Unexpected results and challenges when using mixture priors for Bayesian borrowing.

Darren Scott

1. Effective sample size (ESS) - simple

- In the Bayesian borrowing context we require a metric to tell us how much "information" is in the prior.
- ESS: - "***measure of information in the prior, before the target study data have been realised***".
- In the simple setting;

$$Y|\theta \sim Binomal(n, \theta)$$

$$\theta \sim Beta(a, b)$$

$$p(\theta|x) \propto \binom{n}{x} \theta^x (1-\theta)^{n-x} \theta^a (1-\theta)^b$$

Is there a problem?

- We use 3 different general methods to estimate the ESS of our prior $p(\theta)$.
- We get the following;
    - Moment approach ESS = 1.7,
    - Morita[1] approach ESS = 878,
    - ELIR[2] approach ESS = 86.
- Why?

How could you calculate an ESS – general approach?

- Prior $p(\theta)$
- Pseudo-prior $q(\theta)$
- Likelihood $L(y_{1\times m}|\theta, \boldsymbol{\vartheta})$
- Posterior $q(\theta|y_{1\times m})$
- ESS is m which achieves $p(\theta) \approx q(\theta|y_{1\times m}, \boldsymbol{\vartheta})$

- How do you define $\approx$ as these are distributions?
- What do you do with the data in the posterior, as this is not observed?

Effective sample size - general approach

- The measure of information in the prior is the rate of change of the curvature.

$$I(\theta) = -\frac{\partial^2}{\partial\theta^2}\log p(\theta)$$

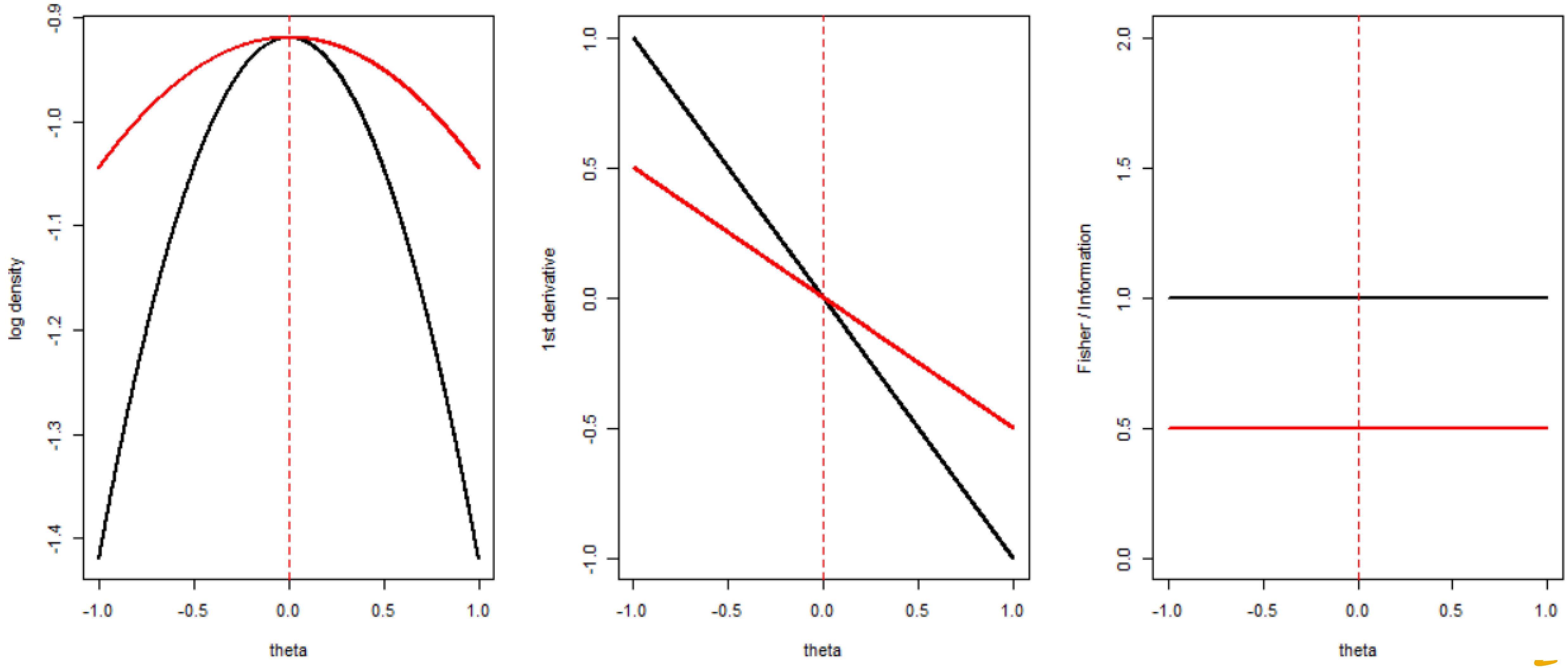
- Information in the likelihood is the expected Fisher information.

$$I_F(\theta) = -E_{y^*}\left[\frac{\partial^2}{\partial\theta^2}\log\ L(y_1|\theta,\boldsymbol{\vartheta})\right]$$

- Quantifies how difficult it is to estimate the parameters -- always free of a location parameter.
- How do approaches differ?

Effective sample size – general approach

$Y_1 \sim N(\theta, 1) \quad \theta \sim N(0, 2)$



Effective sample size – general approach

- What do we do with the data that has not been observed?
  - Take an expectation with respect to the likelihood or marginal likelihood.
- The effective sample size in the Gaussian case

$$Y_1|\theta \sim N(\theta, \sigma^2) \qquad \theta \sim N(\mu_\theta, \tau^2)$$

$$ESS = \frac{\sigma^2}{\tau^2}$$

Key Quantity in Gaussian effective sample size ---- σ^2/m

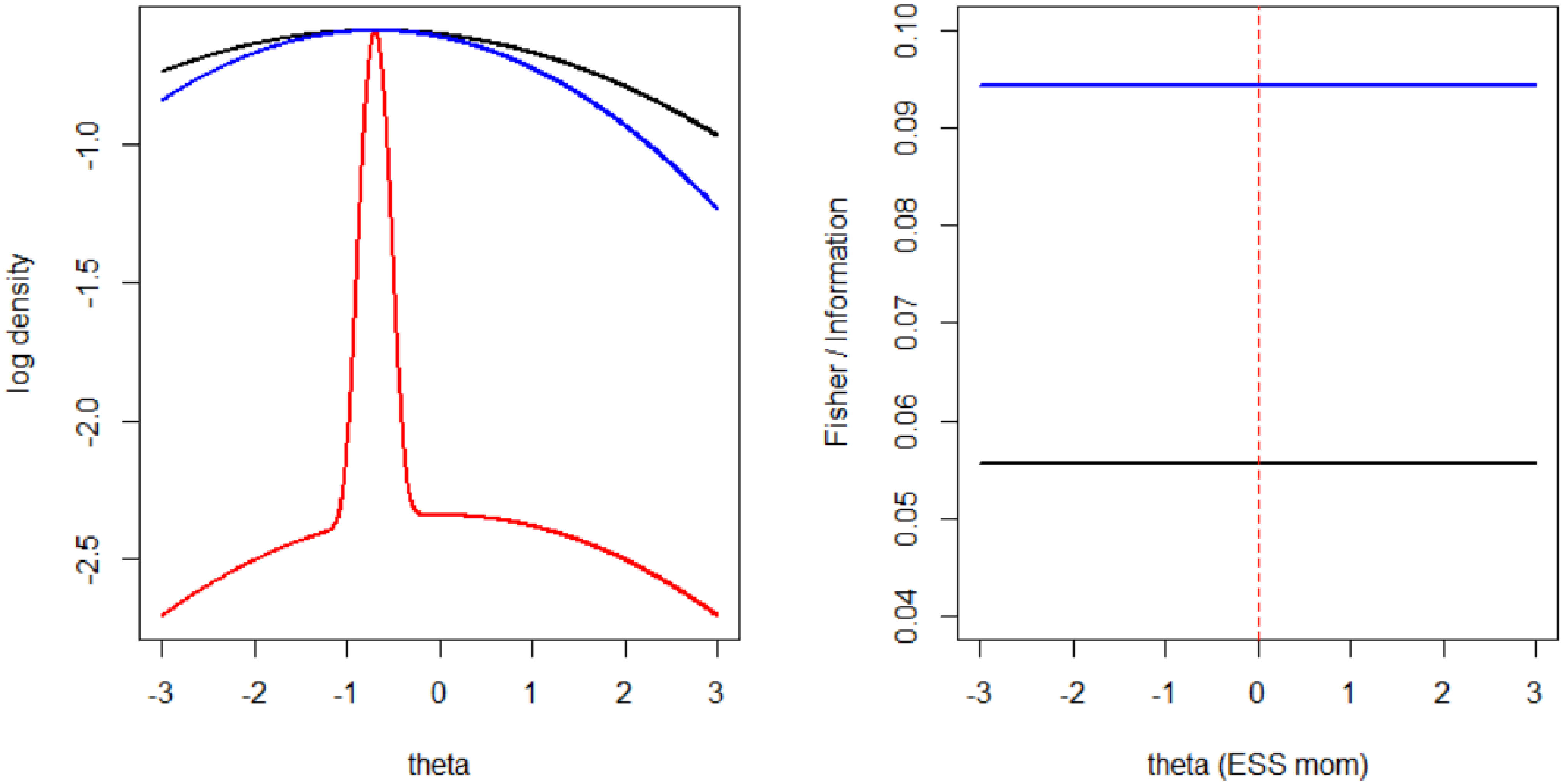
Effective sample size assumptions and mishaps

- Why is my prior[1] getting such different ESS?
- $\theta \sim 0.15N(-0.7, 0.13^2) + 0.85N(0, 3.52^2)$

- Moment -- prior and likelihood both from a Gaussian approximation.
- Morita[2] -- constant information at point of evaluation.
- ELIR[3] -- log-concave prior. Weighted average of the ESS with respect to the prior.

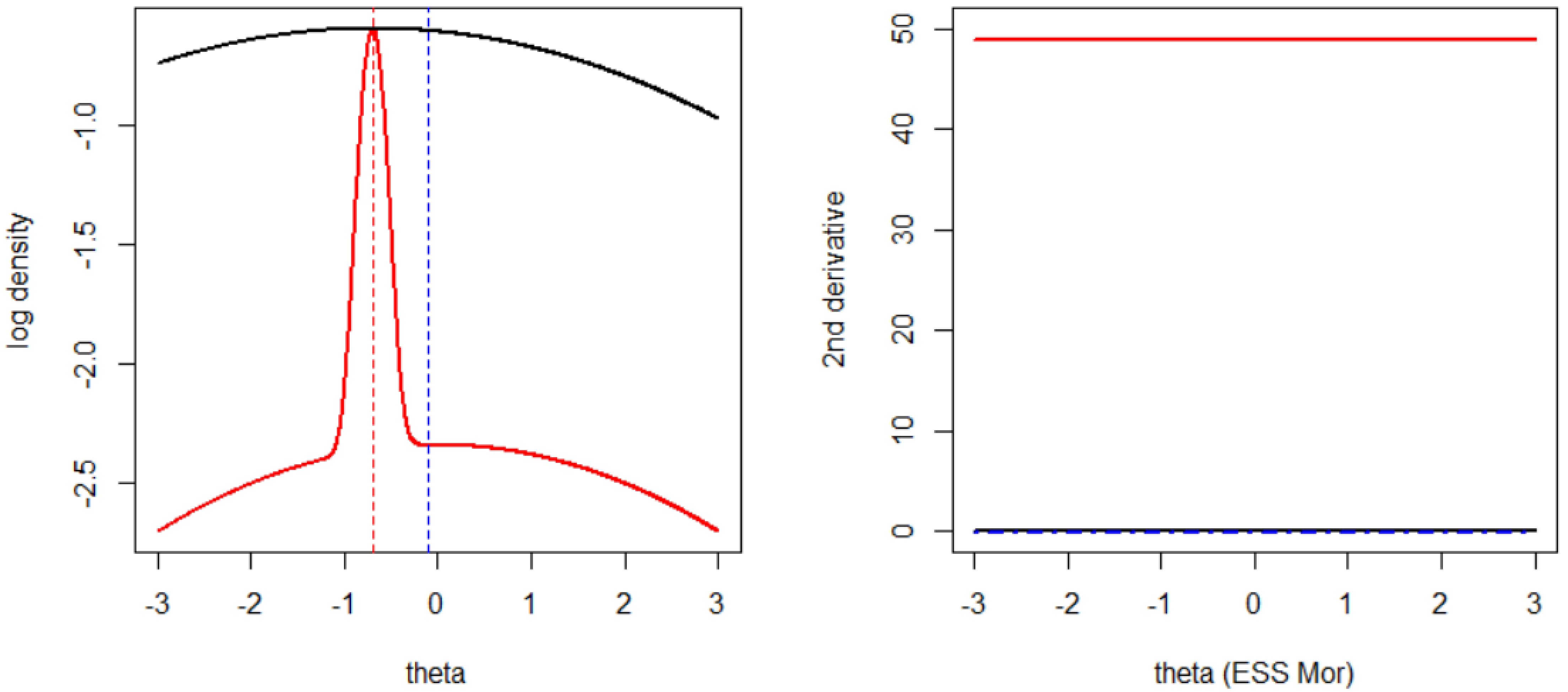
Mixture prior[1] Moment ESS = 1.7

- $Y_1 \sim N(\theta, 18) \quad \theta \sim 0.15N(-0.7, 0.13^2) + 0.85N(0, 3.52^2)$



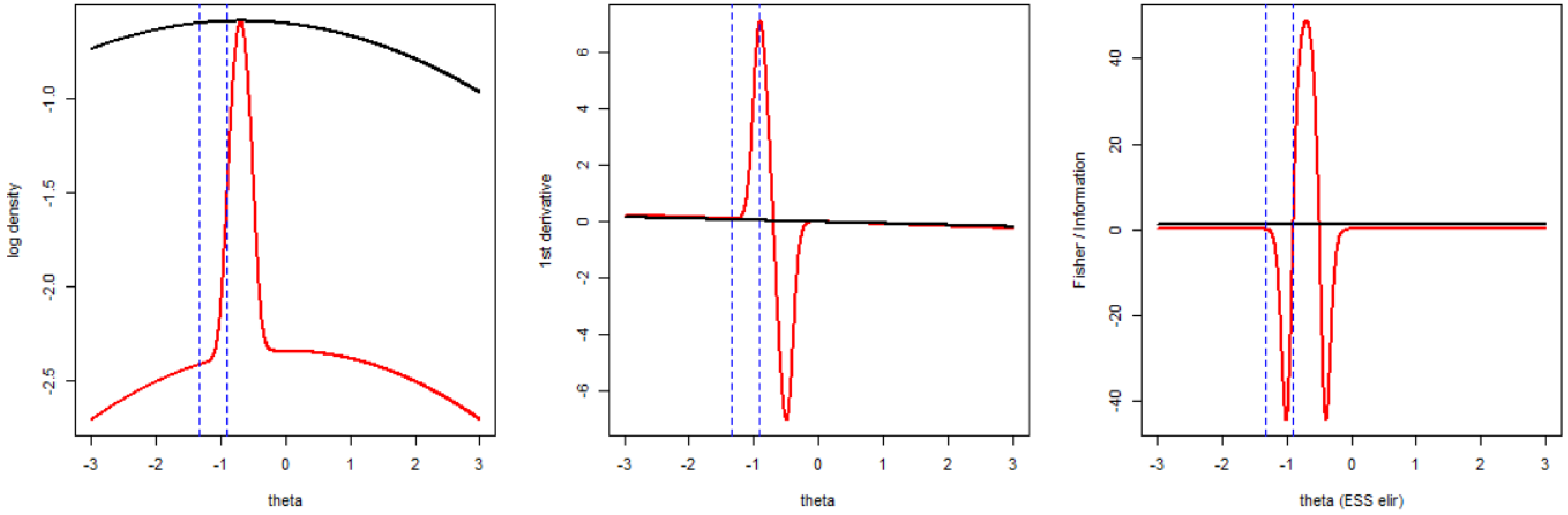
Mixture prior[1]   Morita ESS = 878 or 0

- $Y_1 \sim N(\theta, 18) \quad \theta \sim 0.15N(-0.7, 0.13^2) + 0.85N(0, 3.52^2)$



Mixture prior[1]   Elir = 86

- $Y_1 \sim N(\theta, 18) \quad \theta \sim 0.15N(-0.7, 0.13^2) + 0.85N(0, 3.52^2)$



2. Efficient Bayesian borrowing

- We decide on the nature of the estimand before constructing the model.
- In some cases, the model is collapsible with respect to the treatment effect.
  - Covariates can be included in the model, to increase precision, without changing the interpretation of the treatment effect.
  - $\epsilon_{i1} \sim N(0, \sigma_1^2), \epsilon_{i2} \sim N(0, \sigma_2^2)$

$$C)\ Y_i = \alpha_1 + X_i^T \beta + \lambda_1 Z_i + \epsilon_{i1} \qquad M)\ Y_i = \alpha_2 + \lambda_2 Z_i + \epsilon_{i2}$$

$$\lambda_1 \leftrightarrow \lambda_2$$

Challenge

- I have a sufficient statistic $(\bar{x}_h, s_h)$ on the control group from an historical dataset.

- Dynamic mixture prior $p(\theta_p|\bar{x}_h, s_h)$ robust to prior data conflict

$$w_1 p(\theta_p|\bar{x}_h, s_h) + (1 - w_1)p(\theta_p|c, d).$$

- **How do I use this information to borrow, when I have a conditional model**?

Model

- $Y_i = \alpha + \lambda Treat_i + \beta_1 Region_i + \ \beta_2 Sex_i + \epsilon_i$
- $\epsilon_i \sim N(0, \sigma^2)$

- $\boldsymbol{X}_i \ - discrete\ covariates\ of\ Region\ and\ Sex$
- $Z_i - treatment$

$$\begin{bmatrix} Intercept & Treat & Region & Sex \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

Design matrix construction

- Sum to zero referencing for covariates.
  - $Y_i = \alpha + \lambda Treat_i + \beta_1 Region_i + \beta_2 Sex_i + \epsilon_i$
  - $\theta_p = \alpha$

$$\begin{bmatrix} Intercept & Treat & Region & Sex \\ 1 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & -1 \end{bmatrix}$$

- Place the prior on marginal placebo rate.
- Treatment λ can still be interpreted as our marginal treatment effect.
- Centre any continuous covariates.

Standardisation - $E_X(E(Y|X, Treat = 0))$

$$\begin{bmatrix} Intercept & Treat & Region & Sex \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

- $Y_i = \alpha + \lambda Treat_i + \beta_1 Region_i + \ \beta_2 Sex_i + \epsilon_i$

- Use standardisation where the observed quantities are your weights.
  - $\theta_p = \frac{n_{r_0 s_0}}{n}(\alpha) + \frac{n_{r_0 s_1}}{n}(\alpha + \beta_2) + \frac{n_{r_1 s_0}}{n}(\alpha + \beta_1) + \frac{n_{r_1 s_1}}{n}(\alpha + \beta_2 + \beta_1)$
  - $\theta_t = \theta_p + \lambda$

- Place the prior with the historical information on the marginal placebo rate θ_p.

- In stan this is easily done, treatment has a marginal interpretation.

No Stan – No problem

$$\begin{bmatrix} Intercept & Treat & Region & Sex \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

- Estimate the marginal control using standardisation (empirical Bayes/ MLE).
- $\hat{\theta}_p = \frac{n_{r_0 s_0}}{n}(\hat{\alpha}) + \frac{n_{r_0 s_1}}{n}(\hat{\alpha} + \hat{\beta}_2) + \frac{n_{r_1 s_0}}{n}(\hat{\alpha} + \hat{\beta}_1) + \frac{n_{r_1 s_1}}{n}(\hat{\alpha} + \hat{\beta}_2 + \hat{\beta}_1)$
- Approximately normal with variance calculated using

$$V(a+b) = V(a) + V(b) + 2C(a,b)$$

- $L(\theta_p; Y) = N\left(\theta_p | \hat{\theta}_p, V(\hat{\theta}_p)\right)$
- Treatment $\theta_t | Y \sim N\left(\theta_t | \hat{\theta}_t, V(\hat{\theta}_t)\right)$
- Apply the marginal prior which contains the historical data to the normal likelihood $L(\theta_p; Y)$.

References

1. Best, N., Price, R. G., Pouliquen, I. J., and Keene, O. N. (2021). Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing. Pharmaceutical Statistics, 20:551–562.
2. Morita, S., Thall, P. F., and Muller, P. (2008). Determining the effective sample size of a parametric prior. Biometrics, 64:595–602
3. Neuenschwander, B., Weber, S., Schmidli, H., and O'Hagan, A. (2020). Predictively consistent prior effective sample sizes. Biometrics,76:578–587.