Innovative Approximations for Studentized Multivariate Distributions

John E. Kolassa

December 14, 2023

1 Research Question

My collaborator and I study a class of approximations for probabilities called saddlepoint approximations. My particular expertise for these approximations is in dimensions greater than one. The canonical application of these approximations is to testing with multiple outcomes. Consider a study in which participants are given both a conventional and an experimental treatment, with the aim of determining which treatment is more effective. It may be the case that the difference in performances depends on the order in which treatments are given. In this case, two measures of treatment difference are produced, and one wishes to assess the evidence that the new treatment is more effective in one of the orderings. Various probabilities of joint outcomes for the two orderings are calculated, and our tools are intended to help with this assessment. An example of such an investigation is given by Hao and Kolassa (2016).

These individual assessments often have a distribution with an unknown scale that must be estimated from data. The process of adjusting a variable using an internally-derived scale measure is known as studentizing, and is part of the work that John Robinson and I have in progress.

2 Project Design

This project will continue ongoing work with PhD student Donghyun Lee, and existing work with my collaborator John Robinson, as described in items 4, 8, 9, and 10 on my abbreviated CV.

2.1 The Technical Question

Suppose that random vectors (X_i^1, X_i^2) have a common continuous distribution, for $i \in \{1, \dots, n\}$. Suppose that the common moment generating function

$$M(\tau_1, \tau_2) = \mathrm{E}\left[\exp(\tau_1 X_i^1 + \tau_2 X_i^2)\right]$$

for (X_i^1, X_i^2) is finite for $|\tau_1| < \epsilon$ and $|\tau_2| < \epsilon$, for some $\epsilon > 0$. This paper approximates joint tail probabilities for the mean of an independent and identically distributed set random vectors $\boldsymbol{X}_i = (X_i^1, X_i^2)$, each with this distribution. As noted above in Project Design, these two observations per subject canonically reflect two separate efficacy measures for the same subject. This project explores approximations to the joint distribution of the means $\bar{\boldsymbol{X}}$ of these responses.

Standard Fourier inversion techniques produce an integral representation for these tail probabilities:

$$\mathrm{P}\left[\bar{\boldsymbol{X}} \gtrsim \bar{\boldsymbol{x}}\right] = \frac{1}{(2\pi i)^2} \int_{-i\infty}^{i\infty} \int_{-i\infty}^{i\infty} \frac{\exp(n[K(\tau_1, \tau_2) - \bar{x}^1\tau_1 - \bar{x}^2\tau_2])}{\tau_1\tau_2}\, d\tau_1 d\tau_2. \quad (1)$$

Two competing tools exist for approximation of these probabilities in one dimension are presented by Robinson (1982) and Lugannani and Rice (1980). Of these two approximations, the second is more interpretable, and can be applied din a wider range of conditions. I presented a multivariate analog of Robinson (1982) in Kolassa (2003), but this approximation shares the restricted applicability of Robinson (1982). I and other PhD students Kolassa and Li (2010) and Yaoshi Wu made incomplete progress on this question. Recent work with PhD student Lee and potential collaborator Robinson has the potential to produce more widely applicable and more easily interpreted approximations.

2.2 Timeline

I have sabbatical for calendar year 2025. My plan is to visit Sydney early in January 2025, and spend two weeks to one month. Based on my experience with past collaborations with John Robinson, this should allow us sufficient time to move the project forward. I hope to have a manuscript ready for submission by June 2025.

2.3 List of activities

- Perform preliminary steps in research. This will be performed as time permits during 2024, and is not supported by requested funds.
- Travel to Sydney and collaborate. Requested funds will support this travel.
- Refine manuscript remotely.
- Submit for publication, and revise in light of editorial comments.

2.4 Measurement of Success and Plan for Evaluation

We will measure project success by the quality of journal in which our manuscript is accepted, and by the manuscript's citations.

3 Innovation

The move from one dimension to two dimensions in these probability approximations is surprisingly complicated, because the resulting approximation not only has to adjust for distributions of each variable separately, but to a potentially nonlinear relationship between the variables that is not adequately captured by existing methods. This difficulty arises because the transition from the univariate approach of Robinson (1982) to Lugannani and Rice (1980) involves a re-parameterization of integral (1) which behaves well in one dimension but poorly in multiple dimensions. My previous investigations have not yielded a magic bullet here, but preliminary results indicate that isolating the poor behavior of the integral re-parameterization to a term in the exponent of the integrand which can then be expanded removes this difficulty.

4 Potential for Beneficial Partnership

The University of Sydney hosts numerous scholars, many of whom are John Robinson's proteges, who I will be able to partner with. The University of Sydney is the premier Australian statistical research center.

5 Sustainability Plans

Our research was severely impacted by Covid 19-related travel restrictions. I hope that this project will lead to funding from the Simons Foundation that will support further travel. While John Robinson is in the late stages of his career, this travel might build bridges between his students and me and my students, and may serve as a foundation for further funding.

References

Hao, Y. and Kolassa, J. (2016). Multiple hypotheses in the analysis of a crossover trial. *Statistics in Biosciences*, 8(2):253–263.

Kolassa, J. (2003). Multivariate saddlepoint tail probability approximations. *Annals of Statistics*, 31:274–286.

Kolassa, J. and Li, J. (2010). Multivariate saddlepoint approximations in tail probability and conditional inference. *Bernoulli*, 14(4):1191–1207.

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12:475–490.

Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *Journal of the Royal Statistical Society Series B*, 44:91–101.