PROC. OF JSCE
No. 335, July 1983

DETERMINATION OF DESIGN VALUES FROM STATISTICAL DATA

By Yozo FUJINO*, Akio HASEGAWA**
and Fumio NISHINO***

1. INTRODUCTION

The assurance of structural safety is an important objective in structural design. The reliability of a structure depends on the relative magnitudes of the strength of structures and the magnitude of the loads. Apparently, both have a non-deterministic nature. Probabilistic rationales are widely accepted as appropriate for such circumstances, and much effort has been made to implement probabilistic rationales into structural design to account for uncertainty.

The design format in structural codes generally takes the form

$$\nu \frac{g(S^*)}{R^*} \leq 1 , \qquad (1)$$

where R^* and S^* stand for the *design strength* and the *design load*, respectively, collectively called the "design values", while ν is the so-called *safety factor* accounting for the importance of the structure and for miscellaneous uncertainties not covered explicitly in R^* and S^*. The function $g(\cdot)$ reflects the structural analysis.

The design strength R^*, the design load S^* and the factor ν are to be determined such that a specified safety level is attained. In probability-based design, the design strength R^* and the design load S^* are usually taken to be the *characteristic values* (fractiles) corresponding to certain exceedance probabilities q_R and q_S respectively[1], as shown in **Fig. 1**.

In general, the strength R and the load S are functions of a large number of random variables, but in practice they may be lumped together into a small number of variables. For example, the ultimate strength R of steel members is commonly expressed as a product of three mutually independent *basic* random variables[2]

$$R = MFP , \qquad (2)$$

where M, F and P are random variables associated with material strength, fabrication, and the ultimate strength prediction formula used in the analysis, respectively.

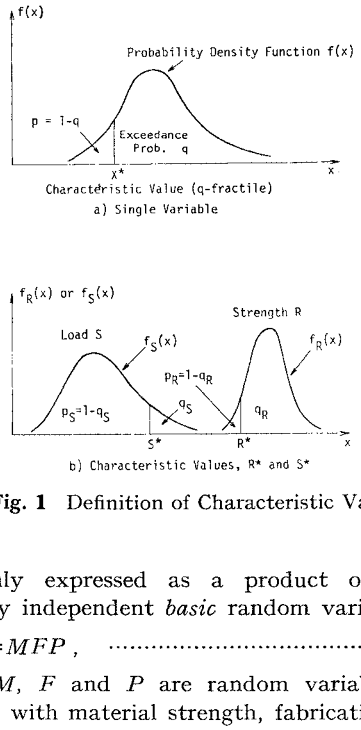


Fig. 1 Definition of Characteristic Values.

If the distributions of the random variables are completely known, R^* and S^* corresponding to the specified exceedance probabilities can be calculated analytically or numerically. However, the information about each random variable is usually limited, available in the form of statistical data of finite size. Furthermore, the size of the statistical data is limited for technical or economical reasons. For example, experimental data of the strength of full scale steel members is very limited, while experimental data on material strength is abundant. Thus the data sizes of the uncertainties M and P in Eq. (2) are generally quite different. Similar facts can also be observed regarding the data sizes of random variables associated with the load S. It should be recognized that the design values R^* and S^*

* Member of JSCE Ph. D., Assoc. Prof. of Civil Eng., Univ. of Tokyo, Tokyo 113
** Member of JSCE Dr. Eng., Assoc. Prof. of Civil Eng., Univ. of Tokyo, Tokyo
*** Member of JSCE Ph. D., Prof. of Civil Eng., Univ. of Tokyo, Tokyo

ought to be determined taking this situation into account.

Most reliability-based design methods, including the second moment design methods proposed thus far, are formulated in a probabilistic (rather than statistical) fashion; the moment values and sometimes the distribution functions of all the associated random variables are assumed to be known. Few works discuss reliability-based design taking into account that the available data is of finite size. Among them, Refs. (3) and (4) have discussed straightforward application of Bayesian frequentist or fiducial statistics to the formulation of reliability-based design. In Ref. (5), a similar attempt was made in a second-moment safety-index based design. All these works place an emphasis on theoretical inclusion of the so-called "sampling uncertainty" into the reliability-based design formulation but they do not particularly consider the practical aspects of reliability-based design, such as the determination of the characteristic values from statistical data.

This paper discusses determination of the characteristic values, used as design values, from sample data.

Consider a continuous random variable X given by a function of basic random variables $X_1, X_2, \cdots, X_m$,

$$X=\psi(X_1, X_2, \cdots, X_m) \ . \qquad (3)$$

When all the probabilistic information of X_i is known, the characteristic value X^* of X having an exceedance probability q is uniquely determined. If m greater than 1 there obviously exist infinitely many sets of $\{X_i^*\}$ which satisfy

$$\psi(X_1^*, X_2^*, \cdots, X_m^*)=X^* \ . \qquad (4)$$

The following equation can be found

$$\psi_q(q_1, q_2, \cdots, q_m)=q \ , \qquad (5)$$

such that $\{X_i^*\}$ corresponding to respective $\{q_i\}$ satisfy Eq. (4). In other words, for any combination of $\{q_i\}$ satisfying Eq. (5), the resulting $\{X_i^*\}$ yield the unique value X^* through Eq. (4).

It is now assumed that statistical information on the basic random variables is available in the form of sample data of finite sizes $n_1, n_2, \cdots, n_m$ for each X_i. A characteristic value denoted by X_S^* determined from those sample data by any statistical method is necessarily subject to error, the so-called "sampling error".

The objective of this paper is to make clear that, in order to accurately determine the characteristic value X_S^*, the difference among sample sizes of basic random variables should be taken into account. It is demonstrated that the magnitude of the error of the characteristic value X_S^* depends on the choice of the exceedance probabilities $\{q_i\}$ even though Eq. (5) is satisfied, and furthermore that there exists a particular set $\{q_i\}$ which minimizes the error of X_S^*. A method to determine the characteristic value X_S^* from given statistical data for a specified q is also proposed. This method is found useful in the Monte Carlo simulations. Finally determination of the optimal sample sizes $\{n_i\}$ of $\{X_i\}$ is presented, considering the cost required to collect the sample data of $\{X_i\}$.

The probabilistic model employed in this paper is very simple, yet sufficient to demonstrate the objective of this study.

2. UNCERTAINTY OF CHARACTERISTIC VALUES DETERMINED FROM FINITE SAMPLE DATA: THE SINGLE VARIABLE CASE

Consider a set of independent observations $\{x_i\}=(x_1, x_2, \cdots, x_n)$ of size n from a population having a probability density function $f(x)$. Let $x_{(1)}$, $x_{(i)}$ and $x_{(n)}$ be smallest, i-th smallest, and largest of $\{x_i\}$ respectively. There are many empirical rules to assign the cumulative probability $p_{(i)}$ $(=1-q_{(i)}$, where $q_{(i)}$ is the exceedance probability) to the quantity $x_{(i)}$. Among them, the most widely used rule may be[6),7)]

$$p_{(i)}=i/(n+1) \ . \qquad (6)$$

For observations from a normal population, the following has been suggested[8)]

$$p_{(i)}=(i-0.375)/(n+0.25) \ . \qquad (7)$$

When one would like to determine the characteristic value X_S^* corresponding to the exceedance probability q on the basis of the data of size n, it is common to adopt the value which is calculated from the cumulative probability given by Eq. (6) or (7), with the aid of interpolation if necessary. Note that this procedure is possible only for the value of q ranging from $1/(n+1)$ to $n/(n+1)$ when Eq. (6) is used and from $0.625/(n+0.25)$ to $(n-0.375)/(n+0.25)$ when Eq. (7) is used.

The quantities $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$, called *order statistics* of the sample n, are random variables. The probability density function $f_{(i)}(x)$ of the i-th order statistic $x_{(i)}$ is given by[6)]

$$f_{(i)}(x)=\frac{n!}{(i-1)!(n-i)!}F^{i-1}(x) \times[1-F(x)]^{n-1}f(x) \ , \qquad (8)$$

where $F(x)$ = the cumulative distribution function of the population. When the function $F(x)$ is known, the moments of $x_{(i)}$ such as the mean and the variance can be calculated by Eq. (8). However, the computations are increasingly

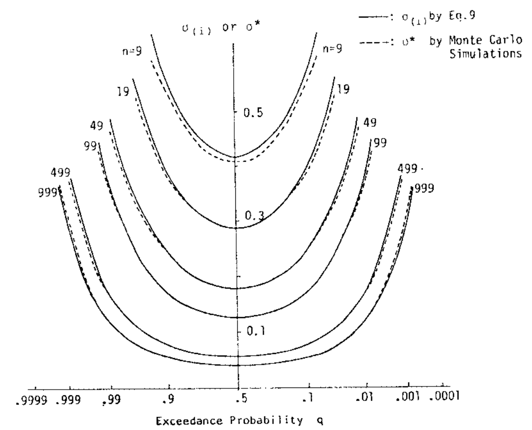


Fig. 2 Values of $\sigma_{(i)}$ and σ^*.

cumbersome for large sample size n. An approximate estimate of the mean square expectation $\sigma^2_{(i)}$ of $x_{(i)}$ about the median, $\tilde{x}_{(i)}$ is given by[6)]

$$\sigma^2_{(i)} \equiv E[(x_{(i)}-\tilde{x}_{(i)})^2] \doteqdot \frac{F(\tilde{x}_{(i)})[1-F(\tilde{x}_{(i)})]}{nf(\tilde{x}_{(i)})} . \qquad \cdots\cdots(9)$$

The mean square $\sigma^2_{(i)}$ in Eq. (9) indicates the variability of the characteristic value X^*_S statistically determined for the exceedance probability $q=1-(i-0.375)/(n+0.25)$ if the population is normally distributed.

The values of $\sigma^2_{(i)}$ for different sample size n and exceedance probabilities q $(=1-p$, where p is the cumulative probability) were calculated for a standardized normal variate. Eq. (7) was used to determine the values of $F(\tilde{x}_{(i)})$ in Eq. (9). The values thus computed by Eq. (9) are presented in **Fig. 2** which also shows the mean square error σ^{*2} of X^*_S, independently calculated by Monte Carlo simulation, for comparison. The abscissa in the figure is the exceedance probability q calculated by Eq. (7). **Fig. 2** indicates that: (1) the value σ^*, magnitude of the error of X^*_S, decreases for larger sample size n; (2) σ^* of X^*_S increases considerably as the exceedance probability q approaches either 0 or 1 and (3) the value of $\sigma_{(i)}$, given by Eq. (9) is a very good estimate of σ^* over a wide range of q unless the sample size n is very small.

3. UNCERTAINTY OF CHARACTERISTIC VALUES: THE CASE OF TWO BASIC RANDOM VARIABLES

The random variables, strength R and load S, are generally expressed as functions of multiple basic random variables, as in Eq. (3). We consider a very simple case where the random variable X can be expressed as the sum of two independent basic random variables X_1 and X_2 as

$$X=X_1+X_2 , \qquad \cdots\cdots(10)$$

and seek to determine the characteristic value X^* of X corresponding to a specified exceedance probability q.

Assume that the variables X_1 and X_2 are normally distributed with the means μ_1 and μ_2 and the variances $\sigma_1{}^2$ and $\sigma_2{}^2$, respectively. The characteristic values $X_1{}^*$ and $X_2{}^*$ are as determined from

$$\left.\begin{array}{l}\text{Prob. }(X_1>X_1{}^*)=q_1 ,\\ \text{Prob. }(X_2>X_2{}^*)=q_2 ,\end{array}\right\} \qquad \cdots\cdots(11)$$

where q_1 and q_2 are the exceedance probabilities. If q_1 and q_2 satisfy the equation

$$[\Phi^{-1}(q_1)\sigma_1+\Phi^{-1}(q_2)\sigma_2]/\sqrt{\sigma_1{}^2+\sigma_2{}^2}=\Phi^{-1}(q) , \qquad \cdots\cdots(12)$$

the quantity X^* defined by

$$X^*=X_1{}^*+X_2{}^* , \qquad \cdots\cdots(13)$$

is the characteristic value of X $(=X_1+X_2)$ corresponding to the exceedance probability q. The function $\Phi(\cdot)$ is the standard normal probabiity integral. Note that Eqs. (12) and (13) correspond to Eqs. (5) and (4) respectively.

Fig. 3 shows Eq. (12) drawn in a two-dimensional standard normal probability plot. To attain the exceedance probability $q=0.99$ in the case of $\sigma_1=\sigma_2$, one can assign 0.50 to 0.99 to q_1 and q_2 respectively, or 0.99 and 0.50, resulting in the same value of X^*; thus there are infinitely many choices of $\{q_1, q_2\}$.

Since the random variable $X=X_1+X_2$ is normally distributed, X^* can be expressed as

$$X^*=\mu_1+\mu_2+k\sqrt{\sigma_1{}^2+\sigma_2{}^2} , \qquad \cdots\cdots(14)$$

in which k is $\Phi^{-1}(q)$. The following linearization is often used in the formulation of probability based design[9)]

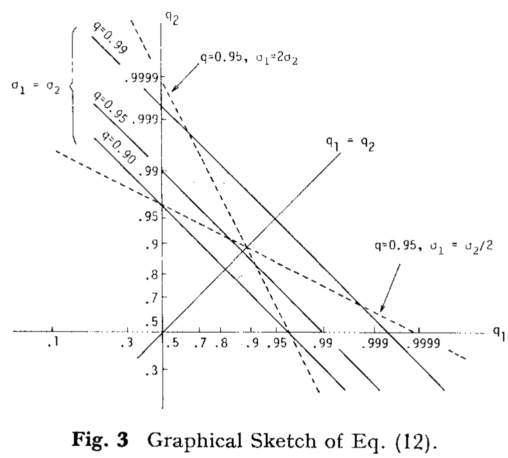


Fig. 3 Graphical Sketch of Eq. (12).

$$\sqrt{\sigma_1^2+\sigma_2^2} \fallingdotseq \alpha(\sigma_1+\sigma_2) \ , \quad \cdots\cdots(15)$$

in which the value of the linearization factor α is usually taken as 0.75. Substituting Eq. (15) into Eq. (14) yields

$$X^* = \mu_1+\mu_2+\alpha k(\sigma_1+\sigma_2) \ , \quad \cdots\cdots(16)$$

i.e.,

$$\left.\begin{array}{l} X_1^* = \mu_1+\alpha k\sigma_1 \ , \\ X_2^* = \mu_2+\alpha k\sigma_2 \ . \end{array}\right\} \quad \cdots\cdots(17)$$

Eq. (17) indicates that $q_1=q_2$ was chosen in the formulation of probability-base design in Ref. 9). For example, q_1 and q_2 both equal approximately 0.88 for exceedance probability $q=0.95$ in the case of $\sigma_1=\sigma_2$ as shown in **Fig. 3**. The two dotted lines in the figure show the reason that q_1 and q_2 were chosen equal. For, as the ratio σ_1/σ_2 varies with constant q, a family of lines is generated. With very good approximation this family is a line bundle through the point (0.88, 0.88).

We herein consider the situation that the information on basic random variables X_1 and X_2 is given in the form of statistical data of size n_1 and n_2, respectively. The question is now raised: Should indeed equal value be assigned to q_1 and q_2 to determine X_S^*, even when the sample size n_1 is very different from n_2?

The characteristic values X_{1S}^* and X_{2S}^* are calculated from the sample data by the procedure described before. Note that X_{1S}^* and X_{2S}^* are random variables. Since X_S^* is given by a sum of X_{1S}^* and X_{2S}^* as given by Eq. (13), the mean square error σ^{*2} of the chracteristic value X^* can be expressed in terms of the respective mean square error σ_1^{*2} and σ_2^{*2} of X_1^* and X_2^* as

$$\sigma^{*2}=\sigma_1^{*2}+\sigma_2^{*2} \ . \quad \cdots\cdots(18)$$

It should be remembered that the value of q_1 ranges from $0.625/(n_1+0.25)$ to $(n_1-0.375)/(n_1+0.25)$, while q_2 ranges from $0.625/(n_2+0.625)$ to $(n_2-0.375)/(n_2+0.25)$ since Eq. (7) is used to assign the cumulative probability. So, the value of σ^{*2} can be calculated for similarly restricted ranges of q_1 and q_2. Eq. (9) is used to estimate σ_1^{*2} and σ_2^{*2} and Eq. (12) is used as the relation between q_1 and q_2. Note that the median $\tilde{x}_{(i)}$ in Eq. (9) is calculated from the following equation

$$\tilde{x}_{(i)}=F^{-1}[(i-0.375)/(n+0.25)] \ . \quad \cdots\cdots(19)$$

The results for some cases are presented in **Fig. 4** a)~d), where the abscissa is the exceedance probability q_1 or q_2 and the ordinate is the non-dimensionalized root mean square error σ_e of X_S^* defined by

$$\sigma_e=\sigma^*/\sqrt{\sigma_1^2+\sigma_2^2} \ . \quad \cdots\cdots(20)$$

Fig. 4 a)~d) correspond to the following cases.

a) $q=0.90$, $\sigma_1=\sigma_2$, $n_1=19$, $n_2=49$, 99, 249, 999
b) $q=0.95$, $\sigma_1=\sigma_2$, $n_1=19$, $n_2=49$, 99, 249, 999
c) $q=0.99$, $\sigma_1=\sigma_2$, $n_1=19$, $n_2=49$, 99, 249, 999
d) $q=0.95$, $\sigma_1=\sigma_2$, $2\sigma_2$, $\sigma_2/2$, $n_1=19$, $n_2=99$, 999

It is interesting to observe in **Fig. 4** that the value of σ_e is a concave function of q_1 and that an optimal set of q_1 and q_2 values exists such that σ_e is minimized. Equal values of q_1 and q_2 do not give the minimum value $\sigma_{e,\min}$ of σ_e, indicating that different exceedance probabilities should be assigned, depending upon sample sizes. For a fixed sample size $n_1(=19)$, as the sample size n_2 increases, the optimal value q_1 decreases and thereby q_2 increases. It appears that in the case of large exceedance probability q, σ_e is rather sensitive to q_1 (**Fig. 4** a)~c)). This is also true in the case where σ_1 is greater than σ_2 (**Fig. 4** d)). On the other hand, the value of σ_e is very insensitive to q_1 when σ_1 is smaller than σ_2 (**Fig. 4** d)).

From **Fig. 4** a)~d) inclusive it can be concluded that, in order to determine a characteristic value having a target exceedance probability q greater than 0.5, the exceedance probability for the basic random variables with abundant statistical data should be larger than that for variables with relatively limited data. It is obvious that a similar argument can be made in the case of q less than 0.5. In other words, the design values based on sample data should be assigned relatively far from the mean of the distributions, while the design value based on the data of small size should be assigned relatively closer to the mean.

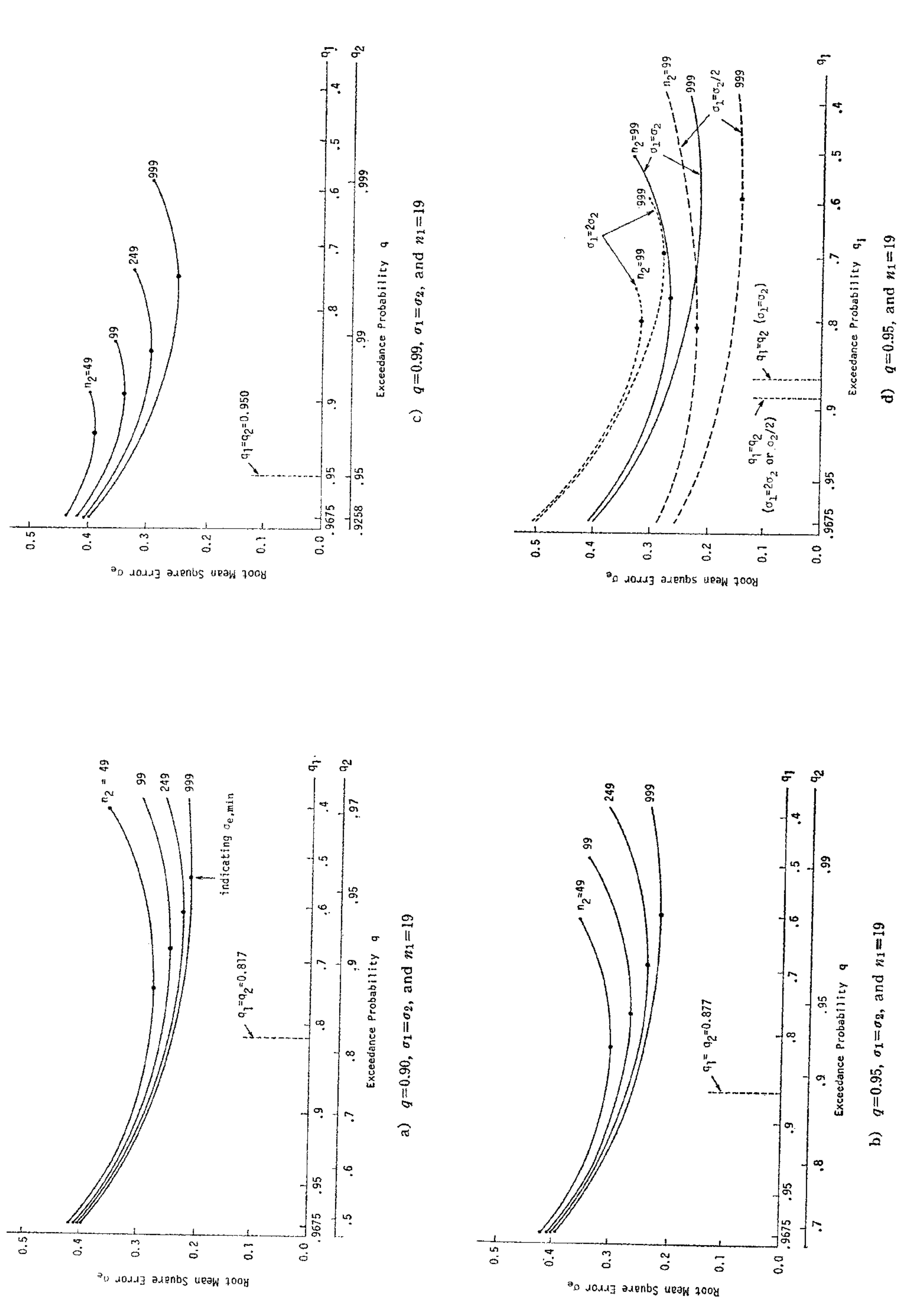
4. A METHOD TO DETERMINE CHARACTERISTIC VALUE FROM STATISTICAL DATA

Using a simple probabilistic model it was shown that weighted exceedance probabilities should be used for the basic random variables when the sizes of the statistical data are significantly different. In the previous sections, the variances σ_1^2 and σ_2^2 were assumed known for analytical purposes. However, these values are generally unknown in real life and must be estimated from sample data. Therefore, Eq. (12) can not be used as it is in order to find the optimal exceedance probabilities q_1 and q_2. The following equation may be used in lieu of Eq. (12)

$$[\Phi^{-1}(q_1)S_1+\Phi^{-1}(q_2)S_2]/\sqrt{S_1^2+S_2^2}=\Phi^{-1}(q) \ , \quad \cdots\cdots(21)$$

where S_1^2 and S_2^2 are the sample variances and unbiased estimators of σ_1^2 and σ_2^2, respectively. Note that replacement of σ_1 and σ_2 in Eq. (12) by S_1 and S_2 yields Eq. (21).

An alternative method to determine the char-



a) q=0.90, $\sigma_1=\sigma_2$, and n_1=19

b) q=0.95, $\sigma_1=\sigma_2$, and n_1=19

c) q=0.99, $\sigma_1=\sigma_2$, and n_1=19

d) q=0.95, and n_1=19

Fig. 4 Values of Root Mean Square Error σ_e.

acteristic value X_S^* from given statistical data $\{x_{1j}\}$ $(j=1-n_1)$ and $\{x_{2j}\}$ $(j=1-n_2)$ can be then described as follows: First, from Eq. (20) the optimal values of q_1 and q_2 for a given exceedance probability q are calculated under the criterion of minimizing σ_e. Next, the respective characteristic values X_{1S}^* and X_{2S}^* corresponding to the optimal values q_1 and q_2 are obtained from the statistical data, resulting in the characteristic value $X_S^*=X_{1S}^*+X_{2S}^*$.

Since the estimators S_1^2 and S_2^2 are random variables, and are generally not equal to σ_1^2 and σ_2^2 respectively, Eq. (21) is not identical to Eq. (12). Consequently, the optimal values q_1 and q_2 calculated with Eq. (21) are not necessarily "true" optimal. Accuracy of the method proposed herein was examined by Monte Carlo simulation.

The simulation procedure is as follows. First, the normally distributed pseudo random realizations of the sample sizes n_1 and n_2 were generated in the computer. The means and the variances were taken as 0.0 and 1.0 respectively for both X_1 and X_2. For a given exceedance probability q, the characteristic value X_S^* was determined from the simulated samples of X_1 and X_2 as explained previously. The true exceedance probability q_S attained by the characteristic value X_S^* can be calculated since the true mean and the true variance of X are known in the simulation. This procedure is repeated over many times. The following measure of closeness to the target exceedance probability q was employed

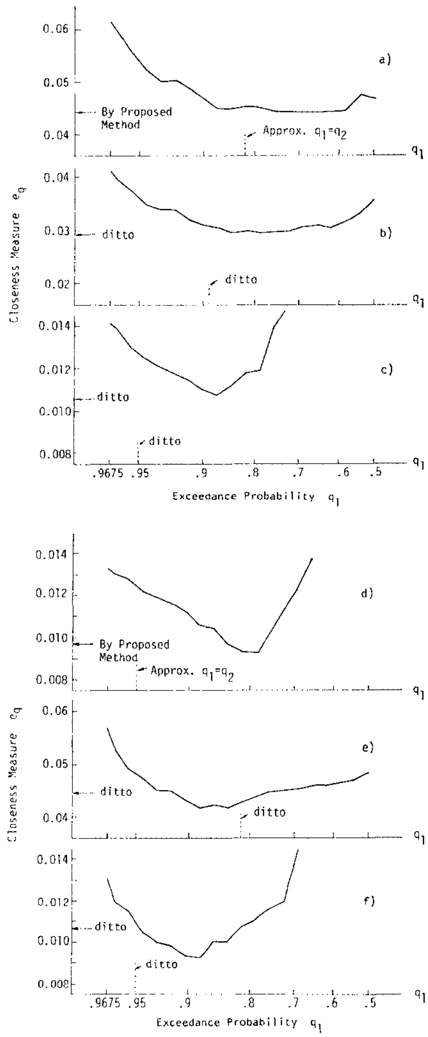
$$e_q=\sqrt{\sum_{l=1}^{n_M}(q_{sl}-q)^2}\ , \qquad \cdots\cdots\cdots(22)$$

where n_M is the number of simulations (=100). The results were

a) $e_q=0.0448$ for $q=0.90$, $n_1=19$, $n_2=99$
b) $e_q=0.0296$ for $q=0.95$, $n_1=19$, $n_2=99$
c) $e_q=0.0107$ for $q=0.99$, $n_1=19$, $n_2=99$
d) $e_q=0.0097$ for $q=0.99$, $n_1=19$, $n_2=199$
e) $e_q=0.0443$ for $q=0.90$, $n_1=19$ $n_2=249$
f) $e_q=0.0107$ for $q=0.99$, $n_1=19$, $n_2=249$

For comparison, the values of e_q are also computed in the simulation, where a fixed value of q_1 is used and then q_2 is determined from Eq. (21) without optimization. The results corresponding to cases a)~f) are shown in **Fig. 5** a)~f) where the values of e_q by the proposed method are also indicated. In **Fig. 5**, the abscissa is the fixed exceedance probability q_1.

In **Fig. 5** a)~c) showing e_q of the cases where the sample size ratio $\gamma_n=n_2/n_1$ is not so large, it is found that e_q by the proposed method is approximately equal to the minimum value of e_q obtained by the method using a fixed exceedance probability q_1, indicating that the proposed method is quite satisfactory. For the cases where the sample size ratio γ_n is relatively large, the results in **Fig. 5** d)~f) are somewhat discouraging; the proposed method appears to be not so effective. However, it should be emphasized that the exceedance probability q_1 which minimizes the value of e_q can not be estimated in advance. Therefore, the proposed method may be considered still useful for larger sample size ratios.



a) $q=0.90$, $n_1=19$, and $n_2=99$
b) $q=0.95$, $n_1=19$, and $n_2=99$
c) $q=0.99$, $n_1=19$, and $n_2=99$
d) $q=0.99$, $n_1=19$, and $n_2=199$
e) $q=0.90$, $n_1=19$, and $n_2=249$
f) $q=0.99$, $n_1=19$, and $n_2=249$

Fig. 5 Values of e_q by Monte Carlo Simulations.

Fig. 5 a)~f) inclusively confirms that the error of the exceedance probability q_S attained by the characteristic value X_S^* from statistical data is a concave function of q_1 on the whole and that the selection of $q_1 = q_2$ may lead to somewhat larger error in the exceedance probability q_S of X_S^*.

5. OPTIMAL SAMPLE SIZE FOR DETERMINATION OF CHARACTERISTIC VALUE

So far, determination of the characteristic value X_S^* from "given" statistical data has been discussed. Here, a problem of optimal sample size $\{n_i\}$ for determination of the characteristic value X_S^* is studied in consideration of the cost required to collect statistical data.

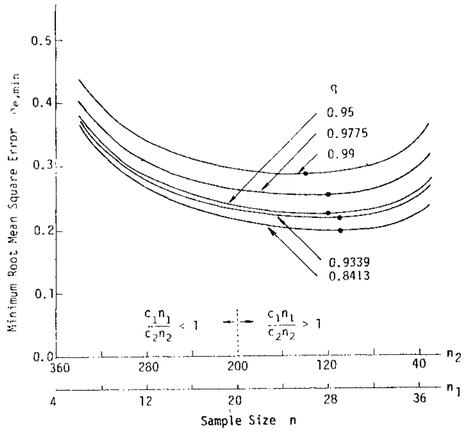
Reliable statistical data (e.g., of strength R) is in most cases provided by testing. In order to collect the data of the strength analysis uncertainty P in Eq. (2), costly testing of real size structural members is often necessary, while testing of material strength for the estimation of the uncertainty M is usually inexpensive. Thus, cost of collecting a data depends on a nature of the basic uncertainty variable X_i. For a limited budget, there must be an optimal set of sample sizes $\{n_i\}_{opt}$ which minimizes the error of the characteristic value X_S^* determined from the collected data of the basic random variables.

Again, consider the case where the random variable X is expressed by a sum of two basic random variables X_1 and X_2 as given by Eq. (10). The optimal sample size problem can be stated in this context as

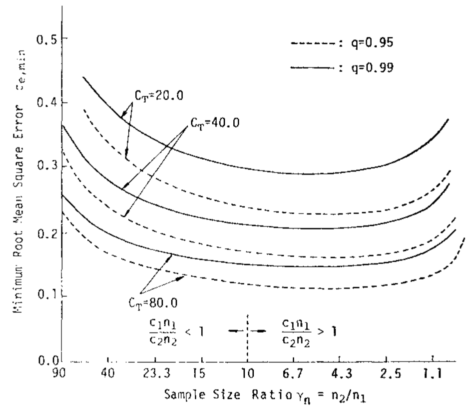
> "To find sample sizes n_1 and n_2 such that the mean square error $\sigma_{e,\min}$ (see **Fig. 4**) of the characteristic value X_S^* determined from statistical data is minimized, under the constraint of $n_1c_1 + n_2c_2 = C_T$"

in which C_T is a total budget and c_1 and c_2 are costs required to obtain a data of the basic random variables X_1 and X_2, respectively.

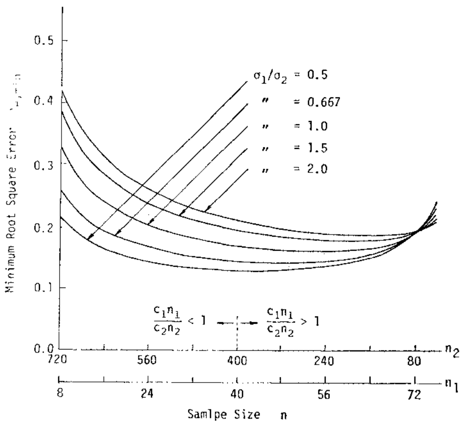
Values of $\sigma_{e,\min}$ are computed according to the procedure described for **Fig. 4**, and the results for the case of $C_T = 20.0$ $c_1 = 0.5$, $c_2 = 0.05$, and $\sigma_1 = \sigma_2$ are presented in **Fig. 6**a), where the abscissa is sample size n_1 or n_2. As expected, it can be seen that there exists an optimal combination of sample sizes which minimizes $\sigma_{e,\min}$. However, it should be also noticed that $\sigma_{e,\min}$ is not sensitive to the sample size ratio $\gamma_n = n_2/n_1$ over a wide range around the optimal point. Further, the optimal ratio $\gamma_{n,opt}$ is not significantly changed for different levels of the exceedance probability q. **Fig. 6** b) shows values of $\sigma_{e,\min}$ computed for different total budgets of $C_T = 20.0$, 40.0 and 80.0. From this figure, it is found that the optimal ratio $\gamma_{n,opt}$ is practically unchanged with respect to the total budget C_T. The optimal ratio $\gamma_{n,opt}$ seems almost



a) $C_T = 20.0$, $c_1 = 0.5$, $c_2 = 0.05$, and $\sigma_1 = \sigma_2$



b) $c_1 = 0.5$, $c_2 = 0.05$, and $\sigma_1 = \sigma_2$



c) $C_T = 40.0$, $c_1 = 0.5$, $c_2 = 0.05$, and $q = 0.95$

Fig. 6 Values of $\sigma_{e,\min}$ for Sample Sizes n_1 and n_2.

independent of the total budget C_T, provided that C_T is reasonably large compared with c_1 and c_2. **Fig. 6** c) shows values of $\sigma_{e,\min}$ for various values of the ratio σ_1/σ_2, indicating that the

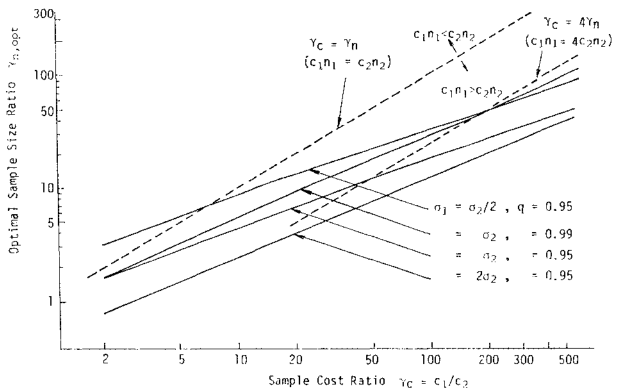


Fig. 7 Optimal Sample Size Ratios $\gamma_{n,\mathrm{opt}}$ for Cost Ratios $\gamma_c(=c_1/c_2)$ in the Case of $C_T=80.0$.

optimal sample size n_1 increases with larger σ_1, although the practical consequence is admissibly small. The optimal ratios $\gamma_{n,\mathrm{opt}}$ for the cost ratios $\gamma_c=c_1/c_2$ in the case of $C_T=80.0$ are presented in **Fig. 7**. This figure also includes the dashed line of $\gamma_c=\gamma_n$ ($c_1n_1=c_2n_2$) which indicates equal distribution of the budget for collecting the two types of data, since c_1n_1 and c_2n_2 are the distributed budgets for respective data collection. It appears rational to allocate more than half of the total budget for the collection of data of an uncertainty variable which needs relatively expensive testing; for example, $\gamma_{n,\mathrm{opt}}$ in the case of $\gamma_c=200$, $\sigma_1=\sigma_2$, and $q=0.99$ is approximately equal to 50, and roughly on the line of $\gamma_c=4\ \gamma_n$, that is $c_1n_1=4\cdot c_2n_2$. However, it should be remarked that a rough estimate of $\gamma_{n,\mathrm{opt}}$ for given γ_c can be allowed since $\sigma_{e,\min}$ is very insensitive to γ_n as seen in **Fig. 6** a)~c).

6. DISCUSSION AND SUMMARY

In general, the strength R and the load S are given by functions of a number of basic random variables. At present, information of these basic random variables is available only in a form of limited statistical data. In view of the fractile-based design[1], accurate determination of the design values R^* and S^* having target exceedance probabilities q_R and q_S respectively is of great importance.

This paper has shown that the difference of the amount of available information of basic random variables should be utilized and reflected to determine the design values R^* and S^*. A remarkable consequence is that, for an exceedance probability q greater than 0.5, relatively large exceedance probabilities of basic random variables with abundant statistical data and small exceedance probabilities of the variables with very limited data lead to less error in the characteristic value (i.e., design values R^* and S^*; determined from the data, and *vice versa* for q less than 0.5.

A very simple probabilistic model is employed in this study as given by Eq. (10). A more general case is

$$X=X_1+X_2+\cdots+X_m\,. \qquad \cdots\cdots(23)$$

Exponetial transformation of Eq. (23) yields

$$Y=Y_1Y_2\cdots Y_m\,, \qquad \cdots\cdots(24)$$

in which $Y=\exp(X)$ and $Y_i=\exp(X_i)$ $(i=1-m)$ In the cases of Eqs. (23) and (24), or in a even more general case, similar results are believed to be obtained.

Two implications in the context of the fractile-based design follow. As given by Eq. (2), the uncertainty of strength R is often expressed by a product of the three uncertainties M, F and P. The statistical data of the material uncertainty M is abundant, whereas the data of the strength uncertainty P is usually very limited. Then, relatively large exceedance probability q_M for the uncertainty M and small q_P for P should be assigned in determining R^*, since specified q_R for strenth R is generally larger than 0.5.

Another implication is as follows. Statistical data of the load S, particularly enviromental loads such as wind and earthquake loads, can be obtained by means of field observation. Since the lifetime of civil engineering structures ranges from 30 to 100 years, reliable estimation of such loads requires data observed for a very long period of time. Currently available data of these loads is an outcome only of 30–40 years observation at best and therefore is not sufficient to accurately describe the upper tail of the distribution of lifetime maxima of the loads[10]. The data of strength is not sufficient either, however, it can be said that strength data is relatively sufficient in comparison with that of the loads. Then, in view of the results of this study, it is appropriate to assign q_R to a somewhat larger value close to 1.0, while q_S to a value rather close to 0.5 (see **Fig. 1**).

Collection of statistical data regarding the strength of structures and the loads by testing or observation is one of the important subjects towards the implementation of reliability-based design. The budget for this purpose is limited and the cost of collecting a statisical data depends on a nature of basic random variables. A problem of determining the optimal set of sample sizes of basic random variables under a limited budget is studied in consideration of the cost required to collect data.

It has been revealed that there exists an optimal set of sample sizes of basic random variables which minimizes the error of the characteristic value, *e.g.* R^* or S^*, determined from the sample data collected. Further, it is found that weighted allocation of the total budget for the collection of data of basic variables which need relatively expensive tesing is effective.

7. ACKNOWLEDGMENTS

The authors are grateful to Professor Manabu Ito of the University of Tokyo for his valuable comments and suggestions. This study is supported in part by the Grant-in-Aid for Scientific Research No. 56750298 from the Japanese Ministry of Education, Science and Culture to the first author. Numerical computations of this study are carried out in the Science Information Processing Center of the University of Tsukuba which the first author used to belong to.

REFERENCES

1) Nishino, F., A. Hasegawa, C. Miki and Y. Fujino: A Fractile-Based Reliability Structural Design, Proc. of JSCE, No. 326, pp. 141∼153, October, 1982.
2) Cornell, C. A.: Structural Safety Specifications Based on Second Moment Reliabilty Analysis, IABSE Symposium on Concepts of Safety and Methods of Design, Final Report, London, England, September, 1969.
3) Veneziano, D.: A Theory of Reliability which Includes Statistical Uncertainty, Proc. of 2nd International Conference on Application of Probability and Statistics in Soil and Structural Engineering, Germany, September, 1975.
4) Fujino, Y.: Discussions to Reliability Analysis with Uncertainties of the Measured Values (by Baba, S. et al.), Proc. of JSCE, No. 268, pp. 129∼132. 1977 (in Japanese).
5) Shiraishi, N., H. Furuta and M. Nakano: Some Considerations on Application of Safety Index to Reliability Analysis, Proc. of JSCE, No. 301, September, pp. 13∼22, 1980 (in Japanese).
6) Gumbel, E. J.: Statistics of Extremes, Columbia University Press, 1958.
7) Ang, A. H-S. and W. H. Tang: Probability Concepts in Engineering Planning and Design, Vol. 1. John Wiley & Sons, 1979.
8) Blom, G.: Statistical Estimates and Transformed Beta-Variables, John Wiley & Sons, 1968.
9) Lind, N. C.: Consistent Partial Safety Factors, Journal of the Structural Division, ASCE, Vol. 97, No. ST 6, pp. 1651∼1669, June, 1971.
10) Fujino, Y. and A. Hasegawa: A Concept of Safety Factor in Structural Design, Journal of JSCE, Vol. 65, No. 10, pp. 39∼44, September, 1980 (in Japanese).

(*Received August 26, 1981*)