Efficient computation for Bayesian comparison of two proportions

Mikkel N. Schmidt[a,*], Morten Mørup[a]

[a]*Section for Cognitive Systems, DTU Compute, Technical University of Denmark*

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.

Keywords: Bayesian analysis; Comparison of proportions; Integral of Beta distribution; Hypergeometric function.

1. Bayesian comparison of two proportions

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

$$T = \left[\begin{array}{cc} y_1 & y_2 \\ n_1 - y_1 & n_2 - y_2 \end{array} \right].$$

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

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

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

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

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

$$= \int_0^1 \theta_1^{\alpha_1 - 1}(1 - \theta_1)^{\beta_1 - 1}\mathrm{B}(\theta_1; \alpha_2, \beta_2)\,\mathrm{d}\theta_1$$

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

$$= \frac{\mathrm{B}(\beta_1, \alpha_1 + \alpha_2)}{\alpha_2}{}_3F_2\left[\begin{matrix}1 - \beta_2, & \alpha_2, & \alpha_1 + \alpha_2 \\ \alpha_2 + 1, & \alpha_1 + \beta_1 + \alpha_2 \end{matrix}; 1\right] \qquad (2)$$

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 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

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

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

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

2

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. $\qquad \square$

Proposition 2. *(Aitchison and Bacon-Shone, 1981). Z can equivalently be written as*

$$Z(\alpha_1, \beta_1, \alpha_2, \beta_2) = \mathrm{B}(\alpha_1, \beta_1)\mathrm{B}(\alpha_2, \beta_2) - Z(\alpha_2, \beta_2, \alpha_1, \beta_1).$$

Proof. This follows directly from $p(\theta_1 > \theta_2 | T) = 1 - p(\theta_1 < \theta_2 | T)$. $\qquad \square$

Applying these two identities to the expression in Eq. (2) leads to four equivalent formulations of Z with different computational properties in terms of their series expansion.

Thomae (1879, Art. 4) shows ten equivalent, but in terms of convergence properties different, parameter transformations for the $_3F_2$ generalized hypergeometric function evaluated at $z = 1$. Applying Thomae's relations to the expression for Z in Eq. (2) yields the ten expressions listed in Table 1, eight of which are convergent. Thomae's relations encompass the relation in Proposition 1, and consist of four pairs of expressions related by the parameter substitution $(\alpha_1, \alpha_2) \leftrightarrow (\beta_2, \beta_1)$ and two expressions that are invariant to this substitution.

The eight convergent series are illustrated in Fig. 1, which shows for nine examples of 2×2 tables the relationship between the number of terms and the relative truncation error for the partial sums as well as the values of the terms. While all series converge to the same solution given enough terms, there is a substantial difference in how quickly the series converge, and the plot of the series terms shows that many of the series are alternating with terms of large magnitude, which can lead to numerical imprecision when computing their partial sum.

1.3. Expression of special interest

Obtained through these equivalence relations, one expression (Table 1, series a)

$$Z(\alpha_1, \beta_1, \alpha_2, \beta_2) = \frac{\Gamma(\alpha_1 + \alpha_2)\Gamma(\beta_1 + \beta_2)}{\beta_1 \alpha_2 \Gamma(\alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 1)} {}_3F_2 \left[\begin{matrix} 1, & 1 - \alpha_1, & 1 - \beta_2 \\ & \beta_1 + 1, & \alpha_2 + 1 \end{matrix} ; 1 \right], \tag{5}$$

is especially interesting for several reasons. Examining the generalized hypergeometric term, in the context of its series expansion in Eq. (3), we make the following remarks:

a	$\dfrac{\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_1+\beta_2)}{\beta_1\alpha_2\Gamma(\alpha_1+\beta_1+\alpha_2+\beta_2-1)}{}_3F_2\left[\begin{matrix}1,\ 1-\alpha_1,\ 1-\beta_2\\ \beta_1+1,\ \alpha_2+1\end{matrix};1\right]$
b	$\dfrac{\mathrm{B}(\alpha_1+\alpha_2,\beta_1+\beta_2)}{\beta_1}{}_3F_2\left[\begin{matrix}1,\ \alpha_1+\beta_1,\ \beta_1+\beta_2\\ \beta_1+1,\ \alpha_1+\beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
c	$\dfrac{\mathrm{B}(\alpha_1+\alpha_2,\beta_1+\beta_2)}{\alpha_2}{}_3F_2\left[\begin{matrix}1,\ \alpha_1+\alpha_2,\ \alpha_2+\beta_2\\ \alpha_2+1,\ \alpha_1+\beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
d	$\dfrac{\Gamma(\alpha_2)\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_1+\beta_2)\Gamma(\beta_1)}{\Gamma(\alpha_1+\beta_1+\alpha_2)\Gamma(\beta_1+\alpha_2+\beta_2)}{}_3F_2\left[\begin{matrix}\beta_1,\ \alpha_2,\ \alpha_1+\beta_1+\alpha_2+\beta_2-1\\ \alpha_1+\beta_1+\alpha_2,\ \beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
e	$\dfrac{\mathrm{B}(\alpha_2,\beta_1+\beta_2)}{\beta_1}{}_3F_2\left[\begin{matrix}1-\alpha_1,\ \beta_1,\ \beta_1+\beta_2\\ \beta_1+1,\ \beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
f	$\dfrac{\mathrm{B}(\beta_1,\alpha_1+\alpha_2)}{\alpha_2}{}_3F_2\left[\begin{matrix}1-\beta_2,\ \alpha_2,\ \alpha_1+\alpha_2\\ \alpha_2+1,\ \alpha_1+\beta_1+\alpha_2\end{matrix};1\right]$
g	$\dfrac{\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_1+\beta_2)\Gamma(\beta_1)}{\alpha_2\Gamma(\alpha_1+\beta_1)\Gamma(\beta_1+\alpha_2+\beta_2)}{}_3F_2\left[\begin{matrix}1-\alpha_1,\ \alpha_2,\ \alpha_2+\beta_2\\ \alpha_2+1,\ \beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
h	$\dfrac{\Gamma(\alpha_2)\Gamma(\alpha_1+\alpha_2)\Gamma(\beta_1+\beta_2)}{\beta_1\Gamma(\alpha_2+\beta_2)\Gamma(\alpha_1+\beta_1+\alpha_2)}{}_3F_2\left[\begin{matrix}1-\beta_2,\ \beta_1,\ \alpha_1+\beta_1\\ \beta_1+1,\ \alpha_1+\beta_1+\alpha_2\end{matrix};1\right]$
i	$\dfrac{\Gamma(\beta_1)\Gamma(\alpha_2)\mathrm{B}(\alpha_1+\alpha_2,\beta_1+\beta_2)}{\Gamma(1-\beta_2)\Gamma(\beta_1+\alpha_2+\beta_2)}{}_3F_2\left[\begin{matrix}\beta_1+\beta_2,\ \alpha_2+\beta_2,\ \alpha_1+\beta_1+\alpha_2+\beta_2-1\\ \beta_1+\alpha_2+\beta_2,\ \alpha_1+\beta_1+\alpha_2+\beta_2\end{matrix};1\right]$
j	$\dfrac{\Gamma(\beta_1)\Gamma(\alpha_2)\mathrm{B}(\alpha_1+\alpha_2,\beta_1+\beta_2)}{\Gamma(1-\alpha_1)\Gamma(\alpha_1+\beta_1+\alpha_2)}{}_3F_2\left[\begin{matrix}\alpha_1+\beta_1,\ \alpha_1+\alpha_2,\ \alpha_1+\beta_1+\alpha_2+\beta_2-1\\ \alpha_1+\beta_1+\alpha_2,\ \alpha_1+\beta_1+\alpha_2+\beta_2\end{matrix};1\right]$

Table 1: Ten equivalent expressions for Z defined in Eq. (1) found by applying Thomae's relations as listed by Milgram (2010, Appendix A). The expressions are listed in arbitrary order, but grouped in four pairs of expressions (b,c; e,f; g,h; i,j) which are related by the parameter substitution $(\alpha_1,\beta_1)\leftrightarrow(\beta_2,\alpha_2)$, and two single expressions (a, d) that are invariant under this substitution. The generalized hypergeometric series in i and j are not convergent. The expression in Eq. (2) from which we started appears here a series f.

Remark 1. *The parameter $a_1 = 1$ in the numerator cancels with the term $k!$ in the denominator, and thus the series reduces to*

$$_3F_2\left[\begin{matrix} 1, & a_2, & a_3 \\ & b_1, & b_2 \end{matrix}; 1\right] = \sum_{k=0}^{\infty} \frac{(a_2)_k (a_3)_k}{(b_1)_k (b_2)_k}. \tag{6}$$

Remark 2. *The two remaining parameters in the numerator are negative,*

$$a_2 = 1 - \alpha_1 = 1 - y_1 - a_1^0, \qquad a_3 = 1 - \beta_2 = 1 - (n_2 - y_2) - b_2^0,$$

except in the atypical case when the table has a zero entry and the corresponding parameter of the prior is less than one. Therefore, the ratio of consecutive terms in the numerator will, in the typical case, initially be absolutely decreasing. Since the parameters in the denominator,

$$b_1 = \beta_1 + 1 = n_1 - y_1 + \beta_1^0, \qquad b_2 = \alpha_2 + 1 = y_2 + \alpha_2^0,$$

are always positive, the ratio of consecutive terms in the denominator is increasing. Amongst the ten expressions in Table 1, series a is the only one with this property, which in part explains its superior convergence properties.

Remark 3. *If α_1 or β_2 is a positive integer, the corresponding rising factorial in the numerator will eventually reach zero, terminating the series. Thus, in this case the series is finite and the number of terms is equal to the smaller integer of the two, $\min(\{\alpha_1, \beta_2\} \cap \mathbb{Z}^+)$. Using Proposition 2 we note that $Z(\alpha_1, \beta_1, \alpha_2, \beta_2)$ has a finite series if it has at least one positive integer parameter, and the smallest attainable number of terms is given by the smallest positive integer parameter, $\min(\{\alpha_1, \beta_1, \alpha_2, \beta_2\} \cap \mathbb{Z}^+)$.*

Proposition 3. *If $(1 - \alpha_1)(1 - \beta_2) < (1 + \alpha_2)(1 + \beta_1)$ and $\alpha_1, \beta_2 > 1$ the series in Eq. (6) is initially decreasing in magnitude. The series is positive for $k < k_\wedge$, alternating for $k_\wedge < k < k_\vee$ and positive or negative for $k > k_\vee$, where $k_\wedge = \min(\alpha_1, \beta_2)$ and $k_\vee = \max(\alpha_1, \beta_2)$.*
Proof. Let z_k denote the k'th term of the series given recursively by

$$z_{k+1} = z_k \frac{(1 - \alpha_1 + k)(1 - \beta_2 + k)}{(1 + \beta_1 + k)(1 + \alpha_2 + k)}.$$

We have $z_0 = 1$, and $|z_1| < |z_0|$ follows from the requirement $(1 - \alpha_1)(1 - \beta_2) < (1 + \alpha_2)(1 + \beta_1)$. The denominator in the recursive formulation is always positive, and when $k < k_\wedge$, both parenthesized terms in the numerator are negative, thus the series is initially positive. When $k_\wedge < k < k_\vee$, the two parenthesized terms in the numerator will have different sign, so the series is alternating. Finally, when $k > k_\vee$ both terms in the numerator are positive, and the series then remains either positive or negative. $\square$

Remark 4. *In the case that the requirement $(1 - \alpha_1)(1 - \beta_2) < (1 + \alpha_2)(1 + \beta_1)$ in Proposition 3 is not satisfied, the identity in Proposition 2 can be invoked to make the substitution $(\alpha_1, \beta_1) \leftrightarrow (\alpha_2, \beta_2)$ such that the result holds for any table.*

Proof. Noting that α_i and β_i are positive, we have

$$(1 - \alpha_1)(1 - \beta_2) \geq (1 + \alpha_2)(1 + \beta_1) \qquad\qquad \Leftrightarrow$$
$$(1 + \alpha_1)(1 + \beta_2) - 2(\alpha_1 + \beta_2) \geq (1 - \alpha_2)(1 - \beta_1) + 2(\alpha_2 + \beta_1) \qquad \Rightarrow$$
$$(1 + \alpha_1)(1 + \beta_2) \geq (1 - \alpha_2)(1 - \beta_1).$$

□

2. Related results

Altham (1969) shows that $p(\theta_1 > \theta_2|T)$ corresponds to Fisher's exact test (FET) for the table

$$\begin{bmatrix} \alpha_1 - 1 & \alpha_2 \\ \beta_1 & \beta_2 - 1 \end{bmatrix}.$$

Hence FET coincides with the Bayesian calculation under the very conservative priors $\theta_1 \sim$ Beta$(1, 0)$ and $\theta_2 \sim$ Beta$(0, 1)$. Thus, the results derived here also apply to the computation of FET with a generalization to non-integer parameters.

Altham (1969) derives a related result for the exact Bayesian posterior probability of negative association in comparison of two proportions in the analysis of a 2×2 table. In our notation, Altham's result (1969, Eq. 2) can be written as

$$\frac{\Gamma(\alpha_1 + \alpha_2)\Gamma(\beta_1 + \beta_2)}{\alpha_2(\alpha_1 + \beta_1 - 1)\Gamma(\alpha_1 + \beta_1 + \alpha_2 + \beta_2 - 1)}{}_3F_2\begin{bmatrix} 1, & 1 - \alpha_1, & \alpha_2 + \beta_2 \\ & \alpha_2 + 1, & 2 - \alpha_1 - \beta_1 \end{bmatrix};1\begin{bmatrix} \end{bmatrix}.$$

This expression does not coincide with any of the expressions derived in this work, and is only valid for integer α_1.

Aitchison and Bacon-Shone (1981) derive an exact expression for the posterior distribution of the risk ratio $p\left(\beta = \frac{\theta_1}{\theta_2}\right)$. Latorre (1985) derives an exact expression for the density of the odds ratio, $p\left(\lambda = \frac{\theta_1}{1-\theta_1}\frac{1-\theta_2}{\theta_2}\right)$, and the distribution function $p(\lambda < t)$. His expression for the special case $t = 1$ (Latorre, 1985, Sec. 4. a) corresponding to $p(\theta_1 < \theta_2|T)$ is identical to our series c in Table 1. Nurminen and Mutanen (1987) extend these results to a general comparative parameter beyond odds ratio, risk difference, and risk ratio. Nadarajah and Kotz (2007) derive the exact distribution for the difference of two proportions $p(\theta_2 - \theta_1)$. Similarly, Kawasaki and Miyaoka (2010) derive the distribution for the difference of proportions as well as the highest posterior density credible interval. Kawasaki and Miyaoka (2012) derive an exact result for $p(\theta_1 < \theta_2|T)$ for two proportions, and their result (2012, Theorem 1) coincides with our series f in Table 1.

3. Discussion

In the comparison of two proportions, the exact Bayesian calculations involve the evaluation of a particular generalized hypergeometric function. Using Thomae's equivalence relations, we have shown that several similar formulations can be arrived at, and that one particular formulation which has not previously been discussed in the literature has superior computational properties in terms of convergence of its series expansion.

6

References

Aitchison, J. and Bacon-Shone, J. (1981) Bayesian risk ratio analysis. *The American Statistician*, **35**, 254–257.

Altham, P. M. E. (1969) Exact Bayesian Analysis of a 2×2 Contingency Table, and Fisher's "Exact" Significance Test. *Journal of the Royal Statistical Society: Series B*, **31**, 261–269.

Bateman, H. (1954) *Tables of integral transforms*, vol. II. McGraw-Hill Book Company, New York.

Kawasaki, Y. and Miyaoka, E. (2010) A posterior density for the difference between two binominal proportions and the highest posterior density credible interval. *Journal of the Japan Statistical Society*, **40**, 265–275.

— (2012) A Bayesian Inference of $P(\pi_1 > \pi_2)$ for Two Proportions. *Journal of biopharmaceutical statistics*, **22**, 425–37.

Latorre, G. (1985) Bayesian comparison of two proportions. *Rivista di Statistica Applicata*, **18**, 3–11.

Milgram, M. (2010) On Hypergeometric 3F2(1) A Review. *ArXiv preprint arXiv:1011.4546*.

Nadarajah, S. and Kotz, S. (2007) Statistical distribution of the difference of two proportions. *Statistics in medicine*, **26**, 3518–3523.

Nurminen, M. and Mutanen, P. (1987) Exact Bayesian analysis of two proportions. *Scandinavian Journal of Statistics*, **14**, 67–77.

Thomae, J. (1879) Ueber die Functionen, welche durch Reihen von der Form dargestellt verden $1 + \frac{p}{1}\frac{p'}{q'}\frac{p''}{q''} + \frac{p}{1}\frac{p+1}{2}\frac{p'}{q'}\frac{p'+1}{q'+1}\frac{p''}{q''}\frac{p''+1}{q''+1} + \cdots$. *Journal für die reine und angewandte Mathematik*, **87**, 26–73.
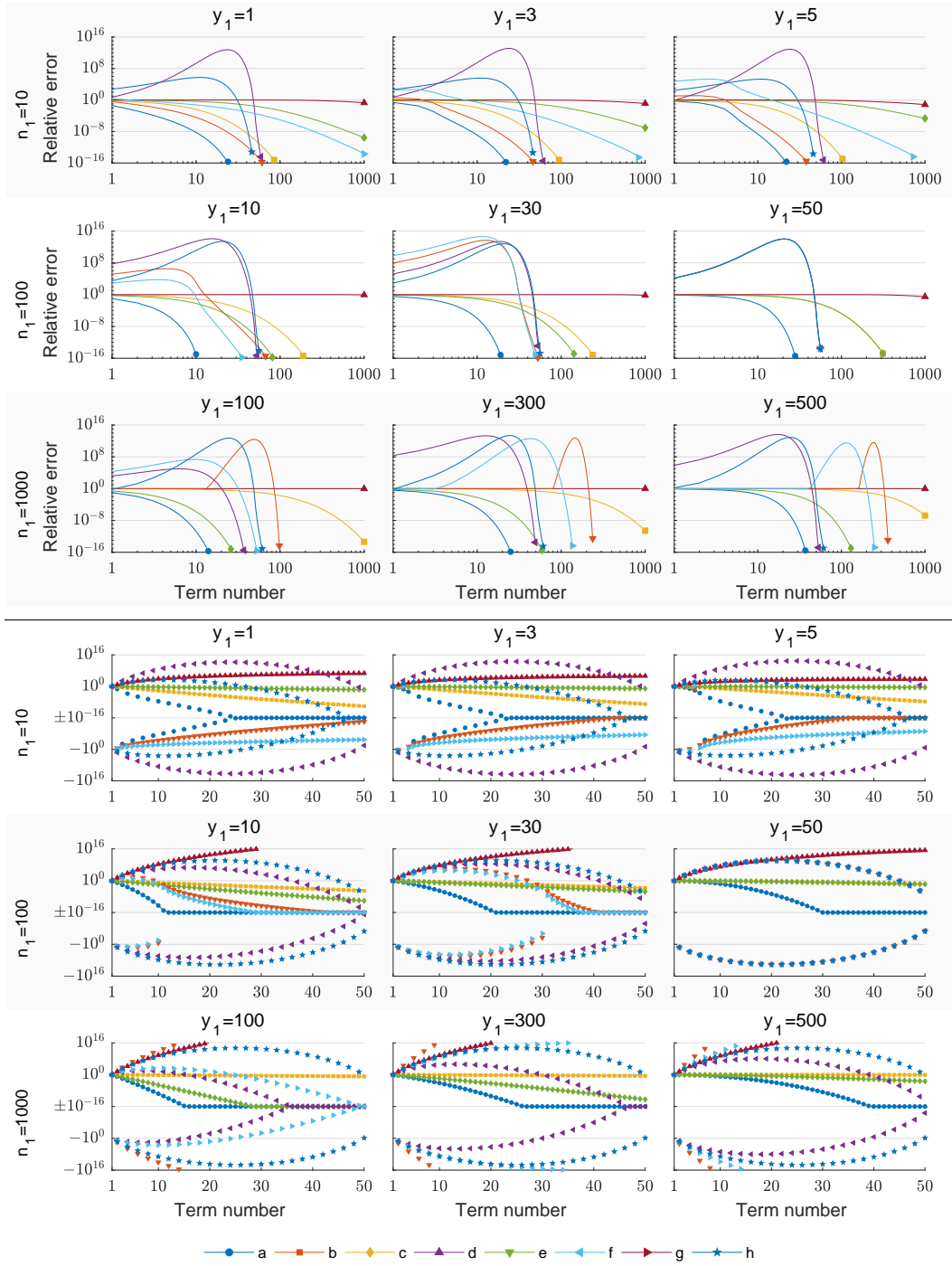
Figure 1: Relative absolute truncation error (top three panels) and values of series terms (bottom three panels) for nine examples of tables: n_1, y_1 are varied, and $n_2 = 100$, $y_2 = 50$, $\alpha_i^0 = \beta_i^0 = 0.5$, for the eight convergent series a–e in Table 1. In all examples series a converges most quickly.