Statistics and Probability Letters 145 (2019) 57–62

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Efficient computation for Bayesian comparison of two proportions

Mikkel N. Schmidt *, Morten Mørup

Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark

ARTICLE INFO

Article history:
Received 23 April 2018
Received in revised form 15 August 2018
Accepted 19 August 2018
Available online 4 September 2018

Keywords:
Bayesian analysis
Comparison of proportions
Integral of beta distribution
Hypergeometric function

ABSTRACT

In Bayesian comparison of two proportions, the exact computation of the evidence involves evaluating a generalized hypergeometric function. Several agreeing, but not identical, expressions for the evidence have been derived in the literature; however, their practical computation (by summing the truncated hypergeometric series) can be troubled by slow convergence or catastrophic cancellation. Using a set of equivalence relations for the generalized hypergeometric function, we derive ten equivalent expressions for the evidence: We show that one of these formulations, which has not previously been studied, is superior in terms of its computational properties. We recommend that this be used instead of existing formulations, and provide an efficient software implementation.

1. Bayesian comparison of two proportions

Let T denote a 2×2 contingency table with fixed column totals n_1 and n_2,

$$T = \begin{bmatrix} y_1 & y_2 \\ n_1 - y_1 & n_2 - y_2 \end{bmatrix}.$$

In the usual Bayesian analysis of such tables, it is assumed that y_1 and y_2 are independent with Binomial distribution

$$p(y_i) = \binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i - y_i}, \quad i \in \{1, 2\}.$$

Here, θ_1 and θ_2 are two hypothetical proportions, and the problem we will address is inference regarding their relative magnitude; in particular determining the probability that θ_1 is greater than θ_2 given the observed data, $p(\theta_1 > \theta_2 | T)$. With the assumption of separate Beta priors

$$p(\theta_i) = \frac{1}{\mathrm{B}(\alpha_i^0, \beta_i^0)} \theta_i^{\alpha_i^0 - 1} (1-\theta_i)^{\beta_i^0 - 1},$$

the independent posterior distributions of θ_1 and θ_2 are given by

$$p(\theta_i | y_i, n_i) = \frac{1}{\mathrm{B}(\alpha_i, \beta_i)} \theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1},$$

* Corresponding author.
E-mail addresses: mnsc@dtu.dk (M.N. Schmidt), mmor@dtu.dk (M. Mørup).

https://doi.org/10.1016/j.spl.2018.08.011
0167-7152/

where $\alpha_i = y_i + \alpha_i^0$ and $\beta_i = n_i - y_i + \beta_i^0$. We can write the joint posterior conditioned on the event $\theta_1 > \theta_2$ as

$$p(\theta_1, \theta_2 | T, \theta_1 > \theta_2) = \frac{{\theta_1}^{\alpha_1-1}(1-\theta_1)^{\beta_1-1}{\theta_2}^{\alpha_2-1}(1-\theta_2)^{\beta_2-1}}{Z(\alpha_1, \beta_1, \alpha_2, \beta_2)} \mathbb{I}[\theta_1 > \theta_2],$$

where the normalizing constant is given by

$$Z(\alpha_1, \beta_1, \alpha_2, \beta_2) = \int_0^1 \int_0^{\theta_1} {\theta_1}^{\alpha_1-1}(1-\theta_1)^{\beta_1-1}{\theta_2}^{\alpha_2-1}(1-\theta_2)^{\beta_2-1} \mathrm{d}\theta_2 \mathrm{d}\theta_1. \tag{1}$$

The posterior probability of the event $\theta_1 > \theta_2$ is then given by

$$p(\theta_1 > \theta_2 | T) = \frac{Z(\alpha_1, \beta_1, \alpha_2, \beta_2)}{\mathrm{B}(\alpha_1, \beta_1)\mathrm{B}(\alpha_2, \beta_2)}.$$

The practical computation of the integral in Eq. (1) is the focus of this paper. It is well known that it can be evaluated in terms of a generalized hypergeometric function (Altham, 1969; Latorre, 1985; Kawasaki and Miyaoka, 2012). To show this, we can write

$$\begin{aligned} Z(\alpha_1, \beta_1, \alpha_2, \beta_2) &= \int_0^1 \theta_1^{\alpha_1-1}(1-\theta_1)^{\beta_1-1} \left(\int_0^{\theta_1} \theta_2^{\alpha_2-1}(1-\theta_2)^{\beta_2-1}\, \mathrm{d}\theta_2 \right) \mathrm{d}\theta_1 \\ &= \int_0^1 \theta_1^{\alpha_1-1}(1-\theta_1)^{\beta_1-1} \mathrm{B}(\theta_1; \alpha_2, \beta_2)\, \mathrm{d}\theta_1 \\ &= \int_0^1 \theta_1^{\alpha_1-1}(1-\theta_1)^{\beta_1-1} \frac{\theta_1^{\alpha_2}}{\alpha_2} {}_2F_1\left[\begin{matrix} \alpha_2, & 1-\beta_2 \\ & \alpha_2+1 \end{matrix} ; \theta_1 \right] \mathrm{d}\theta_1 \\ &= \frac{\mathrm{B}(\beta_1, \alpha_1+\alpha_2)}{\alpha_2} {}_3F_2\left[\begin{matrix} 1-\beta_2, & \alpha_2, & \alpha_1+\alpha_2 \\ \alpha_2+1, & \alpha_1+\beta_1+\alpha_2 & \end{matrix} ; 1 \right] \end{aligned} \tag{2}$$

where $\mathrm{B}(\theta_1; \alpha_2, \beta_2)$ is the incomplete Beta function, ${}_pF_q$ is the generalized hypergeometric function, and where we have used (Bateman, 1954, 20.2.5) in the final step.

Although this provides an analytic expression for Z, this particular formulation is not optimal in terms of the computational properties of its series. Furthermore, while this expression is identical to the formulation derived by Kawasaki and Miyaoka (2012, Theorem 1), it does not coincide with other formulations derived in the literature (Altham, 1969; Latorre, 1985). This leads us to ask which formulation is most favorable from a practical, computational perspective, and whether there exist other equivalent but superior formulations.

1.1. Evaluating the generalized hypergeometric function

The ${}_3F_2$ generalized hypergeometric function is defined by the series

$${}_3F_2\left[\begin{matrix} a_1, & a_2, & a_3 \\ b_1, & b_2 & \end{matrix} ; z \right] = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k (a_3)_k}{(b_1)_k (b_2)_k} \frac{z^k}{k!}, \tag{3}$$

where $(x)_k$ denotes the Pochhammer rising factorial,

$$(x)_k = \frac{\Gamma(x+k)}{\Gamma(x)} = \begin{cases} 1, & \text{if } k = 0 \\ x(x+1)\cdots(x+k-1), & \text{if } k = 1, 2, \ldots \end{cases}$$

If the sequence decreases rapidly, the generalized hypergeometric function can be computed by truncating the summation when sufficient numerical accuracy has been reached, possibly adding an approximation of the remainder. If the sequence decreases slowly, this approach might not be practical, and if the sequence contains terms of opposite sign with similar magnitudes, catastrophic cancellations may lead to loss of numerical precision.

1.2. Equivalence relations

Examining the definition of Z in Eq. (1), we note the two following symmetries in its four arguments, which leads to a set of equivalence relations.

Proposition 1. *Z is invariant to the substitution $(\alpha_1, \alpha_2) \leftrightarrow (\beta_2, \beta_1)$ of its arguments,*

$$Z(\alpha_1, \beta_1, \alpha_2, \beta_2) = Z(\beta_2, \alpha_2, \beta_1, \alpha_1). \tag{4}$$

Proof. This can be shown by making the above substitution in Eq. (1), substituting the parameters $\theta_1 \leftrightarrow 1-\theta_2$, and changing the order of integration. □