# Bayes factors for research workers

Ly, A.

**Publication date**
2018

**Document Version**
Other version

**License**
Other

*Chapter 11*

# Analytic Posteriors for the Binomial Rate Parameters, and the Odds Ratio

**Abstract**

We present analytic posteriors for a binomial rate parameter and the odds ratio. Both expressions involve hypergeometric functions and can be used to derive Bayes factors for these scenarios.

*Keywords:* Bayesian inference, hypergeometric functions.

## 11.1 Introduction

This chapter contains derivations of analytic posteriors for the rate of a binomial distribution and the odds ratio.

## 11.2 Binomial distribution

### 11.2.1 A localised prior for the binomial rate parameter

**Definition 11.2.1** (Localised beta prior)**.** We say that $\theta$ has a beta distribution localised at $\theta_0$ if its density is given by

$$\pi_\eta(\theta) = \underbrace{\tfrac{1}{\mathcal{B}(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}_{\text{Beta}(\theta\,;\,\alpha,\beta)}\left(\tfrac{1-\theta_0}{\theta_0}\right)^{\alpha}\left(1-\left[2-\tfrac{1}{\theta_0}\right]\theta\right)^{-(\alpha+\beta)}, \tag{11.2.1}$$

where $\eta$ is shorthand for the parameter vector $\eta = (\alpha, \beta, \theta_0)$ and where $\text{Beta}(\theta\,;\,\alpha,\beta)$ refers to the (standard) two-parameter beta distribution. ◇

With $\theta_0 = 1/2$ we retrieve the (standard) beta $\text{Beta}(\theta\,;\,\alpha,\beta)$. We choose to write the last term as $[1-\theta+\theta(\frac{1-\theta_0}{\theta_0})] = (1-[2-\frac{1}{\theta_0}]\theta)$ due to its relation with the hypergeometric function.

**Theorem 11.2.1** (Marginal likelihood of a binomially distributed random variable with the beta prior localised at $\theta_0$)**.** *The localised beta prior has the following marginal likelihood*

$$p_\eta(d) = \binom{n}{y} \frac{\mathcal{B}(\alpha + y, \beta + n - y)}{\mathcal{B}(\alpha, \beta)} \times (\tfrac{1-\theta_0}{\theta_0})^{\alpha} {}_2F_1(\alpha + \beta, \alpha + y\,;\, \alpha + \beta + n\,;\, 2 - \tfrac{1}{\theta_0}), \tag{11.2.2}$$

*where $d$ refers to the data $y$ and $n$ and where*

$$_2F_1(u, v\,;\, w\,;\, z) = \sum_{k=0}^{\infty} \frac{(u)_k (v)_k}{(w)_k k!} z^k, \tag{11.2.3}$$

*is Gauss' hypergeometric function (Oberhettinger, 1972, Section 15), where $(u)_k = \frac{\Gamma(u+k)}{\Gamma(u)}$ denotes Pochhammer's raising factorial.* ◇

*Proof.* Writing $p_\eta(\emptyset) = \mathcal{B}(\alpha, \beta)(\frac{1-\theta_0}{\theta_0})^{-\alpha}$ for the normalisation constant of the prior combined with $u_1 = y + \alpha$, $u_2 = n - y + \beta$, $v = \alpha + \beta$, and by definition of the marginal likelihood, we have

$$p_\eta(\emptyset) p_\eta(d) = \binom{n}{y} \int_0^1 \theta^{u_1 - 1} (1 - \theta)^{u_2 - 1} (1 - \theta[2 - \tfrac{1}{\theta_0}])^{-v} \mathrm{d}\theta, \tag{11.2.4}$$

$$= \binom{n}{y} \mathcal{B}(u_1, u_2) {}_2F_1(v, u_1\,;\, u_1 + u_2\,;\, 2 - \tfrac{1}{\theta_0}). \tag{11.2.5}$$

The last equality follows from Euler's integral representation of the hypergeometric function (Abramowitz and Stegun, 1964, p. 558). □

**Corollary 11.2.1** (Localised beta posterior and its characterisation)**.** *The posterior is*

$$\pi_\eta(\theta \,|\, d) = \pi_{\alpha,\beta}(\theta \,|\, d) \frac{(1 - [2 - \frac{1}{\theta_0}]\theta)^{-(\alpha+\beta)}}{{}_2F_1(\alpha + \beta, \alpha + y\,;\, \alpha + \beta + n\,;\, 2 - \frac{1}{\theta_0})}, \tag{11.2.6}$$

*where*

$$\pi_{\alpha,\beta}(\theta \,|\, d) = \frac{\theta^{y + \alpha - 1} (1 - \theta)^{n - y + \beta - 1}}{\mathcal{B}(\alpha + y, \beta + n - y)}, \tag{11.2.7}$$

*is the posterior based on the standard beta prior. The last term of the localised posterior $\pi_\eta(\theta \,|\, d)$ can be thought of as a "skewness" term due to the localisation. The $k$th posterior moment is*

$$E_\eta(\theta^k \,|\, d) = \frac{\mathcal{B}(\alpha + y + k, \beta + n - y)}{\mathcal{B}(\alpha + y, \beta + n - y)} \times \frac{{}_2F_1(\alpha + \beta, \alpha + y + k\,;\, \alpha + \beta + n + k\,;\, 2 - \frac{1}{\theta_0})}{{}_2F_1(\alpha + \beta, \alpha + y\,;\, \alpha + \beta + n\,;\, 2 - \frac{1}{\theta_0})}, \tag{11.2.8}$$

$$= \frac{(\alpha + y)_k}{(\alpha + \beta + n)_k} \frac{{}_2F_1(\alpha + \beta, \alpha + y + k\,;\, \alpha + \beta + n + k\,;\, 2 - \frac{1}{\theta_0})}{{}_2F_1(\alpha + \beta, \alpha + y\,;\, \alpha + \beta + n\,;\, 2 - \frac{1}{\theta_0})},$$

*where $E_\eta(\cdot \,|\, d)$ refers to the expectation with respect to the posterior $\pi_\eta(\theta \,|\, d)$.* ◇

*Proof.* The statements follow directly from the proof given above with $u_1 = y + \alpha + k$. □

**Remark 11.2.1.** *The normalisation constant $p_\eta(\emptyset)$ can be retrieved from $p_\eta(d)$ by taking $y = n = 0$, that is, the normalisation constant can also be expressed as a hypergeometric function. Consequently, the localised prior can be viewed as a partially conjugate prior for the binomial distribution by which we mean that the prior and posterior are of the same form, but that only some of its parameters need updating. More specifically, to update the prior to a posterior, only the exponents of $\theta$ and $(1-\theta)$ need to be changed, and, subsequently, one of the upper and the lower terms of the hypergeometric function ${}_2F_1$.* ◇

*Proof.* The definition of the normalisation constant of the prior and Eq. (11.2.5) implies that

$$p_\eta(\emptyset) = \int \theta^{\alpha-1}(1-\theta)^{\beta-1}(1-\theta[2-\tfrac{1}{\theta_0}])^{-(\alpha+\beta)}\mathrm{d}\theta, \tag{11.2.9}$$

$$= \mathcal{B}(\alpha,\beta)_2F_1(\alpha,\alpha+\beta\,;\,\alpha+\beta\,;\,2-\tfrac{1}{\theta_0}), \tag{11.2.10}$$

which should suffice for the proof. Note that one of the upper and the lower term of the hypergeometric function are the same, which implies that it can be simplified. Indeed, by Eq. (15.4.6) of Olver et al. (2010, p. 386), or just the definition of Gauss' hypergeometric function, we have

$${}_2F_1(\alpha,\alpha+\beta\,;\,\alpha+\beta\,;\,2-\tfrac{1}{\theta_0}) = (1-[2-\tfrac{1}{\theta_0}])^{-\alpha} = (\tfrac{1-\theta_0}{\theta_0})^{-\alpha}, \tag{11.2.11}$$

which completes the proof. □

**Corollary 11.2.2** (Min-sided prior, marginal likelihood, posterior and its characterisation)**.** *By construction, the prior associated to the min-sided hypothesis $\mathcal{H}_- : \theta \in (0, \theta_0)$ is*

$$\pi_\eta^{(-)}(\theta) = \frac{(\frac{1-\theta_0}{\theta_0})^\alpha}{\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}(1-\theta[2-\tfrac{1}{\theta_0}])^{-(\alpha+\beta)}\mathbf{1}_{(0,\theta_0]}(\theta), \tag{11.2.12}$$

*where $\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)$ is the incomplete beta integral evaluated at a half. The min-sided marginal likelihood is*

$$p_\eta^{(-)}(d) = \frac{\theta_0^y(1-\theta_0)^\alpha}{\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)(y+\alpha)}\binom{n}{y} \tag{11.2.13}$$
$$\times AF_1(y+\alpha\,;\,1-\beta-n+y,\alpha+\beta\,;\,y+\alpha+1\,;\,\theta_0,2\theta_0-1),$$

*where*

$$AF_1(u\,;\,v_1,v_2\,;\,w\,;\,x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(u)_{m+n}(v_1)_m(v_2)_n}{m!\,n!\,(w)_{m+n}}x^my^n, \tag{11.2.14}$$

*is known as Appell's hypergeometric function of the first kind. As such, the min-sided posterior is*

$$\pi_\eta^{(-)}(\theta \,|\, d) = \frac{y+\alpha}{\theta_0^{\alpha+y}} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}(1-[2-\tfrac{1}{\theta_0}]\theta)^{-(\alpha+\beta)}\mathbf{1}_{(0,\theta_0]}(\theta)$$
$$\Big/ AF_1(y+\alpha\,;\, 1-\beta-n+y, \alpha+\beta\,;\, y+\alpha+1\,;\, \theta_0, 2\theta_0-1). \quad (11.2.15)$$

*Lastly, the kth posterior moment is*

$$E_\eta^{(-)}(\theta^k \,|\, d) = \frac{y+\alpha}{y+\alpha+k}\theta_0^k \quad (11.2.16)$$
$$\times \frac{AF_1(y+\alpha+k\,;\, 1-\beta-n+y, \alpha+\beta\,;\, y+\alpha+1+k\,;\, \theta_0, 2\theta_0-1)}{AF_1(y+\alpha\,;\, 1-\beta-n+y, \alpha+\beta\,;\, y+\alpha+1\,;\, \theta_0, 2\theta_0-1)}$$

*where $E_\eta^{(-)}(\cdot \,|\, d)$ is the expectation with respect to the min-sided posterior.* ◇

*Proof.* To simplify matters we write $p_\eta^{(-)}(\emptyset) = \mathcal{B}(\frac{1}{2}\,;\, \alpha, \beta)(\frac{1-\theta_0}{\theta_0})^{-\alpha}$ for the normalisation constant of the prior. The integral of interest is then of the form

$$p_\eta^{(-)}(\emptyset)p_\eta^{(-)}(d) = \binom{n}{y}\int_0^{\theta_0} \theta^{u-1}(1-\theta)^{-v_1}(1-[2-\tfrac{1}{\theta_0}]\theta)^{-v_2}\mathrm{d}\theta, \quad (11.2.17)$$

where $u = y+\alpha+k$, $v_1 = 1-n+y-\beta$, $v_2 = \alpha+\beta$. Using the change of variable $t = \theta/\theta_0$, thus, $\int \mathrm{d}\theta = \int \theta_0 \mathrm{d}t$ we can rewrite the integral as

$$p_\eta^{(-)}(\emptyset)p_\eta^{(-)}(d) = \binom{n}{y}\theta_0^u \int_0^1 t^{u-1}(1-\theta_0 t)^{-v_1}(1-[2\theta_0-1]t)^{-v_2}\mathrm{d}t, \quad (11.2.18)$$
$$= \binom{n}{y}\frac{\theta_0^u}{u}AF_1(u\,;\, v_1, v_2\,;\, u+1\,;\, \theta_0, 2\theta_0-1), \quad (11.2.19)$$

where the latter equality is an (Euler) integral representation of Appell's hypergeometric function due to $u = y+\alpha+k$ being positive, see Eq. (3.211) of Gradshteyn and Ryzhik (2007, p. 318) and Bailey (1964, p. 77). Entering the terms $u, v_1, v_2$ with $k = 0$ yields the marginal likelihood, and, subsequently, the posterior and the posterior moments. □

**Remark 11.2.2.** *The normalisation constant $p_\eta(\emptyset)$ can be written as an Appell function $AF_1$, which implies that the min-sided localised prior can be thought of as a partially conjugate prior for the binomial distribution by which we mean that the prior and posterior are of the same form, but that only some of its parameters are updated. More specifically, to update the prior, only the exponents of the $\theta$ and $(1-\theta)$ terms need updating. Similarly, only the Pochhammer coefficients in the Appell series need to be updated for the normalisation constant of the posterior.* ◇

*Proof.* By definition of the normalisation constant of the min-sided prior, the transformation $t = \frac{\theta}{\theta_0}$ and Eq. (11.2.19) we have

$$p_\eta^{(-)}(\emptyset) = \int_0^{\theta_0} \theta^{\alpha-1}(1-\theta)^{\beta-1}(1-[2-\tfrac{1}{\theta_0}])^{-(\alpha+\beta)}\mathrm{d}\theta \quad (11.2.20)$$
$$= \frac{\theta_0^\alpha}{\alpha}AF_1(\alpha\,;\, 1-\beta, \alpha+\beta\,;\, \alpha+1\,;\, \theta_0, 2\theta_0-1). \quad (11.2.21)$$

This should suffice for the statement. As a sanity check we note that the lower term of this Appell function is the sum of two of its upper terms, that is, $\alpha + 1 = \alpha + \beta + 1 - \beta$, which allows us to use Eq. (16.16.1) of Olver et al. (2010, p. 414) resulting in

$$p_\eta^{(-)}(\emptyset) = \tfrac{\theta_0^\alpha}{\alpha}[1 - (2\theta_0 - 1)]^{-\alpha} {}_2F_1(\alpha, 1-\beta\,;\,\alpha+1\,;\,\tfrac{\theta_0 - 2\theta_0 + 1}{1 - 2\theta_0 + 1}), \tag{11.2.22}$$

$$= (\tfrac{1-\theta_0}{\theta_0})^\alpha \underbrace{\tfrac{2^{-\alpha}}{\alpha} {}_2F_1(\alpha, 1-\beta\,;\,\alpha+1\,;\,\tfrac{1}{2})}_{\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)}, \tag{11.2.23}$$

due to the relation between the incomplete beta function and Gauss' hypergeometric function, that is, Eq. (8.17.7) of Olver et al. (2010, p. 183). □

**Corollary 11.2.3** (Plus-sided prior, marginal likelihood and posterior)**.** *By construction the prior associated to the plus-sided hypothesis $\mathcal{H}_+ : \theta \in (\theta_0, 1)$ is*

$$\pi_\eta^{(+)}(\theta) = \frac{1}{p^{(+)}(\emptyset)} \theta^{\alpha-1}(1-\theta)^{\beta-1}(1 - \theta[2 - \tfrac{1}{\theta_0}])^{-(\alpha+\beta)} \mathbf{1}_{(\theta_0,1]}(\theta), \tag{11.2.24}$$

*where $p_\eta^{(+)}(\emptyset) = [\mathcal{B}(\alpha,\beta) - \mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)](\frac{1-\theta_0}{\theta_0})^{-\alpha}$ denotes the normalisation constant of the prior. The plus-sided marginal likelihood is*

$$p_\eta^{(+)}(d) = \binom{n}{y} \frac{\theta_0^{y-1}(1-\theta_0)^{n-y}}{2^{\alpha+\beta}[\mathcal{B}(\alpha,\beta) - \mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)]} (n - y + \beta)^{-1} \tag{11.2.25}$$
$$\times AF_1(1\,;\,1-\alpha-y, \alpha+\beta\,;\,\beta+n-y+1\,;\,\tfrac{\theta_0-1}{\theta_0}, \tfrac{2\theta_0-1}{2\theta_0}).$$

*As such, the plus-sided posterior is*

$$\pi_\eta^{(+)}(\theta \,|\, d) = \frac{2^{\alpha+\beta}(n-y+\beta)}{\theta_0^{y+\alpha-1}(1-\theta_0)^{n-y-\alpha}} \tag{11.2.26}$$
$$\times \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}(1 - \theta[2 - \tfrac{1}{\theta_0}])^{-(\alpha+\beta)} 1_{(\theta_0,1]}(\theta)$$
$$\Big/ AF_1(1\,;\,1-y-\alpha, n-y+\beta-1\,;\,\tfrac{\theta_0-1}{\theta_0}, \tfrac{2\theta_0-1}{,} 2\theta_0).$$

*Lastly, the kth posterior moment is*

$$E_\eta^{(+)}(\theta^k \,|\, d) = \theta_0^k \frac{AF_1(1\,;\,1-y-k-\alpha, \alpha+\beta\,;\,n-y+\beta+1\,;\,\frac{\theta_0-1}{\theta_0}, \frac{2\theta_0-1}{2\theta_0})}{AF_1(1\,;\,1-y-\alpha, \alpha+\beta\,;\,n-y+\beta+1\,;\,\frac{\theta_0-1}{\theta_0}, \frac{2\theta_0-1}{2\theta_0})}, \tag{11.2.27}$$

*where $E_\eta^{(+)}(\cdot \,|\, d)$ is the expectation with respect to the plus-sided posterior.* ◇

*Proof.* To simplify matters we write $p_\eta^{(+)}(\emptyset) = [\mathcal{B}(\alpha,\beta) - \mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)](\frac{1-\theta_0}{\theta_0})^{-\alpha}$ for the normalisation constant of the prior. With $v_1 = 1 - y - k - \alpha$, $w_1 = n - y + \beta$, $v_2 = \alpha + \beta$ the integral of interest is then

$$p_\eta^{(+)}(\emptyset) p_\eta^{(+)}(d) = \binom{n}{y} \int_{\theta_0}^1 \theta^{-v_1}(1-\theta)^{w_1-1}(1 - [2 - \tfrac{1}{\theta_0}]\theta)^{-v_2}. \tag{11.2.28}$$

Using the change of variable $x = (\theta - \theta_0)/(1 - \theta_0)$, thus, $\int \mathrm{d}\theta = \int (1 - \theta_0)\mathrm{d}x$ this integral is then

$$p_\eta^{(+)}(\emptyset)p_\eta^{(+)}(d) = \binom{n}{y} 2^{-v_2}\theta_0^{-v_1}(1-\theta_0)^{w_1 - v_2} \tag{11.2.29}$$
$$\times \int_0^1 (1-x)^{w_1-1}(1 - [\tfrac{\theta_0 - 1}{\theta_0}]x)^{-v_1}(1 - [\tfrac{2\theta_0 - 1}{2\theta_0}]x)^{-v_2}\mathrm{d}x,$$
$$= \binom{n}{y} 2^{-v_2}\theta_0^{-v_1}(1-\theta_0)^{w_1 - v_2} w_1^{-1} \tag{11.2.30}$$
$$\times AF_1(1\,;\, v_1, v_2\,;\, w_1 + 1\,;\, \tfrac{\theta_0 - 1}{\theta_0}, \tfrac{2\theta_0 - 1}{2\theta_0}),$$

where the latter equality is an (Euler) integral representation of Appell's hypergeometric function due to $w_1 = n - y + \beta > 0$. The normalisation constant $p_\eta^{(+)}(\emptyset)$ follows from setting $n = y = k = 0$, the marginal likelihood follows from setting $k = 0$. □

**Corollary 11.2.4** (The two-sided Bayes factor and its relationship to the one-sided Bayes factors)**.** *Let* $f(d \mid \theta_0) = \binom{n}{y}\theta_0^y(1-\theta_0)^{n-y}$ *and define the two-sided Bayes factor as* $\mathrm{BF}_{10\,;\,\eta}(d) = \frac{p_\eta(d)}{f(d \mid \theta_0)}$, *then this two-sided Bayes factor is a convex combination of the one-sided Bayes factors* $\mathrm{BF}_{-0\,;\,\eta}(d) = \frac{p_\eta^{(-)}(d)}{f(d \mid \theta_0)}$ *and* $\mathrm{BF}_{+0\,;\,\eta}(d) = \frac{p_\eta^{(+)}(d)}{f(d \mid \theta_0)}$, *that is,*

$$\mathrm{BF}_{10\,;\,\eta}(d) = \frac{p_\eta^{(-)}(\emptyset)}{p_\eta(\emptyset)}\mathrm{BF}_{-0\,;\,\eta}(d) + \frac{p_\eta^{(+)}(\emptyset)}{p_\eta(\emptyset)}\mathrm{BF}_{+0\,;\,\eta}(d) \tag{11.2.31}$$

*where* $p_\eta(\emptyset), p_\eta^{(-)}(\emptyset), p_\eta^{(+)}(\emptyset)$ *are the normalisation constants of the two-sided, min-sided and plus-sided priors. For the localised prior this implies that*

$$\mathrm{BF}_{10\,;\,\eta}(d) = \frac{\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)}{\mathcal{B}(\alpha,\beta)}\mathrm{BF}_{-0\,;\,\eta}(d) + \Big[1 - \frac{\mathcal{B}(\frac{1}{2}\,;\,\alpha,\beta)}{\mathcal{B}(\alpha,\beta)}\Big]\mathrm{BF}_{+0\,;\,\eta}(d). \tag{11.2.32}$$

*For symmetric priors, thus,* $\alpha = \beta = a$, *the relationship simplifies to*

$$\mathrm{BF}_{10\,;\,a,\theta_0}(d) = \tfrac{1}{2}\mathrm{BF}_{-0\,;\,a,\theta_0}(d) + \tfrac{1}{2}\mathrm{BF}_{+0\,;\,a,\theta_0}(d), \tag{11.2.33}$$

*and this holds for any* $\theta_0$. ◇

*Proof.* Writing $\pi_u(\theta)$ for the unnormalised (two-sided) prior, that is, $\pi_u(\theta) = p_\eta(\emptyset)\pi_\eta(\theta)$ we now have

$$p_\eta(d) = \frac{1}{p_\eta(\emptyset)}\int_0^1 f(d \mid \theta)\pi_u(\theta)\mathrm{d}\theta \tag{11.2.34}$$
$$= \frac{1}{p_\eta(\emptyset)}\Big[\int_0^{\theta_0} f(d \mid \theta)\pi_u(\theta)\mathrm{d}\theta + \int_{\theta_0}^1 f(d \mid \theta)\pi_u(\theta)\mathrm{d}\theta\Big] \tag{11.2.35}$$
$$= \frac{p_\eta^{(-)}(\emptyset)}{p_\eta(\emptyset)}p_\eta^{(-)}(d) + \frac{p_\eta^{(+)}(\emptyset)}{p_\eta(\emptyset)}p_\eta^{(+)}(d) \tag{11.2.36}$$

dividing both sides by $f(d \mid \theta_0)$ now yields the result. □

**Corollary 11.2.5** (Sums of Appell functions of the first kind)**.** *Note that we have shown that certain Gauss' hypergeometric functions can be written as a sum of two Appell functions of the first kind, that is,*

$$g(u_1, u_2, v, \theta_0) = \mathcal{B}(u_1, u_2) {}_2F_1(v, u_1\,;\, u_1 + u_2\,;\, 2 - \tfrac{1}{\theta_0}) \tag{11.2.37}$$

*where*

$$\begin{aligned} g(u_1, u_2, v, \theta_0) = & \tfrac{\theta_0^{u_1}}{u_1} AF_1(u_1\,;\, 1 - u_2, v\,;\, u_1 + 1\,;\, \theta_0, 2\theta_0 - 1) \\ & + 2^{-v} \theta_0^{u_1 - 1} (1 - \theta_0)^{u_2 - v} u_2^{-1} \\ & \times AF_1(1\,;\, 1 - u_1, v\,;\, u_2 + 1\,;\, \tfrac{\theta_0 - 1}{\theta_0}\,;\, \tfrac{2\theta_0 - 1}{2\theta_0}) \end{aligned} \tag{11.2.38}$$

*is the sum of Appell functions of the first kind.* ◇

*Proof.* The assertion follows from the three calculations given above with $u = u_1, v_1 = 1 - u_2, v_2 = v$ in Eq. (11.2.21) and $v_1 = 1 - u_1, w_1 = u_2, v_2 = v$ in Eq. (11.2.30). □

We have shown how to obtain analytic results by using a localised beta prior on $\theta$. This prior is related to the so-called generalised beta prime distribution.

**Definition 11.2.2** (Generalised beta prime distribution)**.** We say that a random variable $\zeta$ has a generalised beta prime distribution and write

$$\zeta \sim \text{genBetaPrime}(\alpha, \beta, u, v), \tag{11.2.39}$$

if the density of $\zeta$ at the outcome $z$ is given by

$$f_\zeta(z) = \frac{v^{-\alpha u}}{\mathcal{B}(\alpha, \beta)} \frac{|u| z^{\alpha u - 1}}{(1 + (\frac{z}{v})^u)^{\alpha + \beta}}, \tag{11.2.40}$$

where $0 < z$ and the parameters $\alpha, \beta, v$ are positive. ◇

**Example 11.2.1** (Localised beta prior and the generalised beta prime distribution)**.** *Let $\theta$ be distributed as a beta distribution $\mathcal{B}(\alpha, \beta)$ localised at $\theta_0$, then the odds form of $\theta$, that is, $\zeta = \frac{\theta}{1-\theta}$ is distributed as a generalised beta prime distribution $\zeta \sim$* genBetaPrime$(\alpha, \beta, 1, \frac{\theta_0}{1-\theta_0})$. ◇

*Proof.* The result follows from first rewriting the localised beta as

$$\int \pi_\eta(\theta) \mathrm{d}\theta = \tfrac{1}{\mathcal{B}(\alpha,\beta)} \int (\tfrac{\theta}{1-\theta})^\alpha (\tfrac{1-\theta_0}{\theta_0})^\alpha [1 + (\tfrac{\theta}{1-\theta})(\tfrac{1-\theta_0}{\theta_0})]^{-\alpha-\beta} \theta^{-1} (1 - \theta)^{-1} \mathrm{d}\theta,$$

and the replacement $\theta = \frac{z}{1+z}$, $z_0 = \frac{\theta_0}{1-\theta_0}$, thus, $\int \mathrm{d}\theta = \int (1 + z)^{-2} \mathrm{d}z$. □

## 11.3 Products of generalised beta prime distributions and the odds ratio

**Theorem 11.3.1** (Products of generalised beta prime distributions). *Let* $\zeta_1 \sim \text{genBetaPrime}(\alpha_1, \beta_1, u, v_1)$ *and* $\zeta_2 \sim \text{genBetaPrime}(\alpha_2, \beta_2, u, v_2)$ *be independent generalised beta prime distributions with common shape* $u$. *The density of the product* $\Omega = \zeta_1\zeta_2$ *is then equivalently given by*

$$f_\Omega(\omega) = \begin{cases} C(\frac{\omega}{v_1 v_2})^{\alpha_2 u - 1}\, {}_2F_1(\alpha_2+\beta_2, \alpha_2+\beta_1\,;\, \alpha_. + \beta_.\,;\, 1 - (\frac{\omega}{v_1 v_2})^u), \\ C(\frac{v_1 v_2}{\omega})^{\beta_1 u + 1}\, {}_2F_1(\alpha_1+\beta_1, \alpha_2+\beta_1\,;\, \alpha_. + \beta_.\,;\, 1 - (\frac{v_1 v_2}{\omega})^u), \\ C(\frac{\omega}{v_1 v_2})^{\alpha_1 u - 1}\, {}_2F_1(\alpha_1+\beta_1, \alpha_1+\beta_2\,;\, \alpha_. + \beta_.\,;\, 1 - (\frac{\omega}{v_1 v_2})^u), \\ C(\frac{v_1 v_2}{\omega})^{\beta_2 u + 1}\, {}_2F_1(\alpha_2+\beta_2, \alpha_1+\beta_2\,;\, \alpha_. + \beta_.\,;\, 1 - (\frac{v_1 v_2}{\omega})^u), \end{cases} \tag{11.3.1}$$

*where* $\alpha_. = \alpha_1 + \alpha_2$, $\beta_. = \beta_1 + \beta_2$, *and* $C = \frac{|u|}{v_1 v_2} \frac{\mathcal{B}(\alpha_2+\beta_1, \alpha_1+\beta_2)}{\mathcal{B}(\alpha_1,\beta_1)\mathcal{B}(\alpha_2,\beta_2)}$. *The four equivalent results can also be derived using Kummer's 24 solutions, thus, Klein's 4-group.* ◇

*Proof.* By the convolution theorem for products of independent random variables we have $f_\Omega(\omega) = \int_{Z_2} \frac{1}{z_2} f_{\zeta_1}(\frac{\omega}{z_2}) f_{\zeta_2}(z_2) \mathrm{d}z_2$, where we have written $Z_i$ for the domain of $\zeta_i$ and $z_i$ for a specific outcome of $\zeta_i$. For $f_{\zeta_i}(\cdot)$ the generalised beta prime density this yields

$$f_\Omega(\omega) = \tilde{C} \int z_2^{(\alpha_2-\alpha_1)u-1}[1 + (\tfrac{\omega}{z_2 v_1})^u]^{-(\alpha_1+\beta_1)}[1 + (\tfrac{z_2}{v_2})^u]^{-(\alpha_2+\beta_2)} \mathrm{d}z_2, \tag{11.3.2}$$

where $\tilde{C} = \frac{v_1^{1-2\alpha_1 u} v_2^{-\alpha_2 u}}{\mathcal{B}(\alpha_1,\beta_1)\mathcal{B}(\alpha_2,\beta_2)} |u|^2 \omega^{\alpha_1 u - 1}$. Writing the linear terms as $(z_2^u + b)^{-c}$ for some $b, c > 0$ we then get

$$f_\Omega(\omega) = \breve{C} \int z_2^{(\alpha_2+\beta_1)u-1}[z_2^u + (\tfrac{w}{v_1})^u]^{-(\alpha_1+\beta_1)}[v_2^u + z_2^u]^{-(\alpha_2+\beta_2)} \mathrm{d}z_2, \tag{11.3.3}$$

where $\breve{C} = \frac{v_1^{-\alpha_1 u} v_2^{\beta_2 u}}{\mathcal{B}(\alpha_1,\beta_1)\mathcal{B}(\alpha_2,\beta_2)} |u|^2 \omega^{\alpha_1 u - 1}$. The change of variable $x = z_2^u$, thus, $\int \mathrm{d}z_2 = \int u^{-1} x^{\frac{1}{u}-1} \mathrm{d}x$ then leads to

$$f_\Omega(\omega) = \hat{C} \int x^{\alpha_2+\beta_1-1}[x + (\tfrac{\omega}{v_1})^u]^{-(\alpha_1+\beta_1)}[x + v_2^u]^{-(\alpha_2+\beta_2)} \mathrm{d}x, \tag{11.3.4}$$

where $\hat{C} = \frac{v_1^{-\alpha_1 u} v_2^{\beta_2 u}}{\mathcal{B}(\alpha_1,\beta_1)\mathcal{B}(\alpha_2,\beta_2)} |u| \omega^{\alpha_1 u - 1}$. To solve this integral we use Eq. (3.197.1) of Gradshteyn and Ryzhik (2007, p. 317) with in their notation $\nu = \alpha_2 + \beta_1, \beta = (\frac{\omega}{v_1})^u, \mu = \alpha_1 + \beta_1, \gamma = v_2^u, \varrho = \alpha_2 + \beta_2$ resulting in

$$\begin{aligned} f_\Omega(\omega) = {} & \hat{C}(\tfrac{\omega}{v_1})^{(\alpha_2-\alpha_1)u} v_2^{-(\alpha_2+\beta_2)u} \\ & \times \mathcal{B}(\alpha_2+\beta_1, \alpha_1+\beta_2)\, {}_2F_1(\alpha_2+\beta_2, \alpha_2+\beta_1\,;\, \alpha_. + \beta_.\,;\, 1 - (\tfrac{\omega}{v_1 v_2})^u), \end{aligned} \tag{11.3.5}$$

and the first equation of the assertion follows after rearranging the terms in $\hat{C}$.

| | A | | |
|---|---|---|---|
| | A | not A | Row total |
| B | $Y_{11}$ | $Y_{12}$ | $Y_{1.}$ |
| not B | $Y_{21}$ | $Y_{22}$ | $Y_{2.}$ |
| Column total | $Y_{.1}$ | $Y_{.2}$ | $Y_{..}$ |

Table 11.1: The focus is on the (in)dependence between $A$ and $B$.

On the other hand, using Eq. (3.197.1) of Gradshteyn and Ryzhik (2007, p. 317) with in their notation $\nu = \alpha_2 + \beta_1, \beta = (\frac{\omega}{v_1})^u, \mu = \alpha_1 + \beta_1, \gamma = v_2^u, \varrho = \alpha_2 + \beta_2$, instead, yields

$$\begin{aligned} f_\Omega(\omega) &= \hat{C}(\tfrac{\omega}{v_1})^{-(\alpha_1+\beta_1)u} v_2^{(\beta_1-\beta_2)u} \\ &\times \mathcal{B}(\alpha_2+\beta_1, \alpha_1+\beta_2) {}_2F_1(\alpha_1+\beta_1, \alpha_2+\beta_1\,;\,\alpha_. + \beta_.\,;\,1-(\tfrac{v_1 v_2}{\omega})^u), \end{aligned} \tag{11.3.6}$$

and the second equation of the assertion follows after rearranging the terms in $\hat{C}$. The third and the fourth equation can be derived analogously by considering the product convolution in terms of $f_\Omega(\omega) = \int_{Z_1} \frac{1}{z_1} f_{\zeta_2}(\frac{\omega}{z_1}) f_{\zeta_1}(z_1) \mathrm{d}z_1$ instead. □

**Example 11.3.1** (Odds ratio)**.** *Let the data $Y$ be arranged in a $2 \times 2$ contingency table, see Table 11.1. When the A assignment is independent on the B assignment, we have $P(A, B) = P(A)P(B)$. If this independence relationship is perfectly mimicked in the data we have $P(Y_{11}) = P(Y_{.1})P(Y_{1.})$. The (sample) odds ratio in a 2-by-2 contingency table is a measure of deviation of independence and defined as $O = \frac{Y_{11}Y_{22}}{Y_{12}Y_{21}}$.*

*In a Bayesian setting we are also interested in the implied deviation of independence on the population level. For this type of inference we have to (1) assume a model that specifies how the observed data are related to the unobserved parameters, and (2) a prior on the parameter, which allows for probabilistic statements about the parameter given the data. A general model for the data is $Y_{ij} \sim \text{Pois}(\lambda_{ij})$ with $Y_{11}, Y_{12}, Y_{21}, Y_{22}$ all independent of each other. A computationally convenient choice would be to take $\lambda_{ij} \sim \text{Gam}(\alpha_{ij}, \beta_{ij})$.*

*Analogous to the sample we define the (population) odds ratio as $\Omega = \frac{\lambda_{11}}{\lambda_{12}} \frac{\lambda_{22}}{\lambda_{21}}$. Note that the two ratios are distributed as generalised beta prime distributions*

$$\tfrac{\lambda_{11}}{\lambda_{12}} \sim \text{genBetaPrime}(\alpha_{11}, \alpha_{12}, 1, \tfrac{\beta_{12}}{\beta_{11}}), \tag{11.3.7}$$

$$\tfrac{\lambda_{22}}{\lambda_{21}} \sim \text{genBetaPrime}(\alpha_{22}, \alpha_{21}, 1, \tfrac{\beta_{21}}{\beta_{22}}), \tag{11.3.8}$$

*respectively. As such, $\Omega$ is distributed according to Eq.* (11.3.1)*. Thus,*

$$f_\Omega(\omega) = \begin{cases} C(\frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}\omega)^{\alpha_{22}-1} {}_2F_1(\alpha_{2.}, \alpha_{.2}\,;\,\alpha_{..}\,;\,1-\frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}\omega), \\ C(\frac{\beta_{12}}{\beta_{11}}\frac{\beta_{21}}{\beta_{22}}\frac{1}{\omega})^{\alpha_{12}+1} {}_2F_1(\alpha_{1.}, \alpha_{.2}\,;\,\alpha_{..}\,;\,1-\frac{\beta_{12}}{\beta_{11}}\frac{\beta_{21}}{\beta_{22}}\frac{1}{\omega}), \\ C(\frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}\omega)^{\alpha_{11}-1} {}_2F_1(\alpha_{1.}, \alpha_{.1}\,;\,\alpha_{..}\,;\,1-\frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}}\omega), \\ C(\frac{\beta_{12}}{\beta_{11}}\frac{\beta_{21}}{\beta_{22}}\frac{1}{\omega})^{\alpha_{21}+1} {}_2F_1(\alpha_{2.}, \alpha_{.1}\,;\,\alpha_{..}\,;\,1-\frac{\beta_{12}}{\beta_{11}}\frac{\beta_{21}}{\beta_{22}}\frac{1}{\omega}), \end{cases} \tag{11.3.9}$$

*where $\alpha_{i.}$ denotes the ith row sum, $\alpha_{.j}$ the jth column sum and $\alpha_{..}$ the total sum of the $\alpha_{ij}$ parameters and where $C = \frac{\beta_{11}\beta_{22}}{\beta_{12}\beta_{21}} \frac{\mathcal{B}(\alpha_{.2}, \alpha_{.1})}{\mathcal{B}(\alpha_{11}, \alpha_{12})\mathcal{B}(\alpha_{22}, \alpha_{21})}$.* ◇

*Proof.* Theorem 11.3.2 below implies that $\lambda_{i1}/\lambda_{i2}$ is indeed a generalised beta prime distribution and the result follows from Theorem 11.3.1. □

**Theorem 11.3.2** (Ratios of gammas)**.** *Let $X \sim \text{Gam}(\alpha_x, \beta_x)$ and $Y \sim \text{Gam}(\alpha_y, \beta_y)$ be independent, then $Z = X/Y \sim \text{genBetaPrime}(\alpha_x, \alpha_y, 1, \frac{\beta_y}{\beta_x})$.* ◇

*Proof.* By the convolution theorem for independent ratios we have $f_Z(z) = \int y f_X(zy) f_Y(y) \mathrm{d}y$, thus, with $C = \frac{\beta_x^{\alpha_x}}{\Gamma(\alpha_x)} \frac{\beta_y^{\alpha_y}}{\Gamma(\alpha_y)}$ and $\alpha_. = \alpha_x + \alpha_y$ we have

$$
\begin{aligned}
f_Z(z) &= C \int y(zy)^{\alpha_x - 1} e^{-\beta_x z y} y^{\alpha_y - 1} e^{-\beta_y y} \mathrm{d}y, && (11.3.10)\\
&= C z^{\alpha_x - 1} \int y^{\alpha_. - 1} e^{-(\beta_x z + \beta_y) y} \mathrm{d}y, && (11.3.11)\\
&= C z^{\alpha_x - 1} \Gamma(\alpha_.) (\beta_x z + \beta_y)^{-(\alpha_.)}, && (11.3.12)\\
&= \frac{(\frac{\beta_y}{\beta_y})^{-\alpha_x}}{\mathcal{B}(\alpha_x, \alpha_y)} z^{\alpha_x - 1} \Big[1 + \frac{z}{\beta_y/\beta_x}\Big]^{-(\alpha_x + \alpha_y)}, && (11.3.13)
\end{aligned}
$$

which is exactly what we wanted to show. □

**Corollary 11.3.1** (Analytic posterior for the odds ratio)**.** *The analytic posterior for the odds ratio is Eq.* (11.3.9) *with $\alpha_{ij}$ replaced by $\alpha_{ij} + y_{ij}$.* ◇

## 11.4 Concluding remarks

We hope that these analytic results provide further insights to posteriors and Bayes factors for the test of two proportions, and the 2-by-2 odds ratio.