Structural Safety & Reliability, Schuëller, Shinozuka & Yao (eds) © 1994 Balkema, Rotterdam, ISBN 90 5410 357 4

Distribution arbitrariness in structural reliability

Ove Ditlevsen
Department of Structural Engineering, Technical University of Denmark, Lyngby, Denmark

ABSTRACT: The author points at the urgent need for code standardizations of distribution types for practical design applications of structural reliability methods. In support of the author's opinion that distribution type standardizations are necessary in order to avoid reliability comparisons on the basis of incommensurable reliability measures, the paper demonstrates the fundamental difficulty of choosing the probability distributions for example for annual extreme loads such as wind and snow loads for which only limited data series are available. Moreover, it is argued that even if several distribution types may pass a statistical test not all of these are reasonable candidates for standardization. The class of standardized distributions in a code of practice for reliability analysis should be looked upon as an internally harmonized entity chosen on the basis of consequence calculations applied to the class of structures for which the code is intended. Otherwise even the imposed ordering with respect to reliability becomes dubious.

JCSS Probabilistic code text example

The distribution tail sensitivity is a well known unavoidable property of structural reliability analysis of highly reliable structures. This fact causes the computed failure probabilities to be of limited informational value except for reliability comparisons made on the basis of the same set of probability distribution types, that is, within the same model universe of probability distributions. Therefore, for the advancement of the use of modern probabilistic reliability analysis to aid rational structural engineering decisions in practice there is an indispensable need for an agreement among competing engineers and the general public on using a standardized distribution model universe as a common reference. In other words, there is an indispensable need for a code of practice for structural reliability analysis.

An attempt to formulate such a code has four years ago been published by the Joint Committee on Structural Safety (JCSS) as a text with the title: Proposal for a Code for the Direct Use of Reliability Methods in Structural Design, Ditlevsen and Madsen (1989). Without achieving general consensus the details of the code text example has been discussed within the committee and the text is published in the form of a Working Document. JCSS is supported by the international associations CEB, CIB, ECCS, FIP, IABSE, IASS, and RILEM.

An important point of this code proposal is that the distribution types to be used in the reliability analysis are standardized. The way in which these standardizations are introduced is best illustrated by direct quotation from the proposal. It is not intended to discuss whether the actual specifications of the code text example are reasonable or not or why the text at some places is ambiguous. Of course, such a text will always be controversial and subject to discussions. In the Working Document's section 9 on reliability models there is the following code type text:

If no specific distribution type is given as standard in the action and material codes this code for the purpose of reliability evaluations standardizes the clipped (or, alternatively, the zero−truncated) normal distribution type for basic load pulse amplitudes. Furthermore, the logarithmic normal distribution type is standardized for the basic strength variables.

Deviations from specific geometrical measures of physical dimensions as length are standardized to have normal distributions if they act at the adverse state in the same way as load variables (increase of value implies decrease of reliability) and to have logarithmic normal distribution if they contribute to the adverse state in the same way as resistance variables (decrease of value implies decrease of reliability).

Further:

In special situations other than the code standardized distribution types can be relevant for the reliability evaluation. Such code deviating assumptions must be well documented on the basis of a plausible model that by its elements generates the claimed probability distribution type. Asymptotic distributions generated from the model are allowed to be applied only if it can be shown that they by application on a suitable representative example structure lead to approximately the same generalized reliability indices as obtained by application of the exact distribution generated by the model.

Experimental verification without any other type of verification of a distributional assumption that deviates strongly from the standard is only sufficient if very large representative samples of data are available.

Distributional assumptions that deviate from those of the code must in any case be tested on a suitable representative example structure. By calibration against results obtained on the basis of the standardizations of the code it must be guaranteed that the real (the absolute) safety level is not changed significantly relative to the requirements of the code.

When arguing within a specific model universe of distributions it is important to ensure that in particular near zero probability value results are used for comparisons only within the model itself. Carrying the results to the outside world and attaching the usual probability interpretation of relative frequency of occurrence in the real world of the considered adverse event will generally be highly misleading even though the model has been carefully calibrated to real world data. This insight is not new. In fact, it has been repeatedly discussed at least during the last quarter of a century of structural reliability theory development. However, it has now become urgent to focus on this point due to the recent maturing of practicable reliability analysis methods as resulting from the almost explosive development of available computational power.

Example: Snow load statistics

The problem of the choice of distribution type is well illustrated in the case of historical data series that by their very nature increase in sample size as slowly as by only one new observation each year. For example, this is the case for yearly extreme ground snow loads at a given location.

Ellingwood and Redfield (1983) report that the lognormal distribution fits several data sets of 28 yearly extreme ground snow loads given in terms of water equivalents, each load data set measured at a different weather station in USA. It is reported that for most weather stations the lognormal distribution fits better than a Gumbel

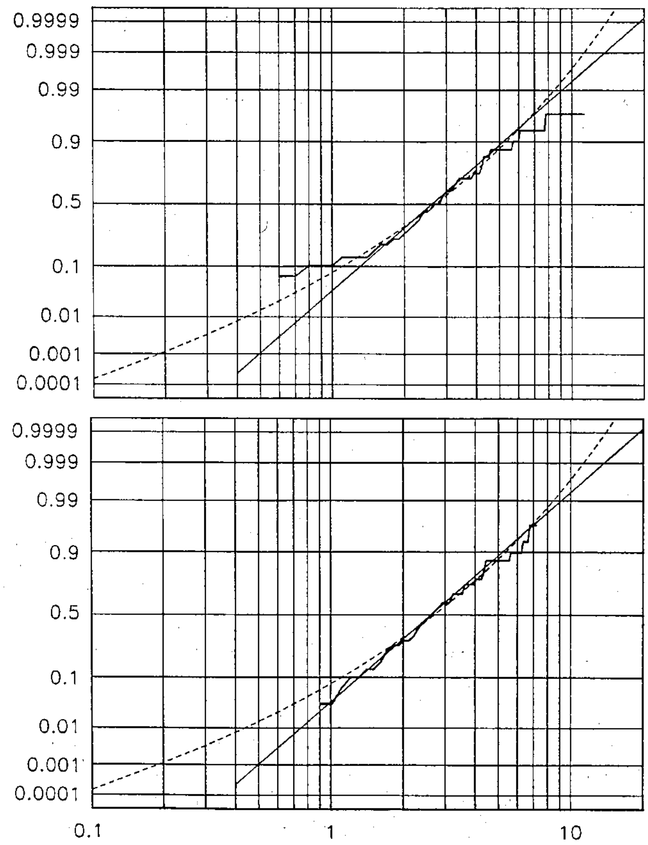

Fig. 1. Lognormal distribution function (straight line) and gamma distribution function (dotted line) with the same mean and coefficient of variation both compared to an empirical distribution function obtained by simulation of two samples of size 30 *from the gamma distribution. The gamma distribution gives the best fit to the empirical distribution function in the top diagram while the lognormal distribution gives the best fit to the empirical distribution function in the bottom diagram.*

distribution (Fisher–Tippet Type I extreme value distribution).

Luy and Rackwitz (1978) report that samples of 30 yearly extreme snow depth observations in Germany are for the larger part better fitted by the gamma distribution than by the lognormal or the Gumbel distribution.

It is a well–known consequence of the nature of statistical uncertainty that several distribution types seem to be reasonable candidates as models for the population from which the actual and only known sample of moderate size is drawn. This is easily demonstrated by a simulation experiment. Let a random variable X be distributed according to a gamma density, that is, a density proportional to $x^{k-1} e^{-ax}$, $x \in \mathbb{R}_+$, with parameters $k = 3$, $a = 1$. Then X has the mean and coefficient of variation $E[X] = k/a = 3$, $V_X = 1/\sqrt{k} \simeq 0.577$ respectively. These parameters are close to be representative for the ground snow load populations reported by Ellingwood and Redfield (1983). The graph of the corresponding distribution function is shown in dotted line in Fig. 1. The abscissa scale is logarithmic while the ordinate scale corresponds to fractiles of the normal distribution. Thus any lognormal distribution function appears as a straight line in this diagram. The straight line shown in Fig. 1 is the graph of the lognormal distribution function $F_{log}(x)$ corresponding to the mean 3 and the coefficient of variation $1/\sqrt{3}$. Fig. 1 also shows the graph of the empirical distribution function $S_n(x)$ = relative number of observations at most equal to x in the simulated sample $x_1,...,x_n$ of size $n = 30$ from the considered gamma distribution.

From the visual impression of the graphs it is difficult to draw any conclusions about whether or not $F_{gam}(x)$ fits the empirical distribution function $S_n(x)$ better than $F_{log}(x)$. Therefore it is necessary to compute some test statistics in order to try to be objective about the matter. Three different standard measures (test statistics) of the goodness of the fit are considered, Kendall and Stuart (1961,Vol.2, 450–452):

Kolmogorov–Smirnov:

$$A = \sqrt{n}\sup|S_n(x) - F(x)| = \sqrt{n}\max_{r=0}^{n}\left\{\left|y_r - \frac{r}{n}\right|\right\} \qquad (1)$$

Cramér–von Mises:

$$B = n\int_{-\infty}^{\infty}[S_n(x) - F(x)]^2 dF(x) = \sum_{r=1}^{n}\left[y_r - \frac{2r-1}{2n}\right]^2 + \frac{1}{12n} \qquad (2)$$

Anderson–Darling:

$$C = n\int_{-\infty}^{\infty}\frac{[S_n(x)-F(x)]^2}{F(x)[1-F(x)]}dF(x) = n\sum_{r=1}^{n}\left[\left[\frac{r-n}{n}\right]^2\log\left[\frac{1-y_r}{1-y_{r+1}}\right] + \left[\frac{r}{n}\right]^2\log\left[\frac{y_{r+1}}{y_r}\right]\right] + n\left[-1-\log(1-y_1)-\log y_n\right] \qquad (3)$$

in which $F(x)$ is the tested distribution function and $y_1 = F(x_1) \le y_2 = F(x_2) \le ... \le y_n = F(x_n)$ correspond to the ordered sample $x_1 \le x_2 \le ... \le x_n$. With $F = F_{gam}$ or $F = F_{log}$ we denote A as A_{gam} or A_{log} respectively, and we consider the indicator random variable I_A where $I_A = 1$ if $A_{log} < A_{gam}$, $I_A = 0$ otherwise. The indicator random variables I_B and I_C are defined correspondingly. Using the joint information from these statistics a reasonable decision rule might be to choose the gamma distribution if at least two of the three variables I_A, I_B, I_C take the value zero. Otherwise the lognormal distribution is chosen.

Even though it is known here that the data are drawn from the gamma distribution, it is an event of considerable probability to choose the lognormal distribution . The empirical distribution function in the top diagram in Fig. 1 leads to the choice of the gamma distribution while the empirical distribution function in the bottom diagram in Fig. 1 leads to the choice of the lognormal distribution. Getting a sample as in the bottom diagram of Fig. 1 may therefore erroneously lead to the adoption of a wrong distribution model for X. Of course, this is a trivial fact in the theory of mathematical statistics, but in the field of structural reliability this quite probable event of choosing the wrong distribution model can have severe consequences. For example, the 0.999 fractile in the gamma distribution of Fig. 1 is about 11.3 while the 0.999 fractile in the lognormal distribution is about 13.8 which is about 22% larger than the first. For the 0.9999 fractile the numbers are about 13.8 and 18.9 respectively with the last being about 37% larger than the first.

In order to evaluate the probability of making the wrong choice, the probability distribution of (I_A,I_B,I_C) has been calculated by repeated simulations of independent samples of X of size n = 30.

Table 1 shows the 8 probabilities obtained in the case where the exact values of E[X] and V_X are used, that is, when no parameters are estimated from the data (simple hypothesis), as well as the 8 probabilities obtained when E[X] and V_X are estimated sample by sample from the simulated data by the method of moments (composite hypothesis). These estimates are used to define the two alternative distribution functions from which the ordered samples $y_1 \le y_2 \le ... \le y_n$ are calculated. Usually it is the last situation with estimated parameters that is relevant in case of samples of data related to natural phenomena.

For the simple hypothesis it is seen that C gives the smallest error probability 0.20 while A gives the largest error probability 0.32 ,

Table 1. Probability distributions of (I_A,I_B,I_C) *for a sample of size* n = 30 *from the gamma distribution corresponding to the parameter values* $E[X] = 3$, $V_X = 1/\sqrt{3}$.

(I_A,I_B,I_C) =	(0,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(1,1,1)	(0,1,1)	(1,0,1)	(1,1,0)
prob.1)	0.61	0.08	0.02	0.03	0.14	0.02	0.01	0.09
	P(choice of gamma) = 0.74				P(choice of lognormal) = 0.26			
	$P(I_A=1) = 0.32$, $P(I_B=1) = 0.27$, $P(I_C=1) = 0.20$							
prob.2)	0.55	0.07	0.01	0.02	0.25	0.07	0.01	0.02
	P(choice of gamma) = 0.65				P(choice of lognormal) = 0.35			
	$P(I_A=1) = 0.35$, $P(I_B=1) = 0.35$, $P(I_C=1) = 0.35$							

1) exact parameters, 2) estimated parameters (method of moments)

results that seem to fit with the intuition when observing that C puts much more weight to the deviations in the tail regions than A does. However, in the usual situation of a composite hypothesis the three test statistics surprisingly give approximately the same error probability of 0.35 .

All in all it can be concluded that in more than 3 out of 10 cases the wrong distribution model will be chosen. In case of a modest sample of data of a natural phenomenon as for example the maximal yearly ground snow load we therefore run a considerable risk of choosing the wrong distribution model when the choice is made solely on the basis of a best fit criterion.

It is also interesting to note that if by bad luck the sample is such that the best fit criterion points at the wrong distribution model, then it takes a considerable increase of sample size before it becomes evident that the model is wrong. This is illustrated in Fig. 2 which shows the simulation estimate of the conditional probability of choosing the wrong distribution as a function of the additional sample size N given that the wrong distribution was chosen for the sample size n = 30. For N = 30 the probability of maintaining the wrong model is about 50% and it is still as high as about 20% for N = 150.

The Ellingwood–Redfield distribution investigation on snow loads

Ellingwood and Redfield (1983) make similar simulation investigations in order to evaluate the probability of choosing the wrong distribution model among the two alternatives they consider. Their choice among the two alternative distributions are based on the maximum probability plot correlation coefficient criterion, Filliben (1975). When samples of size n = 28 are generated from a suitably representative lognormal distribution, about 21% of the samples show a better fit to the Gumbel distribution. Alternatively, simulation from a Gumbel distribution gives that about 25% of the samples are better fitted by the lognormal distribution.

Thus the same difficulty of choosing among the two alternatives shows up. However, these investigators have another seemingly strong argument. They have samples not only from one weather station but from 76 (38) weather stations. Among these samples about 66% (76%) are fitted best by the lognormal distribution and the remaining 34% (24%) best by the Gumbel distribution. The numbers in parenthesis correspond to the subset of weather stations that all have 28 years record of observations with non–zero water equivalent in each year. The observed similarity between the results of simulation from the lognormal distribution and the actual findings among the weather stations seems to point at the lognormal distribution as being closer to the "true" distribution model of ground snow loads than the Gumbel distribution is. Indeed, if the data from the different weather stations can be considered as statistically independent, then it is an easy probability computation to see that the Gumbel distribution should be rejected as a tenable candidate.

This decisive independence assumption is the disputable point in Ellingwood's and Redfield's argumentation. One could as well argue that the mutual dependence is very strong. Cold winters and mild winters are usually not experienced as local phenomena (relatively) but are common to larger geographical regions.

For illustration of the effect of dependency between samples consider the random vector **Z** $= (X+Y_1, X+Y_2, \ldots, X+Y_m)$ where $X, Y_1, \ldots, Y_m$ are mutually independent, X has the gamma density with parameters $k = k_X$, $a = 1$ while $Y_1, \ldots, Y_m$ all have the gamma density with parameters $k = k_Y = 3 - k_X$, $a = 1$. Then the elements in **Z** all have the gamma density with parameters $k = 3$, $a = 1$. The vector **Z** could be envisaged to represent the simultaneous measurements at m stations. A sample of size n of Z gives m empirical distribution functions $S_{1n}(x),...,S_{mn}(x)$. Each of these empirical distribution functions defines a value of the indicator random variable $I = I_A$, say. Thus we get the m–vector $(i_1,...,i_m)$ of zeros and ones and we have that $p = (i_1+...+i_m)/m$ is the fraction of the m measuring stations for which the lognormal distribution is chosen on the basis of the test statistic A . By repeated simulations a large sample of outcomes of p is finally generated and from this sample a distribution function of p is estimated.

A simulation investigation can also be made in the case where the gamma distribution assumption is replaced by a Gumbel distribution assumption allowing another type of dependence model than above. The observations at the m stations are now most conveniently represented by the vector $\mathbf{Z} = (Z_1,...,Z_m)$ where $Z_i = [\max\{X,Y_i\} \mid \max\{X,Y_i\} > 0]$, $i = 1,...,m$, has the truncated Gumbel distribution $[F(x)-F(0)]/[1-F(0)]$ with $F(x) = \exp\{-\exp[-\alpha(x-\beta)]\}$, $x \in \mathbb{R}$. This distribution is obtained by letting X and Y_i be mutually independent random variables that have Gumbel distributions with parameters α, β_x for X and α, β_y for Y_i such that $\exp[\alpha\beta] = \exp[\alpha\beta_x] + \exp[\alpha\beta_y]$. Giving α and β those values that make the mean and variance of Z_i the same as in the gamma distribution case, and also giving β_x such a value that $E[X] = k_x$ we get a reasonable case for comparisons of the two different dependence models.

Under the independence assumption (i.e. if $X \equiv 0$) only a small influence of the simulation distribution type should be expected. Any difference between the two distribution functions for p is an effect of the estimation uncertainty of the moments. If the exact moments are used (simple hypothesis) the results are identical because then the values $y_1 \leq ... \leq y_n$ in both cases are observations of an ordered sample from the uniform distribution.

As in Ellingwood and Redfield (1983) the simulation experiments are run for n = 28 and m = 38 . Using the moment estimates of the parameters of the two alternative distribution functions gamma or lognormal (Gumbel or lognormal), the distribution functions of p corresponding to $k_x = 0$ ($X \equiv 0$, i.e. independence) and $k_x = 2.5$ are shown in Fig. 3. It is interesting that the two models also in the dependence case give almost the same distribution functions for p .

For the Ellingwood and Redfield snow data the hypothesis that the data population is lognormal implies that the fraction of cases for which the Gumbel distribution gives the best fit should be distributed approximately as in Fig. 3. It is seen that 0.24 is about the 0.05–fractile in

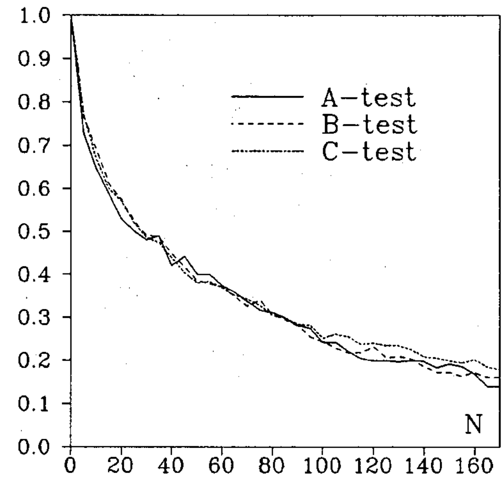

Fig. 2. Simulated estimation of the conditional probability that the lognormal distribution fits better a sample of size 30+N *than the gamma distribution given that the lognormal distribution fits better the subsample consisting of the first* 30 *observations. All sample values are generated from the gamma distribution.*

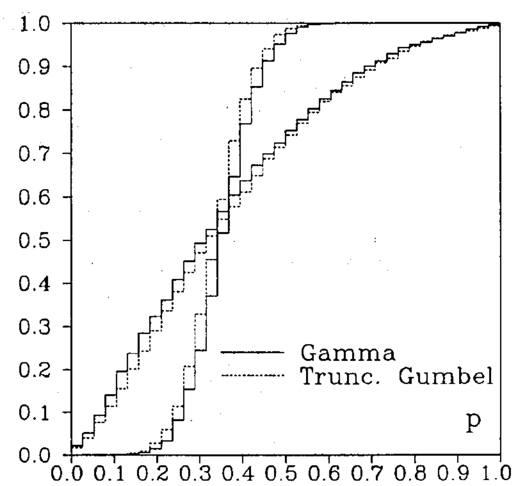

Fig. 3. Distribution functions for the fraction p *of* m = 38 *measuring stations at which the lognormal distribution is chosen instead of the correct gamma distribution (or the correct Gumbel distribution). The distribution choice is made on the basis of a sample of size* n = 28 *using the Cramer–Von Mises test statistic with estimated parameters of the two alternative distributions. The steep curves correspond to independence between measuring stations while the other curves correspond to a specific degree of dependence (in the gamma case the equicorrelation coefficient is 2.5/3 ≃ 0.83).*

case of independence and about the 0.40–fractile in the considered case of dependence. This observation increases the doubt about the validity of the independence assumption. Clearly, the opposite hypothesis of having a Gumbel population is totally ruled out under the independence assumption by observing that 0.76 is far out in the upper tail of the distribution of the fraction of wrong distribution choices. However, if the independence assumption is removed the situation is different. For the considered case of dependence it is seen that 0.76 is about the 0.92–fractile.

Choice of snow load distribution type for structural reliability analysis

Without further considerations Ellingwood's and Redfield's investigation makes it reasonable to choose the lognormal distribution as the standard distribution for ground snow load in the considered region. However, there is yet another consideration to be made. As a formal load distribution model for reliability analysis the fat upper tail of the lognormal distribution may cause that the corresponding loads always dominate over loads that have been assigned distribution types with less fat upper tails. This strong and more or less arbitrary weight on some of the combining load types may be judged to be unreasonable from an engineering point of view. Besides neutralizing the tail sensitivity problem the purpose of the standardization is also to ensure that no single load type arbitrarily gets an overweight of influence on the calculated structural reliability.

In order to justify that there is an empirical evidence of a fat upper tail extending far beyond the range of the available sample of measured values, more than just the limited sample is needed. At least support should be obtained from some reasonable model of a mechanism that generates the fat tail. In this respect the lognormal distribution has a weak position in relation to most physical processes. The lognormal distribution family is closed with respect to multiplication of the random variables but not with respect to addition.

In a continental climate it is reasonable to consider the ground snow load as a result of an additive accumulation process. Taking the skewness of the observed empirical distribution into account the additive mechanism suggests the gamma distribution family as a reasonable candidate. This complies with the findings of Luy and Rackwitz (1978).

Also the Gumbel distribution type may be mechanistically defended. In particular this is the case for marine climates with isolated not accumulating snowfalls. If the snowfalls occur as the pulses of a homogeneous Poisson process with an exponentially distributed maximal snow load at each pulse, the extreme snow load during any period will have a Gumbel distribution function on the positive axis.

Independent of the JCSS there is a CIB Commission W 81: "Actions on Structures" with the task of writing reports on stochastic models for actions which are mutually consistent and which can be used both in probabilistic design and analysis, and as a basis for deterministic models of actions. Until now three reports have been published, CIB, W 81 (1989a, 1989b, 1991) on "self–weight loads", "live loads in buildings", and "snow loads" respectively. On the basis of the arguments given here it is in the snow loads report suggested to standardize the gamma distribution type (or mixtures of gamma distributions) rather than the lognormal distribution type for ground snow loads to be used in structural reliability analyses.

Conclusions

It follows from the simulation demonstrations herein that a best fit criterion is not sufficient as the basis for choosing distribution models for reliability analysis. In particular it is important to be aware of this fact if the reliability analysis is used as the basis for design decisions in competing consulting engineering companies. If there is a free choice of distribution types, competition about material savings, say, can easily end up being based on conflicting "false" information from arbitrary and practically nonverifiable modeling, that is, modeling that carries no empirical evidence. The practical answer to this dilemma is the use of codified internally harmonized standardizations of distribution types imposed on all competitors by some authorized code committee.

Acknowledgement

The computer programming has been made by Johannes M. Johannesen at the author's department. The work has been financially supported by the Danish Technical Research Council.

References

CIB, W 81 (1989a). "Actions on structures. Selfweight loads". CIB Report, Publication 115.

CIB, W 81 (1989b). "Actions on structures. Live loads in buildings". CIB Report, Publication 116.

CIB, W 81 (1991). "Actions on structures. Snow loads". CIB, W81 (1991), Publication 141.

Ditlevsen, O. and Madsen, H.O. (1989). "Proposal for a code for the direct use of reliability methods in structural design". Working document, Joint Committee on Structural Safety, ed. IABSE–AIPC–IVBH, ETH–Hönggerberg, Zürich, Switzerland. Reprinted in slightly revised form in Ditlevsen, O. and Madsen, H.O. (1994). "Structural Reliability Methods" (translation of SBI–rapport 211: Bærende konstruktioners sikkerhed (1990)).

Ellingwood, B. and Redfield, R. (1983). "Ground snow loads for structural design". *J. Struct. Engrg.*, ASCE, 109 (4), 950 – 964.

Filliben, J.K. (1975). "The probability plot correlation coefficient test for normality". *Technometrics*, 17(1), 111–117.

Kendall, M.G. and Stuart, A. (1961). *The advanced theory of statistics*, Vol. 2: Inference and relationship, Charles Griffin, London.

Luy, H. and Rackwitz, R. (1978). "Darstellung und Auswertung von Schneehöhenmessungen in der Bundesrepublik Deutschland". *Berichte zur Zuverlässigkeitstheorie der Bauwerke, Heft 31*, Technische Universität München.