Methods of Comparing Digital PCR Experiments

Michał Burdukiewicz[1], Piotr Sobczyk[2], Paweł Mackiewicz[1], Stefan Rödiger[3]

[1]University of Wrocław, Department of Genomics, Poland

[2]Wrocław University of Technology, Institute of Mathematics and Computer Science, Poland

[3]Faculty of Natural Sciences, Brandenburg University of Technology Cottbus–Senftenberg, Germany

Introduction

The outcome of digital PCR (dPCR) experiments are mean copies per partition ($\boldsymbol{\lambda}$). Results are derived from the measured data, an ordered (in one or two dimensions) sequence of positive partitions. The usual analysis involves assumption the template molecules are Poisson distributed among partitions. On this premise, already proposed approaches, based on the confidence intervals (Dube et al., 2008) or uncertainty quantification (Bhat et al., 2009), allow a comparison of experiments.

GLM testing framework

Generalized Linear Models (GLM) are linear models for data, where the response variables may have non-normal distributions (as binomially distributed positive partitions in digital PCR experiments). We propose a model reflecting relationships in results of digital PCR as given by:

$$\log \boldsymbol{Y} = \beta^{\boldsymbol{T}} \boldsymbol{X} \tag{1}$$

where $\boldsymbol{Y}$ are counts, $\boldsymbol{X}$ are experiments names (categorical data) and $\boldsymbol{\beta}$ are linear model coefficients for every experiment. Moreover, $\exp \boldsymbol{\beta} = \boldsymbol{\lambda}$. Estimated means copies per partitions obtained from the model are compared each other using multiple t-test (Bretz et al., 2010). The linear model above can be used only when the concentration of template molecules in samples is small (positive partitions contain very rarely more than 1 template particle). Higher concentrations basically requires binomial regression.

Multiple testing (MT) framework

The dPCR experimentes are compared pairwise using the uniformly most powerful (UMP) ratio test (Fay, 2010). Furthermore, computed p-values are adjusted using Benjamini Hochberg correction (Benjamini and Hochberg, 1995) to control family-wise error rate.

The UMP ratio test has following null-hypothesis:

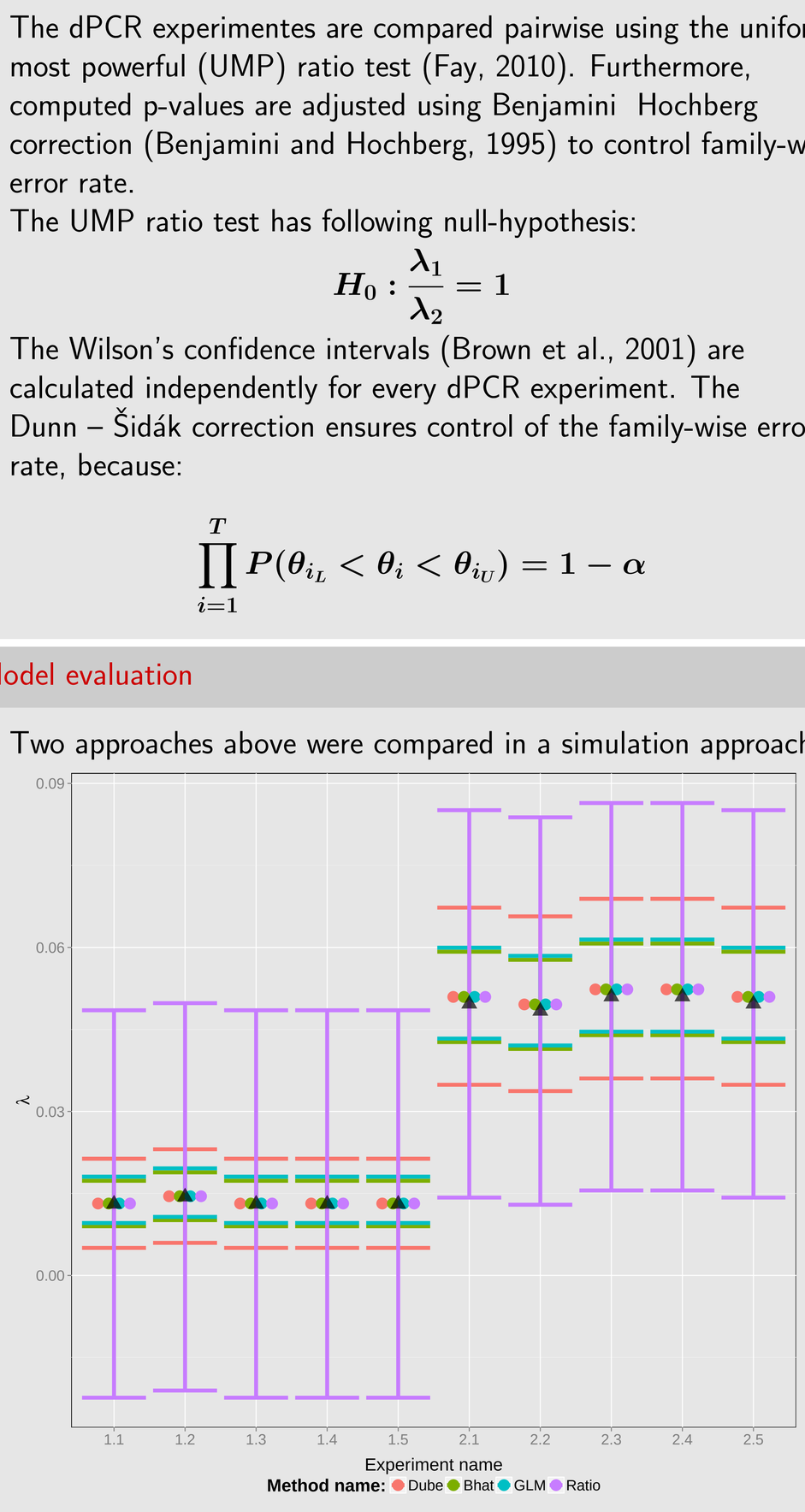
$$\boldsymbol{H_0 : \frac{\lambda_1}{\lambda_2} = 1} \tag{2}$$

The Wilson's confidence intervals (Brown et al., 2001) are calculated independently for every dPCR experiment. The Dunn – Šidák correction ensures control of the family-wise error rate, because:

$$\prod_{i=1}^{T} \boldsymbol{P(\theta_{i_L} < \theta_i < \theta_{i_U}) = 1 - \alpha} \tag{3}$$

Model evaluation

Two approaches above were compared in a simulation approach.

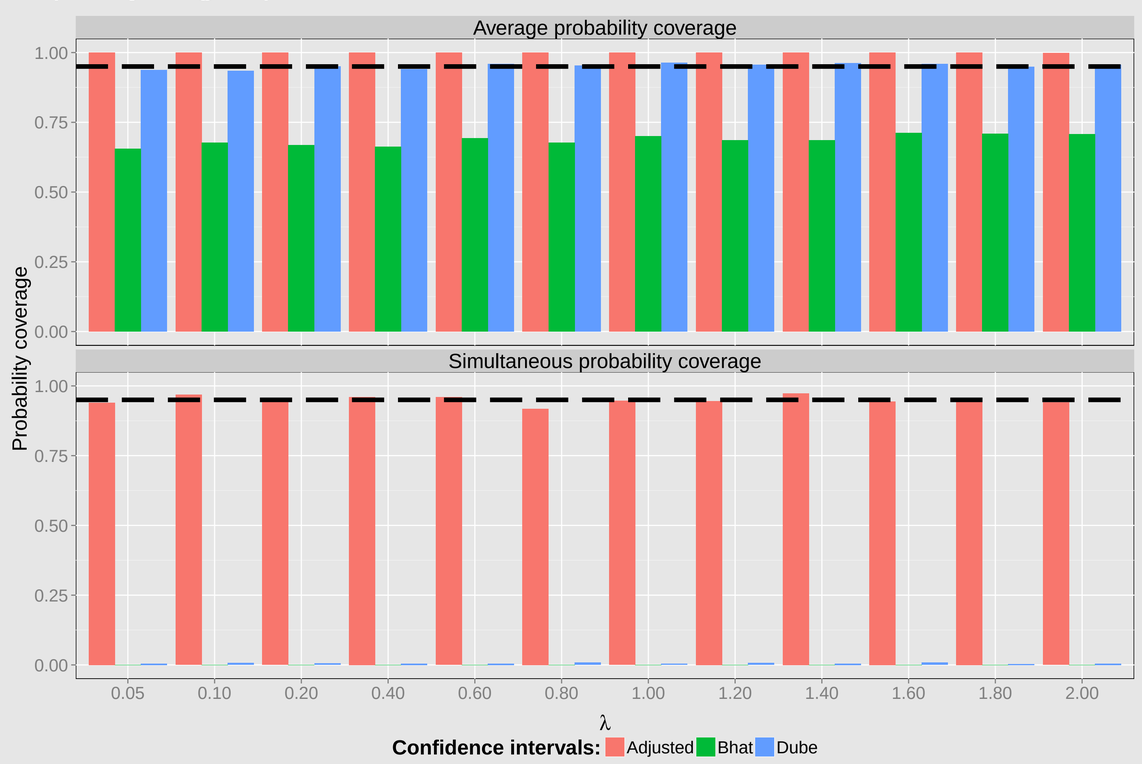


Probability coverage of confidence intervals

Average covarage probability is the proportion of the time that the interval contains the true value of $\boldsymbol{\lambda}$.

In the example below, we simulated $1 \times \mathbf{10^6}$ droplet dPCR experiments ($2 \times \mathbf{10^4}$ droplets each) for each level of $\boldsymbol{\lambda}$ ($1.2 \times \mathbf{10^7}$ experiments total). We computed average probability coverage of CI obtained by three methods: Dube's (Dube et al., 2008), Bhat's (Bhat et al., 2009) and by MT.

To assess simultaneous coverage probability, we randomly divided experiments into 2000 groups (500 experiments each) for each possible value of $\boldsymbol{\lambda}$. We counted frequency of groups in which all confidence intervals contain the true value of $\boldsymbol{\lambda}$.



The dashed black line marks 0.95 border.

Method name	Type of coverage	Value
Adjusted	Average probability coverage	1.00
Bhat	Average probability coverage	0.69
Dube	Average probability coverage	0.95
Adjusted	Simultaneous probability coverage	0.95
Bhat	Simultaneous probability coverage	0.00
Dube	Simultaneous probability coverage	0.01

Summary

- The adjusted confidence intervals computed by MT guarantee 0.95 simultaneous coverage probability.

Avaibility

pcRuniveRsum:

`http://michbur.github.io/pcRuniveRsum/`

dpcR R package:
`http://cran.r-project.org/web/packages/dpcR/`

Simulations

We performed all in-silico experiments using the accurate digital PCR data simulations (Jacobs et al., 2014).

Bibliography

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Bhat, S., Herrmann, J., Armishaw, P., Corbisier, P., and Emslie, K. R. (2009). Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number. *Analytical and bioanalytical chemistry*, 394(2):457–467.

Bretz, F., Hothorn, T., and Westfall, P. (2010). *Multiple Comparisons Using R*. Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.*, 16(2):101–133.

Dube, S., Qin, J., and Ramakrishnan, R. (2008). Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. *PloS one*, 3(8):e2876.

Fay, M. (2010). Two-sided exact tests and matching confidence intervals for discrete data. *Proceedings of the National Academy of Sciences of the United States of America*, 2(1):53–58.

Jacobs, B. K., Goetghebeur, E., and Clement, L. (2014). Impact of variance components on reliability of absolute quantification using digital pcr. *BMC bioinformatics*, 15(1):283.