The effect of informative stopping rules on properties of Bayes estimators

Sergey Tarima
(University of Kentucky)

and
Nancy Flournoy
(University of Missouri)

Compiled on June 18, 2026

Introduction

- This work is motivated by the Bayes practice that advocates ignoring informative interim stopping rules.
- This work formalizes the concept of channeled dependencies for sequential sampling densities with interim stopping rules.

  It is with channeled dependent data, that Bayesian posterior analyses depend on informative interim decisions only through the second stage sample size.
- We study conditional posterior densities for channeled dependent data as an alternative.
- In contrast to the Bayesian posterior, the conditional posterior densities are sensitive to interim decision rules.
- Simulations show
  - the mean conditional posterior *often* outperforms the mean posterior in terms of frequentist MSE and Integrated MSE (Bayes risk).
  - both estimators are suboptimal for Bayes risk.

Two-Stage Sampling

- $\theta \in (-\infty, +\infty)$
- $(\mathbf{X_1}, \mathbf{X_2}) = (X_1, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_{n_1+n_2})$ are rvs with joint p.d.f. $f(\mathbf{x_1}, \mathbf{x_2}|\theta)$
- $(\mathbf{x_1}, \mathbf{x_2}) = (x_1, \ldots, x_{n_1}, x_{n_1+1}, \ldots, x_{n_1+n_2})$ is a realization of $(\mathbf{X_1}, \mathbf{X_2})$.

Density and Channeled Dependence

Density including an interim decision $D = \mathrm{I}\{\text{stop at stage } 1\} = d$ is

$$f(\mathbf{x_1}, d, \mathbf{x_2}|\theta) = f(d|\theta)\ f(\mathbf{x_1}|d,\theta)\ f(\mathbf{x_2}|\mathbf{x_1}, d, \theta).$$

Definition: (channeled dependence)

- Suppose after $\mathbf{X_1} = \mathbf{x_1}$ is observed, an interim decision is made that is fully determined by $\mathbf{x_1}$: $D(\mathbf{x_1}) = d$.
- We say that the *dependence on previous data is channeled* through D if $f(\mathbf{x_2}|\mathbf{x_1}, d, \theta) = f(\mathbf{x_2}|d,\theta)$ and $f(d|\mathbf{x_1},\theta) = 1$

If the dependence of $\mathbf{X_2}$ on $\mathbf{X_1}$ is channeled through $D = D(\mathbf{X_1}) = \mathrm{I}\{\text{stop at stage } 1\}$, then $f(d|\theta)\ f(\mathbf{x_1}|d,\theta) = f(\mathbf{x_1}, d|\theta) = f(\mathbf{x_1}|\theta)$ and $f(\mathbf{x_2} \mid \mathbf{x_1}, d, \theta) = [f(\mathbf{x_2} \mid \theta)]^{d-1}$

Density with channeled through an interim stopping decision $D = d$ is

$$f(\mathbf{x_1}, d, \mathbf{x_2}|\theta) = f(\mathbf{x_1}|\theta)[f(\mathbf{x_2} \mid \theta)]^{d-1}.$$

Likelihood and Channeled Dependence

Likelihood $\mathcal{L}(\theta \mid \mathbf{x_1}, \lceil, \mathbf{x_2}) = \{(\mathbf{x_1}, \lceil, \mathbf{x_2}|\theta) = \{(\mathbf{x_1}|\theta)[\{(\mathbf{x_2} \mid \theta)]^{\lceil - \infty}$.

Example (gaussian data, $X_i \sim N(\theta, \sigma^2)$):

Fixed sample size (n)

- MLE ($\bar{X}_n$) is unbiased,
- $\bar{X}_n$ is sufficient and $\bar{X}_n \sim N(\theta, \sigma^2/n)$
- Information matrix ($\mathcal{I} = n/\sigma^2$)

With informative stopping rule D (efficacy stopping, random sample size $N = n_1 + (d-1)n_2$)

- MLE ($\bar{X}_N$) is biased; Bias is $O(n_1^{-1/2})$
- $\bar{X}_N$ is not sufficient; $(D, \bar{X}_N)$ is sufficient; $f_{D,\bar{X}_N}(d, \bar{x}_N \mid \theta)$ belongs to a curved exponential family
- $\bar{X}_N$ is not normally distributed; Armitage's recursive formula from 1969 is used to find this distribution
- information matrix depends on θ and decomposes as

$$\mathcal{I}(\theta) = EN/\sigma^2 = \mathcal{I}_D(\theta) + \int \mathcal{I}_{MLE|D=d}(\theta) f_D(d \mid \theta) dd$$

Posterior with Channeled Dependence

The posterior of Θ is

$$f(\theta|\mathbf{x_1}, d, \mathbf{x_2}) = \frac{f(\theta) f(\mathbf{x_1}, d, \mathbf{x_2}|\theta)}{\int f(\theta) f(\mathbf{x_1}, d, \mathbf{x_2}|\theta) d\theta} \propto f(\theta) f(d|\theta)\ f(\mathbf{x_1}|d, \theta)\ f(\mathbf{x_2}|\mathbf{x_1}, d, \theta). \quad (1)$$

The posterior with channeled dependence and $D = \mathrm{I}\ \{\text{stop at stage 1}\}$.

- The posterior is

$$\begin{aligned} f(\theta|\mathbf{x_1}, d, \mathbf{x_2}) &\propto f(\theta) f(d|\theta) f(\mathbf{x_1}|d, \theta) f(\mathbf{x_2}|\mathbf{x_1}, d, \theta) && \text{original} \\ &= f(\theta) f(\mathbf{x_1}|\theta) f(\mathbf{x_2}|d, \theta) && \text{channeled dependence} \\ &= f(\theta) f(\mathbf{x_1}|\theta) [f(\mathbf{x_2}|\theta)]^{d-1} && \text{channeled dependence via } D = I. \end{aligned}$$

- The posterior above does not depend on the distribution of D; it only depends on its realization d.
- Non-mathematically speaking, the posterior "does not know" if a coin was flipped to stop the study or a legitimate futility or efficacy stopping rule was used.
- this invariance/insensitivity of the posterior to interim adaptation rules stems out of the controversy of relying on the likelihood principle in sequential settings

Conditional Posterior with $D = \mathrm{I}\{\text{stop at stage } 1\}$.

We induce dependence on interim decisions by considering conditional sampling distributions, rather than the full sampling distribution.

Specifically, instead of the density

$$f(\mathbf{x_1}, d, \mathbf{x_2}|\theta) = f(d|\theta)\ f(\mathbf{x_1}|d,\theta)\ f(\mathbf{x_2}|\mathbf{x_1}, d, \theta)\ \left(= f(\mathbf{x_1}|\theta)[f(\mathbf{x_2}|\theta)]^{d-1}\right),$$

we consider only the conditional on $D = d$ sampling density

$$f(\mathbf{x_1}, \mathbf{x_2}|d,\theta) = f(\mathbf{x_1}|d,\theta)\ f(\mathbf{x_2}|\mathbf{x_1}, d, \theta)\ \left(= f(\mathbf{x_1}|d,\theta)[f(\mathbf{x_2}|\theta)]^{d-1}\right).$$

Definition: ***The conditional posterior is***

$$f^{c}(\theta|\mathbf{x_1}, d, \mathbf{x_2}) \propto \begin{cases} f(\theta)f(\mathbf{x_1}|1,\theta) & d = 1; \\ f(\theta)f(\mathbf{x_1}|0,\theta)f(\mathbf{x_2}|\theta). & d = 0. \end{cases}$$

Three Estimators of θ

At $\Theta \sim N(\theta_0, n_0^{-1})$, $\bar{X}_1 \sim N(\theta, n_1^{-1})$, $\bar{X}_2 \sim N(\theta, n_2^{-1})$, $D = I(\sqrt{n_1}\bar{X}_1 > c_1)$ we consider

- **The MLE**

$$\bar{x} = \begin{cases} \bar{x}_1 & \text{if } D = 1 \\ \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} & \text{if } D = 0, \end{cases}$$

- **Bayes estimator θ^* which minimizes the expected posterior loss**:

$$\theta^* = \begin{cases} \dfrac{n_0\theta_0 + n_1 x_1}{n_0 + n_1} & \text{if } D = 1 \\ \dfrac{n_0\theta_0 + n_1\bar{x}_1 + n_2\bar{x}_2}{n_0 + n_1 + n_2} & \text{if } D = 0, \end{cases}$$

- **Conditional Bayes estimator**

$$\theta^{**} : \mathrm{E}_{\Theta|\bar{X}_N|D}[L(\theta^{**}, \theta)|d, \bar{x}_N] \leq \mathrm{E}_{\Theta|\bar{X}_N|D}[L(\widehat{\theta}, \theta)|d, \bar{x}_N] \qquad \text{given } D = d,$$

where $\mathrm{E}_{\Theta|\bar{X}_N|D}(\cdot|\cdot)$ indicates that the expectation is takes with respect to the conditional posterior $f^c(\theta|\mathbf{x}, d)$

Performance Metrics with Quadratic Loss: $L(\widehat{\theta}, \theta) = (\theta - \widehat{\theta})^2$

Frequentist Risk:

$$FMSE(\widehat{\theta}|\theta) = \int \left(\theta - \widehat{\theta}(\mathbf{x})\right)^2 f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)d\mathbf{x}.$$

Bayes Risk:

$$IMSE(\widehat{\theta}, \Theta) = \int \left[\int \left(\theta - \widehat{\theta}(\mathbf{x})\right)^2 f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)d\mathbf{x}\right] f_\Theta(\theta)d\theta.$$

Frequentist and Integrated MSE with Early Stopping Indicator D

For completeness (you don't need to read), we define

$$
\begin{aligned}
FMSE(\widehat{\theta}|\theta) = {} & \Pr(D=1|\theta)\int_{\mathbf{x_1}:D=1}\left(\theta-\widehat{\theta}(\mathbf{x_1})\right)^2 f(\mathbf{x_1}|D=1,\theta)d\mathbf{x_1} \\
& + \Pr(D=0|\theta)\int_{\mathbf{x_1},\mathbf{x_2}:D=0}\left(\theta-\widehat{\theta}(\mathbf{x_1},\mathbf{x_2})\right)^2 f(\mathbf{x_1},\mathbf{x_2}|D=0,\theta)d\mathbf{x_1}d\mathbf{x_2}.
\end{aligned}
$$

and

$$
\begin{aligned}
IMSE(\widehat{\theta},\theta) = \int_{-\infty}^{\infty} \Bigg[& \Pr(D=1|\theta)\int_{\mathbf{x_1}:D=1}\left(\theta-\widehat{\theta}(\mathbf{x_1})\right)^2 f(\mathbf{x_1}|D=1,\theta)d\mathbf{x_1} \\
& + \Pr(D=0|\theta)\int_{\mathbf{x_1},\mathbf{x_2}:D=0}\left(\theta-\widehat{\theta}(\mathbf{x_1},\mathbf{x_2})\right)^2 f(\mathbf{x_1},\mathbf{x_2}|D=0,\theta)d\mathbf{x_1}d\mathbf{x_2}\Bigg] f(\theta)d\theta.
\end{aligned}
$$

where $\widehat{\theta} \in \{\theta^{**}, \theta^{*}, \bar{x}\}$.

Simulations comparing IMSEs of θ^{**}, θ^{*} and MLE.

- We plot $\dfrac{IMSE(\theta^*)}{IMSE(\bar{x})}$ (red) and $\dfrac{IMSE(\theta^{**})}{IMSE(\bar{x})}$ (blue) as functions of a critical value c_1.
- We consider $\Theta \sim N(\theta_0, n_0^{-1})$ with $\bar{X}_j | \{\Theta = \theta\} \sim N(\theta, n_j^{-1}),\ j = 1, 2$
- Here $D(\mathbf{X_1}) = \mathrm{I}\{\sqrt{n_1}\bar{X}_1 > c_1\}$ and c_1 is the *cutoff* value.

We used Monte-Carlo simulations ($M = 5000$) to approximate IMSE with $M^{-1}\sum_{s=1}^{M}(\widehat{\theta}_s - \theta)^2$ with random $\theta \sim \Theta$, where $\widehat{\theta} \in \{\bar{x}, \theta^*, \theta^{**}\}$

Simulations (Strong Correct Prior, $n_0 = 10,000$)

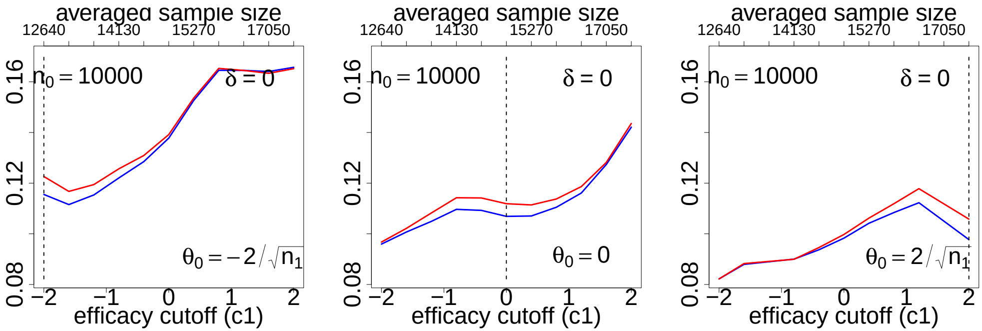

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 10000$; $n_1 = n_2 = 1000$.

Simulations (Strong Misspecified Prior, $n_0 = 10,000$)

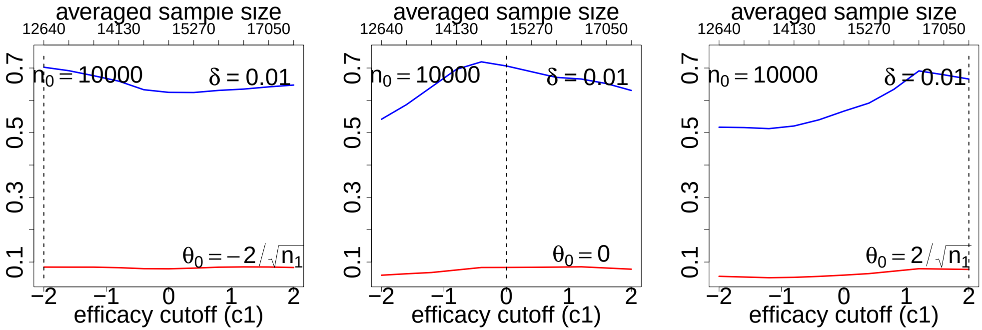

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 10000$; $n_1 = n_2 = 1000$.

Simulations (Moderate correctly specified Prior, $n_0 = 1,000$)

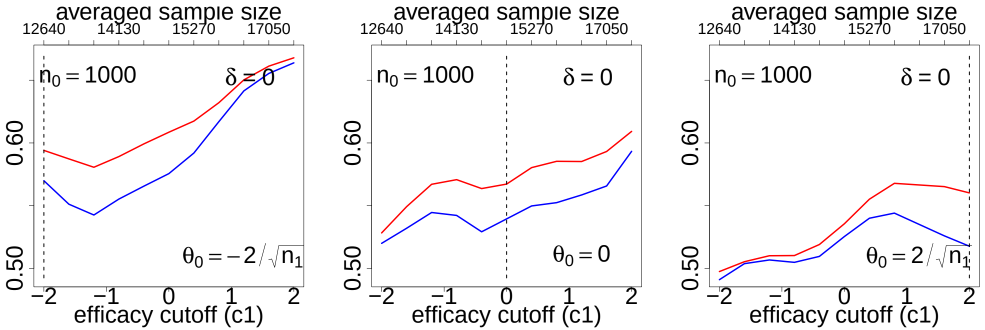

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 1000$; $n_1 = n_2 = 1000$.

Simulations (Moderate misspecified Prior, $n_0 = 1,000$)

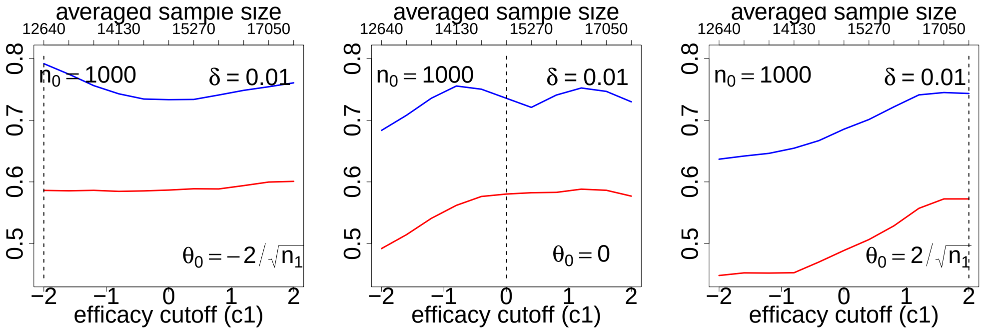

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 1000$; $n_1 = n_2 = 1000$.

Simulations (Weak correctly specified Prior, $n_0 = 100$)

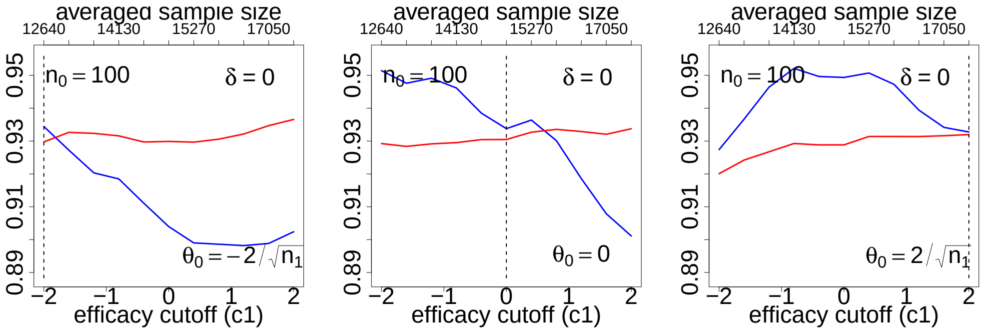

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 100$; $n_1 = n_2 = 1000$.

Simulations (Weak misspecified Prior, $n_0 = 100$)

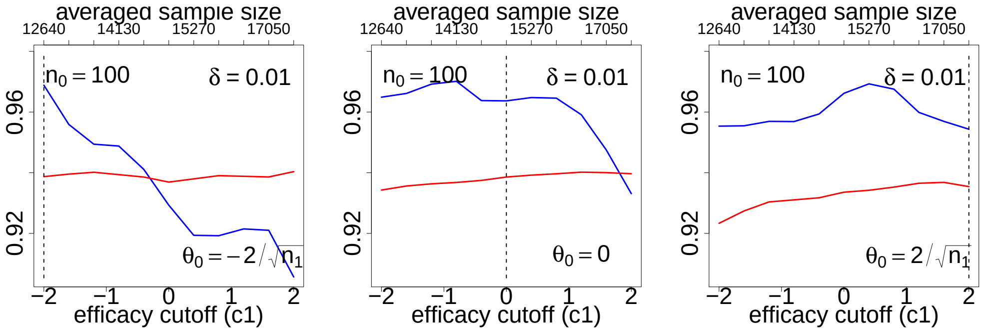

Figure: $IMSE(\theta^*)/IMSE(\bar{x})$ (red) and $IMSE(\theta^{**})/IMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0 + \delta, n_0^{-1})$ with $n_0 = 100$; $n_1 = n_2 = 1000$.

SUMMARY

- We formalized the concept of "channeled dependence" among sequential sampling densities.
- We explored the channeled posterior, which differs from the one currently used in practical applications.
- The channeled posterior is sensitive to interim adaptation rules, whereas the Bayesian posterior commonly used in sequential settings is not.
- Performance of conditional on a stopping decision and unconditional posteriors were compared using IMSE.
  - With a strong correctly specified prior, performance was uniformly better when the posterior was created with a sampling density that included the decision.
  - With a weak correctly specified prior, performance is mixed. We speculate why.
  - With a misspecified prior unconditional posterior typically outperforms the conditional posterior
- Minimum Bayes risk estimation continues to be an open problem in this context.

Simulations FMSE (Strong Correct Prior, $n_0 = 10,000$)

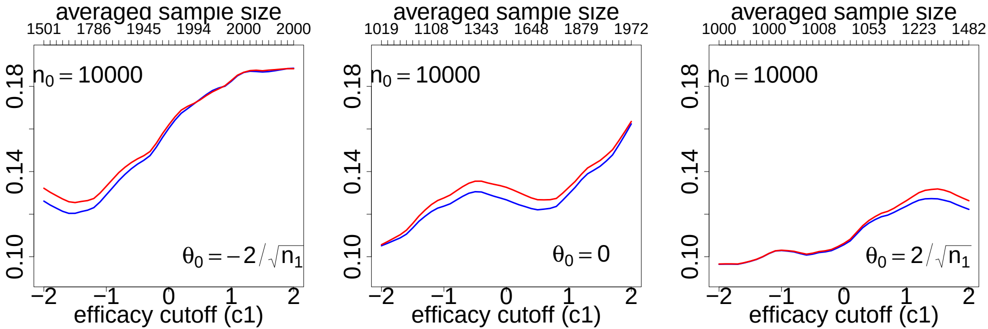

Figure: $FMSE(\theta^*)/FMSE(\bar{x})$ (red) and $FMSE(\theta^{**})/FMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0, n_0^{-1})$ with $n_0 = 10000$; $n_1 = n_2 = 1000$.

Simulations FMSE (Moderate Correct Prior, $n_0 = 1,000$)

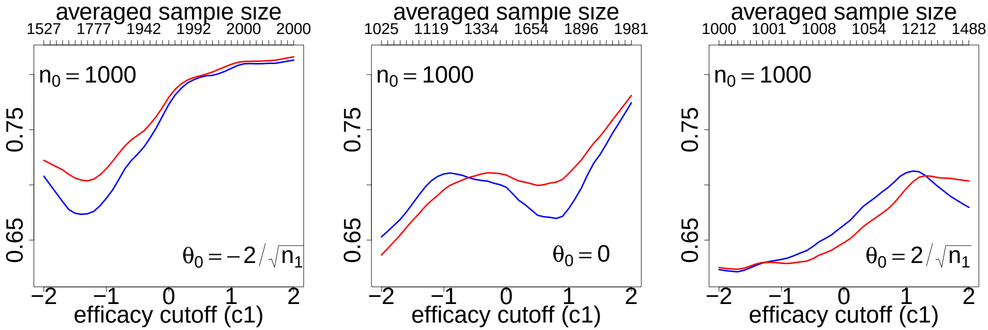

Figure: $FMSE(\theta^*)/FMSE(\bar{x})$ (red) and $FMSE(\theta^{**})/FMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0, n_0^{-1})$ with $n_0 = 1000$; $n_1 = n_2 = 1000$.

Simulations FMSE (Weak Correct Prior, $n_0 = 100$)

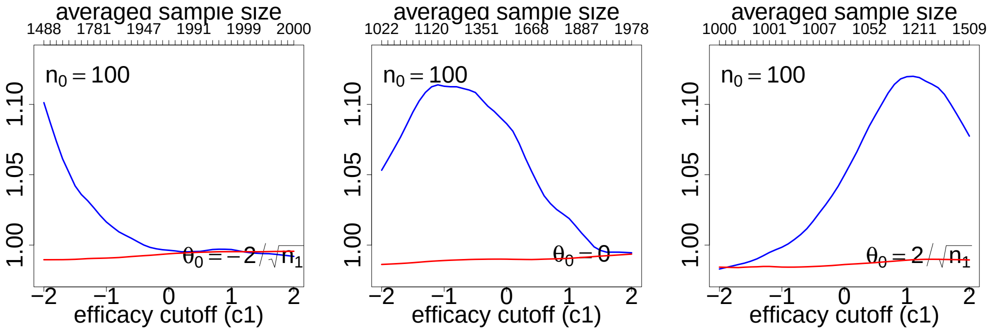

Figure: $FMSE(\theta^*)/FMSE(\bar{x})$ (red) and $FMSE(\theta^{**})/FMSE(\bar{x})$(blue), θ^* as a function of critical values c_1 for $\theta_0 = -2/\sqrt{n_1}$ (left panel), $\theta_0 = 0$ (center panel); $\theta_0 = 2/\sqrt{n_1}$ (right panel) plotted by the vertical dashed line. $\Theta \sim N(\theta_0, n_0^{-1})$ with $n_0 = 100$; $n_1 = n_2 = 1000$.