

A semiparametric approach to the stress-strength problem

Un approccio semiparametrico al problema sollecitazione-resistenza

Gianfranco Adimari, Monica Chiogna¹

Dipartimento di Scienze Statistiche

Università di Padova

[adimari, monica]@stat.unipd.it

Riassunto: In questo lavoro, il problema dell'inferenza sulla probabilità che la resistenza di un certo materiale sia superiore ad una sollecitazione ad esso applicata è affrontato in un contesto semiparametrico. In particolare, si propone una tecnica per la costruzione di intervalli di confidenza per tale probabilità, che presuppone l'assunzione di un modello parametrico solo per la variabile che rappresenta la resistenza del materiale. L'utilità del metodo proposto viene saggiata mediante uno studio di simulazione, di cui vengono riportati alcuni risultati.

Keywords: confidence interval, empirical likelihood, maximum likelihood estimation, normal approximation approach.

1. Introduction and motivation

The statistical problem in many engineering situations is to evaluate the probability that a component stands a stress applied to it. Take X to be a random variable (rv) corresponding to a stress potentially affecting some component, and take Y to be a rv measuring the strength at failure of the component, i.e. the stress at which the component will fail. The interest lies in evaluating the quantity:

$$\rho = \Pr\{Y > X\},$$

i.e. the probability that the component will not fail.

In the reliability literature, this problem is known as the stress-strength problem and it involves two aspects: (i) deriving theoretical expressions for ρ under distributional assumptions about X and Y ; (ii) making inference about ρ in light of sample data. In applications, the data for the strength could be actual data indicative of the strength of the material, i.e. maximum applied stress to cause failure, and the stress data could be actual stress data of the material under use conditions.

A wide literature has developed on the problem since the mid 50's, addressing both the points (i) and (ii) above mentioned. It is practically impossible to mention all the contributors to the topic; we refer the interested reader to the exhaustive review of Kotz *et al.* (2003), which provides a comprehensive overview of the existing literature. Most contributions addressing inference about ρ take a parametric point of view, although some work has been developed also in nonparametric settings.

In this paper, we develop a procedure to handle problems in which distributional assumptions can be reasonably formulated for one of the two rv's, whereas no safe probabilistic assertions can be expressed for the second rv. In reliability studies, this is

¹Via Cesare Battisti 241, 35121 Padova

justified by the fact that the stress variable is usually difficult to model accurately, due to the lack of sufficient knowledge about the stress in use of a component, whereas the strength variable might be more easily elicited via expert opinion. In such a semiparametric framework, we shall consider in particular the problem of building confidence intervals for ρ .

2. Main results

Consider a general parametric model $\{F(y; \theta); \theta \in \Theta\}$ for the variable Y representing the strength. Here, $F(y; \theta) = \Pr\{Y \leq y; \theta\}$ denotes the cumulative distribution function (cdf) which depends on an unknown parameter θ belonging to some set $\Theta \subseteq \mathbb{R}^q$, $q \geq 1$. Let $S(y; \theta) = 1 - F(y; \theta)$ denotes the survival function corresponding to $F(y; \theta)$. We do not make any parametric assumption about the distribution of the variable X representing the stress. We only assume that X is independent of Y .

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from X and $Y_1, Y_2, \dots, Y_m$ a random sample of size m from Y . In the following, we shall assume that the ratio n/m converges to some positive and finite constant κ , as n and m go to $+\infty$. Let θ_0 and ρ_0 indicate the true values of θ and ρ , respectively. Since

$$\Pr\{Y > X\} = E\{\Pr\{Y > X|X\}\} = E\{S(X; \theta_0)\},$$

an obvious estimator for ρ_0 is given by

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n S(X_i; \hat{\theta}),$$

where $\hat{\theta}$ denotes the maximum likelihood estimator (mle) for θ_0 based on the sample $Y_1, Y_2, \dots, Y_m$. Under standard regularity conditions, which ensure the consistency and the asymptotic normality of the mle $\hat{\theta}$ and legitimate some Taylor expansions, we can write

$$\frac{1}{n} \sum_{i=1}^n S(X_i; \hat{\theta}) = \frac{1}{n} \sum_{i=1}^n S(X_i; \theta_0) + \left(\frac{1}{n} \sum_{i=1}^n \dot{S}(X_i; \theta_0) \right)' (\hat{\theta} - \theta_0) + o_p(n^{-1/2}),$$

where $\dot{S}(\cdot; \cdot)$ denotes the derivative of $S(y; \theta)$ with respect to θ . Then

$$\sqrt{n}(\hat{\rho} - \rho_0) = \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n S(X_i; \theta_0) - \rho_0 \right\} + \sqrt{\kappa m} \beta' (\hat{\theta} - \theta_0) + o_p(1),$$

being $\beta = E\{\dot{S}(X; \theta_0)\}$. It follows that $\sqrt{n}(\hat{\rho} - \rho_0)$ is asymptotically normal with mean zero and variance $\omega^2 = \omega_S^2 + \kappa \beta' \Omega \beta$, where ω_S^2 is the variance of $S(X; \theta_0)$ and Ω denotes the asymptotic covariance matrix of $\sqrt{m} \hat{\theta}$. This suggests that confidence intervals for ρ_0 may be obtained by the classical normal approximation approach. Clearly, we need to estimate the asymptotic variance ω^2 . However, despite its usefulness, there are some drawbacks associated with this approach. Firstly, this method does not always work well for small samples, yielding confidence intervals with poor accuracy. Secondly, the method is not range-preserving, so that confidence intervals for ρ_0 could contain values outside its range. Finally, the method imposes a predetermined symmetry constraint on the shape of the confidence intervals, which may be inappropriate in several circumstances.

To overcome the drawbacks encountered by the normal approximation approach, we explore an alternative way. If θ_0 were known, one could construct confidence intervals for ρ_0 , the mean of $S(X; \theta_0)$, using the empirical likelihood method (Owen (1988), Owen (1990)). The empirical log-likelihood function for the mean of $S(X; \theta_0)$, evaluated at some possible candidate ρ , is defined as

$$l(\rho) = -2 \sup_{p_1, p_2, \dots, p_n} \sum_{i=1}^n \log(np_i),$$

where the supremum is taken subject to the constraints

$$p_i \geq 0 \text{ for all } i, \quad \sum_{i=1}^n p_i = 1 \quad \text{and} \quad \sum_{i=1}^n S(X_i; \theta_0) p_i = \rho. \quad (1)$$

It follows from Owen (1988) that, when evaluated at the true value ρ_0 , the empirical log-likelihood has an asymptotic chi-square distribution with one degree of freedom, that is $l(\rho_0) \xrightarrow{d} \chi_1^2$.

Of course, $l(\rho)$ cannot be computed since it depends on θ_0 which is unknown. A natural solution is then to replace θ_0 in (1) by its mle $\hat{\theta}$. This leads to an estimated empirical log-likelihood defined as

$$\hat{l}(\rho) = -2 \sup \left\{ \sum_{i=1}^n \log(np_i) : p_i \geq 0 \forall i, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n S(X_i; \hat{\theta}) p_i = \rho \right\}.$$

A Lagrangian argument gives an explicit expression for $\hat{l}(\rho)$. Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ be the ordered X -values. When $S(X_{(1)}; \hat{\theta}) < \rho < S(X_{(n)}; \hat{\theta})$, we have

$$\hat{l}(\rho) = 2 \sum_{i=1}^n \log\{1 + \lambda(S(X_i; \hat{\theta}) - \rho)\},$$

where $\lambda = \lambda(\rho)$ is the solution of the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{S(X_i; \hat{\theta}) - \rho}{1 + \lambda(S(X_i; \hat{\theta}) - \rho)} = 0.$$

Due to the substitution of θ_0 with $\hat{\theta}$ in (1), $\hat{l}(\rho_0)$ no longer has the usual asymptotic chi-square distribution. However, we show that, under weak regularity conditions,

$$\hat{l}(\rho_0) = n \frac{(\hat{\rho} - \rho_0)^2}{\hat{\omega}_S^2} + o_p(1) \quad \text{as } n, m \rightarrow +\infty,$$

where $\hat{\omega}_S^2 = n^{-1} \sum_{i=1}^n (S(X_i; \hat{\theta}) - \hat{\rho})^2$ is a consistent estimator for ω_S^2 . Therefore, if $\hat{\Omega}$ denotes a consistent estimator for the asymptotic covariance matrix of $\sqrt{m} \hat{\theta}$, we have that $\hat{\omega}^2 = \hat{\omega}_S^2 + (n/m) \hat{\beta}' \hat{\Omega} \hat{\beta}$, with $\hat{\beta} = n^{-1} \sum_{i=1}^n \dot{S}(X_i; \hat{\theta})$, is a consistent estimator for the asymptotic variance of $\sqrt{n}(\hat{\rho} - \rho_0)$ and

$$\tilde{l}(\rho_0) = \frac{\hat{\omega}_S^2}{\hat{\omega}^2} \hat{l}(\rho_0) \xrightarrow{d} \chi_1^2.$$

The function $\tilde{l}(\rho) = (\hat{\omega}_S^2 / \hat{\omega}^2) \hat{l}(\rho)$ represents an adjusted empirical log-likelihood function, with a standard asymptotic behaviour. Then, an approximate confidence interval for ρ_0 , with nominal coverage $1 - \alpha$, is the set $\{\rho : \tilde{l}(\rho) \leq c_\alpha\}$, where c_α is such that $\Pr\{\chi_1^2 \leq c_\alpha\} = 1 - \alpha$. Confidence intervals obtained by $\tilde{l}(\rho)$ are indeed intervals, are range-respecting and are not subject to symmetry constraints.

3. Some simulation results

Here, some results of a simulation study are briefly reported, that highlight the behaviour of the procedure based on $\tilde{l}(\rho)$. Let $\theta = (\gamma, \sigma)$. In the simulation work, we shall assume for the strength variable Y a scaled Burr type X distribution (BurrX(γ, σ)), having cdf

$$F(y; \gamma, \sigma) = \left[1 - e^{-(y/\sigma)^2}\right]^\gamma, \quad y > 0, \quad \gamma > 0, \quad \sigma > 0.$$

We refer the interest reader to Surles and Padget (2001) for results about parametric inference on ρ for scaled Burr type X distributions.

To perform the simulations, random samples of size m for the strength values have been generated from a BurrX(γ, σ) and random samples of size n for the stress values have been generated from a BurrX($\gamma_X, 1$). The values chosen for γ , γ_X and σ are such that the true ρ_0 is equal to 0.95 in all covered cases. For each combination of parameters, we have also considered two scenarios: $m = 25, n = 50$ and $m = 50, n = 100$. Each simulation experiment is based on 5000 replications.

Table 1 reports the estimated coverage probabilities of the confidence intervals for ρ_0 . These are consistently close to the considered nominal coverages ($1 - \alpha$), i.e. 0.99, 0.95, 0.90, in particular when sample sizes are higher.

Table 1: Estimated coverage probabilities of the confidence intervals for $\rho_0 = 0.95$.

γ	γ_X	σ	$m = 25, n = 50$			$m = 50, n = 100$		
			$1 - \alpha$			$1 - \alpha$		
			0.99	0.95	0.90	0.99	0.95	0.90
1	1	4.36	0.968	0.933	0.891	0.992	0.960	0.914
10	10	1.62	0.975	0.931	0.880	0.987	0.949	0.903
10	30	1.83	0.973	0.926	0.876	0.988	0.952	0.909
30	10	1.28	0.975	0.928	0.876	0.990	0.946	0.901
50	50	1.37	0.976	0.936	0.879	0.986	0.937	0.888

References

- Kotz S., Lumelskii Y. and Pensky M. (2003) *The Stress–Strenght Model and its Generalizations*, World Scientific.
- Owen A. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237–249.
- Owen A. (1990) Empirical likelihood ratio confidence regions, *Annals of statistics*, 19, 1725–1747.
- Surles J. and Padget W. (2001) Inference for reliability and stress-strength for a scaled burr type x distribution, *Lifetime Data Analysis*, 7, 187–200.