Predictive distribution of strength under control

R. Rackwitz

Dr.-Ing., Research Assoc., Institut für Massivbau, Technical University of Munich, Munich, Germany.

The joint distribution of strength of materials is derived in terms of a set of conditional distributions to be used in studies on structural reliability. Bayes theorem of probability theory is used to update prior distributions for the parameters of Gaussian sequences by direct observations and/or by compliance tests. Maximum-Likelihood estimators are given for the efficient quantification of prior information. The formulae are applied to concrete production judged by standard tests. It is shown that statistical uncertainties must not be ignored in structural reliability studies.

INTRODUCTION

Studies for the reliability of structures require realistic models for all relevant uncertain quantities. Only if a complete set of models, compatible with the reliability method and the mechanical context in which they are to be used, is available sensible conclusions can be drawn from such studies. The need for realistic and operational models is particularly obvious in the area of code making where, at present, serious attempts are made to design a new generation of probability-based structural codes. It appears, however, that still few studies have been directed towards the elaboration of suitable models for the strength of materials although a vast number of statistical investigations exists for various types of material and structural components where data have been collected under various circumstances of production, sampling and testing. Those generally resulted in an overall knowledge of the distribution type which fits best to the observations and about their distribution parameters the latter most frequently being expressed in terms of means and standard deviations. Many of them, however, failed to recognise that in structural reliability it is necessary to distinguish between two basic types of uncertainties — the first which concerns the spatial (and, possibly, time-dependent) variations of material properties in a given structure and which may be modelled by a random field or just by an independent sequence with given probability distribution and given distribution parameters; and the second, which is essentially of statistical nature and simply expresses the state of ignorance about the actual distribution parameters, again in terms of probability distributions. Quite frequently, the latter type of uncertainty is the dominating one and has important consequences not only on the safety level but on the means and methodology of assurance of structural reliability. The mathematical concept in which both types of uncertainties can consistently be treated is that of Bayesian statistics as proposed by Veneziano [13]. Also, the formulations are compatible with modern first-order reliability methods [4] and, therefore, are amenable to numerical treatment which was not feasible in previous reliability methods.

Subsequently, a rather general model for the distribution of the strength of materials including the favourable effect of compliance control is proposed. It rests on the following basic concept. Given a certain economical climate of production of material, e. g. concrete production in concrete factories in a country under the regime of a certain technological standard and pre-specified compliance rules, it is assumed that a randomly selected structure is supplied

0025-5432/1983/259/$ 5.00/© BORDAS-DUNOD

by a randomly selected producer with material of a certain grade chosen previously by the designer. The material as delivered or produced at a given job is assumed to form a stationary, independent Gaussian sequence with fixed but previously unknown mean and variance. The development of a suitable model then generally requires the following steps:

(i) collection and modelling of prior information on the statistical properties of production;

(ii) updating of the prior information if direct observations are available;

(iii) consideration of the filtering effect of potential compliance tests;

(iv) evaluation of the predictive distribution of strength at a given point in a given structure;

(v) if necessary, characterization of the joint distribution of strength in different points within a structure.

Clearly, if one of the basic conditions is changed, the whole modelling procedure must be repeated.

The perhaps most important decision which has been made before is the assumption of a Gaussian distribution. Of course, log-normal distributions or other distributions derivable from the Gaussian distribution by simple transformations then are also included. In fact, the log-normal distribution is perhaps the most realistic one for physical reasons (no negative values). Further, it is positively skewed as is the empirical distribution of most of the data on material strength. A model from the Gaussian family should be selected for various reasons. First of all, it can hardly be rejected on statistical grounds for most types of material, at least not with substantially more justification than any other of the uni-modal distributions which have been proposed. But it is by far the most developed in a statistical sense as will be seen and, thus, has enormous operational advantages. Furthermore, the strength of material as modelled herein is the strength of test specimen of well-defined shape, size and testing procedure. The determination of the strength of cross-sections or structural members, which is of only interest in reliability studies, is still another step which may distort the original distribution anyway but, again, normal distributions facilitate many derivations. Problems of this type are discussed in a separate paper [5]. The selection of a model from the Gaussian family, therefore, is to a large extent pragmatic. Nevertheless, one has to keep in mind that reliability statements always are conditioned on the set of stochastic models used. (For a detailed discussion of the problem of model selection in structural reliability the reader is referred to reference [3]).

The paper first discusses the type of prior information usually available. It then reviews briefly some classical results of Bayesian statistics for normal variables including the notion of a predictive distribution. The results on the effect of compliance control to be derived are believed novel. Some numerical results indicate the influence of different states of statistical knowledge on the predictive distribution of strength.

PRIOR INFORMATION ON MATERIALS' PRODUCTION

Figure 1 displays the observed means and standard deviations of concrete of a given grade at a number of jobs. Each point represents the statistics of a random sample of at least 50 tests. Therefore, there is only a small statistical error when interpreting the pair $(\bar{x}', s')$ as the true mean and the true standard deviation at that job. The variations in those values rather reflect different production strategies meaning that mean and standard deviation are uncertain quantities as well. For the data shown in figure 1 the specified strength corresponds to the 5%-fractile of the individual populations. The dashed line, therefore, separates those pairs which, by definition, have sufficient quality from those which are insufficient. The arrows indicate those jobs where at least one negative control decision should have been taken according to a certain compliance rule. Finally, the dotted line is the line of linear regression between s' and $\bar{x}'$ demonstrating that inspite of the definition of the specified strength as a 5%-fractile which demands for larger means with larger variability, only weak correlation exists between standard deviations and means. A more detailed discussion of the data shown in figure 1 is given in [9].

A quantification of the prior knowledge on production as given in figure 1 can best be made in the framework of Bayesian statistics. For example, let $f'(\boldsymbol{\theta})$ be the prior probability density of an uncertain parameter vector $\boldsymbol{\Theta}=\{\Theta_1, \Theta_2, \ldots, \Theta_n\}$ and $L'(A(\mathbf{z})|\boldsymbol{\theta})$ the probability of an event $A(\mathbf{z})$, defined in the Θ-space and being a function of the observation vector $\mathbf{z}=\{z_1, z_2, \ldots, z_m\}$, given $\boldsymbol{\Theta}=\boldsymbol{\theta}$. Then, Bayes theorem states that the posterior probability density of $\boldsymbol{\Theta}$ is given by:

$$f''(\boldsymbol{\theta}|\mathbf{z}) \propto L(A(\mathbf{z})|\boldsymbol{\theta})\, f'(\boldsymbol{\theta}). \tag{1}$$

In particular, if $L(\mathbf{z}|\boldsymbol{\theta})$ is the likelihood function of $\mathbf{z}$ the usual form of Bayes theorem in statistics is obtained. For normal sequences with parameter vector $\boldsymbol{\Theta}=\{M, \Sigma\}$, a natural choice of the distribution type for the prior distribution then is the so-called Normal-Gamma distribution whose probability density is given by [12]:

$$\begin{aligned} f_{N\Gamma}(\mu, h|\bar{x}', s', n', v') &= f_N(\mu|\bar{x}', hn')\, f_\Gamma(h|s', v') \\ &= \frac{\sqrt{hn'}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mu-\bar{x}'}{1/\sqrt{hn'}}\right)^2\right] \\ &\times \frac{((1/2)\,v' s'^2 h)^{(1/2)\,v'-1} . \exp[-(1/2)\,v s'^2 h]}{2 . \Gamma(v'/2)} v s'^2, \end{aligned} \tag{2}$$

where $\bar{x}'$ is the mean of an equivalent sample of size n' and s' is the standard deviation of an equivalent sample of size $v'+1$. Note that the uncertain variability is expressed by the precision $h=1/\sigma^2$. Equation (2) is denoted by the natural conjugate of the likelihood function of samples with respective sizes n' and $v'+1$

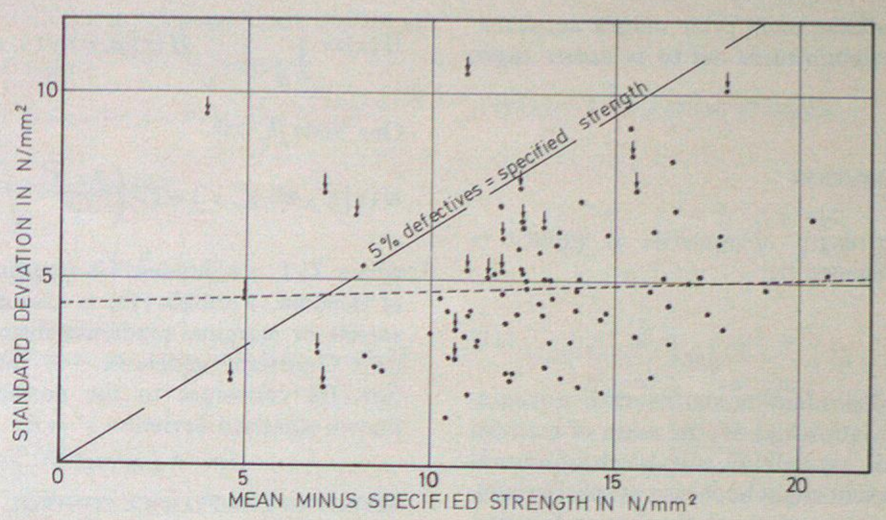

Fig. 1. — Observed mean minus specified strength and standard deviation of standard cube strength of 88 production units of concrete grade B 35
(Arrows indicate lots with at least one negative compliance decision).

taken from the same process. If direct observations of the process are available with statistics $(\bar{x}, n, s, v)$, then, the posterior density is of the same type as equation (2) but with parameters given subsequently. Note that information on μ or h may be obtained from the same (then, $v=n-1$), different or overlapping samples. Furthermore, the two special cases of known precision or known mean and the other respective parameter unknown are determined by obvious limiting operations:

$$\bar{x}'' = \frac{\bar{x}' n' + \bar{x} n}{n''}, \tag{3}$$

$$n'' = n' + n, \tag{4}$$

$$s''^2 = \frac{1}{v''}[v' s'^2 + n' \bar{x}'^2] + [v s^2 + n\bar{x}^2] - n'' \bar{x}''^2, \tag{5}$$

$$v'' = [v' + \delta(n')] + [v + \delta(n)] - \delta(n'') \tag{6}$$

with:

$$\delta(x) \equiv \begin{cases} 0 & \text{for} \quad x=0, \\ 1 & \text{for} \quad x>0. \end{cases}$$

It appears to be important to recognize the nature of the prior distribution in our case. In contrast to the statistical model which leads to formula (2) where, starting from a non-informative prior distribution, two samples are taken sequentially from the same population to yield the density (2) with parameters $(\bar{x}'', n'', s'', v'')$, the prior density with statistics $(\bar{x}', n', s', v')$, again starting from a non-informative prior, is evaluated from a sample of pairs $(\bar{x}, s)$ each representing a different population. We, thus, have to determine appropriate estimators for the parameters of equation (2).

Unfortunately, simple moment estimators as given in [12] are rather inefficient in a statistical sense so that Maximum-Likelihood estimators have been derived in the usual way [8]. With $m_i=\bar{x}_i$ and $h_i=1/s_i^2$ and the abbreviations:

$$\bar{h} = \frac{1}{k}\sum_1^k h_i; \qquad \tilde{h} = \frac{1}{k}\sum_1^k \ln h_i;$$

$$\hat{h} = \frac{1}{k}\sum_1^k h_i m_i; \qquad \check{h} = \frac{1}{k}\sum_1^k h_i m_i^2;$$

one can show that:

$$\bar{x}' = \frac{\hat{h}}{\bar{h}}, \tag{7}$$

$$n' = \left(\check{h} - \frac{\hat{h}^2}{\bar{h}}\right)^{-1}, \tag{8}$$

$$s' = \bar{h}^{-1/2}, \tag{9}$$

$$\tilde{h} = \psi\left(\frac{v'}{2}\right) - \ln\left(\frac{v'}{2\bar{h}}\right) \tag{10}$$

with:

$$\psi\left(\frac{v'}{2}\right) = \ln\left(\frac{v'}{2}\right) - \frac{1}{v'} - \frac{1}{3v'^2} + \frac{2}{15v'^4} - \frac{16}{63v'^6} + \ldots$$

The determination of v' requires the numerical solution of equation (10). For sufficiently large v' the last expansion can be truncated after the second term so that approximately:

$$v' \approx (\ln \bar{h} - \tilde{h})^{-1}.$$

As an example, the same data as for figure 1 are evaluated by the Maximum-Likelihood method yielding $\bar{x}'_{ML}=47.00$, $n'_{ML}=1.37$, $s'_{ML}=3.69$, $v'_{ML}=2.69$ while the moment estimators are: $\bar{x}'_M=47.95$, $n'_M \approx 0$, $s'_M=2.38$, $v'_M=1.55$.

Clearly, if direct observations are available in a particular case, e. g. by trial tests, formulae (3) to (6) may be used to update the prior information. In fact, the information collected in figure 1 may substantially be improved by even small direct samples since, as indicated by the relative small prior sample weights n' and v', prior information turns out to be rather vague in this case.

PREDICTIVE DISTRIBUTION

Now, let the strength of material at point i in structure j be represented by:

$$X_{ij}=U_i\Sigma_j+M_j \tag{11}$$

in which U_i is an independent standard normal sequence, Σ_j the standard deviation and M_j the mean of material strength at job j (or, production unit, batch, structural section or whatever unit might be chosen as appropriate). It follows, that X_{ij} has conditional distribution function (the index j is now being dropped):

$$H_{i+2}(x\,|\,\mu,\sigma)=\Phi\left(\frac{x-\mu}{\sigma}\right) \tag{12}$$

while, from equation (2):

$$H_1(\sigma\,|\,s'',v'')=\int_0^{\sigma}\left\{\frac{2\exp[-(1/2)\,v''(s''/t)^2]\,[(1/2)\,v''(s''/t)^2]^{(v''+1)/2}}{(\Gamma(v''/2)\,(1/2\,v''s''^2)^{1/2})}\right\}dt$$

$$=1-F_\Gamma\left(\frac{1}{2}v''s''^2/\sigma^2\,\middle|\,\frac{1}{2}v''\right) \tag{13}$$

with $F_\Gamma(z\,|\,r)=1/\Gamma(r)\int_0^z e^{-t}\,t^{r-1}\,dt$ the incomplete Gamma-function.

Similarly, the conditional distribution of the mean is given by:

$$H_2(\mu\,|\,\sigma,\bar{x}'',n'')=\Phi\left(\frac{\mu-\bar{x}''}{\sigma/\sqrt{n''}}\right) \tag{14}$$

The set of conditional distribution functions $H_1(.)$, $H_2(.)$ and $H_{i+2}(.)$ completely defines the joint distribution of the sequence of strengths X_i, $i=1, 2, \ldots, r$ within the chosen unit. Note that a rather complex dependence structure exists between different points i within the unit and which must not be ignored in certain reliability applications. The information given by equations (12) to (14) is also sufficient when applying modern transformation techniques within first-order reliability methods [4].

If interest lies only in the marginal distribution of X_i, $i=1, 2, \ldots, r$ (e. g. for reliability calculations of individual cross-sections) and, consequently, stochastic dependence between strength values at different points in the structure has no significance, the conditioning of the parameters (μ, σ) can be removed by integration, i. e.:

$$H(x)=\int_0^{\infty}\int_{-\infty}^{+\infty}H(x\,|\,\mu,\sigma)\,f(\mu,\sigma)\,d\mu\,d\sigma. \tag{15}$$

One finds [1, 15]:

$$H(x\,|\,\bar{x}'',n'',s'',v'')=T_{v''}\left(\frac{x-\bar{x}''}{s''}\sqrt{\frac{n''}{n''+1}}\right), \tag{16}$$

where $T_f(.)$ is Student's-t-distribution with f degrees of freedom. Formula (16) is also denoted by the univariate or marginal predictive distribution of independent Gaussian sequences. For large v'' the distribution (16) converges to the normal, i. e. the case of known standard deviation $s''\to\sigma$.

EFFECT OF COMPLIANCE CONTROL

The information on the distribution of qualities offered might be updated if either actual test data are available or if it is certain that the production process must pass known procedures of compliance testing. In case of rejection, the lot covered by a compliance decision will undergo further detailed investigation. Therefore, it is no longer of interest.

Denote $P(A\,|\,d(z), z, \mu, \sigma)$ the conditional probability of an acceptance decision given a certain decision rule $d(z)$, the outcome z of a random sample and the production process having mean μ and standard deviation σ. As a function of μ and σ this probability, e. g. $L(\mu, \sigma)$, is widely known as the operation characteristic of the "test" $(d(z), z)$.

If $f(\mu, \sigma)$ is the probability density of qualities offered, it is easy to see that:

$$f^*(\mu,\sigma)=\frac{f(\mu,\sigma)\,L(\mu,\sigma)}{\int_0^{+\infty}\int_{-\infty}^{+\infty}f(\mu,\sigma)\,L(\mu,\sigma)\,d\mu\,d\sigma} \tag{17}$$

is the probability density of those qualities $(M=\mu$, $\Sigma=\sigma)$ which have passed control. Analytical results for the distribution function of (M, Σ) filtered by compliance control are very few, one of which is presented next in terms of conditional distribution functions.

One may now wish to distinguish between three cases:

(*a*) Standard deviation known and constant;

(*b*) Standard deviation known during testing but varying from unit to unit according to equation (13);

(*c*) Standard deviation unknown during testing and varying from unit to unit according to equation (13).

Case a

Here, equation (17) reduces to [11]:

$$f^*(\mu)=\frac{f''(\mu)\,L(\mu)}{\int_{-\infty}^{+\infty}f''(\mu)\,L(\mu)\,d\mu}, \tag{18}$$

while a suitable decision rule for compliance control is:

$$d(z)=\begin{cases}a\leqq z=x: A=\text{Acceptance},\\ a>z=x: \bar{A}=\text{Rejection}\end{cases} \tag{19}$$

with operating characteristic [2]:

$$P(A\,|\,d(z),z,\mu)=L(\mu)=\Phi\left(\frac{\mu-a}{\sigma/\sqrt{m}}\right), \tag{20}$$

Obviously, the distribution equation (14) must be replaced by:

$$H_2^*(\mu\,|\,\bar{x}'',n'',A)=P(M\leqq\mu\,|\,\bar{x}'',n'',A)$$

$$=\frac{P(A\,|\,M\leqq\mu)}{P(A)}H_2(\mu\,|\,\bar{x}'',n'') \tag{21}$$

with:

$$P(A)=\int_{-\infty}^{+\infty}f_N''(\mu\,|\,\bar{x}'',n'')\,L(\mu)\,d\mu$$

$$=\varphi\left(\frac{(\bar{x}''-a)/\sigma/\sqrt{m}}{\sqrt{1+(m/n'')}}\right) \tag{22}$$

$$P(A\,|\,M\leqq\mu)=\int_{-\infty}^{\mu}f_N''(\mu\,|\,\bar{x}''.n'')\,L\,\mu)\,d\mu$$

$$=\Phi\left(\frac{\mu-\bar{x}''}{\sigma/\sqrt{n''}},\frac{(\bar{x}''-a)/\sigma/\sqrt{m}}{\sqrt{1+(m/n'')}};-\frac{1}{\sqrt{1+(n''/m)}}\right), \tag{23}$$

where $\Phi(., .; .)$ is the standard bivariate normal integral defined by:

$$\Phi(h,k;\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{h}\int_{-\infty}^{k}\times\exp\left[-\frac{1}{2(1-\rho^2)}(u^2-2uv+v^2)\right]dv\,du.$$

In the derivation, the table of normal integrals given by Owen [6] has been used. Also, in evaluating the binormal integral a series expansions in terms of a function $T(a, b)$, given by the same author in an earlier paper, has been proved to be efficient [8].

Further, it is possible to determine the predictive distribution corresponding to equation (18). In this case, we have:

$$H(x)=\int_{-\infty}^{x}f(x\,|\,\mu)\,f^*(\mu)\,d\mu$$

$$=\frac{1}{N}\int_{-\infty}^{x}\int_{-\infty}^{+\infty}\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)\frac{\sqrt{n''}}{\sigma}\times\varphi\left(\frac{\mu-\bar{x}''}{\sigma/\sqrt{n''}}\right)\varphi\left(\frac{\mu-a}{\sigma/\sqrt{m}}\right)d\mu\,dx$$

$$=\frac{1}{N}\int_{-\infty}^{+\infty}\varphi\left(\frac{x-\mu}{\sigma}\right)\frac{\sqrt{n''}}{\sigma}\varphi\left(\frac{\mu-\bar{x}''}{\sigma/\sqrt{n''}}\right)\times\varphi\left(\frac{\mu-a}{\sigma/\sqrt{m}}\right)d\mu$$

$$=\frac{1}{N}\Phi\left(\frac{x-\bar{x}''}{\sigma\sqrt{1+(1/n'')}},\frac{(\bar{x}''-a)/\sigma/\sqrt{m}}{\sqrt{1+m/n''}};-\frac{1}{[(n''+1)\,((n''/m)+1)]^{1/2}}\right), \tag{24}$$

with:

$N=P(A)$ in equation (22).

Case c

Next we consider case *c*. Unfortunately, no exact analytical solution appears possible but good approximations can be developed. Obviously, given the distributions equations (13) and (14) we now have:

$$H_1^*(\sigma\,|\,s'',v'',A)=P(\Sigma\leqq\sigma\,|\,s'',v'',A)$$

$$=\frac{P(A\,|\,\Sigma\leqq\sigma)}{P(A)}H_1(\sigma\,|\,s'',v'') \tag{25}$$

and:

$$H_2^*(\mu\,|\,\sigma,\bar{x}'',n'',A)=P(M\leqq\mu\,|\,\bar{x}'',n'',\sigma,A)$$

$$=\frac{P(A\,|\,\sigma,M\leqq\mu)}{P(A\,|\,\sigma)}H_2(\mu\,|\,\sigma,\bar{x}'',n''). \tag{26}$$

In order to specify the probabilities $P(.)$, the following, widely used decision rule is introduced:

$$d(z)=\begin{cases}a\leqq z=x+\lambda s: A=\text{Acceptance},\\ a\geqq z=x+\lambda s: \bar{A}=\text{Rejection},\end{cases} \tag{27}$$

where a is a given acceptance limit, λ a given acceptance factor (which sometimes is taken as $\lambda=-1.645$), $\bar{x}$ the sample mean and s the sample standard deviation of a random sample of size m. The distribution of the random quantity Σ is known to be related to the non-central-t-distribution [2], but, in first approximation, one can evaluate $P(A \mid \mathrm{d}(z), z, \mu, \sigma)$ by:

$$P(A \mid \mathrm{d}(z), z, \mu, \sigma) = L(\mu, \sigma) \approx \Phi\left(\frac{\mu-(a-\lambda\sigma)}{\sigma\sqrt{(1/m)+(\lambda^2/2m)}}\right) \tag{28}$$

making use of the fact that sample standard deviations are asymptotically normally distributed with mean $E[S]=\sigma$ and standard deviation $D[S]=\sigma/\sqrt{2m}$. A somewhat better approximation to the non-central-t-distribution has been used in [11]. Then,

$$\begin{aligned} P(A) &= \int_0^\infty \int_{-\infty}^\infty f_{N\Gamma}(\mu, \sigma \mid \bar{x}'', n'', s'', v'') \times L(\mu, \sigma)\, \mathrm{d}\mu\, \mathrm{d}\sigma \\ &= \int_0^\infty f_\Gamma(\sigma \mid s'', v'')\, \Phi\left(\frac{\alpha}{\gamma\sigma}+\frac{\beta}{\gamma}\right) \mathrm{d}\sigma \\ &= \Pr(T'_f \leqq t \mid \delta) \approx \Pr(T_f \leqq t-\delta), \end{aligned} \tag{29}$$

$$\begin{aligned} P(A \mid \Sigma \leqq \sigma) &= \int_0^\sigma \int_{-\infty}^\infty f_{N\Gamma}(\mu, \sigma \mid \bar{x}'', n'', s'', v'') \times L(\mu, \sigma)\, \mathrm{d}\mu\, \mathrm{d}\sigma \\ &= \int_0^\sigma f_\Gamma(\sigma \mid s'', v'')\, \Phi\left(\frac{\alpha}{\gamma\sigma}+\frac{\beta}{\gamma}\right) \mathrm{d}\sigma \\ &= \int_{s/\sigma\sqrt{f}}^\infty \varphi\left(\frac{tu}{\sqrt{f}}-\delta\right) u^{f-1} \varphi(u)\, \mathrm{d}u \\ &= Q_f\left(t, \delta; \frac{s''}{\sigma}\sqrt{v''}, \infty\right) \\ &\geqq \Pr(T'_f \leqq t \mid \delta)\, \Pr\left[\chi_f^2 > \frac{s^2}{\sigma^2} f\right] \\ &\approx \Pr(T_f < t-\delta)\, \Pr\left(\chi_f^2 > \frac{s^2}{\sigma^2} f\right), \end{aligned} \tag{30}$$

$$P(A \mid \sigma) = \int_{-\infty}^{+\infty} f_N(\mu \mid \bar{x}'', n''\sigma)\, L(\mu, \sigma) = \Phi\left(\frac{\alpha}{\gamma\sigma}+\frac{\beta}{\gamma}\right) \tag{31}$$

$$\begin{aligned} P(A \mid \sigma, M \leqq \mu) &= \int_{-\infty}^{\mu} f_N(\mu \mid \bar{x}'', n'', \sigma) \circ L(\mu, \sigma)\, \mathrm{d}\mu \\ &= \Phi\left(\frac{\alpha}{\gamma\sigma}+\frac{\beta}{\gamma}, \frac{\mu-\bar{x}''}{\sigma/\sqrt{n''}}; -\frac{1}{(1+(n''(1+\lambda^2/2)/m)^{1/2}}\right) \end{aligned} \tag{32}$$

which, if inserted into equations (25) and (26), produce the required *a posteriori* distributions (the distributions updated by the fact that certain compliance control will take place). Again, use has been made of reference [6] in order to reduce the integrals in equations (29) and (30). The abbreviations used are: $t=\alpha/(\gamma s'')$, $\delta=-\beta/\gamma$, $f=v''$ with $\alpha=(\bar{x}''-a)$, $\beta=\lambda$ and $\gamma^2=(1/n'')+(1/m)+(\lambda^2/2m)$. Further, T'_f is a non-central-t-variate with f degrees of freedom and non-centrality parameter δ and T_f is the corresponding central-t-variate while χ_f^2 denotes the chi-square variate. $Q_f(\,.\,,\,.\,;\,.\,,\,.\,)$ is a special bivariate non-central-t-distribution introduced in [6] which is easily evaluated in terms of normal densities, normal integrals and bivariate normal integrals. Also, simple numerical integration or the approximate bound in equation (30) is sufficiently accurate. Finally, certain reductions are possible for the univariate predictive distribution. In analogy to equation (24) we have:

$$H(x) = \frac{1}{N}\int_0^\infty \int_{-\infty}^\infty F_N(x \mid \mu, \sigma) \times f_{N\Gamma}(\mu, \sigma \mid \bar{x}'', n'', s'', v'')\, L(\mu, \sigma \mid a, \lambda, m)\, d\mu\, d\sigma \tag{33}$$

which after some obvious substitutions and using again [6] becomes:

$$H(x) = \frac{1}{N}\int_0^\infty f_\Gamma(\sigma \mid s'', v'') \times \Phi\left(\frac{x-x''}{\sigma\sqrt{1+(1/n'')}}, \frac{(\bar{x}''-a+\lambda\sigma)/\sigma\sqrt{1/m+\lambda^2/(2m)}}{\sqrt{1+(m/n'')(1+\lambda^2/2)}}; -\frac{1}{[(n''+1)(n''/m(1+(\lambda^2/2))+1)]^{1/2}}\right) \mathrm{d}\sigma \tag{34}$$

with $N=P(A)$ in equation (29). No further sensible reduction appears possible.

Case b

If, during testing, the standard deviation σ is known, e. g. from previous production records and, consequently, replaces the sample standard deviation s in equation (27), the distribution of Σ remains as in equation (13). Formula (26) is still valid but with the λ's in the parameter γ, in the correlation coefficient in equation (32) and in equation (34) be set equal to zero.

NUMERICAL EXAMPLES

The prior information contained in data collections as represented by figure 1 is, as mentioned, rather vague. This can best be seen in evaluating the predictive distribution equation (16) for various levels of information. For example, assume that n trial tests yielded exactly the parameter $(\bar{x}', s')$ found for figure 1. Then, figure 2 demonstrates the significant effect of more and direct information. Note also the relatively fast convergence of the distribution towards the normal distribution with parameters $(\bar{x}''=\bar{x}', s''=s')$.

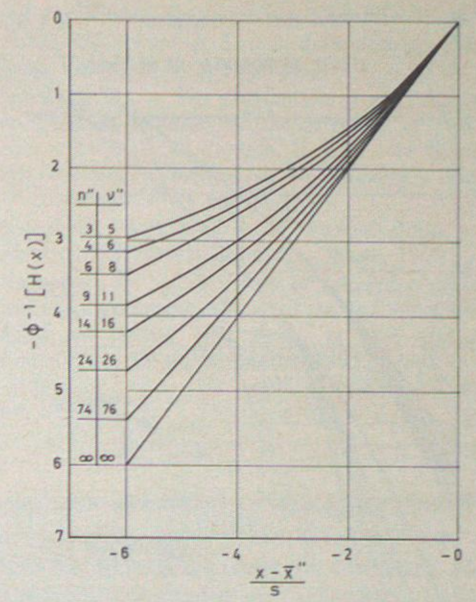

Fig. 2. — Predictive distribution of strength with increasing size of direct sample (underlying normal distribution).

Table I is an example how prior information could be given. It is the result of a large data collection on concrete production in Southern Germany. Concrete cube strength is assumed to be log-normally distributed. Therefore, the parameters given are for a log-normal-gamma distribution. Some obvious interpretations of changes in the parameters with concrete grade and type concrete are omitted for brevity of presentation. Naturally, due to different production policies and different control standards other values are to be expected in other countries but, as far as former statistical surveys in the international literature can be re-evaluated as required, the prior weights n' and v' as well as the numerical values of the parameters $\bar{x}'$ and s' do not differ substantially from those given in table I.

A similar situation holds for the yield strength of reinforcing bars. In this case we assume that the material of given diameter to be built in a specific building is supplied by a randomly selected steel mill from a randomly selected batch. A large survey on Middle-European steel production for hot-rolled, medium diameter bars showed an almost constant and relatively small within batch (=melt) variability of $\sigma \approx 8\ \mathrm{N/mm^2}$, but relatively large variability of the mean strength of batches of any (particular) steel mill and among steel mills. In fact, one determines for a certain steel grade a global mean $\bar{x}' \approx 480\ \mathrm{N/mm^2}$ and standard deviation $\sigma_{\bar{X}} \approx 28\ \mathrm{N/mm^2}$ while the within-mill standard deviation of the means may be about $15\ \mathrm{N/mm^2}$. If one applies the formulae for known standard deviation one obtains an equivalent prior sample size of $n'=0.08$ if the mill is unknown and $n'=0.28$ if it is known. Again, very few trial tests may improve the state of knowledge considerably.

TABLE I

PRIOR PARAMETERS FOR CONCRETE STRENGTH DISTRIBUTION

		Concrete grade				
		C 15	C 25	C 35	C 45	C 55
Site mixed concrete	$\bar{x}'$	3.40	3.65	3.85	-	-
	n'	1.0	2.0	3.0	-	-
	s'	0.15	0.12	0.09	-	-
	v'	3.0	4.0	4.5	-	-
Ready mixed concrete	$\bar{x}'$	3.40	3.65	3.85	3.98	-
	n'	1.5	1.5	1.5	1.5	-
	s'	0.14	0.12	0.09	0.07	-
	v'	6.0	6.0	6.0	6.0	-
Concrete for precast elements	$\bar{x}'$	-	3.80	3.95	4.08	4.15
	n'	-	2.0	2.5	3.0	3.5
	s'	-	0.09	0.08	0.07	0.05
	v'	-	4.5	4.5	5.0	5.5

For the effect of compliance control on the lower tail of the predictive distribution, we expect the filtering to be significant for relatively diffuse prior information, to increase with increasing acceptance limit and/or decreasing acceptance factor and, of course, with the sample size of acceptance testing. One can show, however, that generally small samples produce the main effect and only little is gained if the sample size is increased beyond, say, $m=5$ to 10. As an example, compare figure 3, where the predictive distribution of yield strength given the previous prior information is calculated for various m $(a=435\ \mathrm{N/mm^2})$. Also, the distribution of qualities in terms of the batch means as produced and accepted is given. The percentage of batches with negative acceptance decisions is roughly 5%.

A last demonstration (*fig.* 4), shows that given the relatively vague prior information as collected in table I (mean and standard deviation unknown), probabilities of exceedance can easily be reduced by an order of magnitude in the lower tail of the distribution if compliance control is considered.

In practical applications or detailed reliability studies one, therefore, has to compromise between the simple model as, for example, described by equation (16) and the more complex but also more realistic models as given by equation (24) or even equation (34). Nevertheless, it should be emphasised that direct observations are much more efficient in updating vague prior information than large sample compliance tests. In a particular case one, therefore, might think of carrying out few trial tests well in advance or of controlling production processes with respect to previously selected targets rather than to rely on relatively inefficient

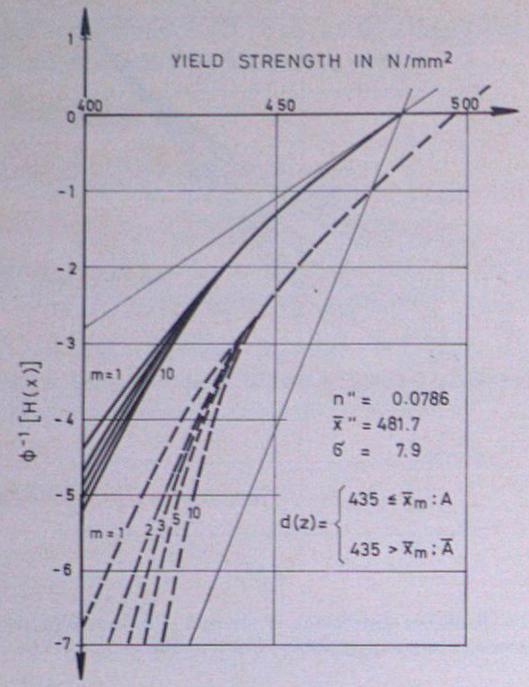

Fig. 3. — Predictive distribution of yield strength of reinforcing steel under control (upper thin line: prior distribution, lower thin line: prior distribution with perfect information, dotted lines: distribution of batch means under control).

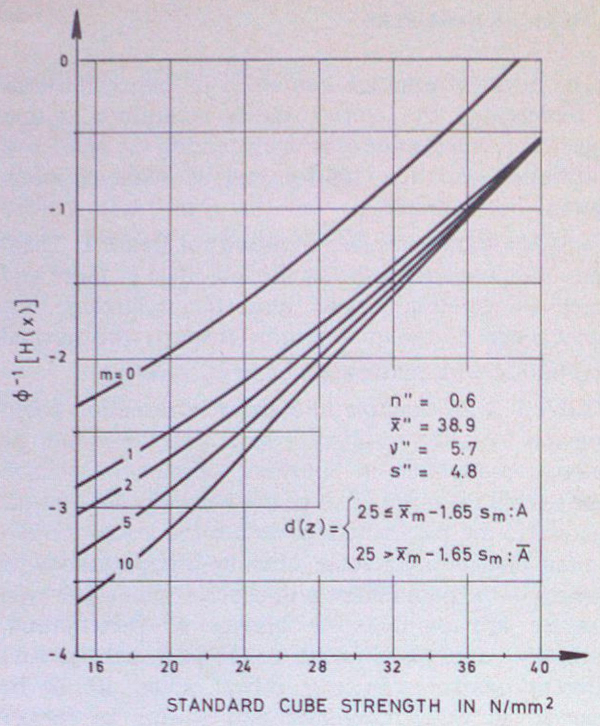

Fig. 4. — Predictive distribution of concrete cube strength under control with prior information equivalent to figure 1.

statistical compliance testing of the end product. This insight is not very new but it appears to be proved here for the first time numerically.

SUMMARY AND CONCLUSIONS

The strength of materials as tested by specimens of standard type, shape and age is modelled as a Gaussian sequence with partially known mean and standard deviation. Information on these parameters must be gained from previous production units in terms of a prior distribution. This can be updated through Bayes' theorem if either direct observations are available or future production is known to undergo certain compliance tests. The sequence of strength values to be expected can then be represented by a set of conditional distribution functions or, if interest lies only in the marginal distribution, by a predictive distribution. Closed formulae are derived for most cases involving the univariate normal, bivariate normal, central and non-central *t*-distributions which all are either tabulated or can be computed by appropriate series expansions available in the statistical literature. The results are demonstrated at an example from concrete production indicating that, generally, uncertainties in the parameters of the strength distribution can be significant in reliability studies. Therefore, it is recommended to consider statistical uncertainties as reflected by the appropriate (possibly updated) prior distribution throughout in reliability studies. The favourable filtering effect of compliance control can additionally be taken into account at the expense of slightly more numerical effort. For most types of material the initial distribution of elemental strength values might sufficiently realistically be taken as lognormal in which case all results remain valid with the necessary straightforward modifications. The models proposed herein can be applied at least to concrete, reinforcing and structural steel.

ACKNOWLEDGEMENTS

The concept and formulae presented in this paper are the result of the author's interest in such problems for more than 5 years. A number of persons, particularly my colleagues K. Breitung and M. Hohenbichler and several students made helpful suggestions or carried out supporting numerical examples which is highly appreciated.

REFERENCES

[1] Aitchison J., Sculthorpe D. — *Some problems of statistical prediction*, Biometrika, Vol. 52, No. 3 and 4, 1965, pp. 469-483.

[2] Fisz M. — *Probability theory and mathematical statistics*, 3d, ed., J. Wiley, New York, 1963.

[3] Grigoriu M., Veneziano D., Cornell A. C. — *Probabilistic modeling as decision making*, Eng. Mech. Div., A.S.C.E., 105, EM4, 1979, pp. 585-596.

[4] Hohenbichler M., Rackwitz R. — *Non-normal vectors in structural reliability*, J. Eng. Mech. Div., A.S.C.E., 107, EM6, 1981, pp. 661-676.

[5] Hohenbichler M., Rackwitz R. — *On structural reliability of parallel systems*, Reliability Engineering, Vol. 2, 1981, pp. 1-6.

[6] Owen D. B. — *A table of normal integrals*, Communications Statist. Simula. Computa., B9, 4, 1980, pp. 389-419.

[7] Owen D. B. — *Tables for computing bivariate normal probabilistics*, Ann. Math. Statist., Vol. 27, 1956, pp. 1075-1090.

[8] Pöhlmann S., Rackwitz R. — *Zwei Schätzprobleme bei Gauß'schen Folgen, Berichte zur Zuverlässigkeitstheorie der Bauwerke*, Sonderforschungsbereich 96, Technische Universität München, Heft 53, 1980.

[9] Rackwitz R., Müller K. F. — *Zum Qualitätsangebot von Beton II*, Beton, Vol. 10, 1977, pp. 391-393.

[10] Rackwitz R. — *Über die Wirkung von Abnahmekontrollen auf das Verteilungsgesetz von normalen Produktionsprozessen bei bekannter Standardabweichung*, Materialprüfung, Vol. 4, 1979, S. 122-124.

[11] Rackwitz R. — *The distribution of qualities offered and accepted as a basis for the selection of partial safety factors for material strength*, II. European Symposium of the E.O.Q.C. "Safeguards in Building", Madrid, 12-14 November 1979, pp. 53-59.

[12] Raiffa H., Schlaifer R. — *Applied statistical decision theory*, MIT Press, Cambridge, Mass., 1969.

[13] Veneziano D. — *A theory of reliability which includes statistical uncertainty*, Proc. of the 2nd Int. Conference of Appl. of Statist. and Prob. in Soil and Structural Engineering, 15-18, September, 1975 in Aachen, Deutsche Gesellschaft für Erd- und Grundbau, Essen, 1975, pp. 231-250.

RÉSUMÉ

Distribution prévisionnelle de la résistance sous contrôle. — *La résistance de matériaux ainsi vérifiée sur des spécimens de type, forme et âge normalisés est modélisée comme une progression gaussienne avec une valeur moyenne et un écart-type partiellement connus. Pour obtenir une information sur ces paramètres, il faut se servir d'unités de production précédentes en s'appuyant sur une distribution* a priori. *La mise à jour peut se faire par le théorème de Bayes, si des observations directes sont disponibles, ou s'il est connu que la production à venir va être soumise à certains essais de conformité. La progression des valeurs de résistances escomptées peut alors être représentée par un ensemble de fonctions de distribution conditionnelle ou, si l'on s'intéresse seulement à la distribution marginale, par une distribution prévisionnelle. Dans la plupart des cas, les formules sont dérivées en incluant la loi de distribution normale, la loi de distribution bi-normale, les distributions t centrales et non centrales, qui figurent toutes sous forme de tableau, ou peuvent être calculées par des expansions de séries appropriées, disponibles dans la littérature statistique. Les résultats sont illustrés par un exemple de production de béton, qui indique que, en général, des incertitudes dans les paramètres de la distribution de résistance peuvent avoir leur signification dans les études de fiabilité. Il est par conséquent tout indiqué de considérer que les incertitudes statistiques se reflètent dans la distribution* a priori *appropriée, si possible mise à jour par le moyen des études de fiabilité. De plus, l'effet filtrant favorable de contrôle de conformité peut être pris en considération avec peu d'effort numérique supplémentaire. Pour la plupart des types de matériaux, la distribution initiale des valeurs de résistances élémentaires peut être admise comme log-normal avec suffisamment de réalisme. Dans ce cas, tous les résultats restent valables avec les modifications nécessaires. Les modèles proposés ici peuvent être appliqués au moins au béton, béton armé et acier structurel.*