Stress-strenght model for skew-normal distributions

A. Azzalini, M. Chiogna

2002.4

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Aprile 2002

Stress–strength model for skew-normal distributions

A. Azzalini & M. Chiogna

Dipartimento di Scienze Statistiche, Università di Padova
Working paper 2002-04 (April 2002)

1 Introduction

The stress-strength problem The stress–strength problem, as it is called in statistical quality control, is concerned with evaluation of the probability that the strength of a given material is larger than the stress which is applied to that material. Phrased in probability terms, the previous question leads to evaluation of

$$\rho = \mathbb{P}\{Y_2 < Y_1\}. \tag{1}$$

for two random variables Y_1 and Y_2, representing strength and stress, respectively.

The problem comes actually in two parts: (i) given some assumptions on the distribution of the variables, provide an expression for ρ; (ii) given a set of data which are supposed to fulfill these assumptions, estimate ρ, both in the form of point estimation and of interval estimation.

Early literature in this area has focussed on the normal assumption for Y_1 and Y_2. This includes Church & Harris (1970) and Downton (1973) who have provided basic results. A number of other distributions and other variants of the problem have been discussed since then. A recent introductory account on this topic is given by Blischke & Prabhakar Murthy (2000, p. 276–8).

The purpose of this note is to examine the above problem when all or some of the underlying probability distributions are of skew-normal type; a brief summary of the skew-normal distribution is provided below.

The skew-normal distribution A random variable Z is said to have a skew-normal distribution if it is continuous and its density function is

$$\phi(z;\lambda) = 2\phi(z)\,\Phi(\lambda z), \quad (z \in \mathbb{R}), \tag{2}$$

for some parameters $\lambda \in \mathbb{R}$; here $\phi(z)$ and $\Phi(z)$ denote the N(0,1) density and distribution function, respectively. The shape of (2) is skewed to the right or to the left, according to the sign of λ; for $\lambda = 0$ we obtain a standard normal density. In fact, (2) refers to the 'standard' skew-normal distribution; if a linear trasform $Y = \xi + \omega Z$ is considered, we shall then say that Y has distribution $SN(\xi, \omega^2, \lambda)$.

Various results about (2) are given by Azzalini (1985); see also Azzalini and Capitanio (1999) and references therein. For our purposes, we only recall that its distribution function $\Phi(z;\lambda)$ enjoys the following properties

$$\begin{aligned} \Phi(-z;\lambda) &= 1-\Phi(z;-\lambda),\\ \Phi(z;\lambda) &= \Phi(z)-2\,T(z,\lambda),\\ \Phi(0;\lambda) &= \frac{1}{2}-\frac{1}{\pi}\arctan\lambda,\\ \Phi(z;0) &= \Phi(z) \end{aligned}$$

where $T(z,\lambda)$ is the function studied by Owen (1956). A computer routine to evaluate $T(z,\lambda)$ has been given by Young and Minder (1974), subsequently improved by various authors.

2 Probability results

Assume that the observed variables Y_1, Y_2 are of type

$$Y_i = \xi_i + \omega_i Z_i, \qquad Z_i \sim SN(\lambda_i) \qquad (i=1,2) \tag{3}$$

and Y_1 is independent of Y_2. Under these assumptions,

$$\begin{aligned} \rho &= \mathbb{P}\{\xi_2+\omega_2 Z_2 < \xi_1+\omega_1 Z_1\}\\ &= \mathbb{E}_{Z_1}\{\mathbb{P}\{\xi_2+\omega_2 Z_2 < \xi_1+\omega_1 Z_1 | Z_1\}\}\\ &= \mathbb{E}_{Z_1}\left\{\mathbb{P}\left\{Z_2 < \frac{\xi_1-\xi_2+\omega_1 Z_1}{\omega_2}\right\}|Z_1\right\}\\ &= \mathbb{E}_{Z_1}\left\{\Phi\left(\frac{\xi_1-\xi_2+\omega_1 Z_1}{\omega_2};\lambda_2\right)|Z_1\right\}. \end{aligned} \tag{4}$$

An explicit expression of this quantity in the general case does not seems feasible. However, if one relaxes the assumption of skew-normality for one component, then computation follows easily from Proposition 2 of Chiogna (1998), which for convenience we reproduce here, up to inessential modifications.

Proposition 1 *If* $Z \sim SN(\lambda)$ *and* $U \sim N(0,1)$, *then*

$$\begin{aligned} \mathbb{E}\{\Phi(hU+k;\lambda)\} &= \Phi(k/\sqrt{1+h^2};\ m(h,\lambda)),\\ \mathbb{E}\{\Phi(hZ+k)\} &= \Phi(k/\sqrt{1+h^2};\ m(1/h,-\lambda)), \end{aligned}$$

where

$$m(h,\lambda) = \frac{\lambda}{\sqrt{1+h^2(1+\lambda^2)}}.$$

If we assume $\lambda_1 = 0$ in (4), we can make use of the first statement of the above proposition and obtain

$$\rho = \Phi(\Delta;\Lambda) = \Phi(\Delta) - 2\,T(\Delta,\Lambda) \tag{5}$$

where

$$\Delta = \frac{\xi_1 - \xi_2}{\sqrt{\omega_1^2 + \omega_2^2}}, \qquad \Lambda = m(\omega_1/\omega_2, \lambda_2).$$

Notice the following facts. If $\lambda_2 = 0$, then also $\Lambda = 0$ and the above expression becomes $\rho = \Phi(\Delta)$, the usual expression for the normal case. If we assume instead that $\lambda_2 = 0$ and λ_1 is unrestricted, then we have the same development as above but (4) is now computed using the second statement of Proposition 1, leading to

$$\rho = \Phi(\Delta, m(\omega_2/\omega_1, -\lambda_1))$$

which is of the same type of the expression for the earlier case. Hence we do not need a special development, and in the following we shall concentrate on the case $\lambda_1 = 0$.

We turn now to consider the case when both λ_1 and λ_2 in (3) are unrestricted, but $\xi_1 = \xi_2$. Hence (4) now becomes

$$\mathbb{E}_{Z_1}\{\Phi\left((\omega_1/\omega_2)Z_1; \lambda_2\right)|Z_1\}$$

and this is readily expressed using Proposition 3 of Chiogna (1998), which again is reproduced here up to inessential modifications.

Proposition 2 *If* $Z \sim SN(\lambda)$*, then*

$$\mathbb{E}\{\Phi(hZ; \beta)\} = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{m(h,\beta) - m(1/h,\lambda)}{m(h,\beta) + m(1/h,\lambda)}. \tag{6}$$

Hence we can write

$$\rho = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{m(h,\lambda_2) - m(1/h,\lambda_1)}{m(h,\lambda_2) + m(1/h,\lambda_1)}.$$

where $h = \omega_1/\omega_2$.

Evaluation of (1) in another special case has been considered by Gupta & Brown (2001); see in particular Section 4. The important difference in the assumptions is that (Z_1, Z_2) in their case follow a bivariate skew-normal distribution with correlated components. The multivariate skew-normal distribution has been studied by Azzalini & Dalla Valle (1996) and subsequently by Azzalini & Capitanio (1999).

3 Likelihood Inference

Consider the problem of statistical inference for ρ under the assumption of independent samples from Y_1 and from Y_2 to make inference on (5).

Since the context is outside the exponential family, UMVU estimation is not feasible and we consider then likelihood inference. Specifically, we replace the parameters in (5) by their MLE's. Notice that, in the normal distribution case, Downton (1973) compares UMVU with other procedures including one which is MLE (although it is not recognised as such); see his r_1^* in the one-sample case and a similar one in the two-sample case. These turn out to compare very favourably with UMVU.

For a generic parameter ψ denote by $\hat{\psi}$ its MLE; similarly $\hat{\Delta}, \hat{\Lambda}$ are obtained by appropriate transformations of Δ, Λ. The estimate of (5) is then given by

$$\hat{\rho} = \Phi(\hat{\Delta}; \hat{\Lambda}) = \Phi(\hat{\Delta}) - 2\,T(\hat{\Delta}, \hat{\Lambda}). \tag{7}$$

Also we shall use the notation $\hat{\psi}_-$ and $\hat{\psi}_+$ to indicate $\hat{\psi} - u_\alpha\,\text{s.e.}(\hat{\psi})$ and $\hat{\psi} + u_\alpha\,\text{s.e.}(\hat{\psi})$ respectively, for an appropriate choice of the normal quantile u_α such that $\Phi(-u_\alpha) = \alpha/2$, for a given choice of the confidence level $1-\alpha$.

Confidence intervals of the form $(\hat{\rho}_-, \hat{\rho}_+)$ are not appropriate, since this may exceed the interval $(0, 1)$. More sensible choices are

I: $(\Phi(\hat{\Delta}_-; \hat{\Lambda}),\ \Phi(\hat{\Delta}_+; \hat{\Lambda}))$,

II: $(\Phi(\hat{\Delta}_-; \hat{\Lambda}_+),\ \Phi(\hat{\Delta}_+; \hat{\Lambda}_-))$,

where the latter form is supported by the fact that $\Phi(z; \lambda)$ is a decreasing function of λ, for any fixed z.

The effectiveness of these choices needs to be assessed by simulation methods. In particular we must consider: (i) bias of (7), (ii) variance of (7), and especially (iii) the associated actual level for procedures I and II, in comparison with the nominal level, To obtain the required standard errors we resort on asymptotic theory. Some points to keep in mind are as follows.

1. Standard errors are obtained via the Fisher information evaluated at the MLE's. Having to choose between observed and expected information, the first one seems preferable to avoid the numerical integrations involved by the other case. Furthermore, the use of observed instead of expected information is keeping with general considerations, such as those given e.g. by Efron & Hinkley (1978).

2. Small samples likelihood inference for the skew-normal distribution poses problems. This fact has emerged from the results of several people, variously focusing on the theoretical and the practical aspects; see Azzalini (1985), Chiogna (1997), Azzalini & Capitanio (1999, section 5), Pewsey (2000). It is then sensible to consider estimation of the parameter only when the sample size is not too small. As a crude guideline, $n = 50$ seems to be about the practical lower bound.

Let $\theta_1 = (\xi_1, \omega_1)$ and $\theta_2 = (\xi_2, \omega_2, \lambda_2)$. The observed Fisher information matrix for the parameter $\theta = (\theta_1, \theta_2)$ has a block structure, because of the independence between the two samples. Denote by $j_{ii}(\theta_i)$ the block corresponding to θ_i, $i = 1, 2$. It is well know that

$$j_{11}(\theta_1) = n \begin{pmatrix} 1/\omega_1^2 & 0 \\ 0 & 2/\omega_1^2 \end{pmatrix}$$

which of course must be evaluated at $\theta_1 = \hat{\theta}_1$. As for $j_{22}(\theta_2)$, the observed information matrix can be obtained from the formulae given in the appendix.

To obtain asymptotic variances (av) for $\hat{\Delta}$ and $\hat{\Lambda}$, it is possible to resort on the multivariate δ-method; for a standard account on this techniques, see for instance Schervish

(1995). By applying the δ-method one obtains:

$$\text{av}(\hat{\Delta}) = \frac{1}{\omega_1^2+\omega_2^2}\left\{\frac{\omega_1^2}{n}\left(1+2\delta^2\omega_1^4\right)+\sigma_{11}-2\sigma_{12}\omega_2\delta+\sigma_{22}\omega_2^2\delta^2\right\},$$

$$\text{av}(\hat{\Lambda}) = \left(\frac{\Lambda}{\omega_2\lambda_2}\right)^6\left\{\frac{2\omega_1^6\omega_2^2\psi^2}{n}+\sigma_{22}\omega_1^4\psi^2+2\sigma_{23}\omega_1^2\psi\bar{\omega}+\sigma_{33}\bar{\omega}^2\right\},$$

where σ_{ij} denotes the (i,j) entry of $j_{22}(\theta_2)^{-1}$, $\delta = \Delta^2/(\xi_2-\xi_1)$, $\psi = \lambda_2(1+\lambda_2^2)$ and $\bar{\omega} = \omega_2(\omega_1^2+\omega_2^2)$. Notice that σ_{ij}'s are terms of order $1/n$.

To perform simulations, it has been assumed that the parameters of Y_1 are known, in analogy with Curch & Harris (1970) and Downton (1973). Without loss of generality, it is assumed that Y_1 is distributed as a standard normal variate, so that $\theta = (0, 1, \xi_2, \omega_2, \lambda_2)$. In this case, it is easily shown that previous expressions reduce to:

$$\text{av}(\hat{\Delta}) = \frac{1}{1+\omega_2^2}\left\{\sigma_{11}-\frac{2\sigma_{12}\xi_2\omega_2}{1+\omega_2^2}+\frac{\sigma_{22}\xi_2^2\omega_2^2}{1+\omega_2^2}\right\},$$

$$\text{av}(\hat{\Lambda}) = \frac{1}{(1+\omega_2^2+\lambda_2^2)}\{\sigma_{22}(\lambda_2(1+\lambda_2^2))^2+2\sigma_{23}(\omega_2(1+\omega_2^2))(\lambda_2(1+\lambda_2^2))+\sigma_{33}(\omega_2(1+\omega_2^2))^2\}$$

Simulation work has been carried out to evalutate the actual level achieved by methods I and II described above when they are used at the nominal level 95%; the outcome is summarised in Table 1. Each entry of the tables has been obtained from 1000 generated samples and the number of cases where the confidence interval covers the actual value of ρ is reported.

Inspection of the table indicates that the agreement to the nominal level is greater for small λ_2 and for large n. Serious discrepancies from 95% are encountered when $n = 50$ and $\lambda_2 = 10$ or even 5. Indications are much more favourable is the other cases, especially for $n = 200$ and $\lambda_2 = 2$, and to some extent $\lambda_2 = 5$.

In interpreting these values, one must bear in mind that, in a number of industrial applications to measurements data, large asymmetries are not to be expected. In this sense, the bad entries for very large λ_2 must be down-weighted, and the overall picture emerging from the table is acceptable comfortable.

Acknowledgments

We would like to thank Samuel Kotz for useful discussions. This research has been carried out within the project 'Extensions of the Gaussian probability distribution and their applications', supported by MIUR (grant PRIN 2000), Italy.

Appendix

Log-likelihood function and derivatives for SN variates

Consider a random sample $y_1, \dots, y_n$ from $SN(\xi, \omega^2, \lambda)$. Write

$$y_i = \xi + \omega\, z_i, \qquad Z \sim SN(0, 1, \lambda).$$

ρ	ω_2	λ_2	n	I	II	ρ	ω_2	λ_2	n	I	II	ρ	ω_2	λ_2	n	I	II
0.90	0.5	2	50	941	974	0.95	0.5	2	50	946	974	0.999	0.5	2	50	918	955
			100	976	998				100	967	999				100	949	989
			200	973	1000				200	952	1000				200	905	991
			500	957	1000				500	967	1000				500	911	998
		5	50	759	859			5	50	747	852			5	50	783	832
			100	871	978				100	874	967				100	906	941
			200	888	999				200	871	996				200	928	983
			500	871	1000				500	879	1000				500	923	993
		10	50	531	627			10	50	528	604			10	50	562	599
			100	720	864				100	756	856				100	785	837
			200	829	982				200	841	976				200	909	963
			500	840	1000				500	894	999				500	923	989
	1	2	50	950	971		1	2	50	927	965		1	2	50	939	948
			100	952	996				100	929	983				100	941	963
			200	940	994				200	949	987				200	932	952
			500	931	997				500	924	988				500	942	959
		5	50	796	841			5	50	805	840			5	50	788	800
			100	905	961				100	910	951				100	921	934
			200	933	982				200	934	970				200	948	964
			500	941	995				500	936	985				500	945	962
		10	50	592	682			10	50	574	596			10	50	569	577
			100	802	841				100	802	828				100	826	838
			200	909	959				200	934	964				200	915	926
			500	931	992				500	929	980				500	950	971
	3	2	50	935	949		3	2	50	938	946		3	2	50	943	961
			100	943	963				100	949	955				100	967	972
			200	958	968				200	936	942				200	958	959
			500	931	942				500	935	937				500	948	948
		5	50	815	838			5	50	806	816			5	50	813	827
			100	927	951				100	932	946				100	916	940
			200	941	958				200	948	959				200	953	968
			500	958	961				500	953	955				500	965	965
		10	50	582	596			10	50	601	606			10	50	608	612
			100	798	816				100	833	852				100	801	821
			200	937	956				200	930	955				200	927	950
			500	966	978				500	955	969				500	956	972

Table 1: Estimated actual confidence levels for ρ using methods I and II described in the text, multiplied by 1000

Then the log-likelihood for $\mathrm{DP} = (\xi, \omega, \lambda)$ is

$$\ell(\mathrm{DP}) = \sum_i \left(-\log\omega - \tfrac{1}{2}z_i^2 + \log\{2\Phi(\lambda z_i)\}\right) = -n\log\omega - \tfrac{1}{2}\sum z_i^2 + \sum \zeta_0(\lambda z_i)$$

with partial derivatives

$$\begin{aligned}
\frac{\partial \ell}{\partial \xi} &= \sum\{-z_i + \lambda\,\zeta_1(\lambda z_i)\}\,(-1/\omega),\\
\frac{\partial \ell}{\partial \omega} &= \sum\{-z_i + \lambda\,\zeta_1(\lambda z_i)\}\,(-z_i/\omega) - n/\omega,\\
\frac{\partial \ell}{\partial \lambda} &= \sum\{z_i\,\zeta_1(\lambda z_i)\}
\end{aligned}$$

Here $\zeta_0(x) = \log\{2\Phi(x)\}$ and $\zeta_r(x)$ is its r-th derivative. The negative second derivatives of ℓ are :

$$\begin{aligned}
-\frac{\partial^2 \ell}{\partial \xi^2} &= \sum\{1 - \lambda^2\zeta_2(\lambda\, z_i)\}/\omega^2,\\
-\frac{\partial^2 \ell}{\partial \xi \partial \omega} &= \sum\left\{[1 - \lambda^2\zeta_2(\lambda\, z_i)]z_i + z_i - \lambda\zeta_1(\lambda z_i)\right\}/\omega^2\\
-\frac{\partial^2 \ell}{\partial \xi \partial \lambda} &= \sum\{\zeta_1(\lambda z_i) + \lambda z_i \zeta_2(\lambda z_i)\}/\omega,\\
-\frac{\partial^2 \ell}{\partial \omega^2} &= \sum\left\{z_i^2[3 - \lambda^2\,\zeta_2(\lambda z_i)] - 2\lambda z_i\,\zeta_1(\lambda z_i) - 1\right\}/\omega^2,\\
-\frac{\partial^2 \ell}{\partial \omega \partial \lambda} &= \sum\{\zeta_1(\lambda z_i) + \lambda\, z_i\,\zeta_2(\lambda z_i)\}\,z_i/\omega\\
-\frac{\partial^2 \ell}{\partial \lambda^2} &= -\sum z_i^2\,\zeta_2(\lambda z_i)
\end{aligned}$$

References

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scand. J. Statist.* **12**, 171–8.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distributions. emphJ. R. Statist. Soc. B **61**, 579–602.

Azzalini, A.& Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–26.

Blischke, W. R. & Prabhakar Murthy, D. N. (2000) *Reliability: modeling, prediction and optimization* Wiley, New York.

Chiogna, M. (1997). Notes on estimation problems with scalar skew-normal distributions. Working paper 1997-15, Dept. Statistical Sciences, University of Padua.

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc* 7, 1-13.

Church, J. D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships (Com: V12 p719-720). *Technometrics* **12**, 49–54.

Downton, F.(1973). The estimation of Pr(Y < X) in the normal case. *Technometrics* **15**, 551–558.

Efron, B. & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65**, 457–487.

Gupta, R. C. & Brown, N. (2001). Reliability studies of the skew-normal distribution and its application to a strength-stress model. *Comm. Statist. – Theory & Methods* **30**,2417–2445.

Owen, D.B. (1956). Tables for computing bivariate Normal probabilities, *Ann. Math. Statist.* **27**, 1075-1090.

Young, J. C. and Minder, Ch. E. (1974). [Algorithm AS 76] An integral useful in calculating non-central t and bivariate normal probabilities (AS R26: 78V27 p379; AS R30: 79V28 p113; Corr: 79V28 p336; AS R55: 85V34 p100-101; AS R65: 86V35 p310-312; AS R80: 89V38 p580-582; AS R89: 92V41 496-497). *Applied Statistics*, **23**, 455–457.

Pewsey (2000). Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist.* **27**, 859–770.

Schervish, M. J. (1995). *Theory of statistics.* Springer, New York.