Rectangular Reference Regions for Multivariate Measurements in Laboratory Medicine

Michael Daniel Lucagbo[1] and Thomas Mathew[2],
[1]University of Maryland, Baltimore County and University of the Philippines Diliman, Email: micluca1@umbc.edu
[2]University of Maryland, Baltimore County, Email: mathew@umbc.edu

1) Introduction

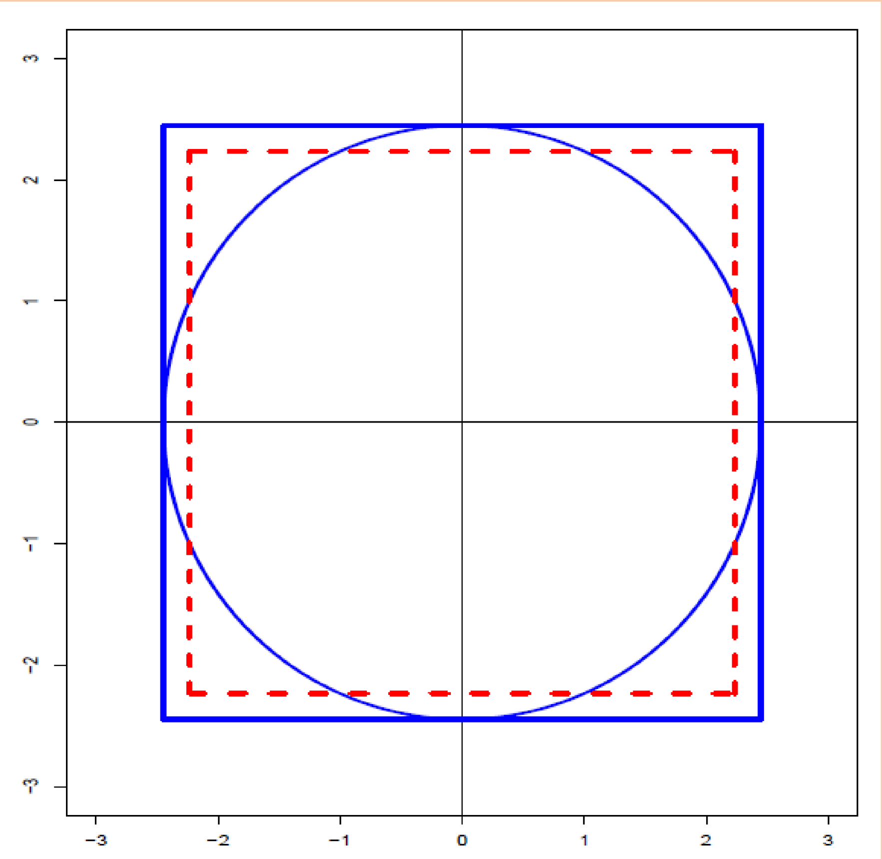
The reference interval is the most widely-used decision making tool. It is indispensable in determining the health of an individual. When there are multiple correlated biochemical analytes to be measured, a multivariate reference region (MRR) is needed.

Ellipsoidal prediction regions are the traditional way to construct such regions, but they cannot detect particular components which are extreme, thus the need for rectangular prediction regions.

2) Objectives

We propose a procedure to compute rectangular prediction regions as reference regions for multivariate normal measurements. In addition, we also propose a procedure to compute nonparametric prediction regions.

3) Rectangular Regions



The figure on the right is from Wolf and Wunderli (2015). The blue circle is a 95% prediction region for the bivariate normal distribution, which is easy to calculate but can not detect component-wise outliers.

We want to find the red square, which is also a 95% prediction region, but is easier to interpret, and can detect component-wise outliers. We do this in both the multivariate normal and nonparametric cases.

4) Multivariate Normal Case

Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be a random sample from the $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, a rectangular prediction region is of the form given below.

$$\left[\bar{X}_1 \pm \kappa\sqrt{S_{11}}\right] \times \cdots \times \left[\bar{X}_p \pm \kappa\sqrt{S_{pp}}\right]$$

The prediction factor κ is unknown and must be estimated. If $\boldsymbol{X} = (X_1,\ldots,X_n)'$ is the observation to be predicted, then κ satisfies:

$$P\left(\bigcap_{i=1}^{p}\left\{\left|\frac{X_i-\bar{X}_i}{\sqrt{S_{ii}}}\right| \leq \kappa\right\}\right) = 1-\alpha$$

With κ now expressed as a quantile, we estimate it via a parametric bootstrap.

5) Simulation Results

The table below shows estimated coverage probabilities. Data are generated from $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = 0.5\boldsymbol{I}_p + 0.5\mathbf{1}_p\mathbf{1}_p'$ and the nominal coverage is 95%. The estimated coverage probabilities are close to 0.95 even for small sample sizes and large dimensions.

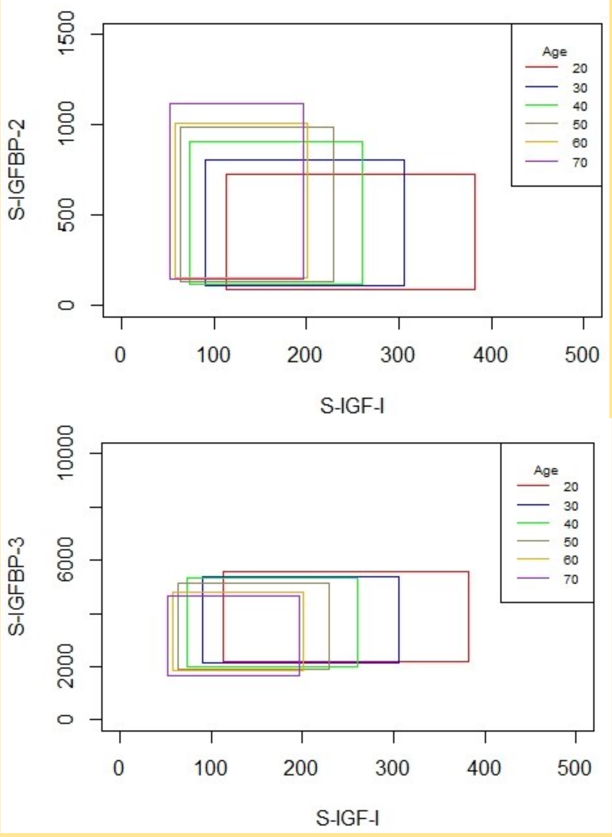
	$p=2$	$p=3$	$p=10$
$n=100$	0.9530	0.9412	0.9458
$n=50$	0.9420	0.9468	0.9460
$n=30$	0.9466	0.9420	0.9490

6) Assessment of Kidney Function

We apply the proposed method in the multivariate normal case using data from Albert and Harris (1987) consisting of measurements on urea, uric acid, and creatinine to assess kidney function. We compare the MRR based on our procedure with standard ranges given in the Wikipedia article *Reference ranges for blood tests*:

Analyte	MRR	Standard range
Urea (mmol/L)	2.63-7.57	3-7
Uric Acid (μmol/L)	161.71-444.29	180-480
Creatinine (μmol/L)	51.81-118.19	68-118 (Males) 68-98 (Females)

7) Insulin-like Growth Factors



The procedure can be extended to incorporate covariates. We use data from Mattsson et al. (2008) to compute a reference region for insulin-like growth factors (S-IGF-I, S-IGFBP2, S-IGFBP3) based on age, sex, and BMI. To the right are plots of the reference regions.

8) Nonparametric Case

Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be a random sample of p-variate measurements. We apply univariate Box-Cox transformation to each component to obtain the transformed measurements $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$. If $\boldsymbol{X}$ is the future observation to be predicted and $\boldsymbol{Y} = (Y_1,\ldots,Y_n)'$ is its transformed version, we want to find the prediction region of the form given below.

$$\left[\bar{Y}_1 \pm \kappa\sqrt{S_{11,y}}\right] \times \cdots \times \left[\bar{Y}_p \pm \kappa\sqrt{S_{pp,y}}\right]$$

The factor κ can be expressed as a quantile (see below) and we estimate it via a nonparametric bootstrap.

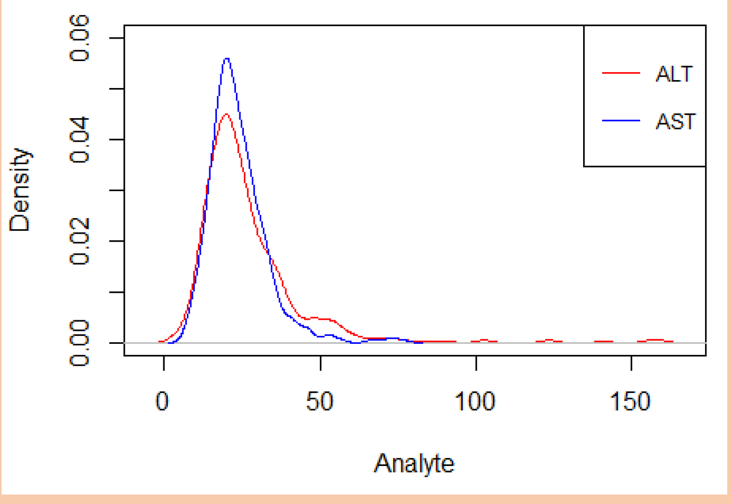
$$P\left(\max_{1\leq j\leq p}\left|\frac{Y_j-\bar{Y}_j}{\sqrt{S_{jj,y}}}\right| \leq \kappa\right) = 1-\alpha$$

9) Simulation Results

For the simulations, data are generated from the p-variate lognormal distribution with log-scale mean and covariance matrix $\mathbf{0}$ and $\boldsymbol{\Sigma} = 0.5\boldsymbol{I}_p + 0.5\mathbf{1}_p\mathbf{1}_p'$, respectively. The coverage probabilities shown below are close to 0.95 for moderate or large n and small p.

	$p=2$	$p=3$	$p=10$
$n=100$	0.9408	0.9396	0.9324
$n=50$	0.9344	0.9396	0.9252
$n=30$	0.9304	0.9224	0.9188

10) Assessment of Liver Function



We apply the proposed nonparametric procedure to data from Harris and Boyd (1995) consisting of measurements on alanine transaminase (ALT) and aspartate transaminase (AST) which are used to assess liver function. The density plots of ALT and AST show they are positively skewed, indicating the aptness of a transformation. Using our procedure the resulting MRR for ALT and AST is: ALT (7.16-84.81 U/L), AST (8.97-54.75 U/L).

11) Contribution to Knowledge

Our work has developed accurate rectangular prediction regions in both the multivariate normal and nonparametric setups that can be used to set reference regions in laboratory medicine.

References

Albert, A. and Harris, E.K. (1987). Multivariate Interpretation of Clinical Laboratory Data. New York, NY: Marcel Dekker, Inc.

Harris, E. K., and Boyd, J. C. (1995). Statistical Bases of Reference Values in Laboratory Medicine. New York, NY: Marcel Dekker, Inc.

Mattsson, A., Svensson, D., Schuett, B., Osterziel, K. J., and Ranke, M. B. (2008). Multidi- mensional Reference Regions for IGF-I, IGFBP-2 and IGFBP-3 Concentrations in Serum of Healthy Adults. Growth Hormone & IGF Research, 18(6), 506-516.

Wolf, M. and Wunderli, D. (2015). Bootstrap Joint Prediction Regions. *Journal of Time Series Analysis* 36(3).

Wikipedia contributors. (2020, July 28). Reference ranges for blood tests. In *Wikipedia, The Free Encyclopedia*. Retrieved 03:02, August 4, 2020, from https://en.wikipedia.org/w/index.php?title=Reference_ranges_for_blood_tests&oldid=969965622.