Supplement of "The Bayes Rule of the Parameter in (0,1) under the Power-Log Loss Function with an Application to the Beta-Binomial Model" ⋆

Ying-Ying Zhang[a,*], Ming-Qin Zhou[a], Yu-Han Xie[a], Wen-He Song[a]

[a] *Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China*

Abstract

This is the supplemental file of the paper.

The proof of $L(\theta + \triangle a|\theta) = o(\triangle a)$. By property (f), the definition of derivative, and assume the validity of interchanging the order of limitation and evaluation, we have

$$\begin{aligned}
0 &= \left[\frac{\partial}{\partial a} L(a|\theta)\right]\bigg|_{a=\theta} = \left[\lim_{\triangle a \to 0} \frac{L(a+\triangle a|\theta) - L(a|\theta)}{\triangle a}\right]\bigg|_{a=\theta} \\
&= \lim_{\triangle a \to 0} \left[\frac{L(a+\triangle a|\theta) - L(a|\theta)}{\triangle a}\bigg|_{a=\theta} \right] \\
&= \lim_{\triangle a \to 0} \frac{L(\theta+\triangle a|\theta) - L(\theta|\theta)}{\triangle a} = \lim_{\triangle a \to 0} \frac{L(\theta+\triangle a|\theta)}{\triangle a},
\end{aligned}$$

that is, $L(\theta + \triangle a|\theta) = o(\triangle a)$. □

The check that the power-log loss function satisfies all the 6 properties listed in Table 1. Now we check that the power-log loss function, $L_{pl}(\theta, a) = L_{pl}(a|\theta) = L_{pl}(x)|_{x=a/\theta}$, satisfies all the 6 properties listed in Table 1. Let

$$g(x) = \frac{\left(\frac{1}{\theta} - 1\right)^2}{\frac{1}{\theta} - x} - \log x \text{ and } g(1) = \frac{1}{\theta} - 1.$$

⋆The research was supported by the Fundamental Research Funds for the Central Universities (CQDXWL-2012-004).

*Corresponding author

Email addresses: robertzhang@cqu.edu.cn or robertzhangyying@qq.com (Ying-Ying Zhang), 1026250991@qq.com (Ming-Qin Zhou), 545552727@qq.com (Yu-Han Xie), wenhesong@163.com (Wen-He Song)

URL: https://zhangyingying319.wordpress.com (Ying-Ying Zhang)

We have

$$L_{pl}(x) = g(x) - g(1) = \frac{\left(\frac{1}{\theta}-1\right)^2}{\frac{1}{\theta}-x} - \log x - \left(\frac{1}{\theta}-1\right)$$

and

$$\begin{aligned} L_{pl}(\theta, a) = L_{pl}(a|\theta) = \left. L_{pl}(x)\right|_{x=a/\theta} &= \frac{\left(\frac{1}{\theta}-1\right)^2}{\frac{1}{\theta}-\frac{a}{\theta}} - \log\frac{a}{\theta} - \left(\frac{1}{\theta}-1\right) \\ &= \frac{\theta\left(\frac{1}{\theta}-1\right)^2}{1-a} - \log a + \log\theta - \left(\frac{1}{\theta}-1\right). \end{aligned}$$

We first show that $L_{pl}(x)$ satisfies all the 6 properties listed in Table 1. First,

$$L_{pl}(1) = g(1) - g(1) = 0,$$

so property (b) is proved. Properties (c) and (d) can be checked directly. For property (f), we have

$$g'(x) = \left(\frac{1}{\theta}-1\right)^2 \frac{1}{\left(\frac{1}{\theta}-x\right)^2} - \frac{1}{x}.$$

Thus,

$$L'_{pl}(1) = g'(1) = \left(\frac{1}{\theta}-1\right)^2 \frac{1}{\left(\frac{1}{\theta}-1\right)^2} - \frac{1}{1} = 0.$$

Now we turn to property (e). We have

$$\begin{aligned} L''_{pl}(x) = g''(x) = \left[g'(x)\right]' &= \left[\left(\frac{1}{\theta}-1\right)^2\left(\frac{1}{\theta}-x\right)^{-2} - x^{-1}\right]' \\ &= \left(\frac{1}{\theta}-1\right)^2 2\left(\frac{1}{\theta}-x\right)^{-3} + x^{-2} > 0, \end{aligned}$$

since $0 < x < \frac{1}{\theta}$ and $0 < \theta < 1$. Therefore, $L_{pl}(x)$ is convex in x and property (e) is proved. With properties (b), (e), and (f), we can prove that

$$L_{pl}(x) \geq 0 \text{ for all } 0 < x < \frac{1}{\theta},$$

that is, property (a) is proved.

Now we prove that $L_{pl}(\theta, a) = L_{pl}(a|\theta)$ satisfies all the 6 properties listed in Table 1. Properties (a), (b), (c), and (d) are satisfied by exploiting

$$L_{pl}(\theta, a) = L_{pl}(a|\theta) = \left. L_{pl}(x)\right|_{x=a/\theta},$$

and that $L_{pl}(x)$ satisfies these properties. For property (f), we have

$$\left[\frac{\partial}{\partial a} L_{pl}(a|\theta)\right]\bigg|_{a=\theta} = \left[\frac{\partial x}{\partial a} \cdot \frac{\partial}{\partial x} L_{pl}(x)\right]\bigg|_{x=1} = \frac{1}{\theta} L'_{pl}(1) = 0.$$

For property (e), we have

$$\begin{aligned}
\frac{\partial^2}{\partial a^2} L_{pl}(a|\theta) &= \frac{\partial}{\partial a}\left[\frac{\partial}{\partial a} L_{pl}(a|\theta)\right] = \frac{\partial}{\partial a}\left\{\left[\frac{\partial x}{\partial a} \cdot \frac{\partial}{\partial x} L_{pl}(x)\right]\bigg|_{x=a/\theta}\right\} \\
&= \frac{1}{\theta}\frac{\partial}{\partial a}\left\{\left[\frac{\partial}{\partial x} L_{pl}(x)\right]\bigg|_{x=a/\theta}\right\} = \frac{1}{\theta}\left\{\left[\frac{\partial x}{\partial a} \cdot \frac{\partial^2}{\partial x^2} L_{pl}(x)\right]\bigg|_{x=a/\theta}\right\} \\
&= \frac{1}{\theta^2} L''_{pl}(x)\big|_{x=a/\theta} > 0.
\end{aligned}$$

Therefore, $L_{pl}(a|\theta)$ is convex in a and property (e) is proved. The check is complete. $\square$

The proof of $\delta_{pl}^{\pi}(\boldsymbol{x}) \le \delta_2^{\pi}(\boldsymbol{x})$. We have

$$\delta_{pl}^{\pi}(\boldsymbol{x}) = \frac{2 + E_1(\boldsymbol{x}) - \sqrt{E_1(\boldsymbol{x})(E_1(\boldsymbol{x}) + 4)}}{2} \le E_4(\boldsymbol{x}) = \delta_2^{\pi}(\boldsymbol{x}) \tag{1}$$

$$\Leftrightarrow\ 2 + E_1(\boldsymbol{x}) - \sqrt{E_1(\boldsymbol{x})(E_1(\boldsymbol{x}) + 4)} \le 2E_4(\boldsymbol{x})$$

$$\Leftrightarrow\ 2 + E_1(\boldsymbol{x}) - 2E_4(\boldsymbol{x}) \le \sqrt{E_1(\boldsymbol{x})(E_1(\boldsymbol{x}) + 4)}$$

$$\Leftrightarrow\ [2 + E_1(\boldsymbol{x}) - 2E_4(\boldsymbol{x})]^2 \le E_1(\boldsymbol{x})(E_1(\boldsymbol{x}) + 4)$$

$$\Leftrightarrow\ 4 + E_1^2(\boldsymbol{x}) + 4E_4^2(\boldsymbol{x}) + 4E_1(\boldsymbol{x}) - 8E_4(\boldsymbol{x}) - 4E_1(\boldsymbol{x})E_4(\boldsymbol{x}) \le E_1^2(\boldsymbol{x}) + 4E_1(\boldsymbol{x})$$

$$\Leftrightarrow\ 4 + 4E_4^2(\boldsymbol{x}) \le 8E_4(\boldsymbol{x}) + 4E_1(\boldsymbol{x})E_4(\boldsymbol{x})$$

$$\Leftrightarrow\ 1 + E_4^2(\boldsymbol{x}) \le 2E_4(\boldsymbol{x}) + E_1(\boldsymbol{x})E_4(\boldsymbol{x})$$

$$\Leftrightarrow\ [1 - E_4(\boldsymbol{x})]^2 \le E_1(\boldsymbol{x})E_4(\boldsymbol{x})$$

$$\Leftrightarrow\ [1 - \mathrm{E}[\theta|\boldsymbol{x}]]^2 \le \mathrm{E}\left[\frac{(1-\theta)^2}{\theta}|\boldsymbol{x}\right]\mathrm{E}[\theta|\boldsymbol{x}]$$

$$\Leftrightarrow\ [\mathrm{E}[1-\theta|\boldsymbol{x}]]^2 \le \mathrm{E}\left[\frac{(1-\theta)^2}{\theta}|\boldsymbol{x}\right]\mathrm{E}[\theta|\boldsymbol{x}]. \tag{2}$$

We need the Covariance Inequality (see Theorem 4.7.9 (p.192) in Casella and Berger (2002)) for the rest of the proof. For convenience, we quote it here.

Theorem (Covariance Inequality) *Let X be any random variable and $g(x)$ and $h(x)$ any functions such that $Eg(X)$, $Eh(X)$, and $E(g(X)h(X))$ exist.*

a. If $g(x)$ is a nondecreasing function and $h(x)$ is a nonincreasing function, then

$$\mathrm{E}(g(X)h(X)) \leq (\mathrm{E}g(X))(\mathrm{E}h(X)).$$

b. If $g(x)$ and $h(x)$ are either both nondecreasing or both nonincreasing, then

$$\mathrm{E}(g(X)h(X)) \geq (\mathrm{E}g(X))(\mathrm{E}h(X)).$$

Let $X = (\theta|\boldsymbol{x})$. It is easy to check that $g(\theta) = \frac{(1-\theta)^2}{\theta}$ is a decreasing function of θ, and $h(\theta) = \theta$ is an increasing function of θ. Thus, by the Covariance Inequality, we have

$$\mathrm{E}\left[\frac{(1-\theta)^2}{\theta}|\boldsymbol{x}\right]\mathrm{E}[\theta|\boldsymbol{x}] \geq \mathrm{E}\left[(1-\theta)^2|\boldsymbol{x}\right] = \mathrm{E}[(1-\theta)(1-\theta)|\boldsymbol{x}].$$

Now we let $g(\theta) = 1-\theta$ which is a decreasing function of θ, and $h(\theta) = 1-\theta$ which is again a decreasing function of θ. Therefore, by the Covariance Inequality, we have

$$\mathrm{E}[(1-\theta)(1-\theta)|\boldsymbol{x}] \geq \mathrm{E}(1-\theta|\boldsymbol{x})\,\mathrm{E}(1-\theta|\boldsymbol{x}) = [\mathrm{E}(1-\theta|\boldsymbol{x})]^2.$$

Therefore, (2) is correct and (1) is proved. $\square$

The calculation of $\pi(\theta|\boldsymbol{x}) \sim Be(\alpha^*, \beta^*)$. Suppose that we observe $X_1, X_2, \ldots, X_n$ from the beta-binomial model:

$$\begin{cases} X_i|\theta \overset{\text{iid}}{\sim} Bin(m,\theta), \ i = 1,2,\ldots,n, \\ \theta \sim Be(\alpha,\beta), \end{cases}$$

where m is a known positive integer, $\alpha > 0$ and $\beta > 0$ are known constants, $\theta \in (0,1)$ is the unknown parameter of interest, $Bin(m,\theta)$ is the binomial distribution, and $Be(\alpha,\beta)$ is the beta distribution. By the Bayes Theorem, we have

$$\pi(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)\,\pi(\theta).$$

It is easy to see that

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, \ \theta \in (0,1), \ \alpha,\beta > 0,$$

and

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \left[\binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i} \right] \propto \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{mn-\sum_{i=1}^{n} x_i}.$$

Therefore,

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{mn-\sum_{i=1}^{n} x_i} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\
&= \theta^{\left(\sum_{i=1}^{n} x_i+\alpha\right)-1} (1-\theta)^{\left(mn-\sum_{i=1}^{n} x_i+\beta\right)-1} \\
&\sim Be(\alpha^*, \beta^*),
\end{aligned}$$

where

$$\alpha^* = \sum_{i=1}^{n} x_i + \alpha \text{ and } \beta^* = mn - \sum_{i=1}^{n} x_i + \beta.$$

The calculation is complete. □

The calculations of $E_i(\boldsymbol{x})$, $i = 1, 2, 3, 4$. The posterior distribution of θ is

$$\pi(\theta|\boldsymbol{x}) \sim Be(\alpha^*, \beta^*).$$

We calculate $E_1(\boldsymbol{x})$, $E_3(\boldsymbol{x})$, and $E_4(\boldsymbol{x})$ first, since the calculations are straightforward. Then we calculate $E_2(\boldsymbol{x})$, which is sophisticated. We have

$$\begin{aligned}
E_1(\boldsymbol{x}) &= \mathrm{E}\left[\theta^{-1}(1-\theta)^2 |\boldsymbol{x}\right] = \int_0^1 \theta^{-1}(1-\theta)^2 \pi(\theta|\boldsymbol{x})\, d\theta \\
&= \int_0^1 \theta^{-1}(1-\theta)^2 \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \theta^{\alpha^*-1}(1-\theta)^{\beta^*-1} d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \int_0^1 \theta^{(\alpha^*-1)-1}(1-\theta)^{(\beta^*+2)-1} d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} B(\alpha^*-1, \beta^*+2) \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \frac{\Gamma(\alpha^*-1)\Gamma(\beta^*+2)}{\Gamma(\alpha^*+\beta^*+1)} \\
&= \frac{(\beta^*+1)\beta^*}{(\alpha^*-1)(\alpha^*+\beta^*)}, \text{ for } \alpha^* > 1,
\end{aligned}$$

$$
\begin{aligned}
E_3(\boldsymbol{x}) = \mathrm{E}\left[\theta^{-1}|\boldsymbol{x}\right] &= \int_0^1 \theta^{-1}\pi(\theta|\boldsymbol{x})\,d\theta \\
&= \int_0^1 \theta^{-1}\frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\theta^{\alpha^*-1}(1-\theta)^{\beta^*-1}\,d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\int_0^1 \theta^{(\alpha^*-1)-1}(1-\theta)^{\beta^*-1}\,d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}B(\alpha^*-1,\beta^*) \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\frac{\Gamma(\alpha^*-1)\Gamma(\beta^*)}{\Gamma(\alpha^*+\beta^*-1)} \\
&= \frac{\alpha^*+\beta^*-1}{\alpha^*-1}, \text{ for } \alpha^*>1,
\end{aligned}
$$

and

$$
E_4(\boldsymbol{x}) = \mathrm{E}[\theta|\boldsymbol{x}] = \frac{\alpha^*}{\alpha^*+\beta^*}.
$$

Now we calculate $E_2(\boldsymbol{x})$. We have

$$
\begin{aligned}
E_2(\boldsymbol{x}) = \mathrm{E}[\log\theta|\boldsymbol{x}] &= \int_0^1 (\log\theta)\,\pi(\theta|\boldsymbol{x})\,d\theta \\
&= \int_0^1 (\log\theta)\frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\theta^{\alpha^*-1}(1-\theta)^{\beta^*-1}\,d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}\int_0^1 (\log\theta)\,\theta^{\alpha^*-1}(1-\theta)^{\beta^*-1}\,d\theta \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)}I_1,
\end{aligned}
$$

where

$$\begin{aligned}
I_1 &= \int_0^1 (\log\theta)\,\theta^{\alpha^*-1}\,(1-\theta)^{\beta^*-1}\,d\theta \\
&= \int_0^1 \frac{d\theta^{\alpha^*-1}}{d\alpha^*}\,(1-\theta)^{\beta^*-1}\,d\theta \\
&= \int_0^1 \frac{d}{d\alpha^*}\left(\theta^{\alpha^*-1}\,(1-\theta)^{\beta^*-1}\right)d\theta \\
&= \frac{d}{d\alpha^*}\int_0^1 \theta^{\alpha^*-1}\,(1-\theta)^{\beta^*-1}\,d\theta \\
&= \frac{d}{d\alpha^*}B\left(\alpha^*,\beta^*\right) \\
&= \frac{d}{d\alpha^*}\left(\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\right) \\
&= \Gamma\left(\beta^*\right)\frac{d}{d\alpha^*}\left(\frac{\Gamma\left(\alpha^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\right) \\
&= \Gamma\left(\beta^*\right)\frac{\Gamma'\left(\alpha^*\right)\Gamma\left(\alpha^*+\beta^*\right)-\Gamma\left(\alpha^*\right)\Gamma'\left(\alpha^*+\beta^*\right)}{\Gamma^2\left(\alpha^*+\beta^*\right)}.
\end{aligned}$$

Note that

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \mathrm{PolyGamma}(z) = \mathrm{PolyGamma}(0,z) = \mathrm{digamma}(z)$$

is the digamma function, and $\Gamma(z)$ is the gamma function. In R software (R Core Team (2017)), the function digamma(z) calculates $\psi(z)$. Thus,

$$\Gamma'(z) = \Gamma(z)\,\psi(z).$$

Therefore,

$$\Gamma'\left(\alpha^*\right) = \Gamma\left(\alpha^*\right)\psi\left(\alpha^*\right)$$

and

$$\Gamma'\left(\alpha^*+\beta^*\right) = \Gamma\left(\alpha^*+\beta^*\right)\psi\left(\alpha^*+\beta^*\right).$$

Consequently,

$$\begin{aligned}
I_1 &= \frac{\Gamma\left(\beta^*\right)}{\Gamma^2\left(\alpha^*+\beta^*\right)}\left[\Gamma\left(\alpha^*\right)\psi\left(\alpha^*\right)\Gamma\left(\alpha^*+\beta^*\right)-\Gamma\left(\alpha^*\right)\Gamma\left(\alpha^*+\beta^*\right)\psi\left(\alpha^*+\beta^*\right)\right] \\
&= \frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\left[\psi\left(\alpha^*\right)-\psi\left(\alpha^*+\beta^*\right)\right].
\end{aligned}$$

Finally,

$$\begin{aligned}E_2\left(\boldsymbol{x}\right) &= \frac{\Gamma\left(\alpha^*+\beta^*\right)}{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\left[\psi\left(\alpha^*\right)-\psi\left(\alpha^*+\beta^*\right)\right]\\ &= \psi\left(\alpha^*\right)-\psi\left(\alpha^*+\beta^*\right).\end{aligned}$$

The calculations are complete. $\square$

The proof of

$$\begin{aligned}\delta_{pl}^{\pi}\left(\boldsymbol{x}\right) &= 1+\frac{\left(\beta^*+1\right)\beta^*}{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}-\frac{1}{2}\frac{1}{\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}\\ &\quad\times\sqrt{\left(\beta^*+1\right)\beta^*\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right)}\\ &< \frac{\alpha^*}{\alpha^*+\beta^*}=\delta_2^{\pi}\left(\boldsymbol{x}\right).\end{aligned} \tag{3}$$

We have, (3)

$$\begin{aligned}&\iff \quad 1+\frac{\left(\beta^*+1\right)\beta^*}{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}-\frac{\alpha^*}{\alpha^*+\beta^*}\\ &< \frac{\sqrt{\left(\beta^*+1\right)\beta^*\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right)}}{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}\end{aligned}$$

$$\begin{aligned}&\iff \quad \frac{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)+\left(\beta^*+1\right)\beta^*-2\left(\alpha^*-1\right)\alpha^*}{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}\\ &< \frac{\sqrt{\left(\beta^*+1\right)\beta^*\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right)}}{2\left(\alpha^*-1\right)\left(\alpha^*+\beta^*\right)}\end{aligned}$$

$$\iff \quad 2\left(\alpha^*-1\right)\beta^*+\left(\beta^*+1\right)\beta^* < \sqrt{\left(\beta^*+1\right)\beta^*\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right)}$$

$$\iff \quad \beta^*\left(2\alpha^*+\beta^*-1\right) < \sqrt{\left(\beta^*+1\right)\beta^*\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right)}$$

$$\iff \quad \beta^*\left(2\alpha^*+\beta^*-1\right)^2 < \left(\beta^*+1\right)\left(4\alpha^{*2}+4\alpha^*\beta^*-4\alpha^*-3\beta^*+\beta^{*2}\right). \tag{4}$$

It is easy to obtain

$$
\begin{aligned}
&4\alpha^{*2} + 4\alpha^*\beta^* - 4\alpha^* - 3\beta^* + \beta^{*2} \\
&= (2\alpha^* + \beta^* - 1)(2\alpha^* + \beta^*) - 2(\alpha^* + \beta^*) \\
&= (2\alpha^* + \beta^* - 1)(2\alpha^* + \beta^* - 1 + 1) - 2(\alpha^* + \beta^*) \\
&= (2\alpha^* + \beta^* - 1)^2 + (2\alpha^* + \beta^* - 1) - 2(\alpha^* + \beta^*) \\
&= (2\alpha^* + \beta^* - 1)^2 - (\beta^* + 1).
\end{aligned}
$$

Therefore, (4)

$$
\begin{aligned}
&\Longleftrightarrow \quad \beta^*(2\alpha^* + \beta^* - 1)^2 < (\beta^* + 1)\left[(2\alpha^* + \beta^* - 1)^2 - (\beta^* + 1)\right] \\
&\Longleftrightarrow \quad \beta^*(2\alpha^* + \beta^* - 1)^2 < (\beta^* + 1)(2\alpha^* + \beta^* - 1)^2 - (\beta^* + 1)^2
\end{aligned}
$$

$$
\Longleftrightarrow \quad (\beta^* + 1)^2 < (2\alpha^* + \beta^* - 1)^2. \tag{5}
$$

Since $\alpha^* > 1$ and $\beta^* > 0$, we obtain

$$
2\alpha^* + \beta^* - 1 > 2 + \beta^* - 1 = \beta^* + 1 > 0.
$$

Therefore, (5) is correct and (3) is proved. □

The derivations of the moment estimators and the MLEs of α and β. Suppose that we observe $X_1, X_2, \ldots, X_n$ from the beta-binomial model:

$$
\begin{cases}
X_i|\theta \overset{\text{iid}}{\sim} Bin(m, \theta), \ i = 1, 2, \ldots, n, \\
\theta \sim Be(\alpha, \beta),
\end{cases}
$$

where m is a known positive integer, $\alpha > 0$ and $\beta > 0$ are unknown hyperparameters determined by the empirical Bayes method, $\theta \in (0, 1)$ is the unknown parameter of interest, $Bin(m, \theta)$ is the binomial distribution, and $Be(\alpha, \beta)$ is the beta distribution.

We first derive the moment estimators of α and β. The expectation and variance of X are respectively given by

$$
\mathrm{E}X = m\frac{\alpha}{\alpha + \beta} \text{ and } \mathrm{Var}(X) = m\frac{\alpha\beta(\alpha + \beta + m)}{(\alpha + \beta)^2(\alpha + \beta + 1)}.
$$

See Examples 4.4.6 and 4.4.8 of Casella and Berger (2002) for details. The moment estimators of α and β are calculated by equating the population moments to the sample moments, that is,

$$\begin{aligned} \mathrm{E}X &= m\frac{\alpha}{\alpha+\beta} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = A_1, \\ \mathrm{Var}\left(X\right) &= m\frac{\alpha\beta\left(\alpha+\beta+m\right)}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = M_2, \end{aligned}$$

where A_1 is the sample first-order moment of X and M_2 is the sample second-order central moment of X. Solving the above equations, we obtain the moment estimators of α and β:

$$\begin{aligned} \alpha_1 &= \tilde{\alpha} = \frac{M_2A_1 - A_1^2\left(m-A_1\right)}{A_1\left(m-A_1\right)-M_2m}, \\ \beta_1 &= \tilde{\beta} = \frac{\left(m-A_1\right)\alpha_1}{A_1}. \end{aligned}$$

Now we derive the MLEs of α and β, α_2 and β_2. The marginal distribution of X is a beta-binomial distribution (BBD) with probability mass function

$$\begin{aligned} m\left(x|\alpha,\beta\right) = P\left(X=x|\alpha,\beta\right) &= \binom{m}{x}\frac{B\left(x+\alpha,m-x+\beta\right)}{B\left(\alpha,\beta\right)} \\ &= \binom{m}{x}\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\frac{\Gamma\left(x+\alpha\right)\Gamma\left(m-x+\beta\right)}{\Gamma\left(m+\alpha+\beta\right)}, \end{aligned}$$

where $B\left(\alpha,\beta\right)$ is the beta function and $\Gamma\left(\alpha\right)$ is the gamma function. Then the likelihood function of α and β is

$$\begin{aligned} L\left(\alpha,\beta|\boldsymbol{x}\right) &= m\left(\boldsymbol{x}|\alpha,\beta\right) = \prod_{i=1}^{n} m\left(x_i|\alpha,\beta\right) \\ &= \prod_{i=1}^{n}\left[\binom{m}{x_i}\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\frac{\Gamma\left(x_i+\alpha\right)\Gamma\left(m-x_i+\beta\right)}{\Gamma\left(m+\alpha+\beta\right)}\right] \\ &= \left[\prod_{i=1}^{n}\binom{m}{x_i}\right]\frac{\Gamma^n\left(\alpha+\beta\right)}{\Gamma^n\left(\alpha\right)\Gamma^n\left(\beta\right)\Gamma^n\left(m+\alpha+\beta\right)}\left[\prod_{i=1}^{n}\Gamma\left(x_i+\alpha\right)\Gamma\left(m-x_i+\beta\right)\right]. \end{aligned}$$

Consequently, the log-likelihood function of α and β is

$$\begin{aligned} \log L\left(\alpha,\beta|\boldsymbol{x}\right) = \log\prod_{i=1}^{n}\binom{m}{x_i} &+ n\log\Gamma\left(\alpha+\beta\right) - n\log\Gamma\left(\alpha\right) - n\log\Gamma\left(\beta\right) - n\log\Gamma\left(m+\alpha+\beta\right) \\ &+ \sum_{i=1}^{n}\left[\log\Gamma\left(x_i+\alpha\right)+\log\Gamma\left(m-x_i+\beta\right)\right]. \end{aligned}$$

Taking partial derivatives with respect to α and β and setting them to zeros, we obtain

$$\frac{\partial}{\partial\alpha}\log L = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\frac{\Gamma'(m+\alpha+\beta)}{\Gamma(m+\alpha+\beta)} + \sum_{i=1}^{n}\frac{\Gamma'(x_i+\alpha)}{\Gamma(x_i+\alpha)} = 0,$$

$$\frac{\partial}{\partial\beta}\log L = n\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)} - n\frac{\Gamma'(\beta)}{\Gamma(\beta)} - n\frac{\Gamma'(m+\alpha+\beta)}{\Gamma(m+\alpha+\beta)} + \sum_{i=1}^{n}\frac{\Gamma'(m-x_i+\beta)}{\Gamma(m-x_i+\beta)} = 0.$$

Since

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d}{dx}\log\Gamma(x) = \text{digamma}(x)$$

which can be directly calculated in R software by digamma(x) (R Core Team (2017)), after some algebra, the above equations reduces to

$$f_1(\alpha,\beta) = \psi(\alpha+\beta) - \psi(\alpha) - \psi(m+\alpha+\beta) + \frac{1}{n}\sum_{i=1}^{n}\psi(x_i+\alpha) = 0, \tag{6}$$

$$f_2(\alpha,\beta) = \psi(\alpha+\beta) - \psi(\beta) - \psi(m+\alpha+\beta) + \frac{1}{n}\sum_{i=1}^{n}\psi(m-x_i+\beta) = 0. \tag{7}$$

The Jacobian matrix of α and β is given by

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial\alpha} & \frac{\partial f_1}{\partial\beta} \\ \frac{\partial f_2}{\partial\alpha} & \frac{\partial f_2}{\partial\beta} \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

where

$$J_{11} = \frac{\partial f_1}{\partial\alpha} = \psi'(\alpha+\beta) - \psi'(\alpha) - \psi'(m+\alpha+\beta) + \frac{1}{n}\sum_{i=1}^{n}\psi'(x_i+\alpha),$$

$$J_{12} = \frac{\partial f_1}{\partial\beta} = \psi'(\alpha+\beta) - \psi'(m+\alpha+\beta),$$

$$J_{21} = \frac{\partial f_2}{\partial\alpha} = \psi'(\alpha+\beta) - \psi'(m+\alpha+\beta),$$

$$J_{22} = \frac{\partial f_2}{\partial\beta} = \psi'(\alpha+\beta) - \psi'(\beta) - \psi'(m+\alpha+\beta) + \frac{1}{n}\sum_{i=1}^{n}\psi'(m-x_i+\beta).$$

Note that

$$\psi'(x) = \frac{d^2}{dx^2}\log\Gamma(x) = \text{trigamma}(x)$$

which can be directly calculated in R software by trigamma(x) (R Core Team (2017)). We can exploit Newton's method to solve the equations (6) and (7) and to obtain the MLEs of α and β, α_2 and β_2. Note that the MLEs of α and β are very sensitive to the initial estimators, and the moment estimators are usually proved to be good initial estimators. $\square$

References

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury, USA. 2nd edition.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.