Confidence Regions for Parameters of Linear Models

Andrew L. Rukhin

rukhin@cam.nist.gov, rukhin@math.umbc.edu

Statistical Engineering Division
National Institute of Standards and Technology
Gaithersburg, MD 20899
Department of Mathematics & Statistics
University of Maryland, Baltimore County (UMBC)
Baltimore, MD 21250

STATISTICS: THEORY and PRACTICE
May 30 , 2008

Heterogeneous Linear Models

General linear model, observe $Y \sim N_p(X\beta, \Sigma)$
Σ diagonal matrix, r-dimensional parameter β, $r \leq p$,
X $p \times r$ matrix of rank r,

$$Y = X\beta + \epsilon.$$

The vector ϵ formed by independent errors
$\epsilon_j, j = 1, \ldots, p$, with zero mean and unknown variances
σ_j^2.
Meta-analysis: p independent but heterogeneous studies, each study produces an unbiased estimate of its linear function of β. The accuracy of this estimate may not be given.

Interlaboratory comparisons

The classical least squares estimate $\hat{\beta}$,

$$X^T \Sigma^{-1} X \hat{\beta} = X^T \Sigma^{-1} Y,$$

not attainable, depends on unknown Σ.
The common mean case, $r = 1$, X p-dimensional vector of ones.
A typical interlaboratory comparisons study involves several laboratories, each of which analyzes its measurements, reports the final result consisting of the estimate of the reference value and the combined uncertainty.
This uncertainty is composed of Type A and Type B components.

Type B uncertainties

Statistical derivations (Type A); scientific judgment (Type B).

ISO GUM (1993) and NIST Guidelines for Evaluating and Expressing Uncertainty (1994).

Type B uncertainties dominate, make the accuracy of the estimators uncertain.

Implication: Σ is not known.

Gold Vapor Pressure

Paule and Mandel (1970): several laboratories performed measurements via different techniques of gold vapor pressure P as a function of the absolute temperature T.

$\log P$ is a linear function of $1/T$. The data on P was collected from 38 runs of ten laboratories, total of 375 different temperature points.

A natural assumption: the error variance depends only on the run within each individual laboratory (not on the temperature value). Fits our model with $p = 375$, only 38 different values of σ_j^2 corresponding to each run. The matrix X: the reciprocals of T employed by each laboratory.

Relevant Work

Carrol and Rupert (1982) investigate different parametric models for heteroscedasticity.

Fuller and Rao (1978) a two stage estimation procedure assuming that observations compose several groups with constant variance within each group. The residuals derived from ordinary least squares are employed to estimate the covariance matrix, subsequent generalized least squares.

Wu (1986) a class of jackknife variance estimators, compared them to the bootstrap method.

The asymptotic behavior of these methods in Shao and Tu (1995) Ch 7.

Linear Estimators

Linear unbiased estimators $\delta = WY$, with $r \times p$ matrix W, $WX = I$, $Var(\delta) = W\Sigma W^T$. If $W = (X^TQX)^{-1}X^TQ$,

$$\delta = (X^TQX)^{-1}X^TQY,$$

$p \times p$ diagonal matrix Q, Q "approximates" Σ^{-1}. If $Q = \Sigma^{-1}$, then $Var(\delta) = (X^TQX)^{-1}$, the commonly used estimator $(X^TQX)^{-1}$ typically underestimates $Var(\delta)$.

To adjust for this bias and to derive a conservative confidence ellipsoid, suggest

$$\widehat{Var(\delta)} = [Y^T(I - XW)^TS(I - XW)Y](X^TQX)^{-1}.$$

The non-negative definite $p \times p$ matrix S defines the quadratic form in the residuals $Y - XWY$.

Confidence Ellipsoid

When $r = 1$,
$S = (X^T Q X)\text{diag}\Big((WW^T)\text{diag}(I - XW)^{-1}\Big)$, this estimator coincides with Horn, Horn and Duncan (1975) method.

A confidence ellipsoid for β based on δ,

$$(\delta - \beta)^T \widehat{Var(\delta)}^{-1} (\delta - \beta) \le t^2.$$

Want to control

$$\sup_{\Sigma} P_{\Sigma} \left((\delta - \beta)^T \widehat{Var(\delta)}^{-1} (\delta - \beta) > t^2 \right).$$

This supremum is 1 for $0 < t \le t(X, W, S)$.

Coverage Probability and Volume

Our goal is to determine the (scale-invariant) coverage probability of this ellipsoid for any Σ.

Have to look at large t.

Issues:

- What t guarantees a (large) confidence coefficient?
- What Σ is the least favorable (minimizes the coverage probability)?
- How to choose S to minimize the volume?

Moments of Quadratic Forms

$Z_1, \ldots, Z_p$ be independent standard normal variables, $\lambda_{r+1}, \ldots, \lambda_p$ fixed positive numbers, $\lambda_1, \ldots, \lambda_r \to 0$,

$$\lim \frac{P\left(\sum_{i=1}^{r} \lambda_i Z_i^2 \geq \sum_{k=r+1}^{p} \lambda_k Z_k^2\right)}{H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)}$$

$$= \frac{\Gamma(p/2)}{\sqrt{\lambda_{r+1} \cdots \lambda_p}\Gamma(r/2)\Gamma((p-r+2)/2)},$$

$$H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r) = \int_{S_r} \left[\sum_i \lambda_i \omega_i^2\right]^{(p-r)/2} d\omega.$$

Tails of the Distribution of the Ratio

When $\lambda_i = t^{-2}, i = 1, \ldots, r$, $\lambda_k = 1, k = r+1, \ldots, p$, the known result for the tail probabilities of $F_{r,p-r}$, $t \to \infty$,

$$P(rF_{r,p-r} > (p-r)t^2) \sim \frac{\Gamma(p/2)}{\Gamma(r/2)\Gamma((p-r+2)/2)t^{p-r}}.$$

$$P\left(\sum_{i=1}^{r} \lambda_i Z_i^2 \geq t^2 \sum_{k=r+1}^{p} \lambda_k Z_k^2\right)$$

$$\sim P\left(\frac{[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)} \sum_{i=1}^{r} Z_i^2}{(\lambda_{r+1} \cdots \lambda_p)^{1/(p-r)} \sum_{k=r+1}^{p} Z_k^2} \geq t^2\right)$$

$$= P\left(\frac{r[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)}}{(p-r)(\lambda_{r+1} \cdots \lambda_p)^{1/(p-r)}} F_{r,p-r} \geq t^2\right).$$

Dirichlet Averages

When approximating the tail probabilities of the ratio of two quadratic forms, the λ_i's in the numerator are to be replaced by their *spherical* average $[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)}$, the λ_k's in the denominator by their geometric mean. (Okamoto inequality implies conservative approximation.)

The function H_q a particular case of Dirichlet averages

$$H_q(\lambda_1, \ldots, \lambda_r) = \int \left[\sum_1^r \lambda_k u_k \right]^q d\mu_b(u),$$

integration over a unit simplex in R^r, μ_b Dirichlet distribution, parameter $b = (1/2, \ldots, 1/2)$. Carlson (1977): a number of useful formulas and transformations for Dirichlet averages.

Main Result

Let $\delta = WY$ be a linear unbiased estimator of β,

$$\mu_i = \lambda_i((X^TQX)^{1/2}(X^T\Sigma^{-1}X)^{-1}(X^TQX)^{1/2}), \quad i = 1, \ldots, r,$$

$$\lim_{t\to\infty} t^{p-r} P_\Sigma \left((\delta - \beta)^T \widehat{Var(\delta)}^{-1} (\delta - \beta) \geq t^2 \right)$$

$$= \frac{H_{(p-r)/2}(\mu_1, \ldots, \mu_r)\Gamma(p/2)}{\det(X^T\Sigma^{-1}XWS^{-1}W^T)^{1/2} \det(\Sigma S)^{1/2}\Gamma(r/2)\Gamma((p-r+2)/2)}$$

$$= \lim_{t\to\infty} t^{p-r} P \left(\frac{rF_{r,p-r}}{p-r} > t^2\kappa \right),$$

$$\kappa = \left[\frac{\det(X^T\Sigma^{-1}XWS^{-1}W^T)\det(\Sigma S)}{H^2_{(p-r)/2}(\mu_1, \ldots, \mu_r)} \right]^{1/(p-r)}$$

Conservative Choice

Our confidence ellipsoid has the approximate confidence coefficient $(1-\alpha)$ for fixed Σ, when

$$t^2 = \frac{rF_{r,p-r}(\alpha)}{p-r}\left[\frac{H^2_{(p-r)/2}(\mu_1,\ldots,\mu_r)}{\det(X^T\Sigma^{-1}XWS^{-1}W^T)\det(\Sigma S)}\right]^{1/(p-r)}.$$

As Σ is unknown, a conservative procedure corresponds to

$$t_0^2 = \frac{rF_{r,p-r}(\alpha)}{(p-r)[\det(WS^{-1}W^T)\det(S)]^{1/(p-r)}}$$

$$\times \sup_{\Sigma}\left[\frac{H^2_{(p-r)/2}(\mu_1,\ldots,\mu_r)}{\det(X^T\Sigma^{-1}X)\det(\Sigma)}\right]^{1/(p-r)}.$$

Example: $r = 1$

For large t the infimum of the coverage probability is attained at F- distribution.
When $r = 1$,

$$H_{(p-r)/2}(\mu) = \mu^{(p-1)/2} = \left(\frac{X^T Q X}{X^T \Sigma^{-1} X}\right)^{(p-1)/2} .$$

For a given value of $\det(\Sigma) = \prod_j \sigma_j^2$, the minimum of $X^T \Sigma^{-1} X = \sum_j X_{j1}^2 \sigma_j^{-2}$ is attained when $\sigma_j^2 \propto X_{j1}^2$,

$$\sup_{\Sigma} \frac{X^T Q X}{(X^T \Sigma^{-1} X)^{p/(p-1)} \det(\Sigma)^{1/(p-1)}} = \frac{X^T Q X}{p^{p/(p-1)} (\prod_1^p X_{j1}^2)^{1/(p-1)}}.$$

$$t_0^2 = \frac{F_{1,p-1}(\alpha)(X^T Q X)}{(p-1) p^{p/(p-1)} W S^{-1} W^T \det(S)]^{1/(p-1)} (\prod_1^p X_{j1}^2)^{1/(p-1)}}.$$

Conservative t-intervals

With T_{p-1} denoting a t-distributed random variable with $p-1$ degrees of freedom,

$$\sup_{\Sigma} \lim_{t\to\infty} t^{p-1} P_{\Sigma}\left((\delta-\beta)^T \widehat{Var(\delta)}^{-1}(\delta-\beta) \geq t^2\right)$$

$$= \frac{(X^T Q X)^{p-1}\Gamma(p/2)}{p^p (W S^{-1} W^T)\det(S)\Gamma(1/2)\Gamma((p+1)/2)\prod_1^p X_{j1}^2}$$

$$= \lim t^{p-1} P\left(T_{p-1}^2 > t^2(p-1)\kappa\right),$$

$$\kappa = \frac{\left[p^p W S^{-1} W^T \det(S)\prod_1^p X_{i1}^2\right]^{1/(p-1)}}{X^T Q X}.$$

Rukhin (2007) Conservative confidence intervals ... *Statist&Probab. Lett.* 77, 1312-21.

Maximization Problem

$A = Q^{1/2}X(X^TQX)^{-1/2}$, so that $A^TA = I$.

Define

$$G(A) =$$

$$\sup_{\Sigma} \left[\frac{H^2_{(p-r)/2}\left(\lambda_1((A^T\Sigma^{-1}A)^{-1}), \ldots, \lambda_r((A^T\Sigma^{-1}A)^{-1})\right)}{\det(A^T\Sigma^{-1}A)\det(\Sigma)} \right]^{\frac{1}{p-r}}$$

$$= \sup_{\Sigma} \left[\frac{H^2_{-p/2}\left(\lambda_1(A^T\Sigma^{-1}A), \ldots, \lambda_r(A^T\Sigma^{-1}A)\right)}{\det(\Sigma)} \right]^{\frac{1}{p-r}} \geq 1.$$

Conservative choice

$$t_0^2 = \frac{rF_{r,p-r}(\alpha)G(A)}{p-r} \left[\frac{\det(Q)}{\det(X^TQXWS^{-1}W^T)\det(S)} \right]^{1/(p-r)}.$$

Solution Space

Any possible maximizer

$$\Sigma = \text{diag}(AFA^T),$$

F is a matrix polynomial in $A^T\Sigma^{-1}A$ of degree $r-1$ (scalar coefficients depend only on the eigenvalues of $A^T\Sigma^{-1}A$.)

diag(Σ) is in the subspace spanned by the vectors with coordinates $a_{\cdot i}a_{\cdot k}, 1 \leq i \leq k \leq r$

diag(Σ) in the column space of the Schur square $A \odot A$.

The solution space has dimension at most $r(r+1)/2$, its orthogonal complement, $\{x : A^T \text{ (diag } x) \, A = 0\}$.

Optimization Sub-Problems

Our maximization problem can be split in two parts, the first consisting of the minimization of

$$\prod_{\jmath} e_{\jmath}^T AODO^T A^T e_{\jmath}$$

over all orthogonal matrices O for a fixed diagonal matrix $D =$diag $(\lambda_1, \ldots, \lambda_r)$, $\lambda_i = \lambda_i(A^T\Sigma^{-1}A)$.
The second sub-problem is maximization in λ_i of

$$\frac{H^2_{-p/2}(\lambda_1, \ldots, \lambda_r)}{\min_O \prod_{\jmath} e_{\jmath}^T AODO^T A^T e_{\jmath}}.$$

Example: $r = 2$

When $r = 2$, to evaluate $G(A)$ one needs

$$H_{(p-2)/2}(\lambda_1, \lambda_2) = (\lambda_1 \lambda_2)^{(p-1)/2} H_{-p/2}(\lambda_1, \lambda_2).$$

Practical to calculate the ratio,

$$R_m(\lambda_1, \lambda_2) = \frac{H_{m+1}(\lambda_1, \lambda_2)}{H_m(\lambda_1, \lambda_2)},$$

$$R_{m+1}(\lambda_1, \lambda_2) = \frac{(2m+1)(\lambda_1 + \lambda_2)}{2(m+1)} - \frac{m \lambda_1 \lambda_2}{(m+1) R_m(\lambda_1, \lambda_2)},$$

$R_1(\lambda_1, \lambda_2) = (\lambda_1 + \lambda_2)/2$, (when $p/2$ is a semi-integer, use the initial value, $R_{1/2}(\lambda_1, \lambda_2) = H_{1/2}(\lambda_1, \lambda_2)/H_{-1/2}(\lambda_1, \lambda_2)$.
$A = (a_{ji})\,, i = 1, 2; j = 1, \ldots, p$

Example Continued

Minimize,

$$\prod_j e_j^T AODO^T A^T e_j$$

$$= 2^p \prod_j (a_{j1}^2 + a_{j2}^2)\left(h_1 \cos^2(\phi_j - \phi) + h_2 \sin^2(\phi_j - \phi)\right)$$

$$= (h_1 + h_2)^p \prod_j (a_{j1}^2 + a_{j2}^2) \prod_j (1 + \rho \cos 2(\phi_j - \phi)),$$

with $\tan \phi_j = a_{j2}/a_{j1}$, $\rho = \frac{h_1 - h_2}{h_1 + h_2}$, over 2×2 orthogonal matrices

$$O = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

Interpretation

Statistical interpretation is maximum likelihood estimation of the rotation parameter ψ,

$$\arg\min_{\psi} \sum_j \log(1 + \rho \cos(\psi_j - \psi)),$$

densities

$$(2\pi)^{-1}\sqrt{1-\rho^2}[1 - \rho\cos(\cdot - \psi)]^{-1}.$$

These densities are popular models for wind directions, the distribution of the polar angle in a bivariate normal vector.

Solution

$\lambda_1 = x \le 1 = \lambda_2,$

$$G(A) = \frac{2^{p/(p-2)}}{p^{p/(p-2)} \prod_j (a_{j1}^2 + a_{j2}^2)^{1/(p-2)}}$$

$$\times \max_{x:0 \le x \le 1} \frac{x^{1/(p-2)} [H_{(p-2)/2}(x,1)]^{2/(p-2)}}{[R_{p/2-1}(x,1)]^{p/(p-2)} [\min_\phi \prod_j (1 + \rho \cos 2(\phi_j - \phi))]^{1/(p-2)}}.$$

When $x \to 0, \rho \to 1,$

$$\frac{x}{1-\rho} \to \frac{p-1}{2},$$

$$\frac{\min_\phi \prod_j (1 + \rho \cos 2(\phi_j - \phi))}{1-\rho} \to 2^{p-1} \min_k \prod_{j:j \ne k} \sin^2(\phi_j - \phi_k),$$

Gold Vapor Pressure Again

$$G(A) = \frac{[\Gamma((p-1)/2)]^{2/(p-2)}}{(p-1)^{(p-1)/(p-2)}\pi^{1/(p-2)}[\Gamma(p/2)]^{2/(p-2)}}$$

$$\times \frac{1}{[\prod_j (a_{j1}^2 + a_{j2}^2)]^{1/(p-2)} \min_k [\prod_{j:j\neq k} \sin^2(\phi_j - \phi_k)]^{1/(p-2)}}$$

$p \geq 5$.

In the interlaboratory studies of gold vapor pressure $\phi_{opt} = 2.19, \lambda_2 = 1, \lambda_1 = 0$, $G(A) = 9.23$, $t_0^2 = 0.1495$.

MATLAB does not work in this problem!

Confidence Ellipsoid for Gold Vapor Pressure Study

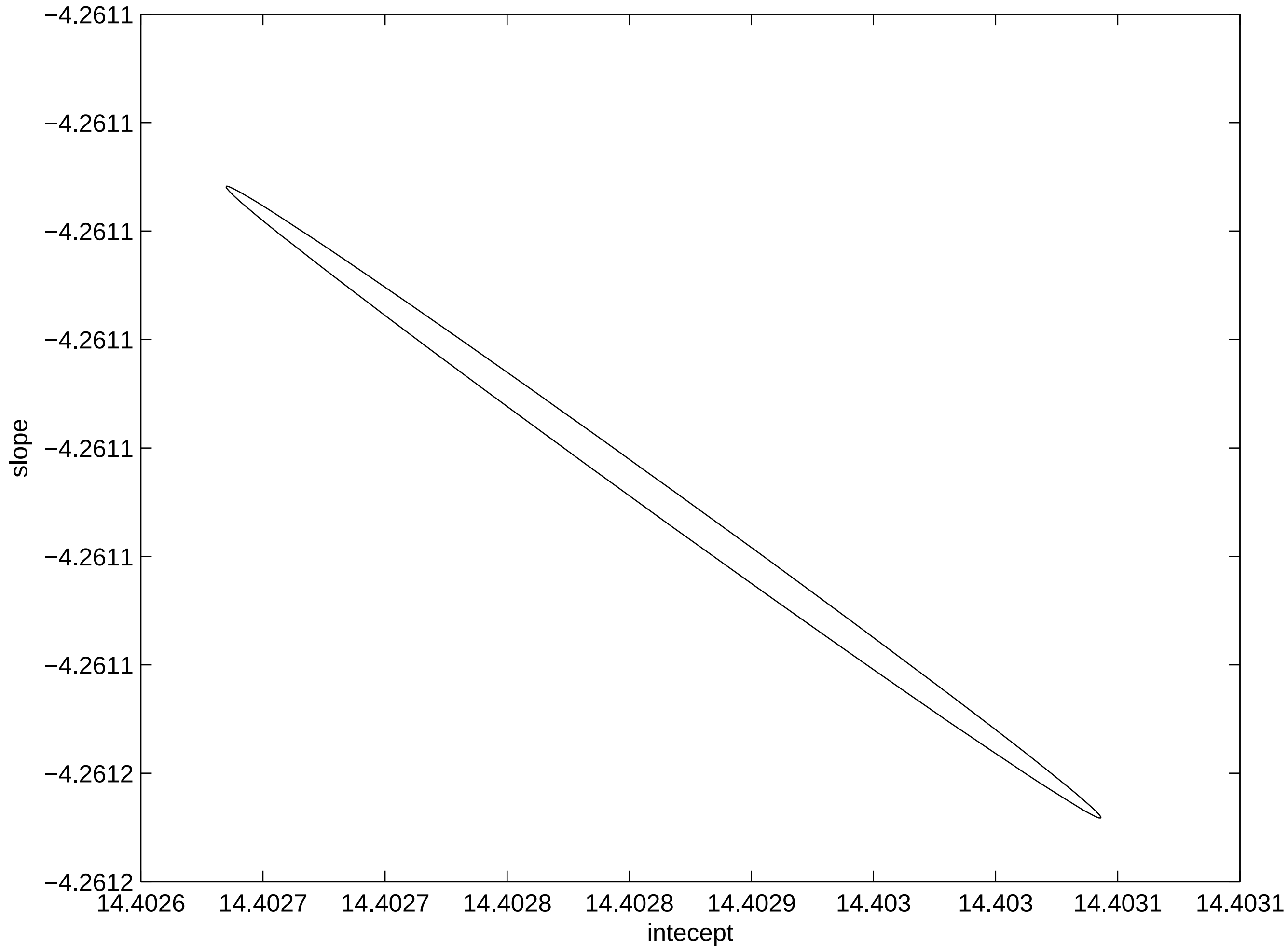

Figure 1: Confidence ellipsoid for β in the gold vapor pressure study

Volume of the Confidence Set

Under error covariance matrix Σ_0, the expected volume

$$\Delta = \frac{t^r \pi^{r/2} E(Z^T \Sigma_0^{1/2} (I - XW)^T S (I - XW) \Sigma_0^{1/2} Z)^{r/2}}{\Gamma((r+2)/2) \sqrt{\det(X^T Q X)}}.$$

The rank of $\Sigma_0^{1/2}(I - XW)^T S (I - XW) \Sigma_0^{1/2}$ is $p - r$,

$$\Delta =$$

$$\left[\frac{r F_{r,p-r}(\alpha) G(A)}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2} \Gamma((p+r)/2)}{\Gamma(p/2)\Gamma((r+2)/2)} \det(X^T Q X)^{\frac{2r-p}{2(p-r)}}$$

$$\times \frac{H_{r/2}(\eta_1, \ldots, \eta_p) \det(Q)^{r/[2(p-r)]}}{[\det(S) \det(W S^{-1} W^T) \det(X^T Q X)^2]^{r/[2(p-r)]}}.$$

Optimal S

For a fixed Q, can determine a matrix S, which minimizes the expected volume, $Q^{-1/2}SQ^{-1/2}$ is the generalized inverse of $Q^{1/2}(I - XW)\Sigma_0(I - XW)^T Q^{1/2}$.

$$\min_S \Delta = \left[\frac{rF_{r,p-r}(\alpha)G(A)}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2}\Gamma(p/2)}{\Gamma((p-r)/2)\Gamma((r+2)/2)}$$

$$\times \frac{\prod_{k=r+1}^{p}[\lambda_k(\Pi\Sigma_0 Q\Pi)]^{r/[2(p-r)]}}{\sqrt{\det(X^TQX)}}.$$

In practice, the matrix Σ_0 is unknown, but Q^{-1} a suitable surrogate.

Final Formulas

If $\Sigma_0 = Q^{-1}$, then $S = Q(I - XW)$,

$$\min_S \Delta =$$

$$\left[\frac{rF_{r,p-r}(\alpha)G(A)}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2}\Gamma(p/2)}{\sqrt{\det(X^TQX)}\Gamma((p-r)/2)\Gamma((r+2)/2)}.$$

For this S with $\Sigma_0 = Q^{-1}$,

$$t_0^2 = \frac{rF_{r,p-r}(\alpha)}{p-r} G(A),$$

Interpret $G(A)$ as the adjustment factor $(G(A) \geq 1)$ to the percentile of $F_{r,p-r}$ distribution needed to obtain a conservative $(1-\alpha)$ confidence region.