Journal of Mathematical Chemistry 21 (1997) 103–109

Calibrated nonparametric confidence sets

Tapon Roy

Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT 06877, USA

Received 23 July 1996

The construction of confidence sets when multivariate normality holds and in the general case where the usual spherical or elliptical structures may not occur is investigated. Calibration is used to correct the coverage probability of the nonparametric sets, and an example involving parameters from a chemical kinetics model in a biological system is used to demonstrate the techniques. Monte Carlo simulations validate the approach.

1. Introduction

In many cases it is necessary to determine a confidence set without assuming a particular distribution for the data. When certain distributional assumptions can be justified, confidence sets can often be easily constructed and have regular forms. For example in the multivariate normal case, one observes the familiar spherical or elliptical structures. There are numerous applications, however, that require confidence set construction when the underlying distribution is not well characterized, and only small samples are available.

An example of this situation is the determination of joint bioequivalence for pharmaceutical formulations. For this problem, one usually constructs confidence intervals for the difference or ratio of test and reference formulation means for various parameters estimated from the concentration-time profiles for each drug formulation. Though the parameters such as area under the curve (AUC) and maximum concentration ($C_{\max}$) are obviously correlated, confidence intervals are typically calculated for each parameter separately and bioequivalence concluded if each interval is completely contained within some specified regulatory boundaries. The literature is vast on this topic (see [8,14,15,17], and the references cited in those articles for a good overview).

In this paper, joint confidence sets assuming multivariate normality using a relatively new approach of Brown, Casella and Hwang [4] will be compared to bootstrap-calibrated corrected confidence sets not requiring distributional assumptions.

These approaches will be compared using a real bioequivalence example, and the corrected confidence set method will be evaluated with regard to the coverage correction required for various simulated distributions and sample sizes.

2. Confidence sets and the limaçon of Pascal

Brown, Casella and Hwang [4] demonstrate that the limaçon of Pascal, a generalization of a cardioid curve, provides an optimal confidence set in the normal case. They show that this set has actual coverage probability equal to the nominal, and smaller volume than the usual elliptical region in the bivariate situation shown. For $C(x)$, a set estimate of a parameter θ, the expected volume, at $\theta = 0$, of the uniformly most accurate set $C^*(x)$ was shown to be

$$\begin{aligned} E_0 \operatorname{vol}\big(C^*(X)\big) &= \int_\Theta \left[\int_{\mathcal{X}} I\big(\theta \in C^*(x)\big) f(x \mid 0)\, \mathrm{d}x \right] \mathrm{d}\theta \\ &= \int_\Theta \left[\int_{(x:\ x'\theta/|\theta| \geqslant |\theta| - a)} f(x \mid 0)\, \mathrm{d}x \right] \mathrm{d}\theta \\ &= \int_\Theta \Phi\big(a - |\theta|\big) \mathrm{d}\theta \end{aligned} \tag{1}$$

since $X \sim N(0, I)$, $X'\theta/|\theta| \sim N(0,1)$ for any nonzero θ. Applying a polar transformation gives

$$\begin{aligned} E_0 \operatorname{vol}\big(C^*(X)\big) &= \frac{\pi^{p/2}}{\Gamma(p/2+1)} \int_0^\infty r^{p-1} \Phi(a - r)\, \mathrm{d}r \\ &= \frac{\pi^{p/2}}{\Gamma(p/2+1)} \int_{-\infty}^a \frac{(a-t)^p}{p} \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}}\, \mathrm{d}t. \end{aligned} \tag{2}$$

Since the volume of a p-sphere of radius 1 is $\pi^{p/2}/\Gamma(p/2+1)$, the pth root of the integral in (2) gives the radius of the set, where $a = \Phi^{-1}(1-\alpha)$ produces a $1-\alpha$ confidence set. In the normal case, the optimal confidence set is then shown to be

$$C^*(x) = \big\{\theta:\ |\theta| \leqslant a + |x| \cos\beta\big\}, \tag{3}$$

where $\cos\beta = x'\theta/|x||\theta|$ and β is the angle between x and θ. The boundary of $C^*(x)$ is the main lobe of the limaçon of Pascal. When normality does not hold, the limaçon is not expected to appear; indeed, even the spherical or elliptical structures or their hybrids do not appear unless star unimodality is present [5].

3. Calibrated confidence sets

When distributional assumptions are untenable, the bootstrap method can be considered for obtaining confidence intervals. It is well-known, however, that the percentile bootstrap method can undercover and that the various bias-corrected, the percentile-t, and the Edgeworth-corrected intervals might not completely correct the coverage problem, or if they do, can give overly long intervals [6,9]. Bootstrap iteration [2] seems to provide a satisfactory solution to the problem of constructing nonparametric confidence intervals with high coverage accuracy and with stable endpoints and

lengths [7]. We will investigate whether this solution holds for joint confidence sets, where one would expect that the undercoverage problem is magnified.

Bootstrap iteration with two levels of bootstrapping for confidence intervals can be described as follows [2,11]. From the estimated distribution $\hat{F}_n$ draw a bootstrap sample of size n, called x_n^*. Let $\hat{F}_n^*$ indicate the recomputed estimate of F from x_n^*. Now, let x_n^{**} be a further bootstrap sample of size n drawn from the estimated distribution $\hat{F}_n^*$. The components of x_n^{**} are conditionally independent given x_n and x_n^*. Then $\hat{\theta}(x_n)$ is an estimate of θ based on x_n and $\hat{\theta}(x_n^*)$ and $\hat{\theta}(x_n^{**})$ are the bootstrap estimates of θ denoted $\hat{\theta}_n^*$ and $\hat{\theta}_n^{**}$.

Correlations among the parameters (and among the repeated measures responses, if applicable) are preserved by vector resampling.

From the first level of bootstrapping, suppose one has obtained B_1 resamples each of size n. Suppressing the n subscript we can obtain from the ordered values $\hat{\theta}_{B_1,1}^* \leqslant \cdots \leqslant \hat{\theta}_{B_1,B_1}^*$, a nominal γ-level percentile-method interval for θ:

$$\big(\hat{\theta}_{B_1,[(1-\gamma)B_1/2]+1}^*, \hat{\theta}_{B_1,[(1+\gamma)B_1/2]+1}^*\big), \tag{4}$$

where [.] indicates the integer function.

For each of the B_1 resamples, resample B_2 times and construct percentile-method intervals for $\hat{\theta}$ at several nominal levels $\gamma_1, \gamma_2, \ldots, \gamma_n$ close to the desired level α, but with enough of a range to correct for possibly substantial miscoverage. The estimate of the coverage probability, $\hat{\eta}(\gamma_i)$ for $i = 1, 2, \ldots, n$ is the proportion of the B_1 intervals of nominal level γ_i which cover $\hat{\theta}$.

We now calibrate the confidence interval by finding the nominal coverage level, β_α, such that the true coverage of the interval is exactly α. Letting β_α solve $\eta(\beta_\alpha) = \alpha$ will provide the needed calibration adjustment. For the bootstrap interval, we obtain an approximate value of $\hat{\beta}_\alpha$ such that $\hat{\eta}(\hat{\beta}_\alpha) = \alpha$. Thus, a percentile-method interval with nominal level $\hat{\beta}_\alpha$ approximates the coverage corrected interval. For multiparameter confidence sets, the natural extension would be to include in the confidence set those cases which are jointly contained in the coverage-corrected (4) for each parameter, that is, for $\hat{\theta}^* = (\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_p^*)$, we require

$$\bigcap_{j=1}^{p} I_j = \big\{\hat{\theta}^* : (\hat{\theta}_1^* \in I_1) \text{ and } (\hat{\theta}_2^* \in I_2) \text{ and } \ldots \text{ and } (\hat{\theta}_p^* \in I_p)\big\},$$

where $I_1, I_2, \ldots, I_p$ are each of the form in the coverage-corrected (4). This general technique, of approaching a complex problem by combining information from less complex problems each having some optimum properties, is due to S.N. Roy [12, 13, p. 12]. The next step is to calibrate the actual set coverage probabilities to the corresponding nominal interval probabilities. The more conventional approach would be to retain the ellipsoidal structure and use bootstrapping to empirically obtain F-type critical points, as Adkins and Carter Hill [1] do in a regression setting.

Bootstrap iteration, even if restricted to two levels, can be quite computationally intensive, since $B_1 B_2$ is recommended to be of the order of 10^6 [3]. There has been

recent research aimed at reducing the computational burden of the iteration procedure, as described in Lee and Young [10].

4. A bioequivalence example

To demonstrate the methods, we will use the data from Sheen et al. [16]. They compare the bioavailability of α-pentyl-3-(2-quinolinylmethoxy)benzenemethanol, a 5-lipoxygenase inhibitor, for tablet and capsule formulations, each with or without food. This is a basic compound with a pK_a value of 3.7 and a solubility of ~ 0.002 mg/mL in water at 37°C at pH of about 6. Eight healthy male volunteers received the treatments according to a randomized four-way crossover design.

For the tablet formulation, we can evaluate the bioequivalence of the compound taken with (subscript 1, below) or without (subscript 2) food. In figure 1, a visual assessment of bivariate normality for AUC and $C_{\max}$ is displayed using a plot of squared Mahalanobis distances vs. $\chi^2(p)$ quantiles, where $p = 2$ variables in this case. There is some suggestion of departure from bivariate normality (nonlinearity

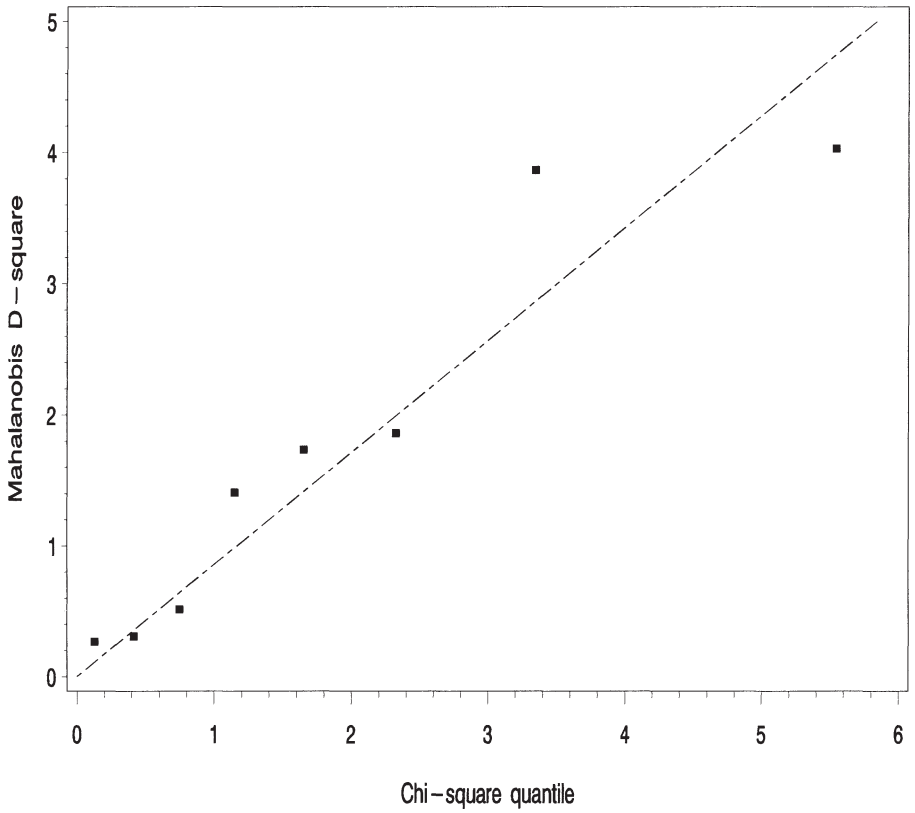

Figure 1. Bivariate normality plot for AUC and $C_{\max}$. A straight line, at 45°, through the origin (the reference line shown) would indicate bivariate normality.

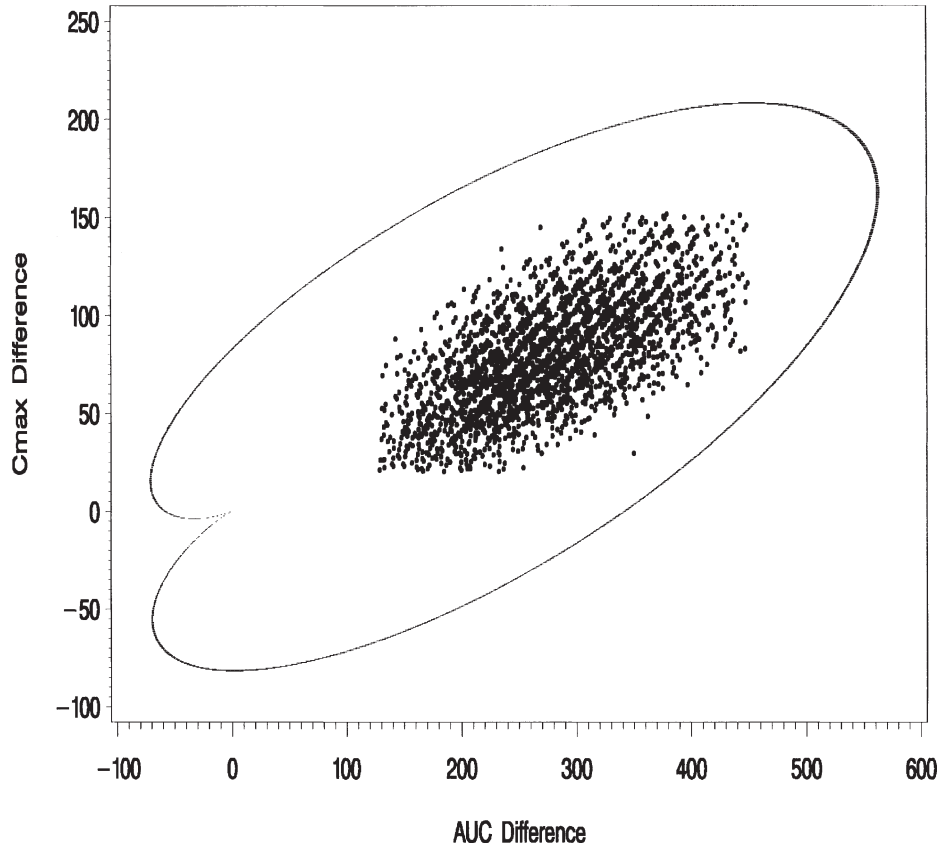

Figure 2. The boundary of the two-dimensional 90% limaçon for the tablet. The points within the limaçon form the coverage-corrected bootstrap confidence set.

of the plot). Each parameter appears normal by the Shapiro–Wilk test (however, the power of the test is low for these small samples), but the limaçon requires multivariate normality.

For the tablet, one can construct a two-dimensional limaçon with AUC difference $\theta_1 = \mu_1 - \mu_2$ and $C_{\max}$ difference $\theta_2 = \tau_1 - \tau_2$. If we assume that the data are bivariate normal with known variance: $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, where $\hat{\theta} \sim N(\theta, \Sigma)$, the $1 - \alpha$ limaçon confidence set is then

$$\left\{\theta\colon\ (\theta'\Sigma^{-1}\theta)^{1/2} \leqslant z_\alpha + \frac{\hat{\theta}'\Sigma^{-1}\theta}{(\theta'\Sigma^{-1}\theta)^{1/2}}\right\} \tag{5}$$

where z_α is the upper α critical value from a univariate standard normal distribution. This set gives a fairly elliptical limaçon as is seen in figure 2 (for $\alpha = 0.10$) which displays the boundary of the limaçon, and within it, the calibrated 90% bootstrap confidence set using (4) represented by the points shown, which looks roughly elliptical. Since this example had a repeated-measures structure, the subject response vectors

were resampled for the bootstrap method. Note that, for this data, the limaçon is completely contained within the usual elliptical confidence region (see [4], figure 5b). Since one would conclude bioequivalence if the confidence set is contained completely within the regulatory boundaries, which would be rectangular or hyperrectangular, the limaçon indicates bioequivalence for smaller differences than the usual confidence ellipse, and the coverage-corrected bootstrap set provides a still sharper inference. The bootstrap can also be used if confidence sets on the ratio of means, or some function of the medians are needed.

5. Simulation results

In previous simulation studies, the iterated bootstrap performed well [6,9] for constructing nonparametric confidence intervals. In this study, we will determine the magnitude of the calibration correction necessary to produce accurate confidence sets when two parameters are considered simultaneously.

Since the example in section 4 had a repeated-measures structure, the simulations were based on correlated normal ($\rho = 0.5$), correlated lognormal ($\rho = 0.378$), and correlated folded normal ($\rho = 0.5$) variates. Results are for 1,000 Monte Carlo samples of sizes $n = 8$ and $n = 24$ for each distribution. Table 1 displays the nominal levels needed to obtain joint bootstrap confidence sets with actual coverage probability 0.9. It is encouraging that for the example in section 4, the bootstrap iteration calibration indicated that a nominal level of 0.976 was required to correct the bootstrap coverage, agreeing closely with the simulation results in table 1. The corrections for each component, separately, were 0.944 and 0.960 for the section 4 example. From table 1, as expected, the magnitude of the correction needed increases as the sample size decreases. If the sample size is quite small, the possible number of distinct resamples becomes limited, rendering iteration less effective.

Table 1
Nominal coverage for bootstrap intervals needed to obtain a joint bootstrap confidence set with actual coverage probability 0.90. Simulations based on 1,000 Monte Carlo samples in each case.

Distribution	Sample size	Nominal coverage
Normal	24 *	0.970
	8 **	0.986
Folded normal	24 *	0.965
	8 **	0.992
Lognormal	24 *	0.973
	8 **	0.995

* $B = 1{,}000$ bootstrap resamples per Monte Carlo sample.
** $B = 4{,}000$ bootstrap resamples per Monte Carlo sample.

6. Discussion

Constructing good nonparametric confidence intervals and sets is a complex task. The results given in this paper indicate that one can obtain satisfactory nonparametric confidence sets, though admittedly, at a high computational cost, and requiring substantial correction to the ordinary percentile sets. With the advances seen in computing power and storage, and the recently developed approximating techniques, the cost issue should become less pronounced. An advantage of the bootstrap is that it can be applied to confidence set estimation for the ratio of means or medians, the mean or median of ratios, and other functionals. If multivariate normality is a reasonable assumption, then the limaçon of Pascal can provide optimal confidence sets in certain circumstances.

References

[1] L.C. Adkins and R. Carter Hill, Journal of Statistical Computation and Simulation 36 (1990) 9.
[2] R. Beran, Biometrika 74 (1987) 457.
[3] J.G. Booth and P. Hall, Biometrika 81 (1994) 331.
[4] L.D. Brown, G. Casella and J.T.G. Hwang, Journal of the American Statistical Association 90 (1995) 880.
[5] A. DasGupta, J.K. Ghosh and M.M. Zen, Annals of Statistics 23 (1995) 1408.
[6] T.J. DiCiccio, M.A. Martin and G.A. Young, Biometrika 79 (1992) 285.
[7] P. Hall, *The Bootstrap and Edgeworth Expansion* (Springer, New York, 1992).
[8] J.C. Hsu, J.T.G. Hwang, H.K. Liu and S.J. Ruberg, Biometrika 81 (1994) 103.
[9] S.M.S. Lee and G.A. Young, Annals of Statistics 23 (1995) 1301.
[10] S.M.S. Lee and G.A. Young, Journal of the Royal Statistical Society, Series B 58 (1996) 235.
[11] M.A. Martin, in: *Computing Science and Statistics: Interface '90 Proceedings of the 22nd Symposium on the Interface*, ed. R. LePage (ASA, East Lansing, Michigan, 1991) pp. 73–78.
[12] S.N. Roy, Annals of Mathematical Statistics 24 (1953) 220.
[13] S.N. Roy, *Some Aspects of Multivariate Analysis* (Wiley, New York, 1957).
[14] T. Roy, Journal of Statistical Computation and Simulation 47 (1993) 89.
[15] D.J. Schuirmann, Journal of Pharmacokinetics and Biopharmaceutics 15 (1987) 657.
[16] P.-C. Sheen, S.-I. Kim, J.J. Petillo and A.T.M. Serajuddin, Journal of Pharmaceutical Sciences 80 (1991) 712.
[17] W.J. Westlake, Journal of Pharmaceutical Sciences 61 (1972) 1340.