OPTIMAL PRIORS FOR THE DISCOUNTING PARAMETER OF THE NORMALIZED POWER PRIOR

Yueqi Shen[1], Luiz M. Carvalho[2], Matthew A. Psioda[3] and Joseph G. Ibrahim[1]

[1] *University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.*

[2]School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, Brazil

[3]GSK, Philadelphia, Pennsylvania, U.S.A

Supplementary Material

S1 Proofs from Section 2

S1.1 Technical conditions for the limit theorems

We start our presentation by stating technical conditions under which the limiting theorems presented in Section 2 hold. Then, we state an important result below (Bayes Central Limit Theorem (Chen (1985))) which gives support to many of the proofs herein. In what follows, we will follow Chen (1985) in establishing the necessary conditions for the limiting posterior density to be normal. Let the parameter space of interest be Θ and a p-dimensional Euclidean space and let $B_r(a) = \{\theta \in \Theta : |\theta - a| \leq r\}$

be a neighbourhood of size r of the point $a \in \Theta$. Also, write $L_n(\theta) := \sum_{i=1}^{n} \log f(x \mid \theta)$.

Theorem S1 (Bayes Central Limit Theorem (Chen (1985)))**.** *Suppose that for each $n > N$ with $N > 0$, L_n attains a strict local maximum $\hat{\theta}_n$ such that $L_n'(\hat{\theta}_n) := \frac{\partial}{\partial\theta} L_n(\hat{\theta}_n) = 0$ and the Hessian $L_n''(\theta) := \frac{\partial^2}{\partial\theta^2} L_n(\theta)$ is negative-definite for all $\theta \in \Theta$.*

Moreover, suppose $\hat{\theta}_n$ converges almost surely to $\theta_0 \in \Theta$ as $n \to \infty$ and the prior density $\pi(\theta)$ is positive and continuous at θ_0. Assume that the following conditions hold:

C1 The largest eigenvalue of $\left[-L_n''(\hat{\theta}_n)\right]^{-1} \to 0$ a.s. as $n \to \infty$;

C2 For $\varepsilon > 0$ there exists (a.s.) $N_\varepsilon > 0$ and $r > 0$ such that for all $n > \max\{N, N_\varepsilon\}$ and $\theta \in B_r(\hat{\theta}_n)$, $L_n''(\theta)$ is well-defined and

$$I_p - A(\varepsilon) \leq L_n''(\theta) \left[L_n''(\hat{\theta}_n)\right]^{-1} \leq I_p + A(\varepsilon),$$

where I_p is the p-dimensional identity matrix and $A(\varepsilon)$ is a $p \times p$ positive semidefinite matrix whose largest eigenvalue goes to zero as $\varepsilon \to 0$.

C3 The sequence of posterior distributions $p_n(\theta \mid x)$ satisfies, as $n \to \infty$,

$$\int_{\Theta \setminus B_r(\hat{\theta}_n)} p_n(t \mid x)\, dt \to 0, a.s.,$$

> *for $r > 0$, i.e., the sequence of posteriors is* consistent *at $\hat{\theta}_n$. Here we have assumed that the support of the posterior distributions is Θ, but this could be replaced by a sequence Θ_n.*

Then we say that the posteriors converge in distribution to a normal with parameters $\hat{\theta}_n$ and $\left[-L_n''(\hat{\theta}_n)\right]^{-1}$.

For notational convenience we will (somewhat informally) write

$$p_n(\theta|x) \to N_p\left(\hat{\theta}_n, \left[-L_n''(\hat{\theta}_n)\right]^{-1}\right),$$

as $n \to \infty$. This should be understood as the posterior density becoming highly peaked and behaving like a normal kernel around $\hat{\theta}_n$) (Chen, 1985, page 541). Since the probability outside $B_r(\hat{\theta}_n)$ is negligible, one needs not to concern oneself with what happens on $\Theta \setminus B_r(\hat{\theta}_n)$ when taking posterior expectations, for instance. See also Theorem 7.89 in Schervish (1995) (page 437).

S1.2 Proof of Theorem 2.1

Now we move on to present a proof for Theorem 2.1 in Section 2, which discusses the concentration of the posterior of a_0 at zero as the sample sizes increase in the case when there is some discrepancy between the historical and current data sets.

Proof. We first employ the Bayes Central Limit Theorem presented above to rewrite the limiting marginal posterior distribution of a_0. Under the regularity conditions as $n \to \infty$,

$$
\begin{aligned}
L_n(\theta|D) &\to N(\hat{\theta}_n, v_n), \quad \text{and} \\
\frac{1}{c(a_0)} L_{n_0}(\theta|D_0)^{a_0}\pi_0(\theta) &\to N(\hat{\theta}_0, v_0(a_0)),
\end{aligned}
$$

where $\hat{\theta}_n = \dot{b}^{-1}(\bar{y})$, $\hat{\theta}_0 = \dot{b}^{-1}(\bar{y}_0)$, $v_n = (n\ddot{b}(\hat{\theta}_n))^{-1}$, and $v_0(a_0) = (a_0 n_0 \ddot{b}(\hat{\theta}_0))^{-1}$. For simplicity of notation, let $v_0 = v_0(a_0)$, $\ddot{b}^{-1} = \ddot{b}^{-1}(\hat{\theta}_n)$ and $\ddot{b}_0^{-1} = \ddot{b}^{-1}(\hat{\theta}_0)$. Then the kernel of the marginal posterior of a_0 becomes

$$
\begin{aligned}
\pi^*(a_0|D_0, D, \alpha_0, \beta_0) \equiv & \int L_n(\theta|D)\frac{L_{n_0}(\theta|D_0)^{a_0}\pi_0(\theta)}{c(a_0)} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\, d\theta, \\
\to & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\int N(\hat{\theta}_n, v_n)N(\hat{\theta}_0, v_0(a_0))d\theta, \\
\propto & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}v_0^{-\frac{1}{2}}\int \exp\left\{-\frac{1}{2v_n}(\theta-\hat{\theta}_n)^2\right\}\exp\left\{-\frac{1}{2v_0}(\theta-\hat{\theta}_0)^2\right\}d\theta, \\
\propto & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left(\frac{v_n+v_0}{v_n}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{v_n\hat{\theta}_0^2 - v_0\hat{\theta}_n^2 - 2v_n\hat{\theta}_n\hat{\theta}_0}{(v_0+v_n)v_n}\right]\right\}, \\
= & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left(\frac{v+v_0}{v_n}\right)^{-\frac{1}{2}}\exp\left\{\frac{v_0\hat{\theta}_n^2 - v_n(\delta^2-\hat{\theta}_n^2)}{2(v_0+v_n)v_n}\right\} \text{ (since } |\hat{\theta}_n - \hat{\theta}_0| = \delta\text{)}, \\
= & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left(\frac{v_n+v_0}{v_n}\right)^{-\frac{1}{2}}\exp\left\{\frac{\hat{\theta}_n^2}{2v} - \frac{\delta^2}{2(v+v_0)}\right\}, \\
= & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left(\frac{v_n+v_0}{v_n}\right)^{-\frac{1}{2}}\exp\left\{\frac{\hat{\theta}_n^2}{2v_n} - \frac{na_0r\delta^2}{2(\ddot{b}_0^{-1}+a_0r\ddot{b}^{-1})}\right\}.
\end{aligned}
$$

Then the marginal posterior of a_0 becomes

$$\pi(a_0|D_0, D, \alpha_0, \beta_0) = \frac{\pi^*(a_0|D_0, D, \alpha_0, \beta_0)}{\int \pi^*(a_0|D_0, D, \alpha_0, \beta_0)\, da_0}, \tag{S1.1}$$

$$\rightarrow \frac{a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}[\ddot{b}^{-1} + (a_0 r \ddot{b}_0)^{-1}]^{-\frac{1}{2}} \exp\left\{-\frac{n a_0 r \delta^2}{2(\ddot{b}_0^{-1} + a_0 r \ddot{b}^{-1})}\right\}}{\int a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}[\ddot{b}^{-1} + (a_0 r \ddot{b}_0)^{-1}]^{-\frac{1}{2}} \exp\left\{-\frac{n a_0 r \delta^2}{2(\ddot{b}_0^{-1} + a_0 r \ddot{b}^{-1})}\right\} da_0}, \tag{S1.2}$$

$$= \frac{a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left[\frac{a_0 r}{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}}\right]^{\frac{1}{2}} \exp\left\{-\frac{n a_0 r \delta^2}{2\ddot{b}_0^{-1}\left(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}\right)}\right\}}{\int a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\left[\frac{a_0 r}{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}}\right]^{\frac{1}{2}} \exp\left\{-\frac{n a_0 r \delta^2}{2\ddot{b}_0^{-1}\left(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}\right)}\right\} da_0}. \tag{S1.3}$$

Let $h(a_0) = \frac{a_0 r \delta^2}{2\ddot{b}_0^{-1}\left(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}\right)}$ and $f(a_0) = \left[\frac{a_0 r}{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}}\right]^{\frac{1}{2}}$. Then the denominator is

$$A = \int_0^1 a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0) \exp\{-nh(a_0)\}\, da_0.$$

Let $A = A_1 + A_2$ where

$$A_1 = \int_0^\epsilon a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0) \exp\{-nh(a_0)\}\, da_0$$

and

$$A_2 = \int_\epsilon^1 a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0) \exp\{-nh(a_0)\}\, da_0.$$

We want to show $\lim\limits_{n\to\infty} \frac{A_2}{A_1} = 0$.

First, we can see that

$$h'(a_0) = \frac{r\delta^2}{2\ddot{b}_0^{-1}} \left(\frac{a_0}{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}} \right)' = \frac{r\delta^2}{2\ddot{b}_0^{-1}} \left(\frac{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}} - a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}}{(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}})^2} \right) = \frac{r\delta^2}{2\ddot{b}_0^{-1}} \left(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}} \right)^{-2} > 0.$$

Then $\inf_{x \in [\epsilon, 1]} h(x) = h(\epsilon)$. We can also see that $h'(a_0)$ is continuous since $1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}$ is nonzero on $(0, 1)$.

We then observe that

$$f'(a_0) = \frac{1}{2} \left[\frac{a_0 r}{1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}}} \right]^{-\frac{1}{2}} \frac{r}{(1 + a_0 r \frac{\ddot{b}^{-1}}{\ddot{b}_0^{-1}})^2} > 0.$$

Thus $\sup_{x \in [\epsilon, 1]} f(x) = f(1)$.

Now we are ready to find the upper bound of A_2. Since, for any $a_0 \in [\epsilon, 1]$, $f(a_0) \le f(1)$ and $\exp(-nh(a_0)) \le \exp(-nh(\epsilon))$, we have

$$\begin{aligned} A_2 &\le f(1) \exp(-nh(\epsilon)) \int_\epsilon^1 a_0^{\alpha_0 - 1} (1 - a_0)^{\beta_0 - 1} da_0 \\ &\le f(1) \exp(-nh(\epsilon)) \int_0^1 a_0^{\alpha_0 - 1} (1 - a_0)^{\beta_0 - 1} da_0, \\ &= f(1) \exp(-nh(\epsilon)) \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)} = C_1 \exp(-nh(\epsilon)), \end{aligned}$$

where $C_1 > 0$ is an integration constant. Now we find the lower bound of A_1. We know that

$$A_1 \ge \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0 - 1} (1 - a_0)^{\beta_0 - 1} f(a_0) \exp\{-nh(a_0)\} \, da_0.$$

Further, $a_0^{\alpha_0 - 1} \ge \min((\frac{\epsilon}{2})^{\alpha_0 - 1}, \epsilon^{\alpha_0 - 1})$, corresponding to $\alpha_0 \ge 1$ and $\alpha_0 < 1$, respectively. Similarly, $(1 - a_0)^{\beta_0 - 1} \le \min((1 - \epsilon)^{\beta_0 - 1}, (1 - \frac{\epsilon}{2})^{\beta_0 - 1})$, corresponding to $\beta_0 \geq 1$ and $\beta_0 < 1$, respectively. Since $h''(a_0) < 0$, $\sup_{x \in [\frac{\epsilon}{2}, \epsilon]} h'(x) = h'(\frac{\epsilon}{2})$. In addition, $\inf_{x \in [\frac{\epsilon}{2}, \epsilon]} f(x) = f(\frac{\epsilon}{2})$. Then we have

$$\begin{aligned}
A_1 &\geq \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0 - 1}(1-a_0)^{\beta_0 - 1} f(a_0) \frac{1}{h'(a_0)} \exp\{-nh(a_0)\} \, h'(a_0) \, da_0, \\
&= \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0 - 1}(1-a_0)^{\beta_0 - 1} f(a_0) \frac{1}{h'(a_0)} \exp\{-nh(a_0)\} \, dh(a_0), \\
&\geq f(\epsilon/2) \times \min((\epsilon/2)^{\alpha_0 - 1}, \epsilon^{\alpha_0 - 1}) \times \min((1-\epsilon)^{\beta_0 - 1}, (1-\epsilon/2)^{\beta_0 - 1}) \times (h'(\epsilon/2))^{-1} \\
&\times \int_{\frac{\epsilon}{2}}^{\epsilon} \exp(-nh(a_0)) \, dh(a_0), \\
&= C_2 \int_{\frac{\epsilon}{2}}^{\epsilon} \exp(-nh(a_0)) \, dh(a_0), \\
&= C_2 \frac{1}{n} [\exp(-nh(\epsilon/2)) - \exp(-nh(\epsilon))],
\end{aligned}$$

where $C_2 > 0$ is again an integration constant. Therefore,

$$0 \leq \frac{A_2}{A_1} \leq \frac{C_1 \exp(-nh(\epsilon))}{C_2 \frac{1}{n} [\exp(-nh(\epsilon/2)) - \exp(-nh(\epsilon))]} = \frac{C_1 n}{C_2 [\exp(-n[h(\epsilon/2) - h(\epsilon)]) - 1]}.$$

Thus, $\lim_{n \to \infty} \frac{A_2}{A_1} = 0$ by L'Hopital's rule. Since $\frac{A_2}{A_1} \geq \frac{A_2}{A}$, $\lim_{n \to \infty} \frac{A_2}{A} = 0$. Hence, $\lim_{n \to \infty} \frac{A_1}{A} = 1$. □

S1.3 Proof for Corollary 2.1

Proof. The result follows by setting $\delta = 0$ and $\ddot{b}^{-1} = \ddot{b}_0^{-1}$ into (S1.3). □

S1.4 Proof for Theorem 2.2

Proof. Let $r = \frac{n_0}{n}$. Since $y_1, \ldots, y_n$ and $y_{01}, \ldots, y_{n_0}$ are i.i.d. normal data, the marginal posterior of a_0 is

$$\pi(a_0|D_0, D) = \frac{\pi(a_0)\left[\frac{a_0 r}{1+a_0 r\frac{\sigma^2}{\sigma_0^2}}\right]^{\frac{1}{2}} \exp\left\{-\frac{n a_0 r d^2}{2\sigma_0^2\left(1+a_0 r\frac{\sigma^2}{\sigma_0^2}\right)}\right\}}{\int \pi(a_0)\left[\frac{a_0 r}{1+a_0 r\frac{\sigma^2}{\sigma_0^2}}\right]^{\frac{1}{2}} \exp\left\{-\frac{n a_0 r d^2}{2\sigma_0^2\left(1+a_0 r\frac{\sigma^2}{\sigma_0^2}\right)}\right\} da_0}.$$

With

$$g_d(a_0) := \pi(a_0)\left[\frac{a_0 r}{1+a_0 r\frac{\sigma^2}{\sigma_0^2}}\right]^{\frac{1}{2}} \exp\left\{-\frac{n a_0 r d^2}{2\sigma_0^2\left(1+a_0 r\frac{\sigma^2}{\sigma_0^2}\right)}\right\},$$

we write

$$F_d(a_0) := \frac{\int_0^{a_0} g_d(x)\, dx}{\int_0^1 g_d(x)\, dx} = \frac{G_d(a_0)}{G_d(1)}.$$

We want to show that

$$\frac{\partial F_d(a_0)}{\partial d} > 0, a_0 \in (0, 1). \tag{S1.4}$$

Using the quotient rule we conclude that (S1.4) holds if and only if:

$$\frac{\partial}{\partial d} G_d(a_0) G_d(1) - \frac{\partial}{\partial d} G_d(1) G_d(a_0) > 0.$$

We note that

$$\frac{\partial}{\partial d} G_d(a_0) = -d\frac{nr}{\sigma_0^2}\int_0^{a_0} h(x) g_d(x)\, dx, \quad \text{with}$$

$$h(a_0) = \frac{a_0}{1+a_0 r\frac{\sigma^2}{\sigma_0^2}}.$$

This in turn means that (S1.4) is equivalent to

$$\int_0^1 h(x)g_d(x)\,dx \int_0^{a_0} g_d(x)\,dx - \int_0^{a_0} h(x)g_d(x)\,dx \int_0^1 g_d(x)\,dx > 0,$$

i.e.,

$$\frac{\int_0^{a_0} g_d(x)\,dx}{\int_0^1 g_d(x)\,dx} > \frac{\int_0^{a_0} h(x)g_d(x)\,dx}{\int_0^1 h(x)g_d(x)\,dx}.$$

We first prove the following lemma.

Lemma 1 (Ratio of truncated expectations)**.** *Let X be a random variable in $(0,1)$ with distribution function F. Take any positive increasing function h. Then*

$$\frac{E[h(X)\mathbb{I}(X \leq a)]}{E[h(X)]} < F(a),$$

for $a \in (0,1)$.

Proof. Start by dividing through by $F(a)$ to get

$$\frac{E[h(X) \mid X \leq a]}{E[h(X)]} < 1.$$

But by the law of total expectation, we have

$$E[h(X)] = E[h(X) \mid X \leq a]F(a) + E[h(X) \mid X > a][1 - F(a)],$$

thus the LHS is

$$\frac{E[h(X) \mid X \leq a]}{E[h(X) \mid X \leq a]F(a) + E[h(X) \mid X > a][1 - F(a)]}.$$

If we let $w = E[h(X) \mid X \leq a]$ and $u = E[h(X) \mid X > a]$, we have that $u = w + \varepsilon$ with $\epsilon > 0$. Putting $\alpha = F(a)$, we have

$$\frac{w}{\alpha w + (1-\alpha)u} = \frac{w}{\alpha(w-u)+u},$$
$$= \frac{w}{w+(1-\alpha)\varepsilon} < 1,$$

which concludes the argument. □

We may assume without loss of generality that g_d is a normalised density. Since $h(a_0)$ is increasing, we apply Lemma 1, which completes the proof. □

S1.5 Proof for Theorem 2.3

Proof. By the Bayes Central Limit Theorem, we know that

$$L_n(\beta|D) \rightarrow N(\hat{\beta}, \Sigma(\hat{\beta})),$$

where $\Sigma(\beta) = -\left[\frac{\partial^2 \log[L_n(\beta|D)]}{\partial\beta_i \partial\beta_j}\right]^{-1}$, and also

$$\frac{1}{c^*(a_0)} L_{n_0}(\beta|D_0)^{a_0}\pi_0(\beta) \rightarrow N(\hat{\beta}_0, \Sigma_0(a_0, \hat{\beta})),$$

where $\Sigma_0(a_0, \beta) = -\left[\frac{\partial^2 \log[L_{n_0}(\beta|D_0)^{a_0}\pi_0(\beta)]}{\partial\beta_i \partial\beta_j}\right]^{-1}$. For simplicity of notation, let $\Sigma = \Sigma(\hat{\beta})$ and $\Sigma_0 = \Sigma_0(a_0, \hat{\beta})$. Then the marginal posterior of a_0 becomes

$$\pi(a_0|D_0, D, \alpha_0, \beta_0) \propto \pi^*(a_0|D_0, D, \alpha_0, \beta_0) \equiv \int L_n(\beta|D)\frac{L_{n_0}(\beta|D_0)^{a_0}\pi_0(\beta)}{c^*(a_0)} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} d\beta,$$

and

$$
\begin{aligned}
\pi^*(a_0|D_0, D, \alpha_0, \beta_0) \to\ & a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\int N(\hat{\beta},\Sigma)N(\hat{\beta}_0,\Sigma_0)d\beta,\\
&\propto a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\int \exp\left\{-\frac{1}{2}(\beta-\hat{\beta})'\Sigma^{-1}(\beta-\hat{\beta})\right\}\times\\
&\det(\Sigma_0)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\beta-\hat{\beta}_0)'\Sigma_0^{-1}(\beta-\hat{\beta}_0)\right\}d\beta,\\
&(\text{Assuming that } \hat{\beta}-\hat{\beta}_0=\delta)\\
&\propto a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}}\\
&\exp\left\{\frac{1}{2}\left[\hat{\beta}'\Sigma^{-1}\hat{\beta}-\delta'(\Sigma_0^{-1}-\Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta\right]\right\},
\end{aligned}
$$

where $\Sigma_n = (\Sigma^{-1}+\Sigma_0^{-1})^{-1}$. Then

$$
\pi(a_0|D_0, D, \alpha_0, \beta_0) \propto \frac{a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\delta'(\Sigma_0^{-1}-\Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta\right\}}{\int a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\delta'(\Sigma_0^{-1}-\Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta\right\}da_0}. \tag{S1.5}
$$

We want to show that, if Σ and Σ_0 are $p\times p$ positive definite matrices,

$$
\lim_{n\to\infty}\frac{\int_0^{\epsilon} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\delta^T(\Sigma_0^{-1}-\Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta\right\}da_0}{\int_0^{1} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}\det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\delta^T(\Sigma_0^{-1}-\Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta\right\}da_0}=1,
$$

for $\delta\neq 0$ and $\epsilon>0$.

We can write $\Sigma = n^{-1}P$ and $\Sigma_0 = (nra_0)^{-1}P_0$ (Fahrmeir and Kaufmann, 1985), where P and P_0 are positive definite and independent of a_0 and n. Then $\Sigma_n = (\Sigma^{-1}+\Sigma_0^{-1})^{-1} = n^{-1}(P^{-1}+ra_0P_0^{-1})^{-1}$.

Now,

$$
I-\Sigma_n\Sigma_0^{-1} = I-(\Sigma^{-1}+\Sigma_0^{-1})^{-1}\Sigma_0^{-1} = (\Sigma^{-1}+\Sigma_0^{-1})^{-1}\Sigma^{-1},
$$

and

$$
\begin{aligned}
\Sigma_0^{-1}(I - \Sigma_n \Sigma_0^{-1}) =& \Sigma_0^{-1} \Sigma_n \Sigma^{-1}, \\
=& nra_0 P_0^{-1} n^{-1} (P^{-1} + ra_0 P_0^{-1})^{-1} n P^{-1}, \\
=& nra_0 P_0^{-1} (P^{-1} + ra_0 P_0^{-1})^{-1} P^{-1}, \\
=& nra_0 (P_0 + a_0 r P)^{-1}, \\
=& nra_0 P^{-1} [P_0 P^{-1} + a_0 r I]^{-1}, \\
=& na_0 P^{-1} [r^{-1} P_0 P^{-1} + a_0 I]^{-1}.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
\det(\Sigma_0)^{-\frac{1}{2}} \det(\Sigma_n)^{\frac{1}{2}}, =& \det((nra_0)^{-1} P_0)^{-\frac{1}{2}} \det(n^{-1}(P^{-1} + ra_0 P_0^{-1})^{-1})^{\frac{1}{2}}, \\
=& \det((ra_0)^{-1} P_0)^{-\frac{1}{2}} \det(P^{-1} + ra_0 P_0^{-1})^{-\frac{1}{2}}, \\
=& \det((ra_0)^{-1} P_0 (P^{-1} + ra_0 P_0^{-1}))^{-\frac{1}{2}}, \\
=& \det(a_0^{-1}(r^{-1} P_0 P^{-1} + a_0 I))^{-\frac{1}{2}}, \\
=& a_0^{\frac{p}{2}} \det(a_0 I - (-r^{-1} P_0 P^{-1}))^{-\frac{1}{2}}.
\end{aligned}
$$

Let

$$h(a_0) = \frac{1}{2n}\delta^T(\Sigma_0^{-1} - \Sigma_0^{-1}\Sigma_n\Sigma_0^{-1})\delta = \frac{1}{2}\delta^T a_0 P^{-1}[r^{-1}P_0P^{-1} + a_0 I]^{-1}\delta.$$

and

$$f(a_0) = \det(\Sigma_0)^{-\frac{1}{2}}\det(\Sigma_n)^{\frac{1}{2}} = a_0^{\frac{p}{2}}\det(a_0 I - (-r^{-1}P_0P^{-1}))^{-\frac{1}{2}}.$$

Then the denominator is

$$A = \int_0^1 a_0^{\alpha_0 - 1}(1-a_0)^{\beta_0 - 1} f(a_0)\exp\big\{-nh(a_0)\big\}da_0.$$

First, we show $h(a_0)$ is differentiable.

Lemma 2. *Let A and B be positive definite matrices of the same dimension. Then, the eigenvalues of AB are positive.*

Proof. By the spectral decomposition, $A = P\Lambda P^T$ where $\Lambda = diag(\lambda_1, \dots, \lambda_p)$ and $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A. Then $A^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P^T$ is symmetric $\Rightarrow v^T A^{\frac{1}{2}} B A^{\frac{1}{2}} v = (A^{\frac{1}{2}}v)^T B (A^{\frac{1}{2}}v) > 0$. So $A^{\frac{1}{2}}BA^{\frac{1}{2}}$ is positive definite. Since $A^{\frac{1}{2}}(A^{\frac{1}{2}}BA^{\frac{1}{2}})A^{-\frac{1}{2}} = AB$, $A^{\frac{1}{2}}BA^{\frac{1}{2}}$ and AB are similar. Then they have the same eigenvalues and the eigenvalues of AB are positive. □

Let $B = a_0 I - (-r^{-1}P_0P^{-1})$. Then

$$h(a_0) = \frac{1}{2}a_0\delta^T P^{-1}B^{-1}\delta = \frac{\frac{1}{2}a_0\delta^T P^{-1} adj(B)\delta}{det(B)},$$

where $adj(B)$ is the cofactor matrix of B. The entries of $adj(B)$ are polynomials in a_0, so $\frac{1}{2}\delta^T a_0 P^{-1} adj(B)\delta$ is a polynomial in a_0 and thus differentiable. Then we show that $\det(B)^{-1}$ is differentiable on $(0,1)$. Since $\det(B)$ is a polynomial of a_0, it suffices to show that it is nonzero on $(0,1)$. Note that $\det(B)$ is the characteristic polynomial of $-r^{-1}P_0P^{-1}$. Since P_0 and P^{-1} are positive definite, $-r^{-1}P_0P^{-1}$ has negative eigenvalues by Lemma 2. So $\det(B)$ is nonzero on $(0,1)$. Thus, we have shown $h(a_0)$ is differentiable.

We then proceed to show that $h'(a) > 0$.

Let $E = P_0 + a_0 rP$. Then $h(a_0) = \frac{1}{2}a_0 r\delta^T E^{-1}\delta$. Therefore,

$$h'(a_0) = \frac{1}{2}r\delta^T E^{-1}\delta + a_0 r\frac{1}{2}\delta^T(E^{-1})'\delta.$$

We know that $(E^{-1})' = E^{-1}E'E^{-1} = E^{-1}PE^{-1}$. Since P is positive definite and E is symmetric, $v^T E^{-1}PE^{-1}v = (E^{-1}v)^T PE^{-1}v > 0$. Thus, $E^{-1}PE^{-1}$ is positive definite. Then $a_0 r\frac{1}{2}\delta^T(E^{-1})'\delta > 0$. Since E^{-1} is positive definite, $\frac{1}{2}r\delta^T E^{-1}\delta > 0$. So $h'(a_0) > 0$.

We also show $h'(a_0)$ is continuous. It suffices to show that $\det(E)$ is nonzero on $[0,1]$. Since $E = rBP$ where P is full rank, $\det(E) = c\det(B)$ where $c \neq 0$. Since $\det(B)$ is nonzero, $\det(E)$ is also nonzero.

Next, we will show that $f(a_0) = a_0^{\frac{p}{2}}\det(a_0 I - (-r^{-1}P_0P^{-1}))^{-\frac{1}{2}} = a_0^{\frac{p}{2}}\det(B)^{-\frac{1}{2}}$ is continuous on $[0,1]$. We have previously proven that $\det(B)$ is nonzero on $[0,1]$. Then $f(a_0)$ is continuous on $[0,1]$, and it will attain its minima and maxima on the closed interval. Let $t_1 = \max_{[\epsilon,1]}(f(a_0))$ and $t_2 = \min_{[\frac{\epsilon}{2},\epsilon]}(f(a_0))$. Since $a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}$ is continuous on $[\frac{\epsilon}{2},\epsilon]$, denote its minimum by t_3.

We write $A = A_1 + A_2$ where

$$\begin{aligned} A_1 &= \int_0^\epsilon a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0)\exp(-nh(a_0))da_0 \quad \text{and} \\ A_2 &= \int_\epsilon^1 a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0)\exp(-nh(a_0))da_0. \end{aligned}$$

Now we want to show that $\lim_{n\to\infty}\frac{A_2}{A_1} = 0$.

First, we find the upper bound of A_2. Since $h(a_0)$ is monotone increasing, $\exp(-nh(a_0)) \le \exp(-nh(\epsilon))$. Since $f(a_0) \le t_1$, we have

$$\begin{aligned} A_2 &\le t_1 \exp(-nh(\epsilon)) \int_\epsilon^1 a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}da_0, \\ &\le t_1 \exp(-nh(\epsilon)) \int_0^1 a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1}da_0, \\ &= t_1 \exp(-nh(\epsilon)) \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0+\beta_0)}, \\ &= C_1 \exp(-nh(\epsilon)). \end{aligned}$$

Next, we find the lower bound of A_1. We have previously shown that $h'(a_0)$ is continuous on $(0,1)$. Then $h'(a_0)$ attains its maximum on $[\frac{\epsilon}{2},\epsilon]$. Let

$t_4 = \max_{[\frac{\epsilon}{2},\epsilon]}(h'(a_0))$. We can write

$$
\begin{aligned}
A_1 &\geq \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} f(a_0)\exp(-nh(a_0))da_0, \\
&\geq \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} \frac{f(a_0)}{h'(a_0)}\exp(-nh(a_0))h'(a_0)da_0, \\
&= \int_{\frac{\epsilon}{2}}^{\epsilon} a_0^{\alpha_0-1}(1-a_0)^{\beta_0-1} \frac{f(a_0)}{h'(a_0)}\exp(-nh(a_0))dh(a_0), \\
&\geq \frac{t_2t_3}{t_4}\int_{\frac{\epsilon}{2}}^{\epsilon} \exp(-nh(a_0))dh(a_0), \\
&= \frac{t_2t_3}{t_4}\frac{1}{n}[\exp(-nh(\epsilon/2) - \exp(-nh(\epsilon))], \\
&= C_2\frac{1}{n}[\exp(-nh(\epsilon/2) - \exp(-nh(\epsilon))].
\end{aligned}
$$

Therefore,

$$
0 \leq \frac{A_2}{A_1} \leq \frac{C_1\exp(-nh(\epsilon))}{C_2\frac{1}{n}[\exp(-nh(\epsilon/2)) - \exp(-nh(\epsilon))]} = \frac{C_1 n}{C_2[\exp(-n[h(\epsilon/2) - h(\epsilon)]) - 1]},
$$

and $\lim\limits_{n\to\infty} \frac{A_2}{A_1} \to 0$ by L'Hopital's rule. Since $\frac{A_2}{A_1} \geq \frac{A_2}{A}$, $\lim\limits_{n\to\infty} \frac{A_2}{A} \to 0$. Then $\lim\limits_{n\to\infty} \frac{A_1}{A} \to 1$. □

S1.6 Proof of Corollary 2.3

Proof. Based on the assumptions, we have $\Sigma = a_0\Sigma_0$. The result follows if we plug $\delta = 0$ into (S1.5). □

S1.7 Proof for Theorem 2.4

Proof. The Laplace approximation for multiple parameters has the form

$$\int \exp(-nf(\beta))d\beta \approx \exp(-nf(\hat{\beta}))\left(\frac{2\pi}{n}\right)^{p/2}|\hat{\Sigma}|^{1/2},$$

where $\hat{\beta}$ is maximizes $-f(\beta)$, and $\hat{\Sigma}_{p\times p} = \left[\frac{\partial^2 f(\hat{\beta})}{\partial\beta_j\partial\beta_k}\right]^{-1}$.

When $X'Y = X_0'Y_0$ and $X \neq X_0$,

$$\begin{aligned}
\pi(a_0|D, D_0) &\propto \int L_n(D|\beta)\frac{L_{n_0}(D_0|\beta)^{a_0}\pi_0(\beta)}{c(a_0)}\pi(a_0)d\beta, \\
&= \int L_n(D|\beta)\frac{\exp\left(a_0\left[\sum_{i=1}^n y_i x_i'\beta - \sum_{i=1}^n b(x_{0i}'\beta)\right]\right)}{\int \exp\left(a_0\left[\sum_{i=1}^n y_i x_i'\beta - \sum_{i=1}^n b(x_{0i}'\beta)\right]\right)d\beta}\pi(a_0)d\beta, \\
&= \pi(a_0)\frac{\int L_n(D|\beta)L_{n_0}^*(D_0|\beta, a_0)d\beta}{\int L_{n_0}^*(D_0|\beta, a_0)d\beta}, \\
&= \pi(a_0)\frac{c_1(a_0)}{c_2(a_0)}.
\end{aligned}$$

Define

$$\begin{aligned}
g_n(\beta) &= -\frac{1}{n}[L_n(D|\beta) + a_0 L_{n_0}^*(D_0|\beta, a_0)] \\
&= -\frac{1}{n}\{\log(Q(Y)) + \sum_{i=1}^n y_i x_i'\beta - \sum_{i=1}^n b(x_i'\beta) + a_0[\sum_{i=1}^n y_i x_i'\beta - \sum_{i=1}^n b(x_{0i}'\beta)]\} \\
&= -\frac{1}{n}\{\log(Q(Y)) + (a_0+1)\sum_{i=1}^n y_i x_i'\beta - \sum_{i=1}^n b(x_i'\beta) - a_0\sum_{i=1}^n b(x_{0i}'\beta)\}.
\end{aligned}$$

Then we have

$$c_1(a_0) \approx \exp(-ng_n(\hat{\beta}))\left(\frac{2\pi}{n}\right)^{p/2}|\hat{\Sigma}_g|^{1/2},$$

where $\hat{\beta}$ maximizes $-g_n(\beta)$. Similarly, define

$$\begin{aligned} k_n(\beta) &= -\frac{1}{n} a_0 l^*(y|x_0, \beta), \\ &= -\frac{1}{n}\left\{a_0 \sum_{i=1}^{n} y_i x_i' \beta - a_0 \sum_{i=1}^{n} b(x_{0i}'\beta)\right\}. \end{aligned}$$

Then we have

$$c_2(a_0) \approx \exp(-nk_n(\tilde{\beta}))\left(\frac{2\pi}{n}\right)^{p/2} |\tilde{\Sigma}_g|^{1/2},$$

where $\tilde{\beta}$ maximizes $-k_n(\beta)$.

We compute the gradients of $g_n(\beta)$ and $k_n(\beta)$ and get

$$\begin{aligned} \nabla g_n(\beta) &= -\frac{1}{n}\{(a_0+1)\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \dot{b}(x_i'\beta)x_i - a_0 \sum_{i=1}^{n} \dot{b}(x_{0i}'\beta)x_{0i}\}, \\ \nabla k_n(\beta) &= -\frac{1}{n}\{a_0 \sum_{i=1}^{n} y_i x_i - a_0 \sum_{i=1}^{n} \dot{b}(x_{0i}'\beta)x_{0i}\}, \\ \nabla g_n(\beta) = 0 &\Rightarrow \sum_{i=1}^{n} \dot{b}(x_i'\hat{\beta})x_i + a_0 \sum_{i=1}^{n} \dot{b}(x_{0i}'\hat{\beta})x_{0i} = (a_0+1)\sum_{i=1}^{n} y_i x_i, \\ \nabla k_n(\beta) = 0 &\Rightarrow \sum_{i=1}^{n} \dot{b}(x_{0i}'\tilde{\beta})x_{0i} = \sum_{i=1}^{n} y_i x_i. \end{aligned}$$

We can see that asymptotically, $\hat{\beta} \neq \tilde{\beta}$. Then we have

$$\frac{c_1(a_0)}{c_2(a_0)} = \frac{|\hat{\Sigma}_g|^{1/2}}{|\tilde{\Sigma}_k|^{1/2}} \exp\{-n[g_n(\hat{\beta}) - k_n(\tilde{\beta})]\}, \tag{S1.6}$$

where

$$
\begin{aligned}
\hat{\Sigma}_g &= \left[\frac{1}{n}\sum_{i=1}^{n}\ddot{b}(x_i'\hat{\beta})x_i x_i' + \frac{a_0}{n}\sum_{i=1}^{n_0}\ddot{b}(x_{0i}'\hat{\beta})x_{0i}x_{0i}'\right]^{-1}, \\
\tilde{\Sigma}_k &= \left[\frac{a_0}{n}\sum_{i=1}^{n_0}\ddot{b}(x_{0i}'\tilde{\beta})x_{0i}x_{0i}'\right]^{-1}, \\
\frac{|\hat{\Sigma}_g|^{1/2}}{|\tilde{\Sigma}_k|^{1/2}} &= \frac{|a_0\sum_{i=1}^{n_0}\ddot{b}(x_{0i}'\tilde{\beta})x_{0i}x_{0i}'|^{1/2}}{|\sum_{i=1}^{n}\ddot{b}(x_i'\hat{\beta})x_i x_i' + a_0\sum_{i=1}^{n_0}\ddot{b}(x_{0i}'\hat{\beta})x_{0i}x_{0i}'|^{1/2}}.
\end{aligned}
$$

The marginal posterior of a_0 is then proportional to S1.6 multiplied by $\pi(a_0)$. $\square$

S2 Additional Figures for Section 2

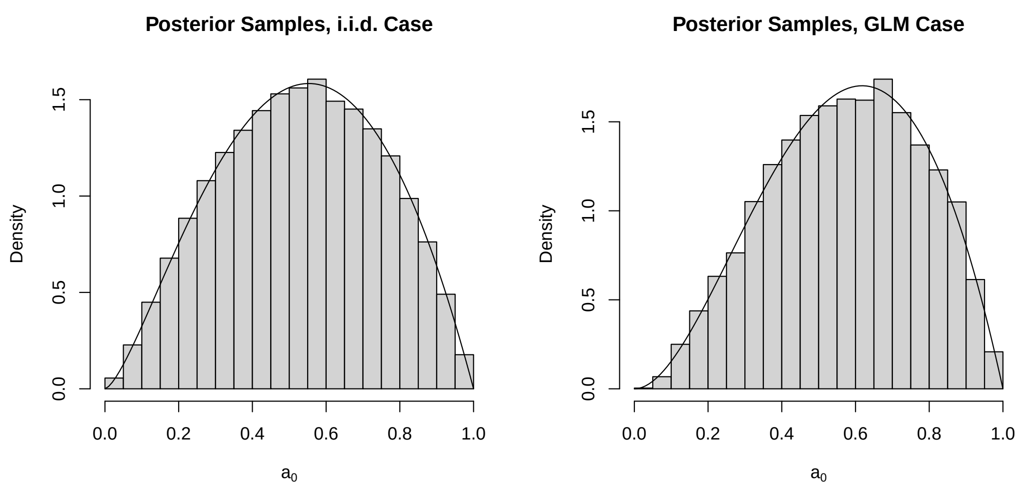

Figure 1: The plot on the left shows the histogram of the posterior of a_0 for i.i.d. Bernoulli data with current and historical mean equal to 0.7, $n = 100$, $n_0 = 200$ and the prior on a_0 is beta(2, 2). The plot on the right shows the histogram of the posterior of a_0 for Bernoulli data with one covariate where the historical and current data are identical. The prior on a_0 is beta(2, 2). The histograms of the posterior samples are produced using R package BayesPPD. The curve represents the theoretical density. We observe that for both i.i.d. and GLM cases, the histograms of posterior samples agree with the theoretical density functions.

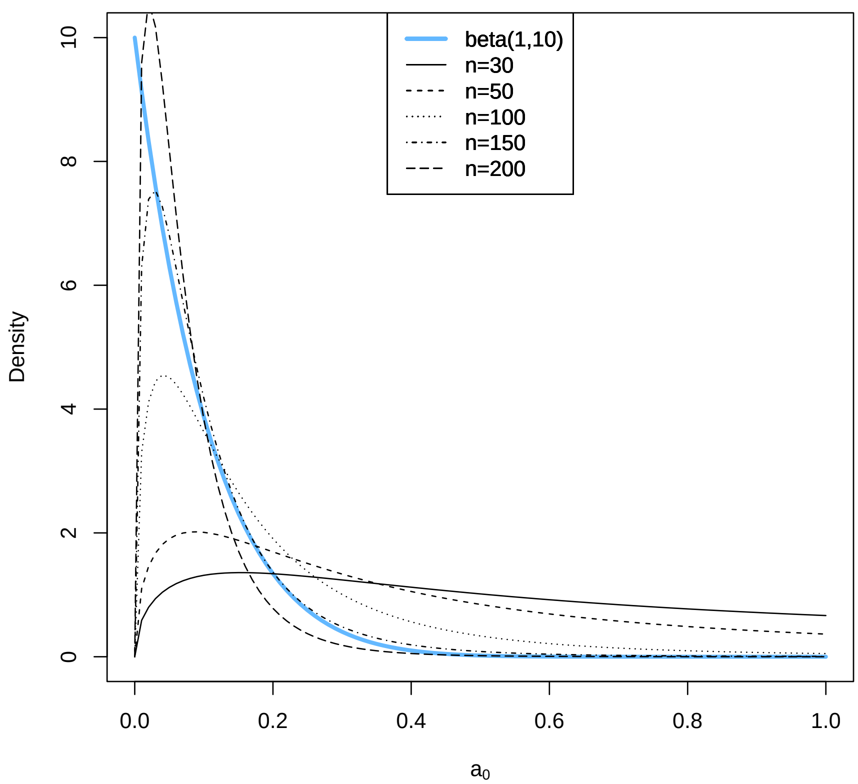

Figure 2: Marginal posterior of a_0 for i.i.d. normal data where $n = n_0$ increases from 30 to 200, the historical data mean is 1.5, the current data mean is 2 and the standard deviations are 1. We observe that when there is some difference between the sufficient statistics of the historical and current data, the marginal posterior of a_0 converge to a point mass at zero quickly.

S3 Numerical Stability of the Optimization Process

We conduct a simple experiment to reproduce the optimal priors derived in Figure 1 when we fix one of α_0 or β_0 and optimize for the other parameter. We can see in Table 1 below that the optimization is stable and reliable.

Table 1: Optimization with one of the hyperparameters fixed

	optimal priors in Fig. 1	optimal priors with fixed α_0	optimal priors with fixed β_0
$d_{MTD} = 0.5$	beta(2.2, 2.3)	beta(2.2, 2.3)	beta(2.2, 2.3)
$d_{MTD} = 1$	beta(1, 0.4)	beta(1, 0.5)	beta(0.9, 0.4)
$d_{MTD} = 1.5$	beta(2.6, 0.5)	beta(2.6, 0.5)	beta(2.6, 0.5)

S4 Bias and Variance Decomposition for the MSE Criterion

Table 2: Bias and variance decomposition for different prior choices

	Optimal Prior	Beta$(1,1)$	Beta$(2,2)$
Bias^2			
$d_{\text{MTD}} = 0.5$	0.011	0.015	0.018
$d_{\text{MTD}} = 1$	0.005	0.012	0.025
$d_{\text{MTD}} = 1.5$	0.003	0.006	0.015
Variance			
$d_{\text{MTD}} = 0.5$	0.043	0.042	0.039
$d_{\text{MTD}} = 1$	0.058	0.057	0.054
$d_{\text{MTD}} = 1.5$	0.049	0.053	0.052

S5 Comparisons with Other Priors

In Figure 3, we generate i.i.d. normal data and compute the MSE based on the posterior mean of the point estimator using three different prior choices, the NPP with the optimal beta prior on a_0 (optimal in the sense of minimizing MSE as defined in the main paper), the NPP with a mixture of two beta priors on a_0, and the robust mixture prior, which is a special case of the robust meta-analytic-predictive prior introduced in Schmidli et al. (2014). The robust mixture prior we use places equal weights on a non-informative normal component and an informative normal component using the historical data. For the NPP with a mixture of beta priors, we use a mixture of beta(1, c) and beta(c, 1) with equal weights, where c ranges from 100 to 1000. As c approaches infinity, the mixture of beta priors on a_0 converges to a mixture of a point mass at zero and a point mass at one, which is equivalent to the robust mixture prior. We vary the difference between the observed current data mean and the historical data mean, i.e., $d_{obs} = \bar{y}_{obs} - \bar{y}_0$. We can see that when the data are compatible ($d_{obs} = 0.5$), the posterior mean based on the NPP with the optimal beta prior produces lower MSE than the estimator based on the robust mixture prior. When the conflict between the data increases, i.e., $d_{obs} = 1$ and $d_{obs} = 1.5$, the NPP with a mixture of beta priors outperforms the robust mixture prior.

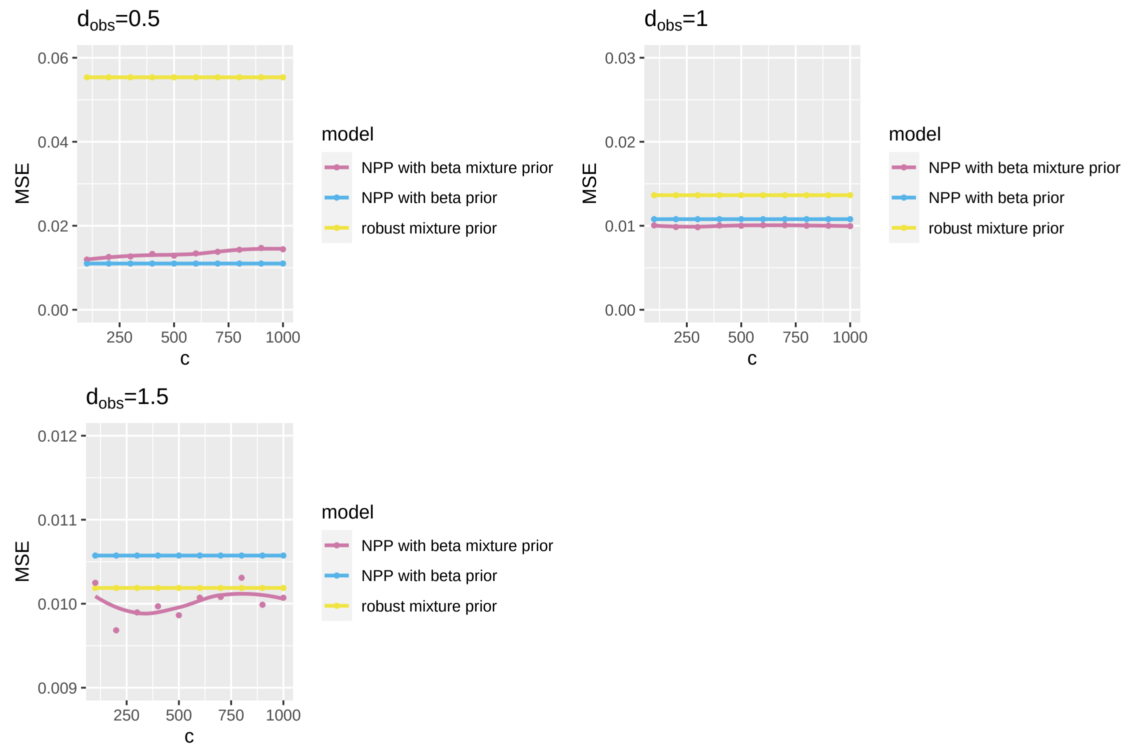

Figure 3: MSE using three different prior choices, the NPP with the optimal beta prior on a_0, the NPP with the mixture of beta priors on a_0 and the robust mixture prior

S6 Additional Simulations for MSE Criterion

Figures 4 and 5 show the MSE as a function of the prior mean of a_0 for increasing ratios of n/n_0 when the total sample size is fixed. We observe that as n/n_0 increases, the model will increasingly benefit, i.e. the MSE is reduced, from borrowing more, but this trend is less prominent when the total sample size is larger.

The total sample size of the PLUTO trials in section 4.1 is about twice the total sample size of the melanoma trials in section 4.2. The total sample size of the melanoma trials is not large enough for the model to criticize the maximal tolerable difference that we chose. Therefore, the optimal prior derived using the MSE criterion encourages borrowing for the melanoma trial.

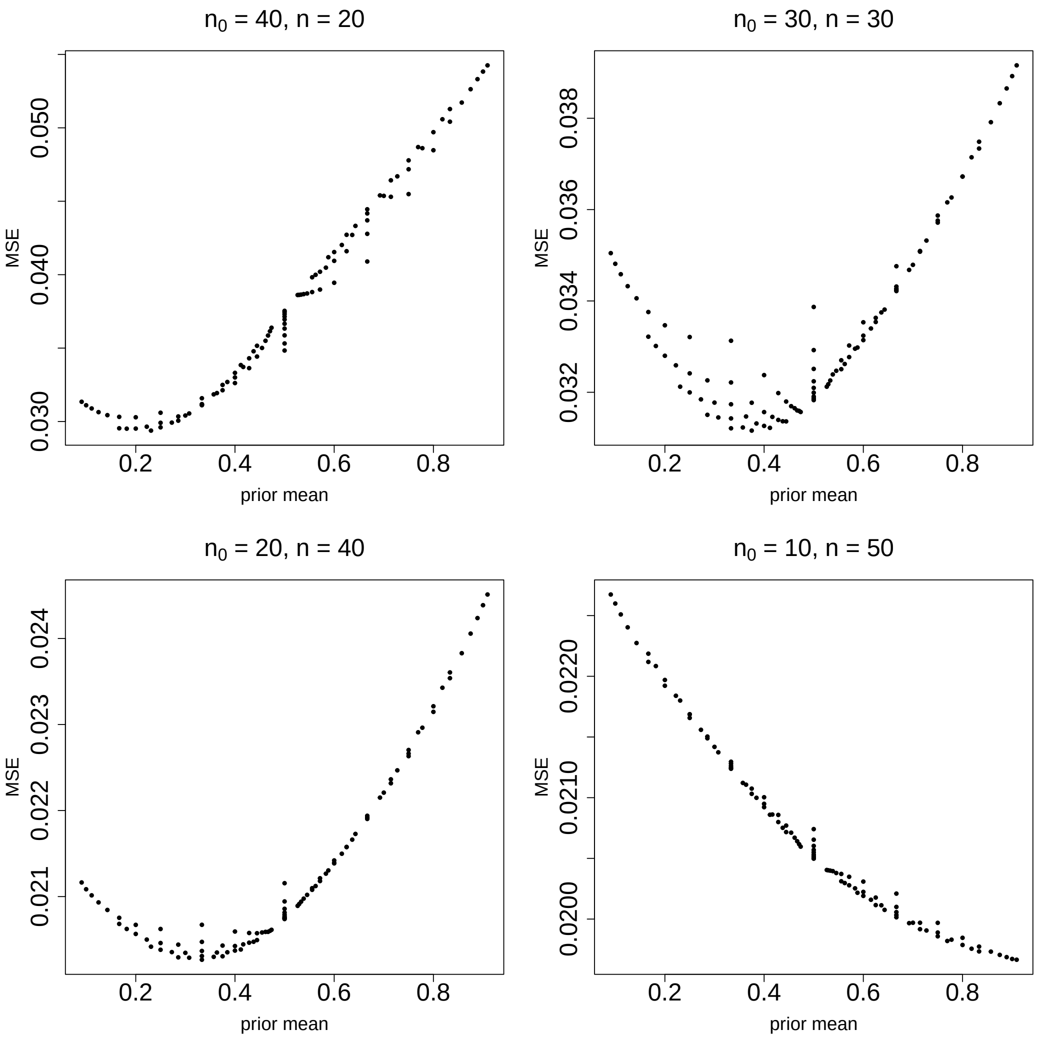

Figure 4: MSE as a function of prior mean of a_0 for increasing ratios of n/n_0 when the total sample size is fixed for the normal *i.i.d.* case.

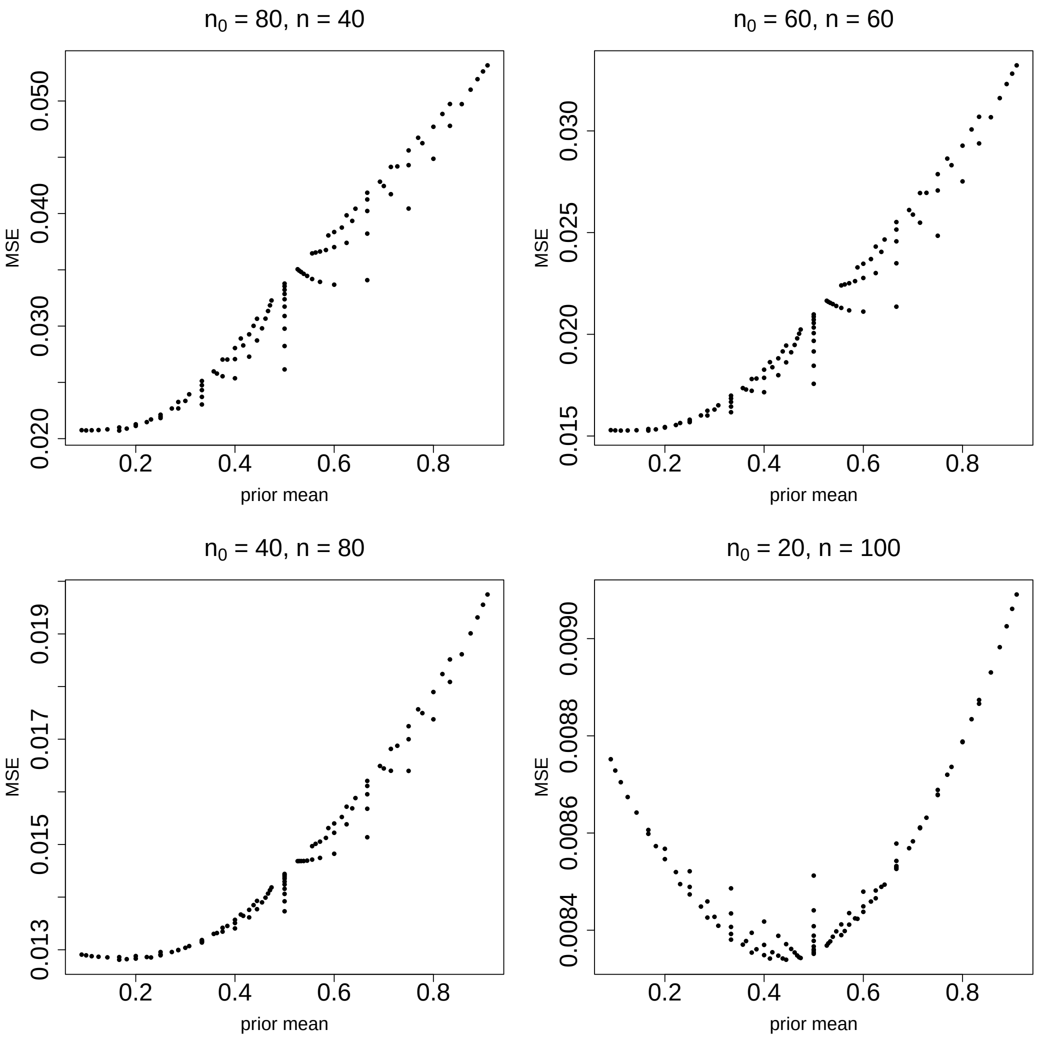

Figure 5: MSE as a function of prior mean of a_0 for increasing ratios of n/n_0 when the total sample size is double the total sample size in Figure 4 for the normal *i.i.d.* case.

S7 Design Application for the Pediatric Lupus Trial

Now we demonstrate using the proposed optimal priors in a clinical trial design application. Suppose we want to design a pediatric trial using data from the adult trials BLISS-52 and BLISS-76. We choose a few sample sizes ranging from 50 to 100 (the actual trial had a sample size of 92) and derive the optimal prior for each sample size using both the KL and MSE criteria. We compute power using the R Package *BayesPPD* which performs Bayesian sample size determination with a simulation-based procedure (Shen et al., 2023). We use the posterior samples given only the historical data as the discrete approximation to the sampling prior (Psioda and Ibrahim, 2019). For the fitting prior, we use a normalized power prior with optimal priors derived for a_0. Figure 6 shows the power curves for three choices of priors on a_0, the optimal prior using the KL criterion, the optimal prior using the MSE criterion, and the uniform prior. Note the optimal prior is derived for each sample size. In this case, the optimal priors do not vary much for different sample sizes due to the small sizes of the current trial relative to the adult trials. We can see that power is the highest when we optimize to minimize KL. Since the optimal prior on a_0 based on the KL criterion is beta(5.5, 5.5) (when $n = 100$), the most amount of historical information is borrowed. Power is the lowest when we optimize to minimize MSE, since the least amount of historical information is borrowed. The two criteria address the problem of how much to borrow from different angles. The KL criterion focuses on how much information one is willing to borrow under two markedly different assumptions about the difference between the prior information and the data generation process for the future study. The KL criterion does not explicitly focus on estimation performance metric. On the other hand, the MSE criterion attempts to ensure that the point estimate of the parameter of interest behaves well in terms of the trade-off between bias and variance. Also note that the two target distributions for the KL criteria used in this application are beta(1, 10) and beta(10, 1). These target distributions should be carefully chosen so that they reflect the desired posterior distributions of a_0 relative to the sample sizes of the historical and current data. For example, by considering $c = 10$ one is targeting borrowing approximately 10% of the prior information when the prior-data conflict is substantial (i.e., in line with d_{MTD}). If 10% of the historical data sample size is large relative to the new study sample size being considered, this choice for c may not be desirable (i.e., a larger c would be warranted).

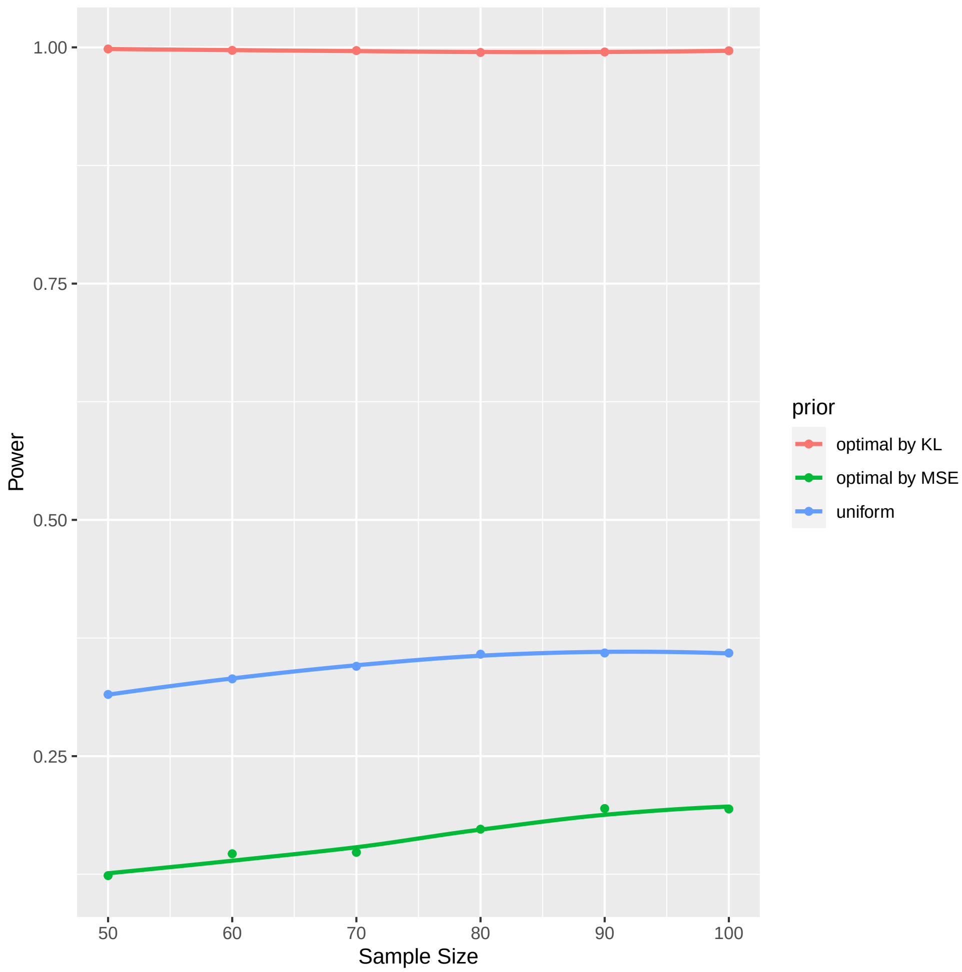

Figure 6: Power curves using three choices of priors on a_0, optimal prior using the KL criterion, optimal prior using the MSE criterion, and the uniform prior. A different optimal prior is derived for each sample size.

Bibliography

Chen, C.-F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *Journal of the Royal Statistical Society. Series B (Methodological) 47*(3), 540–546.

Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics 13*(1), 342–368.

Psioda, M. A. and J. G. Ibrahim (2019, Jul). Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics 20*(3), 400–415.

Schervish, M. J. (1995). *Theory of statistics*. Springer Verlag.

Schmidli, H., S. Gsteiger, S. Roychoudhury, A. O'Hagan, D. Spiegelhalter, and B. Neuenschwander (2014, Dec). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics 70*(4), 1023–1032.

Shen, Y., M. A. Psioda, and J. G. Ibrahim (2023). The R journal: BayesPPD: An R package for bayesian sample size determination using the power and normalized power prior for generalized linear models. *The R Journal 14*, 335–351. https://doi.org/10.32614/RJ-2023-016.