PREDICTOR SORT SAMPLING AND ONE-SIDED CONFIDENCE BOUNDS ON QUANTILES

Steve Verrill, Victoria L. Herian, and David W. Green
USDA Forest Service Forest Products Laboratory

KEY WORDS: tolerance bounds, asymptotics

Introduction

Engineers who design wooden structures must try to ensure that the strengths of the wooden members of their structures exceed the loads to which the members will be subjected. One approach to this problem is to design so that expected loads do not exceed "allowable strength properties" associated with particular structural wood products (e.g., lumber, laminated beams, wooden I-beams). Allowable properties for structural wood products are determined experimentally by taking a sample from the population of wood products, obtaining a lower one-sided confidence bound on the fifth percentile of the distribution of the strength property in question (for example, modulus of rupture), and then dividing this bound by a safety and duration of load factor.

If a normal strength distribution is assumed, wood researchers typically (see, for example, ASTM standard D2915 in volume 4.10 of the Annual Book of ASTM Standards) obtain a parametric one-sided lower confidence bound, CB, on the fifth percentile via the formula

$$\mathrm{CB} = \bar{y}_n - k_{n,\alpha,\beta} s_n \qquad (1)$$

where we want to cover the a quantile with Confidence ß x 100% and we have n replicates. Guttman (1970, table 4.6) provides *k* values for *n* = 2(1)100, (10)300, (25)500, (50)700, (100)1000, a = .01, .05, .10, and .25, and ß = .75, .90, .95, and .99. He credits Owen (1963) for these tables.

Scientists in other areas, e.g., composite materials, groundwater monitoring, and soil remediation, also make use of equation 1 to obtain confidence bounds on quantiles. See, for example, MIL-HDBK-17-1, Gibbons (1994), and Michigan DEQ (1994).

For equation 1 to be valid, the sample of material must be a standard random sample. However, wood strength researchers commonly replace experimental unit allocation via random sampling with allocation via sorts based on non-destructive measurements of strength predictors such as modulus of elasticity and specific gravity. Warren and Madsen (1977) describe the procedure as follows:

> Specifically, then, all the boards in the experiment are ordered from weakest to strongest as nearly as can be judged from their moduli of elasticity, knot size, and slope of grain. To divide the material into *J* equivalent groups the first *J* boards, after ordering, are taken and randomly allocated one to each group. This is repeated with the second, third, fourth, etc., sets of *J* boards. The strength distributions of the resulting groups should then be essentially the same.

This allocation procedure has come to be known as predictor sort sampling.

Predictor sort allocation has long been known to statisticians. In an analysis of variance context, Cox (1957) compared seven procedures that one might use given the availability of a correlated predictor. Cox's calculations showed that the effective variance in these situations is $(1-\rho^2)\sigma_y^2$, a fact also noted by Cochran (1957). Here σ_y^2 is the variance of *y*, and p is the correlation between the predictor *x* and the response *y*.

As noted in Verrill (1993) (also see David and Gunnik (1997)), the correlations among the order statistics of the predictor induce correlations among the responses so that the standard analysis of variance (ANOVA) assumptions are not satisfied for a predictor sort experiment. Verrill demonstrated that blocked ANOVAs are still essentially valid and that simply modified unblocked ANOVAs can also be performed on predictor sort data sets.

Verrill (1999) investigated the effects of predictor sort sampling on standard confidence intervals for the mean in an ANOVA context. He found that confidence intervals for the mean are overly conserva-

tive if an unblocked analysis is performed and non-conservative if a blocked analysis is performed. He obtained asymptotic results that yielded correct confidence interval coverage.

In the current paper we examine the effect of predictor sort sampling on one-sided confidence bounds for normal quantiles. We have found that standard noncentral *t* theory that ignores the predictor sort nature of the sampling leads to $\bar{y}$ - *ks* bounds that are overly conservative. On the other hand, maximum likelihood methods yield non-conservative bounds even for fairly large sample sizes. We provide an asymptotic result that yields the appropriate corrections for the standard noncentral t approach.

In a subsequent paper we will provide methods for calculating correct *k* values for small samples.

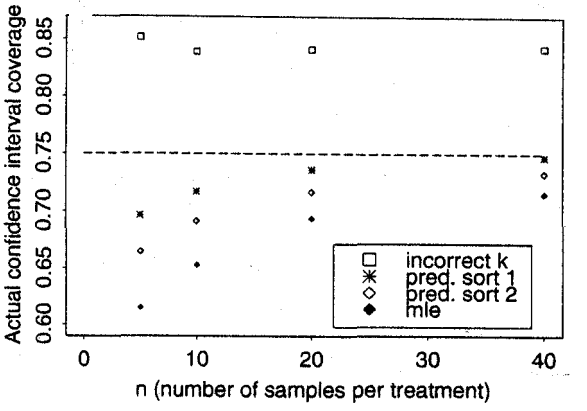

Figure 1: Actual confidence interval coverages, *J* = 4, r = .85, nominal confidence level = .75, quantile = .01

Poor Confidence Interval Coverage of the Standard Approach Given Predictor Sort Sampling

In Tables 1 - 24 of Verrill, Herian, and Green (2002a) we detail the coverages of four kinds of confidence interval for a variety of combinations of r, a, ß, number of treatments, *J*, and number of replicates, *n*. The four approaches that we consider are the incorrect standard approach, two versions of the (correct) predictor sort $\bar{y}$ - *ks* asymptotic approach, and a maximum likelihood approach. The two versions of the predictor sort approach differ in the estimate that is used for the correlation between the predictor and the response. Version 1 uses the naive estimate

$$\hat{\rho} \equiv \frac{\sum_{j=1}^{J}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{..})(y_{ij}-\bar{y}_{.j})}{\sqrt{\sum_{j=1}^{J}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{..})^2\sum_{j=1}^{J}\sum_{i=1}^{n}(y_{ij}-\bar{y}_{.j})^2}},$$

Version 2 uses the maximum likelihood estimate of r.

The tables in Verrill, Herian, and Green (2002a) clearly show that the incorrect approach is overly conservative, that the problem becomes more severe as the correlation between the predictor and the response variable increases, and that the problem does not vanish as sample sizes increase. The tables also show that version 2 of the predictor sort approach dominates the maximum likelihood approach in the sense that the actual coverage always approaches the nominal coverage more rapidly for the version 2 predictor sort approach than for the maximum likelihood approach. For smaller *J*, the version 1 predictor sort approach performs better than the maximum likelihood and version 2 approaches. (See Figure 1. The estimated coverages in each column of four points in the figure are based on separate 4,000 trial simulations.) However, for large *J* and small n, the version 1 approach does not perform as well.

For smaller n, the asymptotic approaches are non-conservative. *k* values that are appropriate for small sample sizes will appear in Verrill, Herian, and Green (2002b).

Sample Size Reductions

In the course of the development of the asymptotic theory we find (see Verrill, Herian, and Green

(2002a)) that the correct k in the appropriate version of $\bar{y}$ - ks is given by

$$k \approx -\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)\sqrt{(\Phi^{-1}(\alpha))^2/(2J) + 1 - \rho^2 + \rho^2/J}/\sqrt{n}$$

where Φ^{-1}denotes the inverse of the standard normal cumulative distribution function. Thus given higher r's we can have smaller n's and still have the same k. In fact if we set

$$\sqrt{(\Phi^{-1}(\alpha))^2/(2J) + 1 - \rho^2 + \rho^2/J}/\sqrt{n}$$

equal to a constant, we obtain

$$n \propto (\Phi^{-1}(\alpha))^2/(2J) + 1 - \rho^2 + \rho^2/J.$$

Thus the (approximate) permissible sample size reduction factor that one obtains by using a predictor sort with a correlation of r between the predictor and the response is

$$\left((\Phi^{-1}(\alpha))^2/(2J) + 1 - \rho^2 + \rho^2/J\right) \div \left((\Phi^{-1}(\alpha))^2/(2J) + 1\right).$$

Plots of this factor as function of r and J are provided in Figure 2. It is clear from the figure that practically significant sample size reductions (e.g., 40%) are attainable for reasonable correlations.

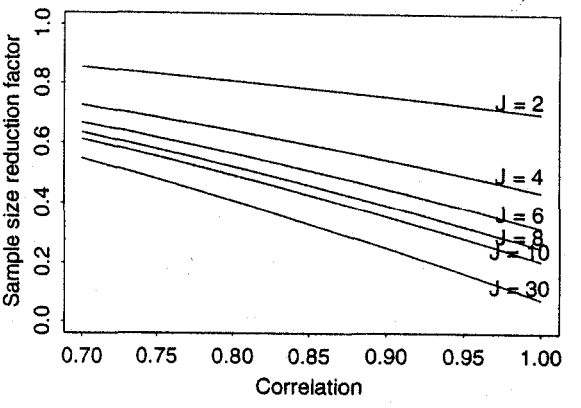

Figure 2: Sample size reduction factor as a function of r and J

Incorrect "Allowable Properties" Given Predictor Sort Sampling and a Non-Predictor Sort Analysis

As noted in the introduction, in lumber strength applications, "allowable properties" are calculated as b/f, where b is a one-sided lower confidence bound on a fifth percentile and f is some "safety and duration of load factor." If b is too low, then the allowable property will be too low. The ratio of the correct to incorrect allowable properties will be approximately equal to

$$r = (1 - k_c\mathrm{COV})/(1 - k_{\mathrm{inc}}\mathrm{COV})$$

where k_C is the correct k value, k_{inc} is the incorrect k value, and COV = s /m for the normal distribution under consideration.

In Figures 3 - 5 we plot r versus r for n = 10, COV = .15, .25, n = 20, COV = .25, and J = 2,4,6,8,10. In these figures we refer to r as the "Confidence bound increase factor.''

Since for small samples k_C must be determined by simulation, there is some irregularity in these curves. However it is clear that r increases as p increases, as COV increases, and as J increases. r decreases as n increases.

In these Figures r is sometimes as high as 1.15, which is is large enough to attract the interest of lumber manufacturers. On the other hand, correlations for solid-sawn lumber (between MOR and MOE, say) are probably not much greater than .70, so actual permissible increases in allowable properties that were calculated on the basis of predictor sort experiments are probably below 5%.

The Theorem

We can think of a predictor sort specimen allocation in the following manner. A response value Y associ-

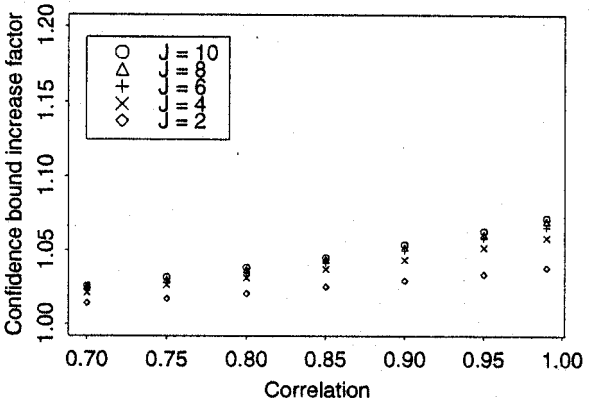

Figure 3: Confidence bound increase factor, *r*, as a function of r and *J*, *n* = 10, COV = .15

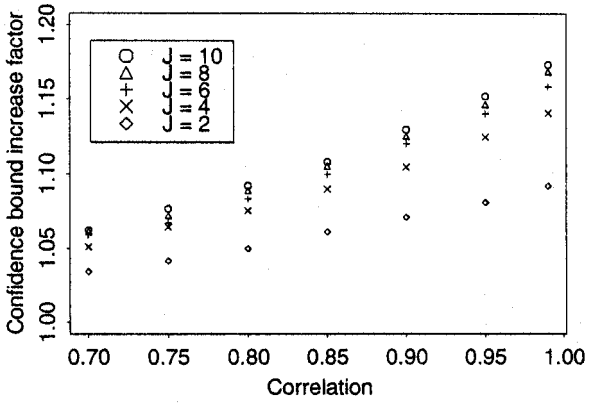

Figure 4: Confidence bound increase factor, *r*, as a function of r and *J*, *n* = 10, COV = .25

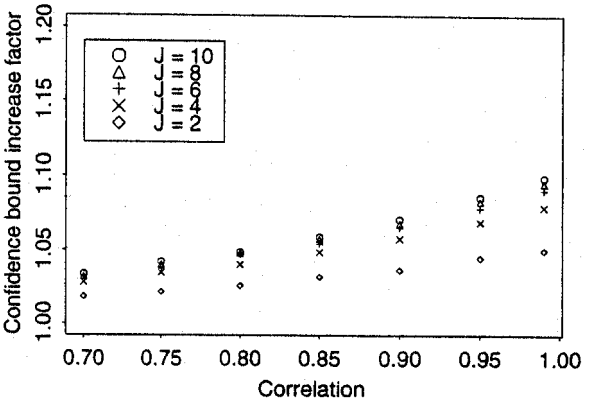

Figure 5: Confidence bound increase factor, *r*, as a function of r and *J*, *n* = 20, COV = .25

ated with a specimen is given by

$$Y = \mu_Y + \sigma_Y \left(\rho \left(X - \mu_X \right) / \sigma_X + \sqrt{1 - \rho^2} Z \right)$$

where $(X - \mu_X)/\sigma_X$ and Z are independent N(0,1)'s, and r is the correlation between X and Y. Prior to the experiment we have values for X. We rank the specimens on the basis of their associated X values and then randomly allocate the top J specimens to the first block, the next J to the second block, and so on.

Then for $1 \leq i \leq n$, $1 \leq j \leq J$ we have

$$Y_{ij} = \mu_j + \sigma_Y (\rho X_{ij} + \sqrt{1 - \rho^2} Z_{ij})$$

where the X_{ij}'s, $1 \leq j \leq J$, are a randomization of the ith group of order statistics from nJ iid N(0,1)'s, the Z_{ij}'s are iid N(0,1), and the X's and Z's are independent.

Define

$$W_{ij} \equiv \rho X_{ij} + \sqrt{1 - \rho^2} Z_{ij}, \tag{2}$$

$$\bar{Y}_{n,1} \equiv \sum_{i=1}^{n} Y_{i1}/n = \mu_1 + \sigma_Y \sum_{i=1}^{n} W_{i1}/n$$

$$= \mu_1 + \sigma_Y \left(\rho \bar{X}_{\cdot 1} + \sqrt{1 - \rho^2} \bar{Z}_{\cdot 1} \right),$$

and

$$s_{n,1}^2/\sigma_Y^2 \equiv \sum_{j=1}^{J}\sum_{i=1}^{n}(W_{ij} - \bar{W}_{\cdot j})^2/(nJ-1).$$

Theorem. Assume that the predictor variable and the variable of interest, *Y*, have a joint bivariate normal distribution with correlation r. Denote the variance of Y by σ_Y^2. Suppose that there are J treatments and n blocks. Let the allocation of samples be as described in the introduction.

Let $\bar{Y}_{n,1}$ and $s_{n,1}$ be defined as in (2), and let $\hat{\rho}$ be any consistent estimator of r. Then

$$\text{Prob}\left(\bar{Y}_{n,1} - \hat{k}_n s_{n,1} \le \mu_1 + \Phi^{-1}(\alpha)\sigma_Y\right) \to \beta$$

as $n \to \infty$, where Φ^{-1} denotes the inverse of a standard normal cumulative distribution function,

$$\hat{k}_n \equiv \sqrt{(1-\hat{\rho}^2+\hat{\rho}^2/J)/n}\ \text{NCT}^{-1}_{\gamma_n(\hat{\rho}),nJ-1}(\beta),$$

$\text{NCT}^{-1}_{\gamma_n(\hat{\rho}),nJ-1}$ denotes the inverse of a noncentral *t* distribution with noncentrality parameter $\gamma_n(\hat{\rho})$ and *nJ* - 1 degrees of freedom, and

$$\gamma_n(\hat{\rho}) \equiv -\Phi^{-1}(\alpha)\sqrt{n}/\sqrt{1-\hat{\rho}^2+\hat{\rho}^2/J}.$$

The proof is provided in Verrill, Herian, and Green (2002a).

Recommendations for Producing Predictor Sort Confidence Intervals

Recommendations for producing predictor sort coddence intervals on the mean appear in Verrill (1999). Recommendations for producing predictor sort confidence intervals on quantiles will appear in Verrill, Herian, and Green (2002b).

Summary

Predictor sort experiments attempt to make use of the correlation between a predictor that can be measured prior to the start of an experiment and the response variable that we are investigating. Properly designed and analyzed, they can reduce necessary sample sizes, increase statistical power, and reduce the lengths of confidence intervals. However, if the non-random nature of the predictor sort is not taken into account, problems can occur.

In particular, standard one-sided lower confidence bounds on quantiles of a normal distribution are overly conservative in a predictor sort situation. For lumber strength applications, this leads to "allowable properties" that are too low. We have developed asymptotic theory that yields the correct *k* value in the $\bar{y}$ - *ks* approach to obtaining a confidence bound. The resulting confidence bounds have coverages that approach the nominal values faster than bounds based on maximum likelihood estimation. In a subsequent paper we will provide *k* values that are appropriate for small sample sizes.

References

Annual Book of ASTM Standards, Volume 4.10, Committee D-7, American Society for Testing and Materials, West Conshohocken, Pennsylvania.

Cochran, W.G. (1957), "Analysis of Covariance: Its Nature and Uses," *Biometrics*, **13**, 261-281.

Cox, D.R. (1957), "The Use of a Concomitant Variable in Selecting an Experimental Design," *Biometrika*, **44**, 150-158.

David, H.A. and Gunnik, J.L. (1997), "The Paired *t* Test Under Artificial Pairing," *The American Statistician*, **51**, 9–12.

Gibbons, R.D. (1994), *Statistical Methods for Groundwater Monitoring*, John Wiley and Sons, New York.

Guttman, I. (1970), *Statistical Tolerance Regions: Classical and Bayesian*, Hafner Publishing Company, Darien, Connecticut.

Michigan DEQ (1994), *Guidance Document, Verification of Soil Remediation*, Hazardous Waste

Program Section, Waste Management Division, Michigan Department of Environmental Quality, Lansing, Michigan.

Owen, D.B. (1963), *Factors for one-sided tolerance limits and for variable sampling plans*, Sandia Corporation Monograph No. SCR-607 (19th edn).

MIL-HDBK-17-1, Composite Materials Handbook, Volume 1, Guidelines for Characterization of Structural Materials, Department of Defense Single Stock Point, Philadelphia, Pennsylvania (http://www.dodssp.daps.mil).

Verrill, S.P. (1993), "Predictor Sort Sampling, Tight T^2s, and the Analysis of Covariance," *Journal of the American Statistical Association*, **88**, 119–124.

Verrill, S.P. (1999), "When Good Confidence Intervals Go Bad: Predictor Sort Experiments and ANOVA," *The American Statistician*, **53**, 38–42.

Verrill, S.P., Herian, V.L., and Green, D.W. (2002a), "Predictor Sort Sampling and Confidence Bounds on Quantiles I: Asymptotic Theory," under review.

Verrill, S.P., Herian, V.L., and Green, D.W. (2002b), "Predictor Sort Sampling and Confidence Bounds on Quantiles II: k's for Small Samples," in preparation.

Warren, W.G., and Madsen, B. (1977), "Computer-Assisted Experimental Design in Forest Products Research: A Case Study Based on Testing the Duration-of-Load Effect," *Forest Products Journal*, **27**, 45-50.