This lecture

Bayes and selection.

- Should Bayesian inference be adjusted for selection?
  - Dawid (1994). "Selection paradoxes of Bayesian inference", *IMS Monograph, Multivariate Analysis and its Applications*.
  - Yekutieli (2012). "Adjusted Bayesian inference for selected parameters", *JRSSB*.
  - Garcia Rasines, Young (2022). "Bayesian selective inference", *Handbook of Statistics*.
- Empirical Bayes:
  - Point estimation: Efron (2011). "Tweedie's formula and selection bias", *JASA*.
  - Confidence intervals: Woody, Hernan Padilla, Scott (2022). "Optimal post-selection inference for sparse signals: a nonparametric empirical-Bayes approach", *Biometrika*.

Bayes and selection

So far we have analysed selection problems from the frequentist perspective.

In this context, the effect of selection on the inference is evident, as it alters the sampling distribution of the statistics.

From the Bayesian viewpoint, however, the matter is not so clear-cut:

- The traditional stance is that Bayesian methods are **immune to selection**, as inference is carried out conditionally on the data and in particular on any hypothetical selection event (Dawid, 1994).

- Recent work, most notably by Yekutieli (2012), challenges this idea, arguing that, in some scenarios, Bayesian inference **needs to be selection-adjusted**.

Bayes and selection

Quoting A. P. Dawid:

> *Since Bayesian posterior distributions are already fully conditioned on the data, the posterior distribution of any quantity is the same, whether it was chosen in advance or selected in the light of the data: that is, for a Bayesian, the face-value approach is fully valid, and no further adjustment for selection is required.*

$\rightarrow$ The fact that the statistician decides to focus on certain aspects of the data after it has been observed does not alter its sampling distribution, which should therefore remain unchanged.

Bayes and selection

The contrast between the Bayesian and the frequentist standpoints is somewhat paradoxical.

In many situations, Bayesian analyses formally match, either exactly or approximately, face-value frequentist analyses.

But it is universally agreed that the latter methods are not valid in the presence of selection.

Why, then, would such results be correct if reached by a Bayesian argument?

Bayes and selection

From the previous point of view, only explanation of the paradox is a **poor prior specification**.

Indeed, if we agree that:

- Whenever a frequentist approach is unreasonable, so is any Bayesian approach that provides similar answers.
- The (unadjusted) likelihood is correct.

Then, necessarily, the problem lies in the prior.

Bayes and selection

We can understand this through a simple example.

Suppose we provide inference for the mean of $Y \sim N(\theta, 1)$ only if $Y > 0$.

Consider the standard class of conjugate priors for θ: $\{N(0, \lambda^2) : \lambda > 0\}$.

The corresponding class of posteriors is

$$\pi_\lambda(\theta \mid y) \sim N\left(\frac{\lambda^2}{1+\lambda^2} y, \frac{\lambda^2}{1+\lambda^2}\right). \tag{1}$$

$\rightarrow$ Small values of λ shrink the posterior around 0.

$\rightarrow$ Large values of λ produce posterior inferences which are very similar to those provided by a face-value frequentist approach.

Bayes and selection

Let's think of the same problem from a frequentist perspective:

- If true $\theta_0 >> 0$, selection is negligible; no adjustment is needed.
- Otherwise, Y will tend to overestimate θ_0.

Thus, by the previous remarks:

- Bayesian analysis with a large λ is appropriate if we expect the true θ is large.
- Otherwise, it will lead us astray if it is not.

Obvious, in a way: we would not adjust for selection if we believed *a priori* that θ is large, and viceversa.

Bayes and selection

Two conclusions:

- Prior choice is always important, *but* the **impact of the prior** on the analysis is stronger under selection.
- Unclear how to carry out **non-informative or weakly informative** analyses in these settings.

For *non-selective* inference on θ in $Y \sim N(\theta, 1)$, lack of prior information about θ can be dealt with by setting $\lambda = \infty$ (or $\lambda >> 0$), which minimises the influence of the prior on the analysis.

In the presence of selection, however, such "non-informative" priors do in fact entail critical information about the parameter: **its likelihood to lie in a region of the parameter space for which a selection adjustment would be appropriate**.

Bayes and selection

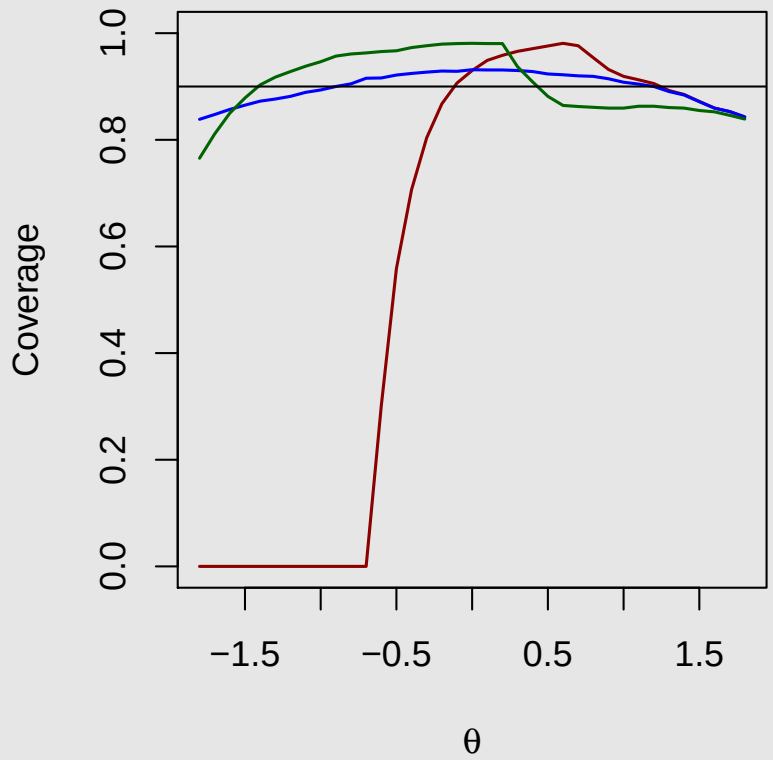
Further problem: even if we are comfortable with the implications of the prior, **repeated-sampling properties** are deteriorated under selection.

Let $\theta \sim N(0, 1)$ and $Y \mid \theta \sim N(\theta, 0.2)$, and suppose we construct equal-tailed 90% credible intervals for θ under two sampling regimes:

- A selective one, where only observations with $Y > 0$ are kept.
- A non-selective one, where intervals are constructed for all samples.

Bayes and selection

Figure: Coverage of CIs as a function of θ with unadjusted posterior (selection; no selection), and with selective posterior.



$\rightarrow$ All coverages have 90% expectation wrt the prior, but conditionally on θ, the coverages under selection are much more sensitive.

Bayes and selection

One solution inspired by the *conditional approach*: get rid of the information used for selection by conditioning on the selection event.

That is, apply Bayes' rule with the likelihood of the conditional model $Y \mid E$, obtaining the posterior

$$\pi(\theta \mid y) \propto \frac{\pi(\theta) f(y \mid \theta)}{\varphi(\theta)}, \quad \varphi(\theta) = P(Y \in E \mid \theta). \tag{2}$$

Denote this the **selective posterior**.

Key idea: inference based on selective posteriors allows the injection of prior information while avoiding potential problems arising from selection.

Bayes and selection

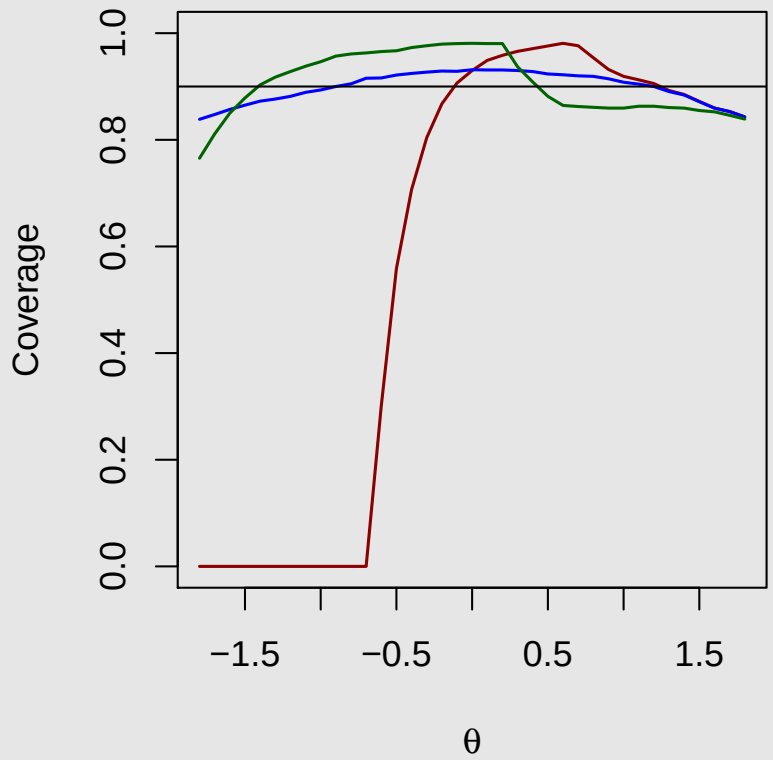
Alternative interpretation: selective posterior follows from a standard Bayesian updating of the modified prior

$$\pi^*(\theta) \propto \frac{\pi(\theta)}{\varphi(\theta)}. \tag{3}$$

$\rightarrow$ Since unadjusted inference is imprecise for values of θ with small selection probability $\varphi(\theta)$, assign a larger prior probability to those values to achieve a higher protection in those regions.

Bayes and selection

Figure: Coverage of CIs as a function of θ with unadjusted posterior (selection; no selection), and with selective posterior.



$\rightarrow$ Robustness of the conditional coverage is recovered with the selective posterior.

Bayes and selection

Yekutieli (2012) provides a different take on the problem.

Suppose a parameter θ is analysed iff $Y \in E$.

In a Bayesian generative model with selection, the sampling and selection mechanisms can interact in different ways, giving two different "regimes":

- *Fixed parameter* regime:
  - Sample $\theta \sim \pi(\theta)$.
  - Sample $Y \mid \theta \sim f(y \mid \theta)$ until $Y \in E$.
- *Random parameter* regime:
  - Sample $(\theta, Y) \sim \pi(\theta) \times f(y \mid \theta)$ until $Y \in E$.

Bayes and selection

With this explicit distinction the "correct" posterior distribution becomes apparent.

- If θ is *fixed*, its marginal distribution is unaffected by selection, i.e.

$$\pi(\theta \mid Y \in E) = \pi(\theta). \tag{4}$$

- If θ is *random*, its marginal distribution is

$$\pi(\theta \mid Y \in E) = \frac{P(Y \in E \mid \theta)\pi(\theta)}{P(Y \in E)} \propto \varphi(\theta)\pi(\theta), \tag{5}$$

  i.e. selection favours observation of θ's with larger selection probability.

Bayes and selection

In both cases, the likelihood is

$$f(y \mid \theta, Y \in E) = \frac{f(y \mid \theta)\mathbf{1}(y \in E)}{\varphi(\theta)}. \tag{6}$$

Therefore, if θ is fixed,

$$\pi(\theta \mid y) \propto \frac{\pi(\theta) f(y;\theta)}{\varphi(\theta)}, \tag{7}$$

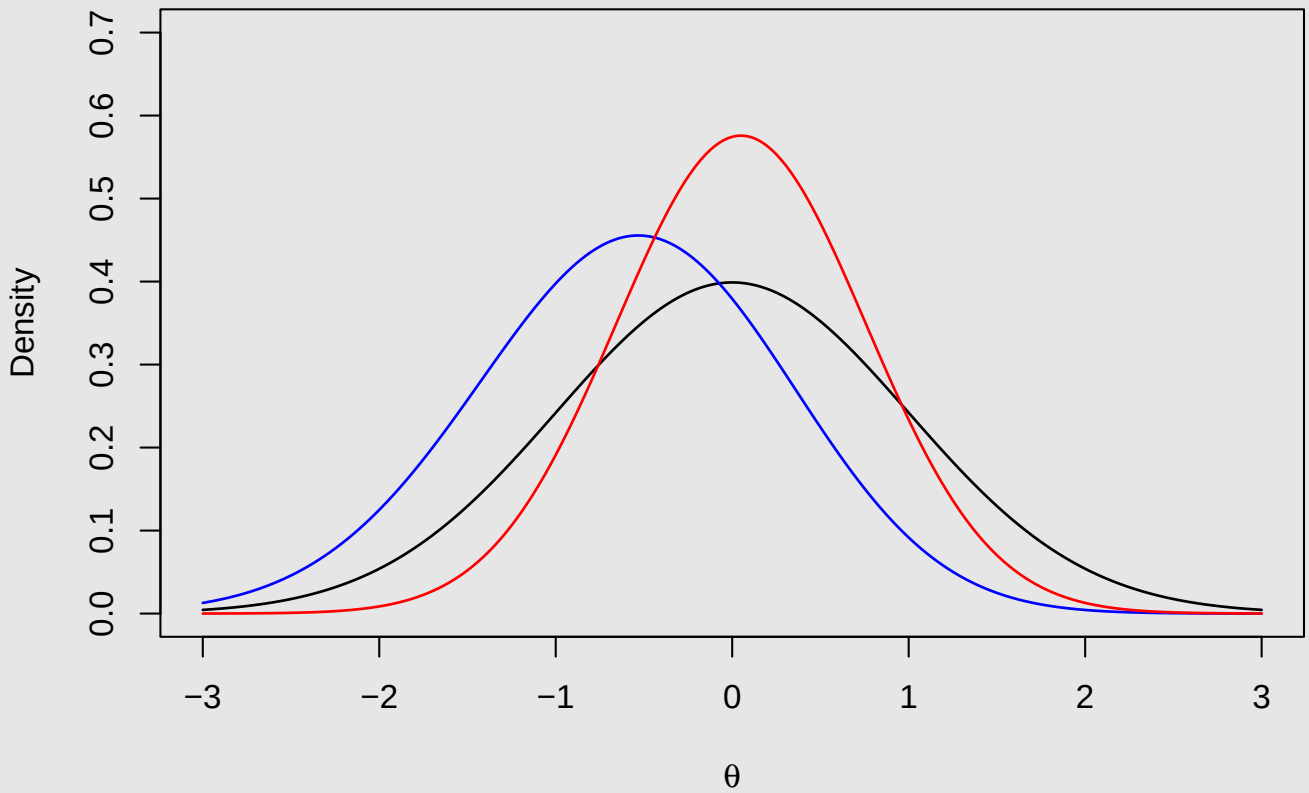
which is the selective posterior, while if θ is random,

$$\pi(\theta \mid y) \propto \frac{\pi(\theta)\varphi(\theta) f(y;\theta)}{\varphi(\theta)} = \pi(\theta) f(y;\theta), \tag{8}$$

which is the standard, unadjusted posterior.

Bayes and selection

Figure: Illustration for model $\theta \sim N(0,1)$, $Y \mid \theta \sim N(\theta,1)$, selection event $Y > 0$ and $y = 0.1$. Prior; Selective posterior; Unadjusted posterior.



Bayes and selection

Thus, whether a correction for selection is needed depends on which regime is appropriate for the given problem. Consider the following examples:

- Let θ be the average academic ability of the students in school.

  Each student takes an exam and gets a grade $Y_i \sim N(\theta, 1)$, and the students for which $Y_i \geq t$ are admitted to university.

  If a person in the university, with access to the grades obtained by the admitted students, wants to estimate θ, they should carry out a *fixed parameter* analysis.

- Suppose instead that θ is student-specific, representing the student's academic ability.

  If the person in the university wants to estimate each of the academic abilities of the admitted students, then each estimated θ should be treated as a *random parameter*.

Bayes and selection

What about **non-informative priors**?

Both in the fixed and random parameter regimes, the appropriate likelihood is constructed from the **conditional distribution given selection**,

$$L(\theta) \propto \frac{f(y \mid \theta)}{\varphi(\theta)}. \tag{9}$$

Since all information about θ in a non-informative analysis comes from the data, the posterior should be constructed with this likelihood.

Also, non-informative priors are designed to have certain properties wrt the sampling model, so they need to be **adjusted for selection**.

Bayes and selection

For example, consider location model $\{f(y;\theta) = g(y-\theta) \colon \theta \in \mathbb{R}\}$.

The standard non-informative prior for θ in absence of selection is $\pi(\theta) \propto 1$, which is

- Jeffreys invariant prior.
- The reference prior: minimises the prior influence.
- Probability matching: produces inferences with a valid frequentist interpretation.

However, these properties don't hold if inference for θ is only provided for certain values of y, as the conditional model

$$\{g(y-\theta)\mathbf{1}(y \in E)/\varphi(\theta) \colon \theta \in \mathbb{R}\} \tag{10}$$

is not a location model.

Bayes and selection

Non-informative priors in Gaussian selection models

Let $Y = (Y_1, \ldots, Y_n)$ be IID $N(\theta, 1)$ and let E be the selection event.

Two (improper) prior densities are available to carry out a non-informative analysis:

- Jeffreys prior from the conditional distribution:

$$\pi_J(\theta) \propto \left\{ n + \frac{\partial^2}{\partial \theta^2} \log \varphi(\theta) \right\}^{1/2}, \quad \varphi(\theta) = P_\theta(Y \in E). \tag{11}$$

- Exact probability-matching prior (PMP):

$$\pi_y(\theta) \propto -\frac{\frac{\partial}{\partial \theta} H(\theta; \bar{y})}{\frac{\partial}{\partial \bar{y}} H(\theta; \bar{y})}, \quad H(\theta; \bar{y}) = P_\theta\left(\bar{Y} \geq \bar{y} \mid E\right). \tag{12}$$

Bayes and selection

The PMP is derived as follows. The probability-matching requirement is

$$\mathbb{P}_\theta\{\theta \leq \Pi^{-1}(\alpha|\bar{Y})\} = \alpha \tag{13}$$

for all α and θ, where $\Pi(\theta \mid \bar{y})$ is the posterior CDF given the sufficient statistic $\bar{Y}$. In particular, this implies that $1-\alpha$ credible intervals are also $1-\alpha$ confidence intervals.

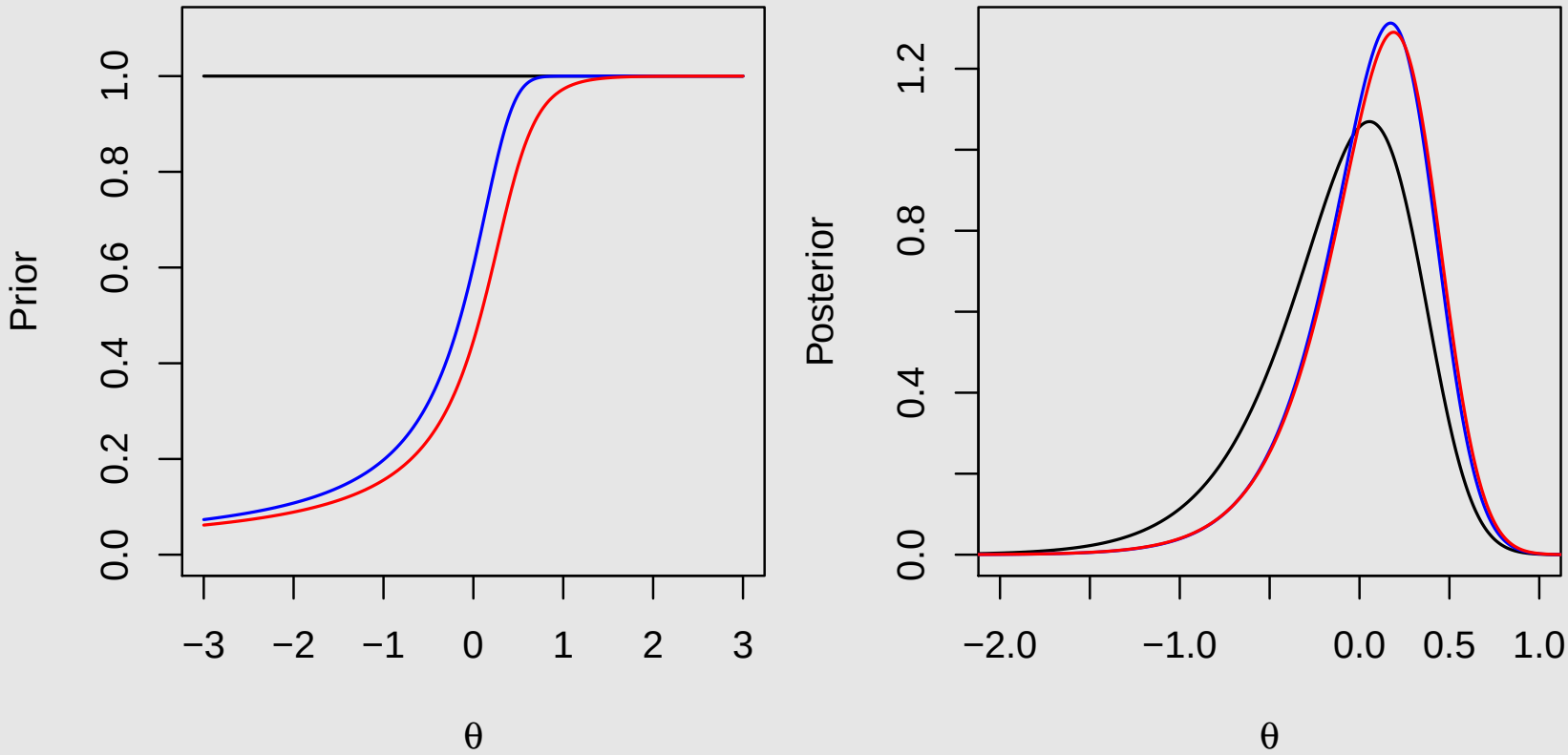
This is in fact equivalent to requiring that $\Pi(\theta \mid \bar{Y}) \sim U(0,1)$ under θ.

By construction, the p-value function $H(\theta; \bar{Y})$ is uniformly distributed under θ, so equating $\Pi(\theta \mid \bar{Y}) = H(\theta; \bar{Y})$ and solving for the prior gives the PMP.

$\rightarrow$ Note that this prior is data-dependent, unlike in the non-selective setting.

Bayes and selection

Figure: Left: uniform prior, probability-matching prior and Jeffreys prior, for $n = 20$, $\bar{y} = 0.2$ and selection $\bar{y} > 0$. Right: corresponding posteriors for $\bar{y} = 0.2$.



Bayes and selection

These priors can be extended to univariate exponential families. Informal argument is as follows.

Let $Y_1, \ldots, Y_n$ from an exponential family, with PDF/PMF

$$f(y_i; \theta) = h(y_i) \exp\{\eta(\theta) s(y_i) - A(\theta)\}. \tag{14}$$

Let $i(\theta) = -\mathbb{E}_\theta[(\partial^2/\partial\theta^2) \log f(Y_i; \theta)]$ be FI of the *non-selective* model, and let $g(\theta)$ satisfy $g'(\theta) = i(\theta)^{1/2}$. This is the **variance-stabilising transformation**, for which

$$n^{1/2}\{g(\hat{\theta}) - g(\theta)\} \xrightarrow{d} N(0, 1). \tag{15}$$

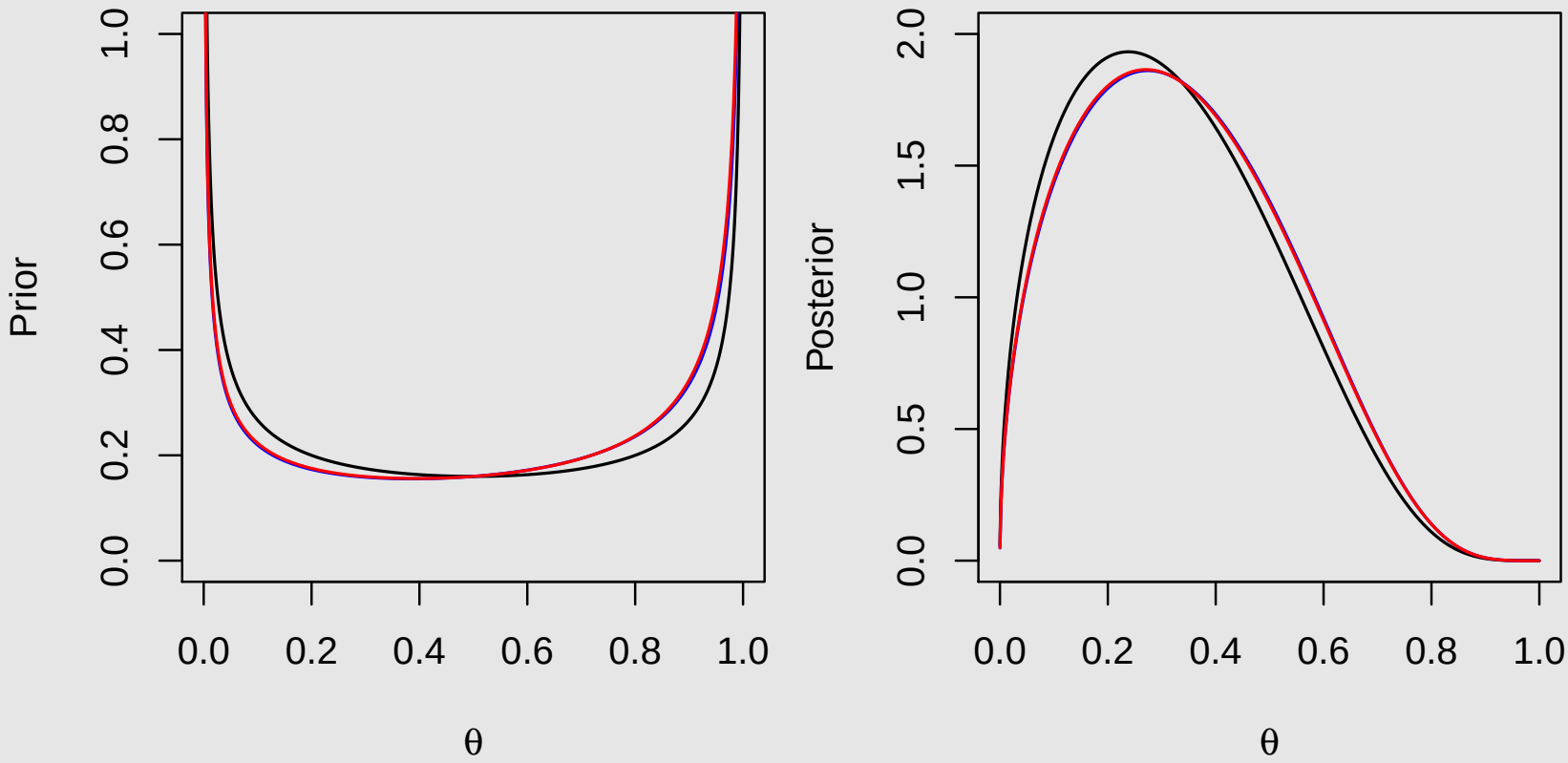
The idea is to impose a prior on $\nu \equiv g(\theta)$ that "works well" in the asymptotic model, so can take

$$\pi_\theta(\theta) \propto i(\theta)^{1/2} \pi_\nu\{g(\theta)\}, \tag{16}$$

where $\pi_\nu(\nu)$ is the Jeffreys/PMP prior in the ν-parametrisation.

Bayes and selection

Figure: Illustration in model $Y_i \sim \text{Bernoulli}(\theta)$, with $n = 10$ and selection $Y_1 + \ldots + Y_8 > 8 \times 0.5$. Left: non-selective Jeffreys prior, selective Jeffreys prior and probability-matching prior, for sample with $\bar{y} = 0.5$. Right: corresponding posteriors.



Bayes and selection

Further remarks:

- Both priors admit natural extensions to general exponential families with nuisance parameters.
- Jeffreys prior is more computationally stable, independent of the data, and produces almost the same results as the PMP.
- Both priors depend on the sample size, unlike in non-selective regimes. This is needed for correct asymptotic calibration: priors independent of n overstate regions of the parameter space with low selection probability.

Empirical Bayes: setup

Within the classical/random parameter view, inferences are immune to selection, but can be very **sensitive to prior misspecification**.

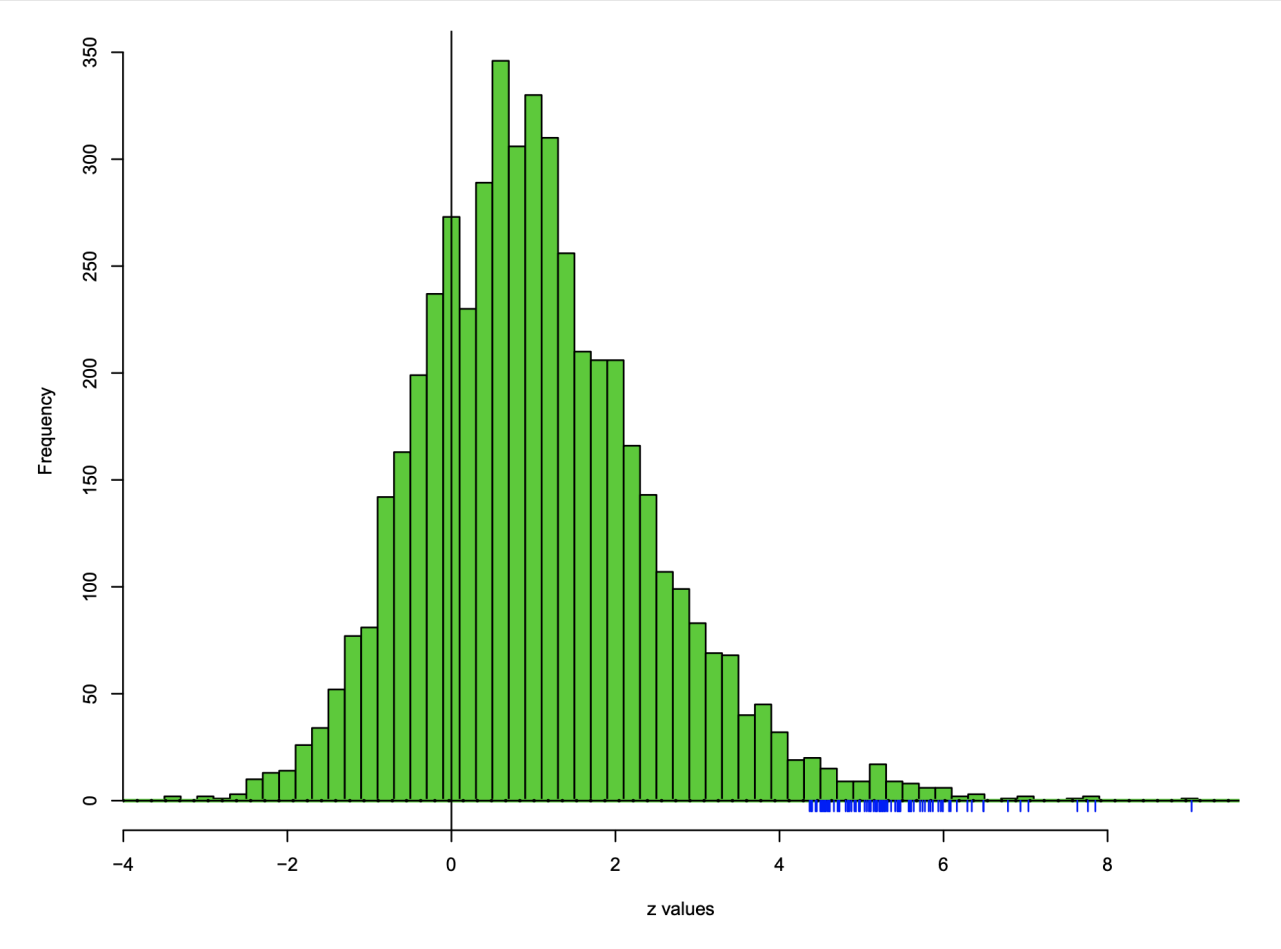
This approach thus lends itself to implementation via empirical Bayes methods, which estimate the true prior distribution using the data.

Suppose we observe $Y_1, \ldots, Y_n$ independently from parametric models $f(y_i; \theta_i)$, where n is *large*.

As usual, inference is assumed to be provided only for a subset of the θ_i's selected with the data; e.g. for the parameters of the top K observations.

Empirical Bayes: setup

Figure: Exponential example: $n = 5000$, $Y_i \sim N(\theta_i, 1)$, $\theta_i \sim \text{Exp}(1)$. Inference sought for the means of the top 100 samples, indicated in blue (Efron, 2011).



Point estimation

For point estimation of the selected effects, **Tweedie's formula** (Robbins, 1956) offers a simple yet powerful solution for exponential family likelihoods.

Let $\theta \sim \pi(\theta)$, and $Y \mid \theta \sim f(y \mid \theta)$, where

$$f(y \mid \theta) = f_0(y) e^{\theta y - A(\theta)}. \tag{17}$$

Bayes' rule gives

$$\pi(\theta \mid y) = \frac{\pi(\theta) f(y \mid \theta)}{f(y)}, \tag{18}$$

where $f(y)$ is the marginal density

$$f(y) = \int_{\Theta} \pi(\theta) f(y \mid \theta) \mathrm{d}\theta. \tag{19}$$

Point estimation

For the exponential family model, we get

$$\pi(\theta \mid y) = e^{y\theta - \lambda(y)} \left[\pi(\theta) e^{-A(\theta)}\right], \quad \lambda(y) = \log\left(\frac{f(y)}{f_0(y)}\right), \tag{20}$$

i.e. the class of posterior distributions is an exponential family with natural parameter y and CGF $\lambda(y)$.

Differentiating the CGF gives

$$\mathbb{E}[\theta \mid y] = \lambda'(y), \quad \mathrm{Var}(\theta \mid y) = \lambda''(y), \quad \text{etc.} \tag{21}$$

In particular, letting $l(y) = \log f(y)$ and $l_0(y) = \log f_0(y)$,

$$\mathbb{E}[\theta \mid y] = l'(y) - l_0'(y). \tag{22}$$

Point estimation

We can easily check that

$$\mathbb{E}[-l_0'(Y) \mid \theta] = \theta, \tag{23}$$

so that $-l_0'(y)$ is an unbiased estimate for θ.

Therefore, we can write the posterior mean as

$$\mathbb{E}[\theta \mid y] = -l_0'(y) + l'(y) = \text{Unbiased estimate} + \text{Bayes correction}. \tag{24}$$

In particular, for the Gaussian distribution with known variance, $Y \mid \theta \sim N(\mu, \sigma^2)$, where $\mu = \sigma^2\theta$, we obtain

$$\mathbb{E}[\mu \mid y] = y + \sigma^2 l'(y). \tag{25}$$

Point estimation

Recall, in the "random parameter" setting, $\mathbb{E}[\theta \mid y]$ is unbiased under selection regardless of the selection criterion, as

$$\mathbb{E}\left[\mathbb{E}[\theta \mid Y] - \theta\right] = 0. \tag{26}$$

However, this estimate is not computable as it requires knowledge of the marginal density $f(y)$, and therefore of the prior.

Empirical Bayes solution: estimate $l'(y) = (d/dy) \log f(y)$ using the data.

Unlike with most empirical Bayes procedures, here we estimate $f(y)$ directly rather than the prior $\pi(\theta)$.

Point estimation

A *smooth and non-parametric estimate* of $l(y)$ can be obtained with **Lindsey's method**.

Assume that $l(y)$ is a J-th degree polynomial:

$$\log f_\beta(y) = \sum_{i=0}^{J} \beta_i y^j. \tag{27}$$

This is equivalent to assuming that Y follows (marginally) a J-dimensional exponential family distribution with natural parameter

$$\beta = (\beta_1, \ldots, \beta_J)^T. \tag{28}$$

Note that β_0 is a function of β through normalisation of $f_\beta(y)$.

Point estimation

The MLE of β can be approximated as follows by partitioning the sample space into T bins, with counts

$$z_t = \#\{y_i\text{'s in } t\text{-th bin}\}, \quad t = 1, \ldots, T, \tag{29}$$

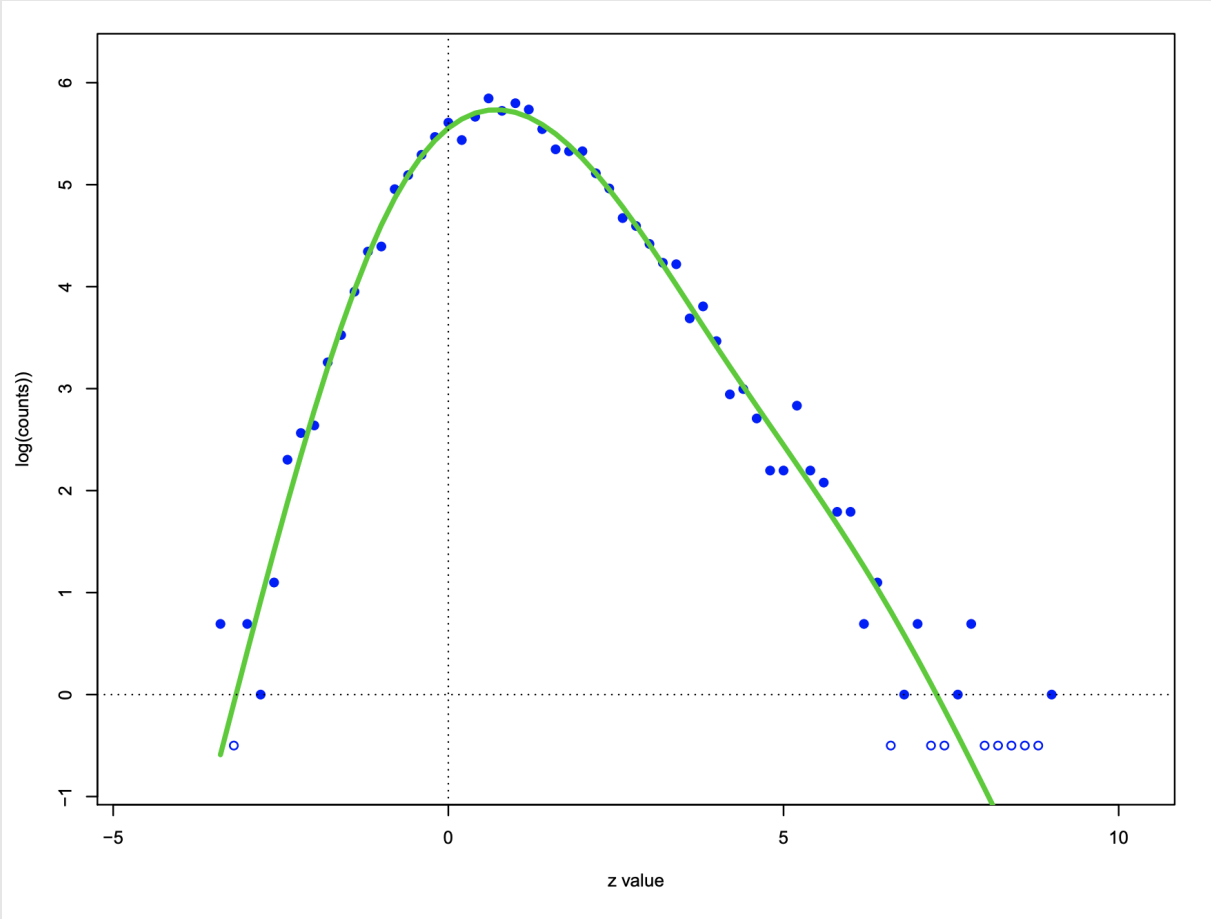
and fitting a Poisson regression model

$$z_t \sim \text{Poisson}(\nu_t) \text{ independently}, \quad t = 1, \ldots, T, \tag{30}$$

where $\nu_t = n \times d \times f_\beta(x_t)$, d is the bin width, and the x_t's are the bin centers.

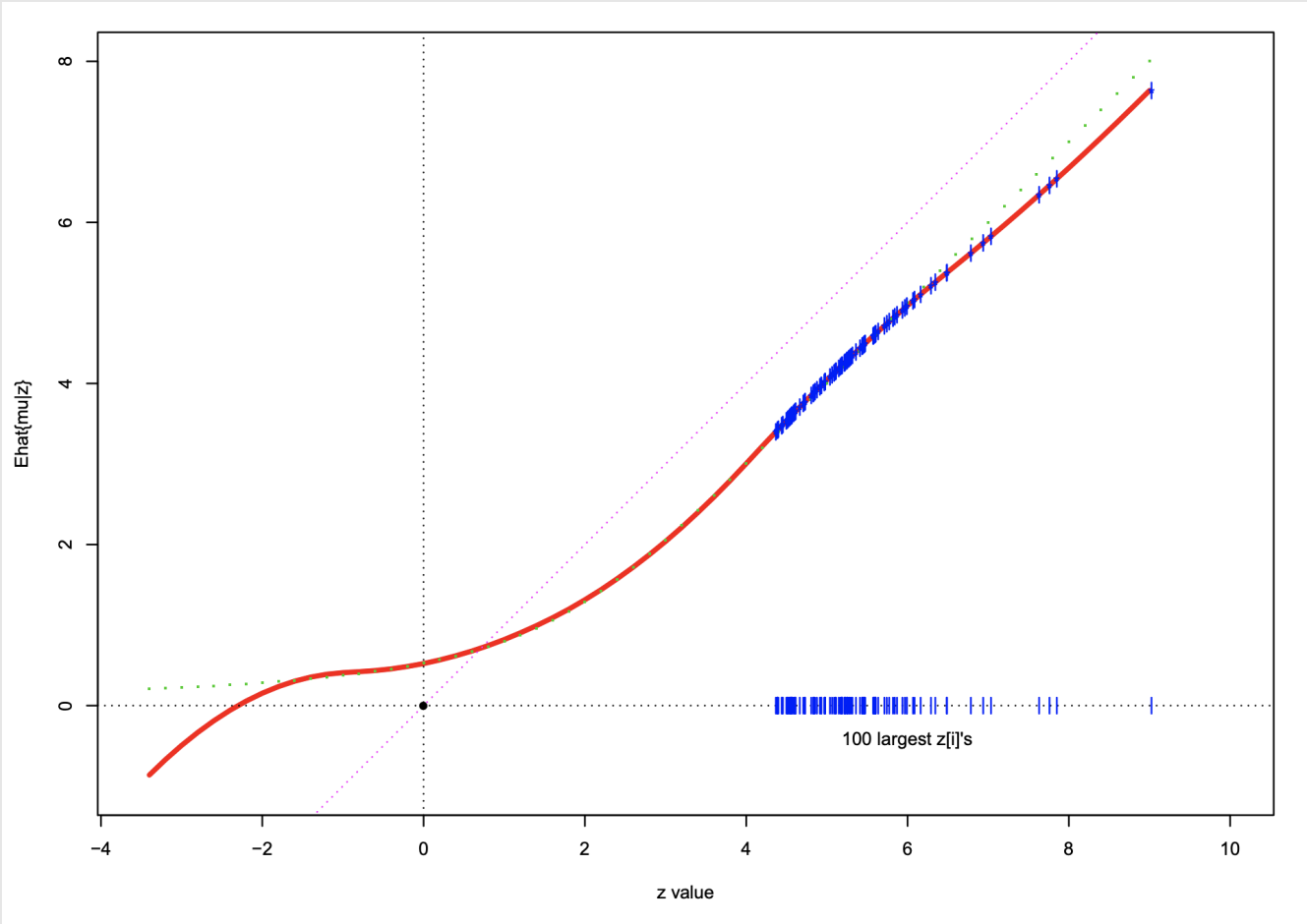
Point estimation

Figure: Blue dots: log bin counts for data in the exponential example vs. bin centers (zero counts indicated with empty dots). Green curve: MLE of $l(y)$ with $J = 5$ (Efron, 2011).



Point estimation

Figure: Empirical Bayes estimate $y + \hat{l}'(y)$ as a function of y, where $\hat{l}(y)$ is the previous estimate (Efron, 2011).



Point estimation

Does empirical Bayes actually correct for selection bias? i.e. does empirical estimation of $l(y)$ break the unbiasedness of the true Bayes estimator?

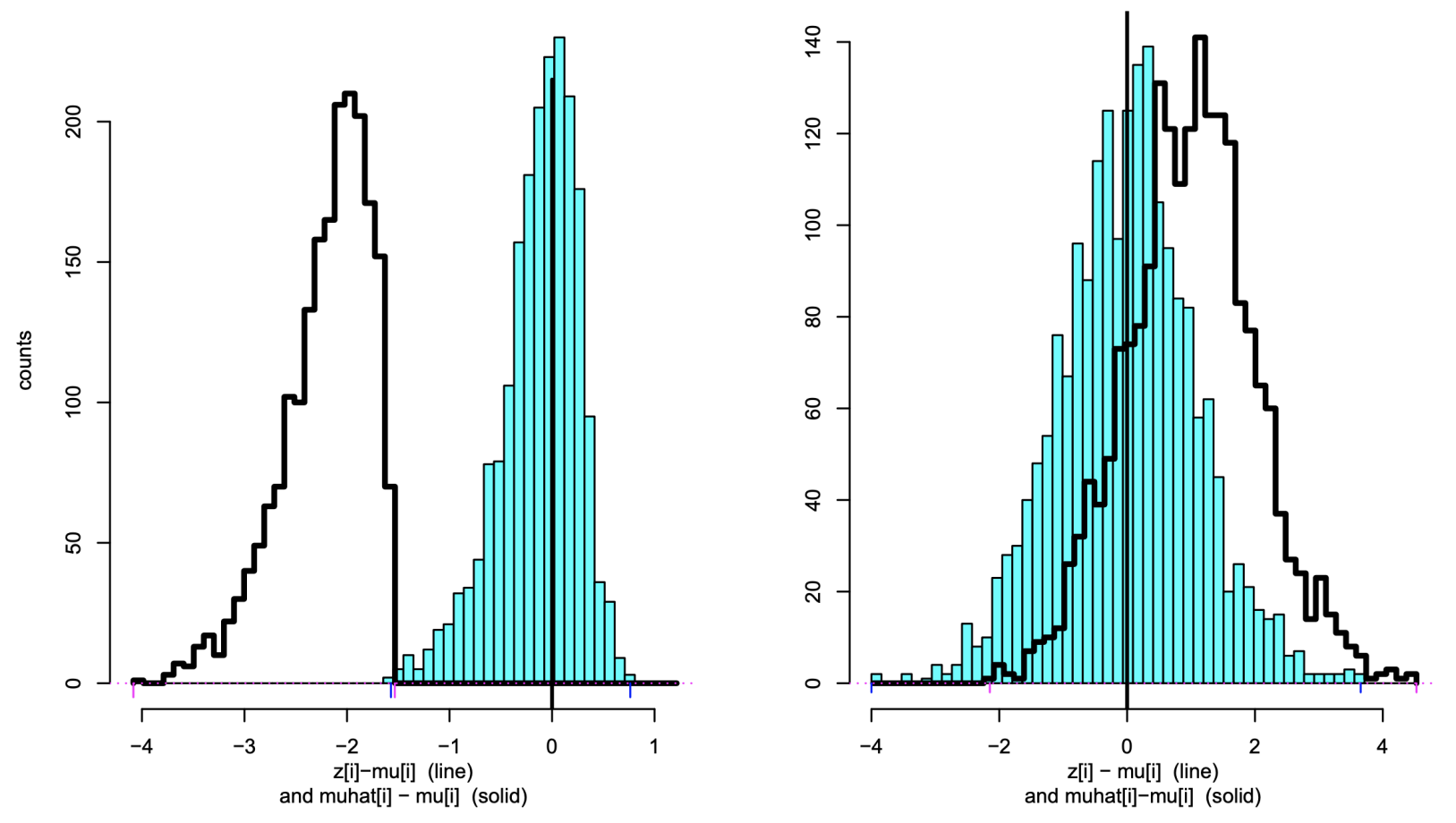
Simulation:

$\rightarrow$ 100 replications of the Exponential-Gaussian example with $n = 1000$; $l(y)$ estimated with $J = 5$.

$\rightarrow$ Corrected estimates $\hat{\mu}_i = y_i + \hat{l}'(y_i)$ computed for 20 largest and 20 smallest y_i's, giving a total of 2000 triplets $(\mu_i, y_i, \hat{\mu}_i)$ for the largest group, and similarly for the smallest group.

Point estimation

Figure: Line histogram: uncorrected differences $y_i - \mu_i$; Solid histogram: corrected differences $\hat{\mu}_i - \mu_i$. Left: smallest observations; Right: largest observations (Efron, 2011).



Confidence intervals

Finally, we present the "frequentist assisted Bayes" framework of Woody et al. (2022) for construction of confidence sets.

Same setting as before, with $Y_1, \ldots, Y_n$ independently from $f(y_i; \theta_i)$.

Inference is provided for θ_i iff $Y_i \in E$.

Recall generic construction of frequentist CIs for the selected effects via the conditional approach:

$$R(y) = \{\theta \colon q_1 \leq F_\theta^E(y) \leq q_2\}, \tag{31}$$

where $q_2 - q_1 = 1 - \alpha$, and $F_\theta^E(y) = P_\theta(Y \leq y \mid E)$.

For **equal-tailed** CIs, take $q_1 = \alpha/2$, $q_2 = 1 - \alpha/2$.

Confidence intervals

However, we need not restrict ourselves to equal-tailed intervals and can instead consider more flexible quantile choices.

Consider the family $[q_1, q_2] = [\alpha w, \alpha w + 1 - \alpha]$, where $w \in [0, 1]$ controls the probability allocation in each tail, and let $R_w(y)$ be the respective CI.

Objective: tune w so as to achieve the **shortest expected length** with respect to a prior distribution $\pi(\theta)$.

Confidence intervals

First, need to find a *spending function* $w\colon \mathbb{R} \to [0,1]$ such that $R_{w(\theta)}(Y)$ has minimum expected length given θ.

Let

$$A_w(\theta_0) = \{y\colon \{F^E_{\theta_0}\}^{-1}(\alpha w) \leq y \leq \{F^E_{\theta_0}\}^{-1}(\alpha w + 1 - \alpha)\} \tag{32}$$

be the acceptance region of $H_0\colon \theta = \theta_0$ obtained by inversion of $R_w(\theta)$.

We can write the expected length risk as

$$r(\theta; w) = \int \int \mathbf{1}\{y \in A_w(\tilde{\theta})\} f(y \mid E; \theta) \mathrm{d}\tilde{\theta} \mathrm{d}y. \tag{33}$$

Confidence intervals

For a given prior $\pi(\theta)$, the Bayes risk is thus given by

$$
\begin{aligned}
r(\pi, w) &= \int r(\theta; w)\pi(\theta)\mathrm{d}\theta && (34)\\
&= \int\int\int \mathbf{1}\{y \in A_w(\tilde{\theta})\} f(y \mid E; \theta)\mathrm{d}y\pi(\theta)\mathrm{d}\theta\mathrm{d}\tilde{\theta} && (35)\\
&= \int P\{Y \in A_w(\tilde{\theta})\}\mathrm{d}\tilde{\theta}. && (36)
\end{aligned}
$$

The integrand $P\{Y \in A_w(\theta)\} \equiv H(w; \theta)$ is the marginal probability that Y belongs to the acceptance region.

Confidence intervals

Letting $F_E(y) = P(Y \leq y \mid E)$, we can write

$$H(w;\theta) = F_E[\{F_\theta^E\}^{-1}(\alpha w + 1 - \alpha)] - F_E[\{F_\theta^E\}^{-1}(\alpha w)]. \tag{37}$$

The *optimal spending function* is then given by

$$w^*(\theta) = \arg \min_{w \in [0,1]} H(w;\theta). \tag{38}$$

Finally, the optimal confidence regions are given by

$$R_{w^*(\theta)}(y) = \{\theta \colon y \in A_{w^*(\theta)}\}. \tag{39}$$

Note:

- In general, w^* needs to be computed numerically.
- w^* is not guaranteed to be monotonic, in which case the corresponding confidence region is a union of disjoint intervals.

Confidence intervals

As we saw before, the marginal distribution of Y depends on whether θ is *fixed* or *random*.

For a fixed parameter,

$$f(y \mid E) = \mathbf{1}(y \in E) \int \pi(\theta) \frac{f(y \mid \theta)}{P(E \mid \theta)} \mathrm{d}\theta. \tag{40}$$

For a random parameter,

$$f(y \mid E) = \frac{\mathbf{1}(y \in E) f(y)}{P(E)}. \tag{41}$$

$\rightarrow$Therefore the optimal spending function depends on how the selection mechanism operates on the joint distribution of (θ, Y).

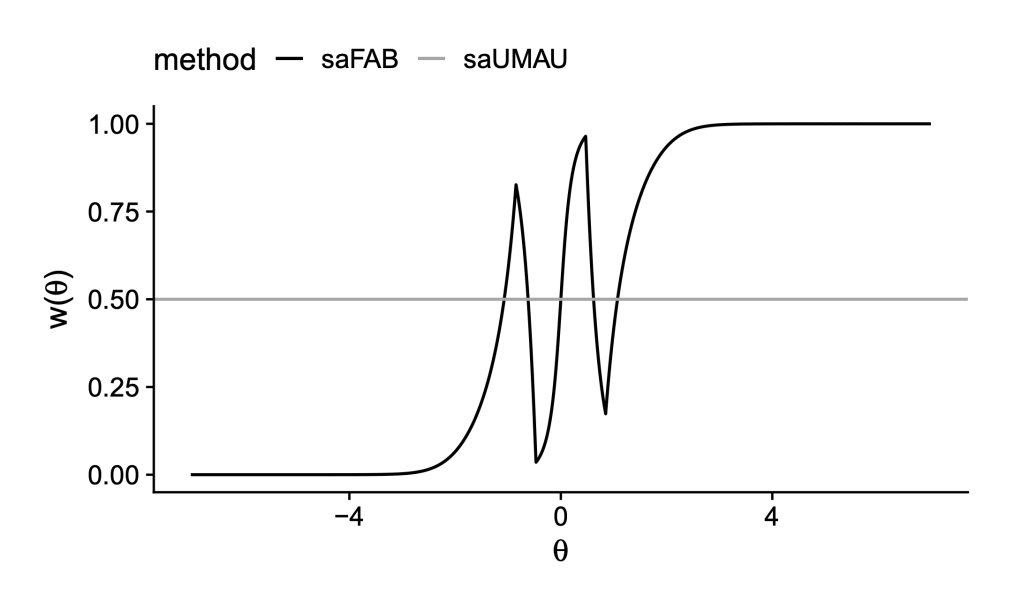
Confidence intervals

As an example, consider the two-groups prior

$$\theta \sim 0.1 \times N(0, 3) + 0.9 \times \delta_0, \quad P(\delta_0 = 0) = 1. \tag{42}$$

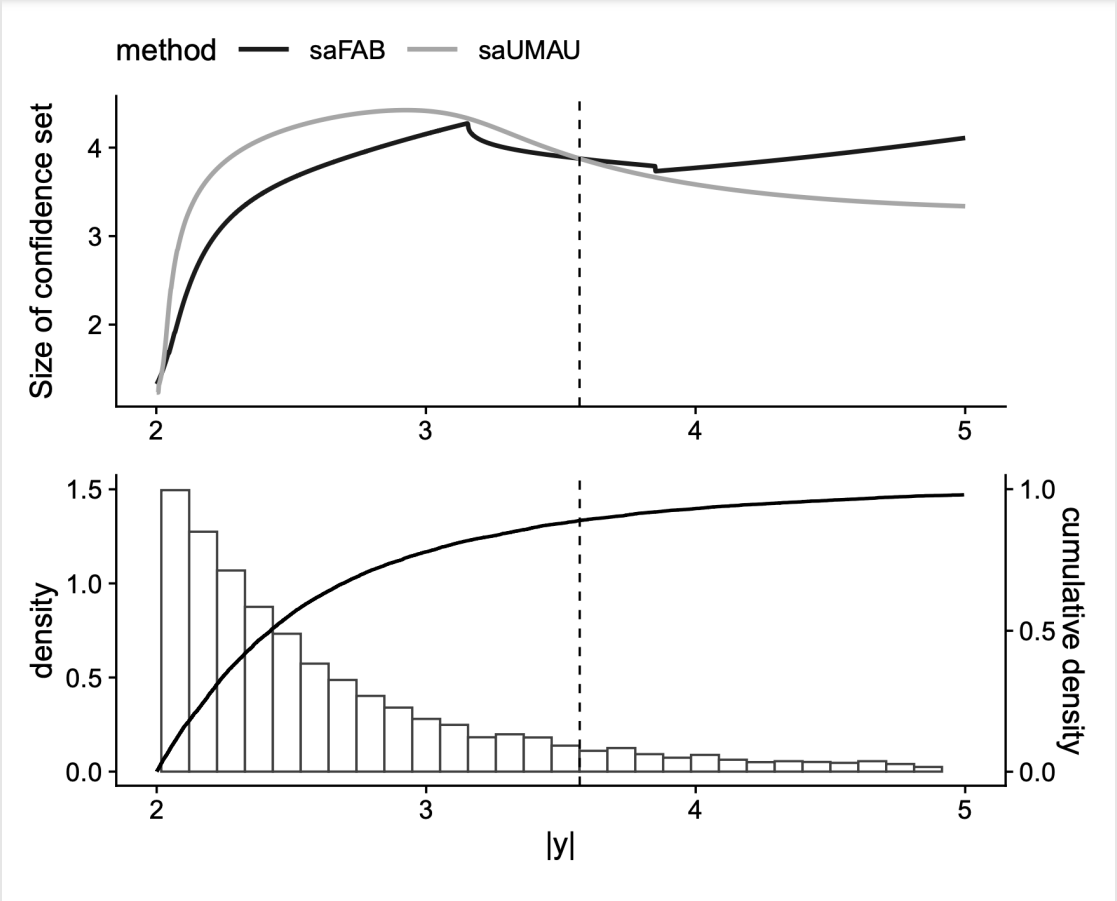
A total of 10000 pairs (θ_i, y_i)'s where sampled from a random-parameter model with $Y_i \mid \theta_i \sim N(\theta_i, 1)$ and selection event $|Y_i| > 2$.

The following figure (Woody et al. 2022). shows the optimal spending function (black), and the spending function for equal-tailed intervals (gray).



Confidence intervals

Figure: First plot: sizes of optimal 90% CIs (black) and equi-tailed CIs (gray) as a function of $|y|$. On average, optimal CIs are 12% smaller. Second plot: marginal distribution of $Y \mid E$ (Woody et al. 2022).



Confidence intervals

To implement this method in practice we need to elicit a prior distribution.

Key result: plugging in an appropriate non-parametric estimator $\hat{\pi}(\theta)$ of the true prior can lead to a consistent approximation of the optimal spending function, and consequently to valid CIs.

But, estimation of the prior is tricky: using the data both for estimation of $\pi(\theta)$ and for construction of the CIs is not valid.

Solution via K-fold data splitting: split observations into K disjoint subsets and construct CIs for selected parameters in a given split using a spending function estimated with data from the remaining splits.

Actual estimation of $\pi(\theta)$ is via a *predictive recursion algorithm* (details in Woody et al., 2022).

Confidence intervals

Now, for a given estimate $\hat{\pi}(\theta)$, the marginal PDF of Y is estimated by

$$\hat{f}(y \mid E) = \mathbf{1}(y \in E) \int \hat{\pi}(\theta) \frac{f(y \mid \theta)}{P(E \mid \theta)} \mathrm{d}\theta \tag{43}$$

for a fixed parameter, or by

$$\hat{f}(y \mid E) = \frac{\mathbf{1}(y \in E)\hat{f}(y)}{\hat{P}(E)}, \tag{44}$$

where

$$\hat{f}(y) = \int f(y \mid \theta)\hat{\pi}(\theta)\mathrm{d}\theta \text{ and } \hat{P}(E) = \int \int \mathbf{1}(y \in E) f(y \mid \theta)\hat{\pi}(\theta)\mathrm{d}y\mathrm{d}\theta, \tag{45}$$

for a random parameter.

This implies an approximate objective function $\hat{H}(\theta; w)$ which can then be minimised to obtain $\hat{w}^*(\theta)$.

Confidence intervals

The following result establishes asymptotic validity of this approximation in the random parameter setting.

Theorem

Assume $E = \{|y| \geq t\}$ for some fixed $t > 0$. In a random parameter setting, under certain regularity conditions, for all fixed θ and every $\varepsilon > 0$,

$$\lim_{n \to \infty} P\left(\sup_{\hat{w}^* \in \hat{\Omega}} \inf_{w^* \in \Omega_0} |\hat{w}^* - w^*| \geq \varepsilon \right) = 0, \tag{46}$$

where $\hat{\Omega}$, Ω_0 are the sets of minimisers of $\hat{H}(\cdot; \theta)$ and $H(\cdot; \theta)$, respectively.

Confidence intervals

Simulation study

– Set $n = 2000$, $\alpha = 0.1$, $Y_i \mid \theta_i \sim N(\theta_i, 1)$, $E = \{|y| > 2\}$, and

$$\theta \sim p \times N(0, \tau^2) + (1 - p) \times \delta_0, \quad P(\delta_0 = 0) = 1, \tag{47}$$

with $p = 0.2$ and $\tau^2 = 3$.

– 1000 batches of n samples generated under fixed and random parameter regimes.

– Three different constructions of the spending function:

1. Oracle, where the prior is fully known.
2. Parametric EB, where the prior is assumed known up to the hyperparameters (p, τ^2), and these are estimated via maximum marginal likelihood estimation using 5-fold data splitting.
3. Nonparametric EB with 5-fold data splitting.

Confidence intervals

Table: Results for fixed parameter.

Method	Coverage	Av. size
Oracle	0.9005	3.5138
PEB	0.9004	3.5113
NPEB	0.8999	3.5123
Eq.-tailed	0.8998	3.7441

Table: Results for random parameter.

Method	Coverage	Av. size
Oracle	0.8990	3.3493
PEB	0.8990	3.3510
NPEB	0.8995	3.3555
Eq.-tailed	0.8993	3.7399

Confidence intervals

Overview:

- Extended "Bayes is unaffected by selection" view questionable in at least some settings.
- In particular, non-informative analyses require a modification of standard, off-the-shelf priors.
- In random parameter problems with many independent observations, empirical Bayes can successfully remove estimation bias regardless of the selection rule.
- Shorter CIs with conditional validity can be computed with EB methods in some circumstances.