^a x^q(a - x)^r dx = \frac{\Gamma(q + 1)\Gamma(r + 1)}{\Gamma(q + r + 2)} a^{q+r+1} \tag{28}$$

We now can easily show that the integration of the multiple integral in equation (26) can be carried out as a succession of integrals of the type appearing in equation (28). Performing the integration over p_{n-1} (corresponds to $q = 0$, $r = 0$, $a = 1 - p_1 - p_2 - \dots - p_{n-2}$) leads to the following integral taken over p_{n-2}:

$$\int_0^{1-p_1-p_2-\dots-p_{n-3}} (1 - p_1 - p_2 - \dots - p_{n-2})dp_{n-2}$$

This integral is of the form given in equation (28) with $q = 0$, $r = 1$, and

$$a = 1 - p_1 - p_2 - \dots - p_{n-3}$$

Hence, application of equation (28) leads to the following integral now taken over p_{n-3}:

$$\frac{\Gamma(1)\Gamma(2)}{\Gamma(3)} \int_0^{1-p_1-p_2-\dots-p_{n-4}} (1 - p_1 - p_2 - \dots - p_{n-3})^2 \, dp_{n-3}$$

which again can be evaluated by applying equation (28). Carrying out the $(n - 1)$ integrations in succession yields the value $(n - 1)! / \sqrt{n}$ for the normalization constant c so that

$$f_0(\vec{p}) = \begin{cases} \dfrac{(n - 1)!}{\sqrt{n}} & \text{for all points on } S \\ 0 & \text{elsewhere} \end{cases} \tag{29}$$

is the prior probability density function for the probabilities p_i.

This distribution $f_0(\vec{p})$ can be modified by using the rule of Bayes which takes into account the experimental data $\left\{m_1, m_2, \ldots, m_n\right\}$. The rule of Bayes is given by the expression

$$f_D(\vec{p}) = \frac{f_0(\vec{p})P(D/\vec{p})}{\int\int\cdots\int f_0(\vec{p})P(D/\vec{p})\,dS} \tag{30}$$

where

$f_D(\vec{p})$ posterior probability density function for the probabilities

$f_0(\vec{p})$ prior probability density function for the probabilities

$P(D/\vec{p})$ probability of the data $\left\{m_1, m_2, \ldots, m_n\right\}$ given that $\vec{p}$ takes on the set of values $\left\{p_1, p_2, \ldots, p_n\right\}$

The expression for the $P(D/\vec{p})$ is merely the multinomial probability distribution:

$$P(D/\vec{p}) = \frac{M!}{m_1!\ m_2!\ \ldots\ m_n!}\ p_1^{\ m_1} p_2^{\ m_2} \ldots p_n^{\ m_n} \tag{31}$$

while that for $f_0(\vec{p})$ is given by equation (29). The denominator of the right side of equation (30) can be evaluated in the same way as was done for the multiple integral of equation (26). We obtain

$$f_D(\vec{p}) = \frac{1}{\sqrt{n}} \frac{(M + n - 1)!}{m_1!\ m_2!\ \ldots\ m_n!}\ p_2^{\ m_1} p_2^{\ m_2} \ldots p_n^{\ m_n} \tag{32}$$

on surface S.

The posterior probability density for the probability p_i alone may be obtained by integrating over the other probabilities:

$$f_D(p_i) = \int\int\cdots\int f_D(\vec{p})\sqrt{n}\ dp_1\ dp_2 \ldots dp_{i-1}\ dp_{i+1} \ldots dp_{n-1}$$

$$= \frac{(M + n - 1)!}{m_i!\ (M - m_i + n - 2)!}\ p_i^{\ m_i}(1 - p_i)^{M-m_i+n-2},\quad 0 \leq p_i \leq 1 \tag{33}$$

The expectation value of p_i based on the posterior probability density function as given by equation (33) is our desired objective estimate $\left(\tilde{p}_D\right)_i$ of p_i. The estimate is then

$$\left(\tilde{p}_D\right)_i = E(p_i/D) = \int_0^1 p_i f_D(p_i) dp_i = \frac{m_i + 1}{M + n} \tag{34}$$

This relation, which holds when the prior probabilities are uniformly distributed, is known as Laplace's rule of succession (refs. 15 and 16). Thus we can say that according to the Bayes-rule estimate the probability that the next measurement will be the i^{th} result is $(m_i + 1)/(M + n)$.

This answer differs from that given by the maximum-likelihood estimate of p_i which is m_i/M. We note that the maximum-likelihood estimate of p_i would be identical to the Bayes-rule estimate if, in the maximum-likelihood case, we supposed that, before any actual measurements were made, there were n hypothetical measurements wherein each of the possible outcomes turned up exactly once.

MODIFICATION OF THE EXPRESSION FOR ENTROPY WITH THE ACQUISTION OF DATA IN THE ABSENCE OF PRIOR INFORMATION

Equation (34) has shown that when we start with a uniform distribution for the probabilities $\vec{p}$ the use of Bayes' rule to incorporate the data $\left\{m_1, m_2, \ldots, m_n\right\}$ leads to the objective estimate $(m_i + 1)/(M + n)$ for p_i. Now we remark that application of equations (16) and (17) with n_i replaced by $(m_i + 1)$ demonstrates that the same result could be obtained by finding those p_i's that maximize the expression.

$$S_D = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{m_i + 1} \tag{35}$$

subject, of course, only to the constraint given by equation (1). Hence, this modification of the entropy expression should be examined in order to see whether it can be interpreted in any sensible way.

First, equation (35) says that the entropy formulation S_D changes with the accumulation of experimental data. If we consider the term $(m_i + 1)$ as the number of measurements resulting in the i^{th} outcome, then the initial state of no data can be interpreted as starting out with a total of n fictitious measurements, one each for every possible outcome i. But this is the same correspondence that we noted in the last section in order to obtain agreement between the maximum-likelihood estimate and the Bayes-rule estimate.

Second, we see that as the number M of measurements increases, the least-biased estimate of p_i based on equation (35) more and more approximates the frequency with which the i^{th} result occurs; that is, $\left(\hat{p}_D\right)_i \to m_i/M$ as both m_i and M become large compared to n. Obviously this should be the case.

Thus the reformulation S_D of the expression for entropy in the presence of data appears to yield sensible answers for the limiting cases of $M = 0$ and $M \to \infty$. Also, the theory as based on equation (35) illustrates that Laplace's principle of insufficient reason and Laplace's rule of succession now rest on a common footing; namely, the initial state corresponding to the situation where no prior knowledge exists consists of the assumption that n measurements have been made, each possible outcome of the n possible outcomes turning up exactly once. Hence, this is equivalent to starting out with the assumption that all outcomes are equally likely.

CONSIDERATION OF THE PRIOR DISTRIBUTION WHEN PRIOR INFORMATION IS AVAILABLE

To repeat, we were able to interpret the term $(m_i + 1)$ appearing in equation (35), the 1 being the number of hypothetical measurements yielding the i^{th} result for the case where no prior knowledge is on hand and the m_i being the actual number of measurements yielding the i^{th} result. By contrast, when we have some initial information, then, by applying Jaynes' method, we can obtain least-biased prior estimates $\left(\hat{\vec{p}}_0\right)$ of $\vec{p}$ and no longer would it be "reasonable" to assume a priori the same number of hypothetical measurements for each of the outcomes. Instead, we can generalize the expression given in equation (35) by writing

$$H = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{m_i + a_i} \tag{36}$$

where a_i is the number of times that outcome i would be expected to occur based on our initial information and Jaynes' method, where equation (17) is employed as the expression for entropy. Thus, we let a_i be given as

$$a_i = \mathscr{C} n\left(\hat{p}_0\right)_i \tag{37}$$

where $\mathscr{C}$ is an adjustable parameter that plays the role of the flattening constant (ref. 16). We note that equation (36) reverts to equation (35) when no prior information is available and $\mathscr{C} = 1$.

By equation (37), the number of hypothetical prior measurements is

$$M_0 = \mathscr{C} n \tag{38}$$

so that equation (37) takes the form

$$a_i = M_0\left(\hat{p}_0\right)_i \tag{39}$$

It may be noted that the larger the value of $\mathscr{C}$ or M_0 the more data is needed to significantly alter the original hypothetical distribution $\vec{a} = \left\{a_1, a_2, \ldots, a_n\right\}$. The greater the degree of credibility in the initial information, the larger the value $\mathscr{C}$ should take on. Caution must be exercised in this regard because the vector $\vec{a}$ corresponds to a complete specification of the distribution and this distribution, even though it represents the least-biased distribution based on initially known true values of averages, could be very far from the appropriate distribution. There might be some cases when it is justifiable to take $\mathscr{C}$ as large as 3 or 4. Even zero might be chosen for $\mathscr{C}$ if it is desired to obtain the maximum likelihood estimate of $\vec{p}$.

At this point, the concepts associated with the generalized method by which we can proceed have been completely established. If there is some initial knowledge as given by the constraints, then, by using Jaynes' method, the least-biased prior estimate $\left(\hat{\vec{p}}_0\right)$, as given by equation (16), is set up. The hypothetical distribution $\vec{a}$ is found and inserted into the expression for H given in equation (36) by using equation (37) (or by eq. (39)) and our degree of credibility as characterized by $\mathscr{C}$ (or M_0). If the experimental data $\left\{m_1, m_2, \ldots, m_n\right\}$ are now incorporated into the problem, then H may be maximized with respect to $\vec{p}$ subject to any constraints that may apply a posteriori. If these posterior constraints are written as

$$\sum_{i=1}^{n} p_i h_q(i) = \langle h_q \rangle, \quad q = 1, 2, \ldots, Q \tag{40}$$

then the expression for $\left(\hat{p}_D\right)_i$ may be given by

$$\left(\hat{p}_D\right)_i = (m_i + a_i)e^{-\nu_0 - \sum_{q=1}^{Q} \nu_q h_q(i)} \tag{41}$$

where $\nu_0 -1, \nu_1, \ldots, \nu_Q$ are Lagrangian multipliers. Thus, this procedure yields least-biased estimates $\left(\hat{p}_0\right)_i$, $i = 1, 2, \ldots, n$, which take into account both prior and posterior information as well as experimental data.

Note that maximization of H with respect to $\vec{p}$ when all $m_i = 0$ yields the least-biased estimates

$$\left(\hat{p}_0\right)_i = \frac{a_i}{\sum_j a_j}, \quad i = 1, 2, \ldots, n$$

so that, in the limit where all $m_i = 0$, the generalized method still gives the same values for the least-biased estimates as those given by the Jaynes' method.

It may be observed at this point that the method as stated may be employed for the case of continuous random variables with very little change in viewpoint. This fact is shown in the subsequent examples illustrating the generalized method.

EXAMPLE INVOLVING DISCRETE VARIABLE

We consider a simple example to illustrate the method. Let the number n of possible outcomes be 3 with i = 1, 2, or 3 designating the three possible outcomes. Let the prior information be given by

$$\sum_{i=1}^{3} ip_i = 1.5 \tag{42}$$

Also, the multiplicity of outcome 3 is believed to be twice that of either outcome 1 or 2. Hence, $n_i = n_2 = 1$ and $n_3 = 2$.

Applying Jaynes' method, we find that the least-biased prior probability estimates are

$$\left(p_0\right)_i = \frac{n_i\left(\frac{1}{3}\right)^i}{\sum_{j=1}^{3} n_j\left(\frac{1}{3}\right)^j} = \begin{cases} 0.643, & i = 1 \\ 0.214, & i = 2 \\ 0.143, & i = 3 \end{cases}$$

Taking $\mathscr{C} = 1$, we write H as

$$H = -\sum_{i=1}^{3} p_i \ln \frac{p_i}{m_i + 3\left(\hat{p}_0\right)_i}$$

Let us assume that six experimental measurements have been made and the data is $\vec{m} = \{3, 1, 2\}$; that is, outcome $i = 1$ has occurred three times, outcome $i = 2$ once, and outcome $i = 3$ twice. Then

$$m_i + a_i = \begin{cases} 4.929, & i = 1 \\ 1.642, & i = 2 \\ 2.429, & i = 3 \end{cases}$$

We shall obtain objective posterior probability estimates for two cases:

(1) Where the constraint as given by equation (42) no longer applies (This is the situation that occurs when the constraint represented an initial guess and we want to relax this constraint a posteriori because the guess might well prove to be wrong.)

(2) Where the constraint as given by equation (42) still applies (In this case, we know definitely that this relation holds and all probability estimates, whether prior or posterior, must conform to this relation.)

We obtain the following posterior probability estimates:

Case	Outcome, i			Constraints
	1	2	3	
1	0.548	0.182	0.270	$\sum_{i=1}^{3} p_i = 1$
2	.670	.160	.170	$\sum_{i=1}^{3} p_i = 1, \sum_{i=1}^{3} ip_i = 1.5$

We see that for the situation where the constraint that the expectation of i remains at 1.5 still holds (i.e., case 2), the constraint acts to increase the posterior probability values for small values of i and lower those for larger values of i. In case 1, the inclusion of the experimental measurements coupled with a relaxation of the constraint of equation (42) served to increase the estimate of the expectation of i from 1.5 at the outset to the value of 1.722.

APPLICATION TO CONTINUOUS RANDOM VARIABLE

We wish to modify the generalized method so that it applies to a continuous random variable. Let the random variable we are considering be denoted by X and let $f(x)$ denote the probability density function of X so that $f(x)dx$ designates the probability that X lies in dx at x, that is, the probability that $x \leq X \leq x + dx$.

To establish the least-biased prior probability density function $\hat{f}_0(x)$ for X, Jaynes' method is used as previously described but, inasmuch as the random variable is now continuous where before it was discrete, the sums become integrals. Hence, $\hat{f}_0(x)$ is found by the calculus of variations technique as that function which maximizes

$$S = -\int_{-\infty}^{\infty} f(x) \ln f(x)dx \tag{43}$$

subject to the constraint

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{44}$$

that always holds and any other constraints written as

$$\int_{-\infty}^{\infty} g_k(x)f(x)dx = \langle g_k \rangle, \; k = 1, 2, \ldots, K \tag{45}$$

that would express the extent of the experimenter's prior knowledge. The result of the maximization is given by

$$\hat{f}_0(x) = \exp\left[-\lambda_0 - \sum_{k=1}^{K} \lambda_k g_k(x)\right] \tag{46}$$

where $\lambda_0 - 1, \lambda_1, \lambda_2, \ldots, \lambda_k$ are Lagrangian multipliers.

Obviously, once the prior distribution $\hat{f}_0(x)$ has been established, the initial number M_0 of initial hypothetical measurements can be chosen in accordance with our degree of credibility. Thus, the prior distribution $\hat{f}_0(x)$ can be said to be based on M_0 measurements.

Now, we arrive at the problem of incorporating in our distribution any measurements of X that have been made as well as any posterior constraints that apply. This problem can easily be resolved if we look at equation (41). There we see that, in the absence of any constraints except that of equation (1), $\left(\hat{p}_D\right)_i$ is proportional to the sum of the number of initial hypothetical measurements that have yielded outcome i and number of actual measurements that have yielded outcome i. The posterior constraints merely serve to modify this sum. Therefore, the corresponding expression for the least-biased posterior probability density function $f_D(x)$ of the continuous random variable X must be

$$\hat{f}_D(x) = \left[M_0 f_0(x) + \sum_{i=1}^{M} \delta(x - x_i)\right] \exp\left[-\nu_0 - \sum_{q=1}^{Q} \nu_q h_q(x)\right] \tag{47}$$

where the data $\left\{x_1, x_2, \ldots x_m\right\}$ are the M values that have been observed in M measurements of X and $\delta(x - x_i)$ is the Dirac delta function with argument $(x - x_i)$. In equation (47), $\nu_0 - 1, \nu_1, \nu_2, \ldots, \nu_Q$ are Lagrangian multipliers and the posterior constrains that apply are

$$\int_{-\infty}^{\infty} h_q(x)\hat{f}_D(x)dx = \langle h_q \rangle, \; q = 1, 2, \ldots, Q \tag{48}$$

It is to be noted that, although $\hat{f}_D(x)$ as written appears to describe a variable of the so-called mixed type, partly discrete and partly continuous, it is not claimed that X is such a variable. Rather it is claimed that $\hat{f}_D(x)$ is an unbiased probability density function for X based on incomplete knowledge about the distribution of the variable X and that from $\hat{f}_D(x)$ one can derive unbiased estimates for the expectations of statistics involving X. It might be possible to devise an objective method for incorporating the measured values of X in the probability density function $\hat{f}_D(x)$ in such a way that, say, $\hat{f}_D(x)$ and its derivative remain continuous throughout the domain of x but such a method will not be attempted herein.

EXAMPLES INVOLVING CONTINUOUS VARIABLE

Bounded Continuous Variable

In this first example, we consider a simple case wherein the sample space for the continuous variable X is the open interval (0, 1). The prior information is given as

$$\int_0^1 x\, f(x)dx = 0.4 \qquad (49)$$

Then, by equation (46), the least-biased prior probability density function is of the form

$$\hat{f}_0(x) = e^{-\lambda_0 - \lambda_1 x}$$

where λ_0 and λ_1 are constants whose values are determined by equations (44) and (49). We obtain

$$\hat{f}_0(x) = 1.74\, e^{-1.23x} \qquad 0 < x < 1 \qquad (50)$$

We take M_0 to be 4; that is, we assume the prior distribution is worth 4 measurements.

Now, suppose two independent measurements of X were made and the values observed were $x_1 = 0.817$ and $x_2 = 0.574$. Let us consider two cases:

(1) There are no posterior constraints. We wish to obtain least-biased estimates of E[X], E[X^2], and the probability that $X < 0.5$.

(2) The constraint given by equation (49) still applies a posteriori. This means that E[X], the expectation value of X, remains fixed. We wish to obtain least-biased estimates of $P(X > 0.5)$ and E[X^2].

Based on the prior distribution $\hat{f}_0(x)$ given in equation (50), we get

$$\hat{P}_0(X > 0.5) = \int_{1/2}^{1} \hat{f}_0(x)dx = 0.351 \tag{50a}$$

$$\hat{E}_0[X^2] = \int_0^1 x^2\, \hat{f}_0(x)dx = 0.237 \tag{50b}$$

The results for the two cases are as follows:

Case (1):

$$\hat{f}_D(x) = \frac{1}{6}\left[4(1.74)e^{-1.23x} + \delta(x - 0.817) + \delta(x - 0.574)\right] \qquad 0 < x < 1 \tag{51}$$

$$\hat{P}_D(X > 0.5) = \int_{1/2}^{1} \hat{f}_D(x)dx = 0.567 \tag{51a}$$

$$\hat{E}_D[X] = \int_0^1 x\, \hat{f}_D(x)dx = 0.499 \tag{51b}$$

$$\hat{E}_D[X^2] = \int_0^1 x^2\, \hat{f}_D(x)dx = 0.324 \tag{51c}$$

Case (2):

$$\hat{f}_D(x) = 2.06\, e^{-2.51x} + 0.296\, e^{-1.28x}\left[\delta(x - 0.817) + \delta(x - 0.574)\right] \qquad 0 < x < 1 \tag{52}$$

$$\hat{P}_D(X > 0.5) = \int_{1/2}^{1} \hat{f}_D(x)dx = 0.413 \tag{52a}$$

$$\hat{E}_D[X^2] = \int_0^1 x^2\, \hat{f}_D(x)dx = 0.236 \tag{52b}$$

Now, suppose that one were interested in establishing a least-biased estimate for the variance ${\sigma_X}^2$ of X for the prior and posterior situations for each of the previous two cases. In order to arrive at these estimates, we need to find out just how our prior and posterior knowledge arises. The reason that such knowledge is required is shown by the following. We have

$$\widehat{\sigma_X^2} = \hat{E}\left[\left(X - \mu_X\right)^2\right] = \hat{E}\left\{\left[\left(X - \hat{E}[X]\right) + \left(\hat{E}[X] - \mu_X\right)\right]^2\right\}$$

$$= \hat{E}[X^2] - \left(\hat{E}[X]\right)^2 + \hat{E}\left[\left(\hat{E}[X] - \mu_X\right)^2\right] \tag{53}$$

where the variance of X is denoted by σ_X^2 and the mean of X by μ_X. Hence, in order to determine $\widehat{\sigma_X^2}$, we not only need to know $\hat{E}[X]$ and $E[X^2]$ but also an estimate for the last term on the right side of equation (53) which depends on how close $\hat{E}[X]$ is to the true mean μ_X of X. Because the constraint (eq. (49)) sets the value for $\hat{E}[X]$, we have to know the reasons underlying equation (49) in order to estimate σ_X^2.

Let us suppose that, in case (1) at the outset, the prior information consists of knowing only the average value 0.4 of 4 independent measurements of X but nothing else about these measurements. Then

$$\hat{E}_0[X] = 0.4$$

represents an estimate based on $M_0 = 4$ measurements and we can write

$$\hat{E}\left[\left(\hat{E}[X] - \mu_X\right)^2\right] = \frac{\sigma_X^2}{M_0} = \frac{\sigma_X^2}{4}$$

Substituting this relation into equation (53), we get for case (1) where we have employed the values in equations (49) and (50b)

$$\left(\widehat{\sigma_X^2}\right)_0 = \frac{\hat{E}_0[X^2] - \left(\hat{E}_0[X]\right)^2}{1 - \frac{1}{4}} = \frac{0.237 - (0.4)^2}{\left(\frac{3}{4}\right)} = 0.103$$

In addition, the estimate $\hat{E}_D[X]$ as given by equation (51b) is based on $M + M_0 = 6$

measurements. Therefore, by employing the values in equations (51b) and (51c), we find

$$\left(\widehat{\sigma_X^2}\right)_D = \frac{\hat{E}_D[X^2] - \left(\hat{E}_D[X]\right)^2}{1 - \frac{1}{6}} = 0.090$$

On the other hand, suppose, in case (2), the prior information as given by equation (49) represents the average of a very large number of measurements so that we can consider 0.4 as being very close to the true mean μ_X. Then the last term in equation (53) is essentially zero so that

$$\left(\widehat{\sigma_X^2}\right)_0 \cong \hat{E}_0[X^2] - \left(\hat{E}_0[X]\right)^2 = 0.077$$

and

$$\left(\widehat{\sigma_X^2}\right)_D \cong \hat{E}_D[X^2] - \left(\hat{E}_D[X]\right)^2 = 0.236 - (0.4)^2 = 0.076$$

It should be fairly obvious how to treat an intermediate situation where the value of the average of X as given in equation (49) is based on, say, 20 measurements. In this instance, the posterior constraint would no longer be equation (49) because, after measurements of 0.817 and 0.574 for X, the value of the average would change from 0.4 to

$$\frac{20\times 0.4 + 0.817 + 0.574}{22} = 0.427$$

Hence, in this instance, the posterior constraint would become

$$\int_0^1 x\, \hat{f}_D(x)dx = 0.427$$

which change would in turn modify $\hat{f}_D(x)$. Also, the value of $\hat{E}_D[X]$ would now be considered as based on 22 measurements of X when calculating $\left(\widehat{\sigma_X^2}\right)_D$.

Time-to-Failure

Here we are interested in the time-to-failure of a given part subjected to a prescribed loading when the material is manufactured according to certain specifications. Suppose we have reason to guess initially that the average time-to-failure is about 10 hours and suppose that five measurements were made, in three of which the failure times were 7, 8, and 12 hours and in the remaining two, for one reason or another, the experiment was stopped before failure at the runout times of 5 and 10 hours. What we wish to find, on the basis of what has been given, is an objective estimate of the expectations of the first and second moments of the time-to-failure plus, say, an estimate of the probability that the survival time is larger than 10 hours. We shall arrive at estimates for three cases, namely, when the initial guess is equivalent to $M_0 = 1, 2,$ or 5 measurements.

We first have to find the least-biased prior distribution for the time-to-failure. Letting $f(t)dt$ be the probability that failure occurs at time t in dt (t in units of hr) we can maximize

$$S = -\int_0^\infty f(t) \ln f(t)dt$$

with respect to the probability density function $f(t)$ subject to the constraints

$$\int_0^\infty f(t)dt = 1, \int_0^\infty t\, f(t)dt = 10$$

to find the least-biased estimate for the prior probability density function

$$\hat{f}_0(t) = 0.1\, e^{-0.1t} \tag{54}$$

According to the least-biased prior probability distribution, the probability that the time-to-failure t lies between times t_a and t_b is equal to

$$P_0(t_a < t < t_b) = e^{-0.1t_a} - e^{-0.1t_b} \tag{55}$$

The average time-to-failure for failures occurring between times t_a and t_b is for the prior distribution

$$\hat{E}_0[t]\bigg|_{(a,b)} = 10\left[\frac{(1 + 0.1t_a)e^{-0.1t_a} - (1 + 0.1t_b)e^{-0.1t_b}}{e^{-0.1t_a} - e^{-0.1t_b}}\right] \tag{56}$$

The average square of the time-to-failure for failures occurring between times t_a and t_b is

$$\hat{E}_0[t^2]\bigg|_{(a,b)} = \frac{10^2\left\{\left[(0.1t_a)^2 + 2(0.1t_a) + 2\right]e^{-0.1t_a} - \left[(0.1t_b)^2 + 2(0.1t_b) + 2\right]e^{-0.1t_b}\right\}}{e^{-0.1t_a} - e^{-0.1t_b}} \tag{57}$$

Also the expectation of the square of the time-to-failure for the least-biased prior distribution is

$$\hat{E}_0[t^2] = 200 \text{ hr}^2$$

Let us follow the method in detail for the case $M_0 = 5$. It is convenient to consider the two runout measurement times of 5 and 10 hours as dividing the time-to-failure axis into the three intervals:

I: $0 \leq t \leq 5$ hr

II: 5 hr $< t \leq 10$ hr

III: $t > 10$ hr

Then, by equation (55), we can find the prior probability estimate of failure occurring in each of these intervals. Also, we can make use of equations (56) and (57) and compile a table based on the prior probability estimates:

	Interval		
	I	II	III
$M_0\left(\hat{p}_0\right)_i$	1.97	1.19	1.84
$\left(\hat{E}_0[t]\right)_i$	2.29	7.29	20
$\left(\hat{E}_0[t^2]\right)_i$	7.31	55.2	499

Thus, before measurements begin, on the basis of our initial knowledge and the assumption that $M_0 = 5$, there are 1.97 measurements in interval I with an average t-measurement of 2.29 hours and an average t^2-measurement of 7.31 hour^2, 1.19 measurements in interval II with an average t-measurement of 7.29 hours and an average t^2-measurement of 55.2 hour^2, and 1.84 measurements in interval III with an average t-measurement of 20 hours and an average t^2-measurement of 499 hour^2.

These figures can first be altered so as to incorporate the three time-to-failure measurements of 7, 8, and 12 hours. The altered probability density function is

$$\hat{f}_A(t) = \frac{1}{8}\left[5\,\hat{f}_0(t) + \delta(t - 7) + \delta(t - 8) + \delta(t - 12)\right] \tag{58}$$

and the table based on this altered probability distribution is

	Interval		
	I	II	III
Number of measurements	1.97	3.19	2.84
$\left(\hat{E}_A[t]\right)_i$	2.29	7.43	17.18
$\left(\hat{E}_A[t^2]\right)_i$	7.31	56.05	375.0

These tabulated values may be calculated in an obvious way using the values from the table based on $f_0(t)$. For example, the 7-hour and 8-hour failure times fall in interval II so that

$$\left(\hat{E}_A[t]\right)_{II} = \frac{1.19\times 7.29 + 7 + 8}{3.19} = 7.43$$

$$\left(\hat{E}_A[t^2]\right)_{II} = \frac{1.19\times 55.2 + (7)^2 + (8)^2}{3.19} = 56.05$$

The number of measurements in each interval now have to be changed in accordance with the two runout time measurements of 5 and 10 hours. The 10-hour measurement increases by one the number of measurements in interval III to 3.84 whereas the 5-hour measurement is to be distributed over intervals II and III. If we use the numbers in intervals II and III to represent a relative probability of a measurement arriving there, then the fraction of the runout time 5-hour measurement ascribed to interval II is $(3.19)/(3.19 + 3.84) = 0.454$ while the fraction ascribed to interval III is $(3.84)/(3.10 + 3.84) = 0.546$. Hence, our table now becomes

	Interval		
	I	II	III
Number of measurements	1.97	3.644	4.386
$(\hat{p}_D)_i$	.197	.3644	.4386
$(\hat{E}_D[t])_i$	2.29	7.43	17.18
$(\hat{E}_D[t^2])_i$	7.31	56.05	375.0

where

$$(\hat{p}_D)_i = \frac{(\text{number of measurements in interval } i)}{M_0 + M}$$

so that, for $M_0 = 5$,

$$(\hat{p}_D)_i = \frac{(\text{number of measurements in interval } i)}{10}$$

The same kind of reasoning also gives the least-biased estimate of the posterior probability density function as

$$\hat{f}_D(t) = \begin{cases} 0.5\,\hat{f}_0(t), & 0 \le t \le 5 \\ 0.114\left[5\,\hat{f}_0(t) + \delta(t-7) + \delta(t-8)\right], & 5 < t \le 10 \\ 0.155\left[5\,\hat{f}_0(t) + \delta(t-12)\right], & t > 10 \end{cases} \tag{59}$$

Thus, for $M_0 = 5$, we have

$$\hat{E}_D[t] = \sum_i \left(\hat{p}_D\right)_i \left(\hat{E}_D[t]\right)_i = (0.197)(2.29) + (0.3644)(7.43) + (0.4386)(17.18) = 10.70 \text{ hr}$$

and

$$\hat{E}_D[t^2] = \sum_i \left(\hat{p}_D\right)_i \left(\hat{E}_D[t^2]\right)_i = 186.5 \text{ hr}^2$$

Also, an objective estimate of the probability that a randomly selected part will last at least 10 hours is given by

$$\left(\hat{p}_D\right)_{III} = 0.44$$

These values enable us to obtain a posteriori an estimate for the variance of the time-to-failure t. Again, we resort to equation (53) where $X \to t$ and see that we have to arrive at an estimate of the value of $\hat{E}\left[\left(\hat{E}[t] - \mu_t\right)^2\right]$ under the conditions that $\hat{E}[t]$ was determined by a total of $M + M_0 = 8$ measurements of the time-to-failure t and $M_r = 2$ measurements of runout times. Obviously, a measurement of runout time would not be expected to be as effective as a measurement of the time-to-failure t in fixing the value of $\hat{E}[t]$ close to μ_t.

As shown in any textbook treating stratified sampling, the variance of t can be written

$$\sigma_t^2 = \sum_i p_i \left(\sigma_t^2\right)_i + \sum_i p_i \left[\left(\mu_t\right)_i - \mu_t\right]^2 \tag{60}$$

where

p_i probability that t lies in interval i

$\left(\sigma_t^2\right)_i$ variance of t if t is constrained to lie in interval i

$\left(\mu_t\right)_i$ mean of t if t is constrained to lie in interval i

It can be demonstrated that the variance of an average value $\bar{t}$ of t based on $(M + M_0)$ measurements of t and M_r measurements of runout times is given by the formula

$$\sigma_{\bar{t}}^2 = \frac{\sum_i p_i \left(\sigma_t^2\right)_i}{M + M_0} + \frac{\sum_i p_i \left[\left(\mu_t\right)_i - \mu_t\right]^2}{M + M_0 + M_r} \tag{61}$$

provided the class boundaries of the intervals are the runout time measurements. Equation (61) can be written

$$\sigma_{\bar{t}}^2 = \frac{\sigma_t^2}{M + M_0} - \frac{M_r}{(M + M_0)(M + M_0 + M_r)} \sum_i p_i \left[\left(\mu_t\right)_i - \mu_t\right]^2 \tag{62}$$

Now

$$\widehat{\sigma_{\bar{t}}^2} \equiv \hat{E}\left[\left(\hat{E}[t] - \mu_t\right)^2\right] \tag{63}$$

so that applying equation (53) we obtain the desired relation

$$\widehat{\sigma_t^2} = \frac{\hat{E}[t^2] - \left(\hat{E}[t]\right)^2 - \dfrac{M_r}{(M + M_0)(M + M_0 + M_r)} \sum_i \hat{p}_i \left[\left(\hat{E}[t]\right)_i - \hat{E}[t]\right]^2}{1 - \dfrac{1}{M + M_0}} \tag{64}$$

Using the values obtained for the $M_0 = 5$ case, we get

$$\left(\widehat{\sigma_t^2}\right)_D = \frac{1}{1 - \frac{1}{8}} \left\{186.5 - (10.70)^2 - \frac{2}{8(10)}\left[0.197(2.29 - 10.70)^2 + 0.3644(7.43 - 10.70)^2 \right.\right.$$

$$\left.\left. + 0.4386(17.18 - 10.70)^2\right]\right\} = 81.3 \text{ hr}^2$$

A tabulation of least-biased estimates of the various quantities is given in the following:

M_0	$\hat{p}$			$\hat{E}[t]$			$\hat{E}[t^2]$			$\hat{E}[t]$	$\hat{E}[t^2]$	$\widehat{\sigma_t^2}$
	I	II	III	I	II	III	I	II	III			
(Prior)	0.3935	0.2386	0.3676	2.29	7.29	20	7.31	55.2	499	10	200	----
5	.1968	.3644	.4386	↓	7.43	17.18	↓	56.05	375	10.70	186.5	81.3
2	.1124	.4217	.4659	↓	7.465	15.38	↓	56.2	295	10.57	162	61.2
1	.0656	.4541	.4803	↓	7.475	14.15	↓	56.25	240	10.35	141	43.5

Fatigue Stress Distribution Based on Runout Data Alone

In this example, we will consider the same problem treated by Shah (ref. 17). He was interested in determining an objective estimate of the probability that a certain manufactured part will last longer than 10^7 cycles under an alternating fatigue stress load of 140.1×10^3 newtons per square meter (20 325 psi). He had the results of four fatigue tests run at different values of the alternating stress σ. In all four of these tests, the part did not fail in 10^7 cycles even though the stress loading in each test was much larger than 140.1×10^3 newtons per square meter. The values of the alternating stresses for the four tests were 266.5×10^3, 296.2×10^3, 377.4×10^3, and 247.9×10^3 newtons per square meter (38 650, 42 960, 54 730, and 35 960 psi). In his paper, Shah used the value of 284.6×10^3 newtons per square meter (41 270 psi) as his prior estimate for the average stress at which failure at 10^7 cycles occurs.

Let $f(\sigma)d\sigma$ be interpreted as the probability that the alternating stress σ in $d\sigma$ will cause fatigue failure in a randomly selected manufactured part after exactly 10^7 cycles. Hence the probability that a part will last longer than 10^7 cycles under a stress load of σ' is given by the integral

$$\int_{\sigma'}^{\infty} f(\sigma)d\sigma$$

which is also the probability that a stress larger than σ' will be needed to cause breakage at exactly 10^7 cycles.

We handle this problem in the same way as we did the previous problem. The σ-axis is divided into six intervals where the runout stress load measurements and the stress we are interested in serve as the boundaries. Thus, the six intervals are the following:

I: $0 \leq \sigma \leq 140.1\times10^3$ N/m^2

II: 140.1×10^3 N/m$^2 < \sigma \leq 247.9\times10^3$ N/m^2

III: 247.9×10^3 N/m$^2 < \sigma \leq 266.5\times10^3$ N/m^2

IV: 266.5×10^3 N/m$^2 < \sigma \leq 296.2\times10^3$ N/m^2

V: 296.2×10^3 N/m$^2 < \sigma \leq 377.4\times10^3$ N/m^2

VI: $\sigma > 377.4\times10^3$ N/m^2

The least-biased probability density function for the stresses at which the part lasts exactly 10^7 cycles based on the initial guess that $\bar{\sigma}_0 = 284.6\times10^3$ newtons per square meter (41 270 psi) is of the same form as that given in equation (39) for the time-to-failure

$$\hat{f}_0(\sigma) = \frac{1}{\bar{\sigma}_0} C^{-\sigma/\bar{\sigma}_0} \tag{65}$$

and again, as we did before, we shall carry out the calculations for several different values of M_0; in this problem, let us take the initial guess as worth $M_0 = 1$, 2, 4, or 10 measurements.

The fact that all the measurements are runout stresses means that the estimated average value of the stress in each of the intervals remains constant as the measurements proceed. Only the probability distribution over the intervals changes. The results of the calculations are tabulated as follows:

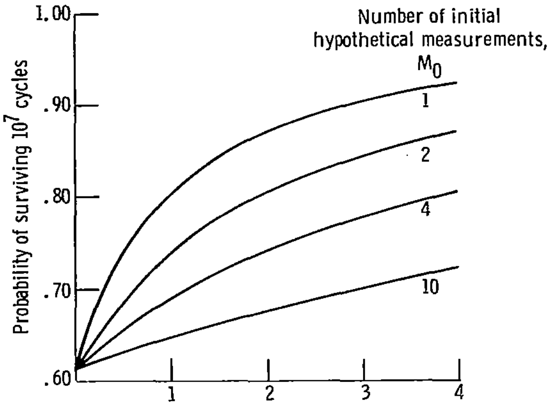
Interval					
I	II	III	IV	V	VI
Estimated average stress for interval, $\hat{E}[\sigma]$, N/m^2					
64.5×10^3	190.0×10^3	256.8×10^3	280.6×10^3	335.1×10^3	661.9×10^3

M_0	$\hat{p}_I$	$\hat{p}_{II}$	$\hat{p}_{III}$	$\hat{p}_{IV}$	$\hat{p}_V$	$\hat{p}_{VI}$	$\hat{E}[\sigma]$
Prior probabilities							
	0.389	0.192	0.027	0.0385	0.088	0.2655	284.6×10^3
$\sigma_{ro} = 266.5\times10^3$ N/m^2 (38 650 psi)							
10	0.354	0.175	0.0246	0.0439	0.100	0.303	308.8×10^3
4	.311	.154	.0216	.0505	.115	.348	337.9
2	.259	.128	.018	.0584	.134	.403	373.6
1	.1945	.096	.0135	.0684	.1562	.4714	417.8
$\sigma_{ro} = 296.2\times10^3$ N/m^2 (42 960 psi) + previous 1							
10	0.324	0.160	0.0225	0.0386	0.113	0.342	331.9×10^3
4	.259	.128	.018	.0357	.139	.420	380.2
2	.195	.096	.0135	.0300	.166	.500	429.3
1	.1297	.064	.0090	.022	.193	.582	479.1
$\sigma_{ro} = 377.4\times10^3$ N/m^2 (54 730 psi) + previous 2							
10	0.299	0.148	0.0208	0.0346	0.105	0.393	357.6×10^3
4	.222	.110	.0154	.0281	.0910	.533	430.7
2	.156	.0768	.0108	.0209	.0759	.660	495.9
1	.097	.048	.00675	.0137	.0543	.780	555.5
$\sigma_{ro} = 247.9\times10^3$ N/m^2 (35 960 psi) + previous 3							
10	0.278	0.137	0.0220	0.0366	0.102	0.424	375.0×10^3
4	.195	.096	.0164	.0299	.0967	.567	450.8
2	.1297	.064	.0113	.0220	.0797	.693	515.5
1	.0778	.0384	.00698	.0141	.0561	.807	570.9

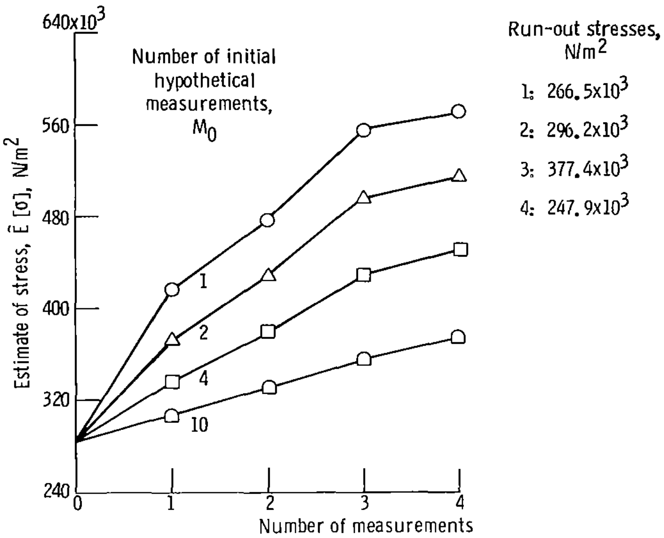
The probability that the part lasts more than 10^7 cycles under an alternating stress of 140.1×10^3 newtons per square meter (20 325 psi) is plotted in figure 1(a) as a function of the number of measurements and M_0. The curves satisfy a simple relation because none of the runout stresses were less than 140.1×10^3 newtons per square meter. Hence, no portion of any of the four measurements is allotted to interval I and

$$\left(\hat{p}_D\right)_I = \frac{M_0\left(\hat{p}_0\right)_I}{M_0 + M} = \frac{0.389\ M_0}{M_0 + M}$$

where M is the number of actual measurements. Obviously the estimated probability



(a) Estimate of probability of surviving 10^7 cycles at stress level of 140.1×10^3 newtons per square meter (20 325 psi).



(b) Estimate of average stress for which failure occurs at 10^7 cycles.

Figure 1. - Fatigue stress example.

that the part lasts longer than 10^7 cycles under a stress of 140.1×10^3 newtons per square meter is

$$1 - \left(\hat{p}_D\right)_I = \frac{M + 0.611\ M_0}{M_0 + M}$$

In figure 1(b), a posterior estimate of the average stress for which failure occurs at 10^7 cycles is plotted as a function of the number of measurements and the value of M_0.

CONCLUDING REMARKS

A method has been devised that enables the calculation of objective and reasonable posterior probability estimates for both discrete and continuous sample spaces. It has been shown that the method agrees with the rule of Bayes and provides a simple interpretation of Laplace's "principle of insufficient reason" and "rule of succession".

The procedure is based on a reformulation of the expression for entropy first suggested by Jaynes for arriving at "least-biased" probability estimates and extends Jaynes' reasoning to the situation wherein experimental data is on hand in addition to any constraints that may apply a posteriori.

Lewis Research Center,
National Aeronautics and Space Administration,
Cleveland, Ohio, October 28, 1971,
132-15.

REFERENCES

1. Cornfield, Jerome: The Bayesian Outlook and its Application. Bimetrics, vol. 25, no. 4, Dec. 1969, pp. 617-657.
2. Raiffa, Howard; and Schlaifer, Robert O.: Applied Statistical Decision Theory. Harvard Univ., Boston, 1961.
3. Discussion to reference 1, Biometrics, vol. 25, no. 4, Dec. 1969, pp. 643-657.
4. Guttman, Irwin: Tolerance Regions: A Survey of its Literature. VI - The Bayesian Approach. Tech. Rept. 127, pt. 6, Univ. of Wisconsin, Aug. 1968. (DDC No. AD-686294.)
5. Watanabe, Satosi: Knowing and Guessing - A Quantitative Study of Inference and Information. John Wiley & Sons, Inc., 1969.
6. Jaynes, E. T.: New Engineering Applications of Information Theory. Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability. J. L. Bogdanoff and F. Kozin, eds., John Wiley & Sons, Inc., 1963, pp. 163-203.
7. Shannon, Claude E.; and Weaver, Warren: The Mathematical Theory of Communication. Univ. Illinois Press, 1949.
8. Jaynes, E. T.: Information Theory and Statistical Mechanics. Phys. Rev., vol. 106, no. 4, May 15, 1957, pp. 620-630.
9. Jaynes, E. T.: Information Theory and Statistical Mechanics. II. Phys. Rev., vol. 108, no. 2, Oct. 15, 1957, pp. 171-190.
10. Tribus, Myron; and Fitts, Gary: The Widget Problem Revisited. IEEE Trans. on Syst. Sci. Cybernetics, vol. SSC-4, no. 3, Sept. 1968, pp. 241-248.
11. Tribus, Myron: The Use of the Maximum Entropy Estimate in the Estimation of Reliability. Recent Developments in Information and Decision Processes. Robert E. Machol and Paul Gray, eds., Macmillan Book Co., 1962, pp. 102-140.
12. Rowlinson, J. S.: Probability, Information and Entropy. Nature, vol. 225, no. 5239, Mar. 28, 1970, pp. 1196-1198.
13. Jaynes, Edwin T.: Prior Probabilities. IEEE Trans. on Syst. Sci. Cybernetics, vol. SSC-4, no. 3, Sept. 1968, pp. 227-241.
14. Abramowitz, Milton; and Stegun, Irene A.: Handbook of Mathematicsl Functions with Formulas, Graphs, and Mathematical Tables. Appl. Math. Ser. 55, National Bureau of Standards, June 1964.

15. Jeffreys, Harold: Theory of Probability. Third ed., Oxford Univ. Press, 1961.

16. Good, Irving J.: The Estimation of Probabilities. MIT Press, 1965.

17. Shah, Haresh C.: Reliability Estimates from Run-Out or Non-Failure Data Using Information Theory. Transections of the 22nd Annual Technical Conference of the American Society for Quality Control, 1968, pp. 425-434.

NASA-Langley, 1972 —— 19 E-6382