Simultaneous Confidence Intervals for Linear Estimates of Linear Functionals

P.B. Stark
Department of Statistics
University of California
Berkeley CA 94720-3860
stark@stat.berkeley.edu

Technical Report 417
August, 1994; revised March, 1995

Abstract

This note presents three ways of constructing simultaneous confidence intervals for linear estimates of linear functionals in inverse problems, including "Backus-Gilbert" estimates. Simultaneous confidence intervals are needed to compare estimates, for example, to find spatial variations in a distributed parameter. The notion of simultaneous confidence intervals is introduced using coin tossing as an example before moving to linear inverse problems. The first method for constructing simultaneous confidence intervals is based on the Bonferroni inequality, and applies generally to confidence intervals for any set of parameters, from dependent or independent observations. The second method for constructing simultaneous confidence intervals in inverse problems is based on a "global" measure of fit to the data, which allows one to compute simultaneous confidence intervals for any number of linear functionals of the model that are linear combinations of the data mappings. This leads to confidence intervals whose widths depend on percentage points of the chi-square distribution with n degrees of freedom, where n is the number of data. The third method uses the joint normality of the estimates to find shorter confidence

intervals than the other methods give, at the cost of evaluating some integrals numerically.

1 Introduction

Backus-Gilbert theory [*Backus and Gilbert,* 1968, 1970; *Backus,* 1970abc] estimates certain weighted spatial averages of some physical property from indirect linear measurements. Often, averages sensitive to different regions in the object of study are compared to try to identify differences in properties between the regions. For example, in both geophysics (*e.g., Backus and Gilbert* [1970], *Parker* [1970], *Johnson and Gilbert* [1972], *Masters* [1979], *Oldenburg* [1979,1981]) and astrophysics (*e.g., Christensen-Dalsgaard et al.* [1990], *Däppen et al.* [1991], *Gough and Toomre* [1991], *Schou* [1991]) estimates of averages centered at different depths are frequently plotted in the same figure, inviting the reader to compare the estimated values at different depths. (A common plotting method is to use "plusses" centered at the depths and values of the estimates, with the lengths of the vertical bars corresponding to the nominal resolutions, and the lengths of the horizontal bars corresponding to the nominal uncertainties.) It does not appear to be appreciated generally that the uncertainties must be magnified somewhat for this comparison to be valid. In this note, I illustrate the issue of simultaneous versus individual confidence intervals, and give three simple methods to compute simultaneous confidence intervals for linear estimates of linear functionals, including "Backus-Gilbert" estimates, that can be used to compare estimates of different averages, such as averages centered at different points.

2 Individual versus Simultaneous Coverage Probability

Consider tossing a (loaded) coin independently 100 times to find a confidence interval for the probability p with which the coin lands heads. The distribution of the number of heads in 100 tosses is, under these assumptions, binomial with 100 trials and probability p of success in each trial. Provided p is not too close to zero or one, this Binomial distribution is approximated well by a normal distribution with mean $100p$ and standard deviation

$10\sqrt{p(1-p)}$. With high probability, $10\sqrt{\hat{p}(1-\hat{p})}$ will be close to the standard deviation, where the $\hat{p}$ is the sample proportion: $\hat{p} = (\#\text{ heads observed})/100$. Again, provided p is not too close to zero or one, the distribution of the sample proportion is approximately normal with mean p and standard deviation $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})}/10$. Thus $P\{|p - \hat{p}| \leq 1.96\hat{\sigma}\} \approx 0.95$, and an approximate 95% confidence interval for p is given by

$$\mathcal{I} = [\hat{p} - 1.96\hat{\sigma}, \hat{p} + 1.96\hat{\sigma}]. \tag{1}$$

The *coverage probability* of the confidence interval $\mathcal{I}$ is the chance that $\mathcal{I}$ contains p. Note that this probability makes sense only before the data are observed: the coverage probability refers to the properties of the procedure applied to random data, not to the ultimate numerical values one computes after observing the data. After observing the data and constructing a specific confidence interval, the interval either does or does not contain p, and there is no more randomness in the problem. At that point, what we have is a "confidence level," which is equal to the coverage probability of the random interval before observing the data.

Suppose now that we have two biased coins, and we wish to know whether there is a difference between p_1 and p_2, the probabilities of heads for coin 1 and coin 2. One way to do this is to compare $\hat{p}_1$ and $\hat{p}_2$, the sample proportions of heads in 100 independent tosses of each coin separately. Intuition and common practice suggest that we can use $\hat{p}_1$ and $\hat{p}_2$ to construct 95% confidence intervals for p_1 and p_2, then conclude that $p_1 \neq p_2$ if the confidence intervals do not overlap. In fact, testing the hypothesis that $p_1 = p_2$ in this way has a higher significance level (lower confidence level, loosely speaking) than is commonly recognized: If $\hat{p}_1$ and $\hat{p}_2$ are independent, then the chance that *both* $\hat{p}_1 - 1.96\hat{\sigma}_1 \leq p_1 \leq \hat{p}_1 + 1.96\hat{\sigma}_1$ *and* $\hat{p}_2 - 1.96\hat{\sigma}_2 \leq p_2 \leq \hat{p}_2 + 1.96\hat{\sigma}_2$ is only about $0.95 \times 0.95 = 0.9025$. Thus if the two 95% confidence intervals do not overlap, we have only about 90% confidence that $p_1 \neq p_2$. (More precisely, we could reject the hypothesis that $p_1 = p_2$ at significance level approximately 0.1, not 0.05.) The essential point is that even though each interval $\mathcal{I}_j$, $j = 1, 2$ contains its corresponding parameter p_j with probability $1 - \alpha$, part of the time that $\mathcal{I}_1$ contains p_1, $\mathcal{I}_2$ will not contain p_2, and *vice versa*. As a result, the intervals both contain their parameters less than $1 - \alpha$ of the time. Figure 1 illustrates this point. If we try to compare the probabilities of heads for n coins in this way, we end up with a *simultaneous* confidence level of about $0.95^n \times 100\%$. Conversely, if we want to end up with simultaneous 95% confidence, we need to begin with $0.95^{1/n} \times 100\%$ confidence intervals for each p_j. The situation is more complicated when the sample proportions $\hat{p}_j$ are not independent, which is more directly analogous to comparing several Backus-Gilbert estimates, since estimates "centered" at different points typically involve correlated linear combinations of the same data.

3 Simultaneous Backus-Gilbert Estimates

The canonical linear inverse problem is to make inferences about the function $x(r)$ from n observations $\{\delta_j\}_{j=1}^n$ that are linearly related to $x(r)$ but contaminated by additive Gaussian noise:

$$\delta_j = < K_j, x > + \epsilon_j, \quad j = 1, \ldots, n. \tag{2}$$

Here $< K_j, x >$ is a linear functional of the function x and may be thought of as an integral $< K_j, x > = \int K_j(r) x(r)$ over the appropriate domain, and the noise terms ϵ_j are assumed to be realizations of independent, zero mean Gaussian random variables with variances σ_j^2. We abbreviate these n equations using vector notation:

$$\boldsymbol{\delta} = < \mathbf{K}, x > + \boldsymbol{\epsilon}. \tag{3}$$

The random vector $\boldsymbol{\epsilon}$ has zero expectation and covariance matrix

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2).$$

The inverse of the covariance matrix is $\Sigma^{-1} = \mathrm{diag}(\sigma_1^{-2}, \ldots, \sigma_n^{-2})$, and the square-root of the inverse is $\Sigma^{-1/2} = (\sigma_1^{-1}, \ldots, \sigma_n^{-1})$. We shall assume henceforth that the functionals $\{K_j\}$ are linearly independent; *i.e.*, if

$$< \sum_{j=1}^{n} \lambda_j K_j, y > = 0$$

for every function y, the constants $\{\lambda_j\}$ must all be zero.

Suppose we wish to estimate the linear functional $< L, x >$ from the data $\boldsymbol{\delta}$. A fundamental result of linear inverse theory is that this is impossible without additional information unless one can write L as a linear combination of the data mappings $\{K_j\}$; *i.e.,* one can estimate $< L, x >$ with finite uncertainty only if

$$< L, y > \; = \; < \sum_{j=1}^{n} \lambda_j K_j, y > \tag{4}$$

for all y [*Backus and Gilbert*, 1968]. The corresponding linear combination of the data, $\sum_{j=1}^{n} \lambda_j \delta_j$, is an unbiased estimate of $< L, x >$. Much of Backus-Gilbert theory consists of considerations and methods for selecting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ so that the functional L is a "localized" average and can be estimated with reasonable uncertainty. Suppose we wish to estimate m such functionals $\{L_k\}_{k=1}^{m}$. Let $L_k = \sum_{j=1}^{n} \Lambda_{kj} K_j$ in the sense that

$$< L_k, y > \; = \; < \sum_{j=1}^{n} \Lambda_{kj} K_j, y > \tag{5}$$

for all y. Nothing in this note depends on how Λ_{kj} are selected, so Backus-Gilbert estimates are included as special cases. Let $\boldsymbol{\Lambda}_k$ be the n-vector $(\Lambda_{k1}, \ldots, \Lambda_{kn})$ and define

$$\boldsymbol{\Lambda}_k \cdot \mathbf{K} \equiv \sum_{j=1}^{n} \Lambda_{kj} K_j \tag{6}$$

and

$$\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} \equiv \sum_{j=1}^{n} \Lambda_{kj} \delta_j, \tag{7}$$

with $\boldsymbol{\Lambda}_k \cdot \boldsymbol{\epsilon}$ defined similarly. For any vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$, define its ordinary two-norm

$$\|\boldsymbol{\gamma}\| \equiv \sqrt{\boldsymbol{\gamma} \cdot \boldsymbol{\gamma}}, \tag{8}$$

and (for a positive-definite matrix Σ) its weighted two-norm

$$\|\boldsymbol{\gamma}\|_\Sigma \equiv \sqrt{\boldsymbol{\gamma} \cdot \Sigma \cdot \boldsymbol{\gamma}}. \tag{9}$$

3.1 Bonferroni simultaneous confidence intervals

Bonferroni's inequality (see, *e.g., Bickel and Doksum* [1977]) says that the probability that at least one of many events $\{A_k\}_{k=1}^{m}$ occurs is no larger than the sum of the chances that each occurs; *i.e.*,

$$P\{ \ A_1 \text{ occurs or } A_2 \text{ occurs or } \ldots \text{ or } A_m \text{ occurs } \} \leq \sum_{k=1}^{m} P\{ \ A_k \text{ occurs } \}. \tag{10}$$

Suppose we have a procedure for producing a set of individual (not necessarily simultaneous) $1 - \alpha'$ confidence intervals for the parameters $< L_k, x >$, $k = 1, \ldots, m$. The chance that at least one of these intervals fails to contain its corresponding parameter is the chance that at least one of the events $\{A_j\}_{j=1}^n$ occurs, where A_j is the event that the jth confidence interval fails to contain its parameter. Each of these events has (by assumption) probability α'. Bonferroni's inequality implies that the chance that one or more of the m intervals fails to contain its corresponding parameter is no larger than $m\alpha'$. Thus the chance that all the confidence intervals simultaneously include their corresponding parameters is at least $1 - m\alpha'$. It follows that if $\alpha' = \alpha/m$, so that the original intervals have coverage probability $1 - \alpha/m$, the set of m confidence intervals has simultaneous coverage probability at least $1 - \alpha$, as desired.

This relatively simple approach to simultaneity can be quite useful if the number of simultaneous confidence intervals one desires is not too large, and the length of a confidence interval required to get $1 - \alpha/m$ coverage probability is not too long. It is particularly simple to use since we need no assumption about the dependence of the estimates.

As an example, suppose we desire to have a set of simultaneous 95% confidence intervals for the expected values of 10 Gaussian random variables with common variance σ^2, based on one observation X_j of each variable, $j = 1, \ldots, m$. Bonferroni's inequality implies that if we start with confidence intervals with individual confidence levels $1 - 0.05/10 = 0.995$, the intervals will have simultaneous coverage probability at least 95%. Referring to a standard normal distribution table shows that the intervals $\{[X_j - 2.81\sigma, X_j + 2.81\sigma]\}_{j=1}^{10}$ have simultaneous 95% coverage probability. A 95% confidence interval for a single parameter would have length $2 \times 1.96\sigma$, so the increase in length required for simultaneity is $(2 \times 2.81\sigma)/(2 \times 1.96\sigma) \approx 1.43$. This ratio increases with the number of intervals, and for Gaussian estimates, decreases with increasing confidence level: for example, for 99% simultaneous confidence in this same problem, the ratio would be about $3.27/2.56 \approx 1.28$.

3.2 Simultaneous confidence intervals based on the chi-square distribution with n degrees of freedom

Since the errors $\{\epsilon_j\}$ are independent zero-mean Gaussian random variables with variances $\{\sigma_j^2\}$, the distribution of $\|\boldsymbol{\epsilon}\|^2_{\Sigma^{-1}}$ is chi-square with n degrees of freedom. If $\chi = \sqrt{\chi^2_{n,1-\alpha}}$ is the square-root of the $(1-\alpha)\times 100\%$ percentage point of the chi-square distribution with n degrees of freedom, by definition

$$P\{\|\boldsymbol{\epsilon}\|_{\Sigma^{-1}} \leq \chi\} = 1 - \alpha. \tag{11}$$

Thus

$$P\{\|\boldsymbol{\delta} - < \mathbf{K}, x > \|_{\Sigma^{-1}} \leq \chi\} = 1 - \alpha. \tag{12}$$

For any $\boldsymbol{\Lambda}_k$, define

$$\ell_k = \inf\{< \boldsymbol{\Lambda}_k \cdot \mathbf{K}, y >: \|\boldsymbol{\delta} - < bfK, y > \|_{\Sigma^{-1}} \leq \chi\} \tag{13}$$

and

$$u_k = \sup\{< \boldsymbol{\Lambda}_k \cdot \mathbf{K}, y >: \|\boldsymbol{\delta} - < bfK, y > \|_{\Sigma^{-1}} \leq \chi\} \tag{14}$$

Then

$$P\{\ell_k \leq \ < L_k, x > \ \leq u_k\} \geq 1 - \alpha; \tag{15}$$

i.e., $[\ell_k, u_k]$ is a $1-\alpha$ confidence interval for $< L_k, x >$. Note that the constraint set for this optimization problem, $\{y : \|\boldsymbol{\delta} - < bfK, y > \|_{\Sigma^{-1}} \leq \chi\}$, does not depend on L_k. This makes these confidence intervals simultaneously valid for all k:

$$P\{\ell_1 \leq \ < L_1, x > \ \leq u_1 \text{ and } \ell_2 \leq \ < L_2, x > \ \leq u_2 \text{ and } \cdots$$
$$\text{and } \ell_m \leq \ < L_m, x > \ \leq u_m\} \geq 1 - \alpha. \tag{16}$$

It is possible to find ℓ_k and u_k explicitly: if $\| < \mathbf{K}, x > - \boldsymbol{\delta}\|_{\Sigma^{-1}} \leq \chi$ (which occurs with probability $1-\alpha$),

$$\begin{aligned}
|\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} - < \boldsymbol{\Lambda}_k \cdot \mathbf{K}, x > | &= |\boldsymbol{\Lambda}_k \cdot (< \mathbf{K}, x > + \boldsymbol{\epsilon}) - < \boldsymbol{\Lambda}_k \cdot \mathbf{K}, x > | \\
&= |\boldsymbol{\Lambda}_k \cdot \boldsymbol{\epsilon}| \\
&\leq \|\boldsymbol{\Lambda}_k\|_{\Sigma} \|\boldsymbol{\epsilon}\|_{\Sigma^{-1}} \\
&\leq \chi \|\boldsymbol{\Lambda}_k\|_{\Sigma}.
\end{aligned} \tag{17}$$

This bound is attained provided $\{K_j\}$ are linearly independent (which we have assumed), so $\ell_k = \boldsymbol{\Lambda} \cdot \boldsymbol{\delta} - \chi \|\boldsymbol{\Lambda}_k\|_\Sigma$ and $u_k = \boldsymbol{\Lambda} \cdot \boldsymbol{\delta} + \chi \|\boldsymbol{\Lambda}_k\|_\Sigma$. See also *Stark* [1992].

This gives an explicit solution to the problem of constructing simultaneous confidence intervals for any number of Backus-Gilbert type averages. However, the confidence intervals obtained this way are unnecessarily wide, especially if one is interested in only a few such linear combinations. The next subsection shows how we can refine the intervals if we are interested in relatively few functionals compared with n, the number of data.

3.3 Confidence intervals based on the m-dimensional normal distribution

As noted previously, the linear combination of data $\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta}$ is an unbiased estimate of the functional $< \boldsymbol{\Lambda}_k \cdot \mathbf{K}, x >$. We see directly from the defining equation (2) that the estimate $\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta}$ is a linear combination of n independent Gaussian random variables with means $\{< K_j, x >\}$ and variances $\{\sigma_j^2\}$. As a result, the estimate is itself Gaussian with mean $< \boldsymbol{\Lambda}_k \cdot \mathbf{K}, x >$ and variance $\tau_k^2 \equiv \boldsymbol{\Lambda}_k \cdot \boldsymbol{\sigma}^2$, where $\boldsymbol{\sigma}^2$ is the vector $(\sigma_1^2, \ldots, \sigma_n^2)$. There are innumerable choices of simultaneous confidence intervals for the collection of functionals $\{< L_k, x >\}$, and we need to narrow the field somehow. A typical way to put all the intervals on an equal footing is to make the widths of all the intervals proportional to the standard deviations of the corresponding estimates, with the same constant of proportionality for all the intervals. That is, we seek a single constant c so that the set of intervals $\{[\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} - c\tau_k, \boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} + c\tau_k]\}_{k=1}^m$ are simultaneous $1 - \alpha$ confidence intervals for $\{< L_k, x >\}$.

If we set $\alpha = 0.05$ for example, we know that for $m = 1$, $c = \approx 1.96$. The first section of this note shows that if $m = 2$ *and the estimates* $\boldsymbol{\Lambda}_1 \cdot \boldsymbol{\delta}$ *and* $\boldsymbol{\Lambda}_2 \cdot \boldsymbol{\delta}$ *are statistically independent*, then c is the $\sqrt{1 - \alpha/2} = \sqrt{0.975} \approx 0.987$ percentage point of the normal distribution, which is about 2.23. The estimates are independent when the coefficients in the linear combinations are orthogonal vectors, for example, when they depend on mutually exclusive subsets of the data. In the typical case, the estimates are not independent. Note that if we wish to compare only two Backus-Gilbert estimates, we should proceed by finding a single confidence interval for the difference functional $< (\boldsymbol{\Lambda}_1 - \boldsymbol{\Lambda}_2) \cdot \mathbf{K}, x >$ instead of carrying out the following prescription. Only when we want to compare more than two averages does it make sense to use simultaneous confidence intervals.

To simplify the notation, let $\mu_k =< \boldsymbol{\Lambda}_k \cdot \mathbf{K}, x >$ be the kth "truth" and let $a_k = \boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta}$ be the kth "average" of the data. We have just seen that a_k is normally distributed with expected value μ_k and variance τ_k^2. We seek the smallest constant c such that

$$P\left\{\max_{k=1}^{m}\left|\frac{\mu_k - a_k}{\tau_k}\right| \leq c\right\} = 1 - \alpha. \tag{18}$$

To find c, we need the joint distribution of $v_k = (a_k - \mu_k)/\tau_k$, $k = 1, \ldots, m$. These m variables are Gaussian with expected value zero and variance 1 by construction, but we need to know the off-diagonal elements of their covariance matrix in order to find c. Note that $v_k = \boldsymbol{\Lambda}_k \cdot \boldsymbol{\epsilon}/\sqrt{\boldsymbol{\Lambda}_k \cdot \boldsymbol{\sigma}^2}$. Now if E is the expectation operator,

$$\begin{aligned}
\mathrm{Cov}(v_k, v_l) &= E\left[(v_k - Ev_k)(v_l - Ev_l)\right] \\
&= E[v_k v_l] \\
&= E\left[\frac{1}{\tau_k \tau_l}(\boldsymbol{\Lambda}_k \cdot \boldsymbol{\epsilon})(\boldsymbol{\Lambda}_l \cdot \boldsymbol{\epsilon})\right] \\
&= \frac{1}{\tau_k \tau_l}\sum_{j=1}^{n}\sum_{i=1}^{n} \Lambda_{kj}\Lambda_{li} E[\epsilon_j \epsilon_j] \\
&= \frac{1}{\tau_k \tau_l}\sum_{j=1}^{n}\sum_{i=1}^{n} \Lambda_{kj}\Lambda_{li}\sigma_j^2 1_{i=j} \\
&= \frac{1}{\tau_k \tau_l}\sum_{j=1}^{n} \Lambda_{kj}\Lambda_{lj}\sigma_j^2.
\end{aligned} \tag{19}$$

This is the (k, l)th element of the matrix

$$\tilde{\Lambda} \cdot \Sigma \cdot \tilde{\Lambda}^T, \tag{20}$$

where

$$\begin{aligned}
\tilde{\Lambda}_{kj} &= \frac{\Lambda_{kj}}{\tau_k} \\
&= \frac{\Lambda_{kj}}{\sqrt{\boldsymbol{\Lambda}_k \cdot \boldsymbol{\sigma}^2}}.
\end{aligned} \tag{21}$$

Thus the covariance matrix of $\{v_k\}$ is

$$\Gamma \equiv \tilde{\Lambda} \cdot \Sigma \cdot \tilde{\Lambda}^T, \tag{22}$$

and c is the solution to

$$(2\pi)^{-m/2}|\Gamma|^{-1/2} \int_{-c}^{c} dv_1 \int_{-c}^{c} dv_2 \cdots \int_{-c}^{c} dv_m \exp\left(-\frac{1}{2|\Gamma|}\mathbf{v}^T \cdot \Gamma \cdot \mathbf{v}\right) = 1 - \alpha, \tag{23}$$

where $|\Gamma|$ is the determinant of the matrix Γ. For a general covariance matrix Γ, c can not be found in closed form. However, it can be found iteratively by performing the integral (23) numerically for trial values of c. Since the integral is monotone in c, a bisection or other search can find the unique value of c relatively efficiently to any desired precision. The FORTRAN subroutine `mulnor`, publicly available through the Web site `statlib`, evaluates this integral numerically. The on-line statistical library `statlib` has a gopher connection:

```
Type=1
Name=StatLib Server (Carnegie Mellon University)
Path=
Host=lib.stat.cmu.edu
Port=70
```

The `mulnor` algorithm is number AS 195 in the Applied Statistics (`apstat`) section of the `statlib` library.

Once c has been found, the intervals

$$[\boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} - c\tau_k, \boldsymbol{\Lambda}_k \cdot \boldsymbol{\delta} + c\tau_k], \quad k = 1, \ldots, m \tag{24}$$

are a set of simultaneous $1 - \alpha$ confidence intervals for the functionals $\{< L_k, x >\}$, as we have sought to construct. If $m < n$, these intervals in general will be shorter than those derived in the previous subsection.

Acknowledgements. I thank Peter Wilson for encouraging me to write this down. This work was supported by NSF Presidential Young Investigator Award DMS-8957573 and Grant DMS-9404276, and NASA Grant NAGW-2515.

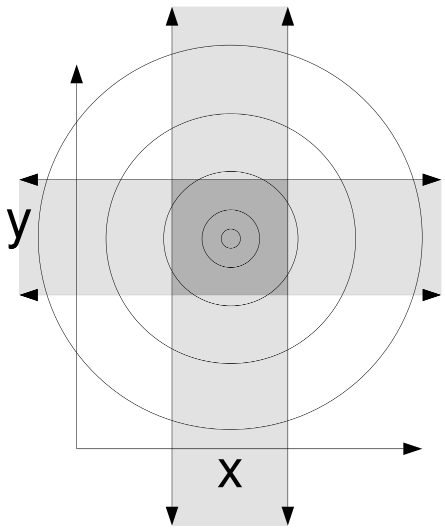

Figure 1: Illustration of the issue of simultaneity. The circles represent contours of the joint probability density function of X and Y. The chance that X is in the (infinite) shaded vertical slice is $1 - \alpha$, as is the chance that Y is in the horizontal slice. However, the chance that X is in the shaded region and Y is in the shaded region is the probability of the small square, which is less than $1 - \alpha$. In this sketch, X and Y are independent, so the probability of the square is $(1 - \alpha)^2$. In general, the probability that both X and Y are in a pair of ranges will depend on their joint distribution. The text gives three ways to adjust the lengths of a set of n confidence intervals so that the chance the intervals simultaneously contain their n associated parameters is at least $1 - \alpha$.

References

[1] G. Backus. Inference from inadequate and inaccurate data, I. *Proc. Natl. Acad. Sci.*, 65:1–7, 1970.

[2] G. Backus. Inference from inadequate and inaccurate data, II. *Proc. Natl. Acad. Sci.*, 65:281–287, 1970.

[3] G. Backus. Inference from inadequate and inaccurate data, III. *Proc. Natl. Acad. Sci.*, 67:282–289, 1970.

[4] G. Backus and F. Gilbert. The resolving power of gross Earth data. *Geophys. J. R. Astron. Soc.*, 16:169–205, 1968.

[5] G.E. Backus and F. Gilbert. Uniqueness in the inversion of inaccurate gross Earth data. *Phil. Trans. Roy. Soc. Lon. A.*, 266:123–192, 1970.

[6] P.J. Bickel and K.A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden Day, San Francisco, 1977.

[7] J. Christensen-Dalsgaard, J. Schou, and M.J. Thompson. A comparison of methods for inverting helioseismic data. *Mon. Not. R. astr. Soc.*, 242:353–369, 1990.

[8] W. Däppen, D.O. Gough, A.G. Kosovichev, and M.J. Thompson. A new inversion for the hydrostatic stratification of the Sun. In D. Gough and J. Toomre, editors, *Challenges to theories of the structure of moderate-mass stars*, pages 111–120. Springer-Verlag, New York, 1991.

[9] D.O. Gough and J. Toomre. Seismic observations of the solar interior. *Ann. Rev. Astron. Astrophys.*, 29:627–685, 1991.

[10] L.E. Johnson and F. Gilbert. Inversion and inference for teleseismic ray data. *Meth. in Comp. Phys.*, 12:231–266, 1972.

[11] T.G. Masters. Observational constraints on the chemical and thermal structure of the Earth's deep interior. *Geophys. J. R. Astron. Soc.*, 57:507–534, 1979.

[12] D.W. Oldenburg. One-dimensional inversion of natural source magnetotelluric observations. *Geophysics*, 44:1218–1244, 1979.

[13] D.W. Oldenburg. Conductivity structure of oceanic upper mantle beneath the Pacific plate. *Geophys. J. R. Astron. Soc.*, 65:359–394, 1981.

[14] R.L. Parker. The inverse problem of electrical conductivity in the mantle. *Geophys. J. R. Astron. Soc.*, 22:121–138, 1970.

[15] P.B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. *J. Geophys. Res.*, 97:14,055–14,082, 1992.