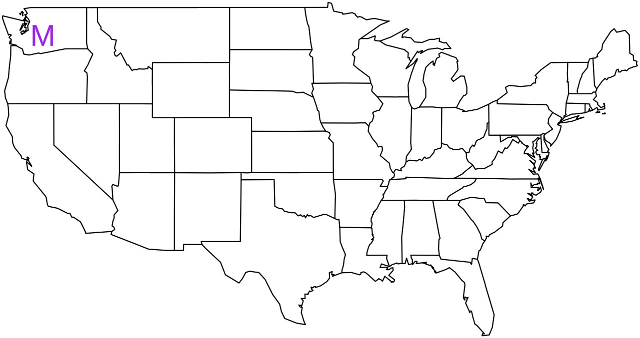
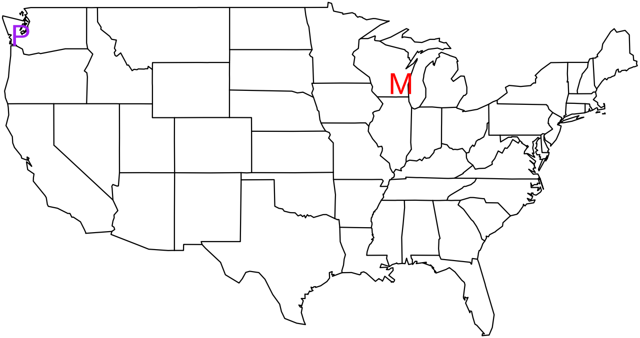
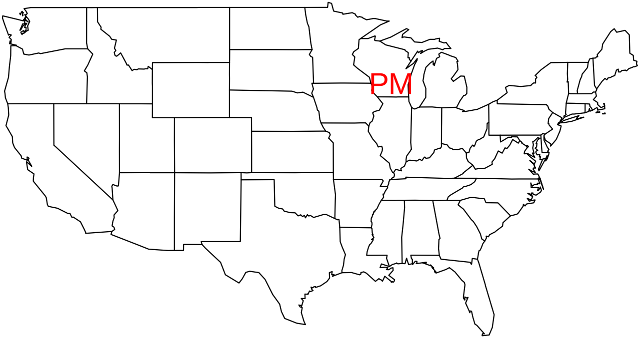
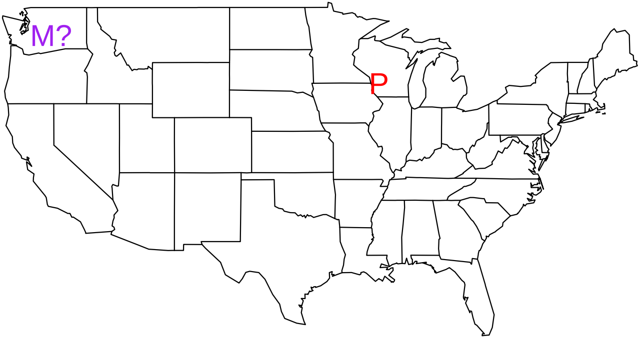
FAB Inference

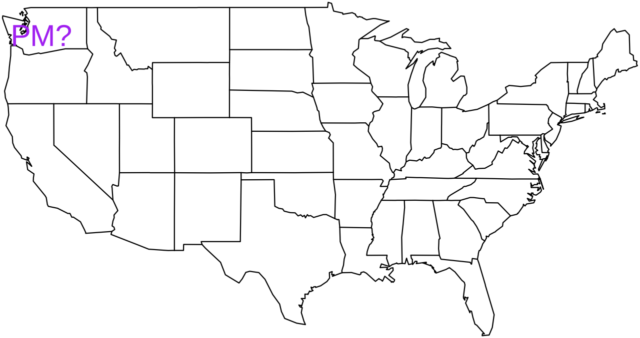
Peter Hoff
Duke University

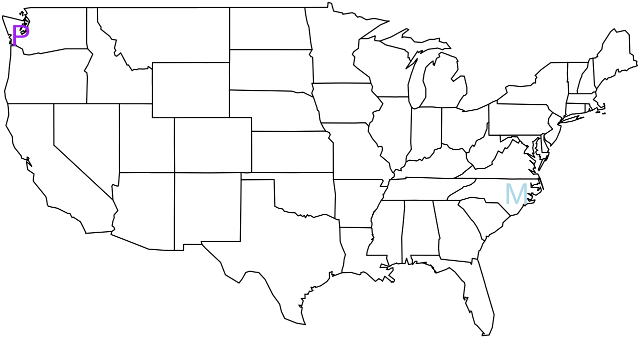

M

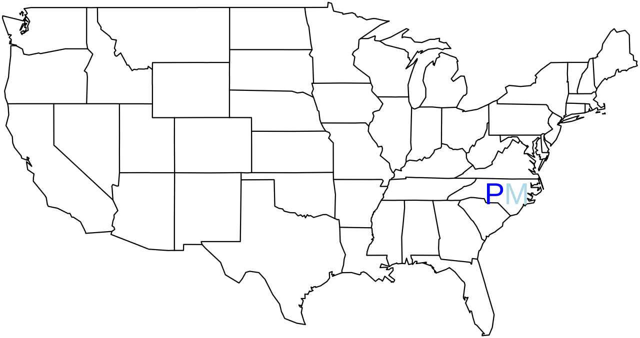

P
M


PM


M?
P


PM?


P
M


PM

Objective objections to Bayesian methods

Computational burden of posterior calculations

- $p(\theta|y)$ is often only available approximately, via simulation.

Lack of frequentist coverage or error rate control

- Posterior CIs lack coverage if $p(\theta)$ inaccurate.

Subjectivity of prior information

- Where does the prior come from? $\theta \sim N(0, 10000)$?

Objective objections to Bayesian methods

Computational burden of posterior calculations

- $p(\theta|y)$ is often only available approximately, via simulation.

Lack of frequentist coverage or error rate control

- Posterior CIs lack coverage if $p(\theta)$ inaccurate.

Subjectivity of prior information

- Where does the prior come from? $\theta \sim N(0, 10000)$?

Objective objections to Bayesian methods

Computational burden of posterior calculations

- $p(\theta|y)$ is often only available approximately, via simulation.

Lack of frequentist coverage or error rate control

- Posterior CIs lack coverage if $p(\theta)$ inaccurate.

Subjectivity of prior information

- Where does the prior come from? $\theta \sim N(0, 10000)$?

Methods that are Frequentist and Bayesian

Bayes methods without posterior calculations

$p(\theta|y)$ not needed.

Proper Bayes methods with exact freqentist coverage/error rate control

Coverage rate does not depend on $p(\theta)$.

Prior distributions from indirect data information

Prior comes from real data.

Methods that are Frequentist and Bayesian

Bayes methods without posterior calculations

$p(\theta|y)$ not needed.

Proper Bayes methods with exact freqentist coverage/error rate control

Coverage rate does not depend on $p(\theta)$.

Prior distributions from indirect data information

Prior comes from real data.

Methods that are Frequentist and Bayesian

Bayes methods without posterior calculations

$p(\theta|y)$ not needed.

Proper Bayes methods with exact freqentist coverage/error rate control

Coverage rate does not depend on $p(\theta)$.

Prior distributions from indirect data information

Prior comes from real data.

Bayes optimal procedure

- Model: $y \sim P_\theta$, some $\theta \in \Theta$.
- Loss: $L(\delta, \theta)$, $\delta \in \mathcal{D}$, $\theta \in \Theta$.

For a procedure $d : \mathcal{Y} \to \mathcal{D}$,

$$R(d, \theta) = \int L(d(y), \theta)\, P(dy|\theta).$$

For a prior $\theta \sim \pi$,

$$R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

The Bayes procedure is the minimizer of the Bayes risk:

$$d_\pi = \arg\min_d \; R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

Bayes optimal procedure

- Model: $y \sim P_\theta$, some $\theta \in \Theta$.
- Loss: $L(\delta, \theta)$, $\delta \in \mathcal{D}$, $\theta \in \Theta$.

For a procedure $d : \mathcal{Y} \to \mathcal{D}$,

$$R(d, \theta) = \int L(d(y), \theta)\, P(dy|\theta).$$

For a prior $\theta \sim \pi$,

$$R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

The Bayes procedure is the minimizer of the Bayes risk:

$$d_\pi = \arg\min_d \; R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

Bayes optimal procedure

- Model: $y \sim P_\theta$, some $\theta \in \Theta$.
- Loss: $L(\delta, \theta)$, $\delta \in \mathcal{D}$, $\theta \in \Theta$.

For a procedure $d : \mathcal{Y} \to \mathcal{D}$,

$$R(d, \theta) = \int L(d(y), \theta)\, P(dy|\theta).$$

For a prior $\theta \sim \pi$,

$$R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

The Bayes procedure is the minimizer of the Bayes risk:

$$d_\pi = \arg\min_d \; R(d, \pi) = \int R(d, \theta)\, \pi(d\theta).$$

Bayes estimators from posterior risk

$$\begin{aligned} R(d,\pi) &= \int R(d,\theta)\,\pi(d\theta) \\ &= \int\int L(d(y),\theta)\,P_\theta(dy)\,\pi(d\theta) \\ &= \int \left(\int L(d(y),\theta)\,\pi(d\theta|y) \right) P_\pi(dy). \end{aligned}$$

For each y, let

$$d_\pi(y) = \arg\min_{d\in\mathcal{D}} \int L(d(y),\theta)\,\pi(d\theta|y).$$

Then $d_\pi(\cdot)$ minimizes $R(d,\pi)$.

Bayes estimators from posterior risk

$$
\begin{aligned}
R(d,\pi) &= \int R(d,\theta)\,\pi(d\theta) \\
&= \int\int L(d(y),\theta)\,P_\theta(dy)\,\pi(d\theta) \\
&= \int \left(\int L(d(y),\theta)\,\pi(d\theta|y) \right) P_\pi(dy).
\end{aligned}
$$

For each y, let

$$d_\pi(y) = \arg\min_{d\in\mathcal{D}} \int L(d(y),\theta)\,\pi(d\theta|y).$$

Then $d_\pi(\cdot)$ minimizes $R(d,\pi)$.

Interpretations

Pragmatic Bayes: Average risk optimality

- π is a "weighting function";
- d_π is the procedure that minimizes the risk on-average over θ values.

Subjective Bayes: A state of mind

- $\pi(\theta|y)$ is "where you think θ is", given y.
- $d_\pi(y)$ is the action that minimizes loss on-average over θ values, given y.

Interpretations

Pragmatic Bayes: Average risk optimality

- π is a "weighting function";
- d_π is the procedure that minimizes the risk on-average over θ values.

Subjective Bayes: A state of mind

- $\pi(\theta|y)$ is "where you think θ is", given y.
- $d_\pi(y)$ is the action that minimizes loss on-average over θ values, given y.

Bayes procedures with no posteriors

What if L doesn't depend on θ?

- $d(y)$ is an interval, $L(d(y))$ is its width.

$$\begin{aligned} R(d,\pi) &= \int R(d,\theta)\,\pi(d\theta) \\ &= \int\int L(d(y))\,P_\theta(dy)\,\pi(d\theta) \\ &= \int \left(\int L(d(y))\,\pi(d\theta|y)\right)\,P_\pi(dy) \\ &= \int L(d(y))\,P_\pi(dy). \end{aligned}$$

- The Bayes-optimal procedure minimizes the prior expected loss.
- A Posterior is not relevant - the "posterior risk" is $L(d(y))$.

Bayes procedures with no posteriors

What if L doesn't depend on θ?

- $d(y)$ is an interval, $L(d(y))$ is its width.

$$
\begin{aligned}
R(d,\pi) &= \int R(d,\theta)\,\pi(d\theta) \\
&= \int\int L(d(y))\,P_\theta(dy)\,\pi(d\theta) \\
&= \int \left(\int L(d(y))\,\pi(d\theta|y) \right) P_\pi(dy) \\
&= \int L(d(y))\,P_\pi(dy).
\end{aligned}
$$

- The Bayes-optimal procedure minimizes the prior expected loss.
- A Posterior is not relevant - the "posterior risk" is $L(d(y))$.

FAB confidence procedures

Task: Based on $y \sim p_\theta(y),\ \theta \in \Theta$, construct $C(y) : \mathcal{Y} \to 2^\Theta$ that has

- constant frequentist coverage,

$$\Pr(\theta \in C(y)|\theta) = 1 - \alpha \ \forall \theta \in \Theta,$$

- optimal prior expected width,

$$\mathrm{E}[|C(y)|] \leq \mathrm{E}[|C'(y)|] \text{ among all } C' \text{ with } 1-\alpha \text{ frequentist coverage.}$$

All frequentist CIPs

$$y \sim N(\theta, \sigma^2), \ \theta \in \mathbb{R}.$$

Standard procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w)} < \theta < y + \sigma z_{1-\alpha w}\}$$

In fact, w may depend on θ: If $w : \mathbb{R} \to [0,1]$ then

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\}$$

satisfies $\Pr(\theta \in C_w(y)|\theta) = 1 - \alpha$

- Examples in Bartholomew [1971], Stein [1962].
- Essentially complete class result in Yu and Hoff [2018].

All frequentist CIPs

$$y \sim N(\theta, \sigma^2), \ \theta \in \mathbb{R}.$$

Standard procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w)} < \theta < y + \sigma z_{1-\alpha w}\}$$

In fact, w may depend on θ: If $w : \mathbb{R} \to [0,1]$ then

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\}$$

satisfies $\Pr(\theta \in C_w(y)|\theta) = 1 - \alpha$

- Examples in Bartholomew [1971], Stein [1962].
- Essentially complete class result in Yu and Hoff [2018].

All frequentist CIPs

$$y \sim N(\theta, \sigma^2), \ \theta \in \mathbb{R}.$$

Standard procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w)} < \theta < y + \sigma z_{1-\alpha w}\}$$

In fact, w may depend on θ: If $w : \mathbb{R} \rightarrow [0, 1]$ then

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\}$$

satisfies $\Pr(\theta \in C_w(y)|\boldsymbol{\theta}) = 1 - \alpha$

- Examples in Bartholomew [1971], Stein [1962].
- Essentially complete class result in Yu and Hoff [2018].

FAB: Bayes-optimal frequentist interval

Task: Find the w-function that minimizes the prior expected width

$$\int\int |C_w(y)|\, p(dy|\theta)\pi(d\theta) < \int\int |C(y)|\, p(dy|\theta)\pi(d\theta)$$

Such an interval will have

- constant coverage, because C_w has constant coverage for any w-function;
- optimal precision on average with respect to π, by construction.

Pratt's (1961) Bayes-optimal 1-α frequentist interval:

$$C_\pi(y) = \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\}$$
$$w(\theta) = g^{-1}(2\sigma(\theta - \mu)/\tau^2)$$
$$g(w) = \Phi^{-1}(\alpha w) - \Phi^{-1}(\alpha(1-w))$$

But where does the prior come from?

FAB: Bayes-optimal frequentist interval

Task: Find the w-function that minimizes the prior expected width

$$\int\int |C_w(y)|\, p(dy|\theta)\pi(d\theta) < \int\int |C(y)|\, p(dy|\theta)\pi(d\theta)$$

Such an interval will have

- constant coverage, because C_w has constant coverage for any w-function;
- optimal precision on average with respect to π, by construction.

Pratt's (1961) Bayes-optimal 1-α frequentist interval:

$$\begin{aligned} C_\pi(y) &= \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\} \\ w(\theta) &= g^{-1}(2\sigma(\theta-\mu)/\tau^2) \\ g(w) &= \Phi^{-1}(\alpha w) - \Phi^{-1}(\alpha(1-w)) \end{aligned}$$

But where does the prior come from?

FAB: Bayes-optimal frequentist interval

Task: Find the w-function that minimizes the prior expected width

$$\int\int |C_w(y)|\, p(dy|\theta)\pi(d\theta) < \int\int |C(y)|\, p(dy|\theta)\pi(d\theta)$$

Such an interval will have

- constant coverage, because C_w has constant coverage for any w-function;
- optimal precision on average with respect to π, by construction.

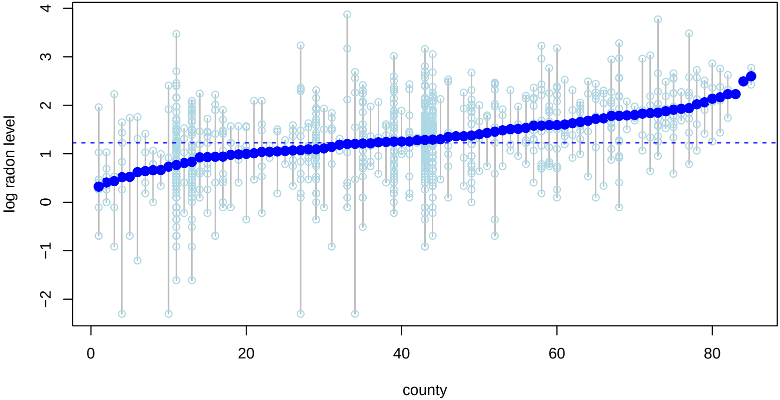
Pratt's (1961) Bayes-optimal 1-α frequentist interval:

$$\begin{aligned} C_\pi(y) &= \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\} \\ w(\theta) &= g^{-1}(2\sigma(\theta - \mu)/\tau^2) \\ g(w) &= \Phi^{-1}(\alpha w) - \Phi^{-1}(\alpha(1-w)) \end{aligned}$$

But where does the prior come from?

Radon data

Log radon levels of 85 Minnesota counties.



Model and inferential goals

$$
\begin{aligned}
y_{1,1},\ldots,y_{n_1,1} &\sim \text{ i.i.d. } N(\theta_1,\sigma_1^2) \\
y_{1,2},\ldots,y_{n_2,2} &\sim \text{ i.i.d. } N(\theta_2,\sigma_2^2) \\
\vdots \qquad & \qquad \vdots \\
y_{1,p},\ldots,y_{n_p,p} &\sim \text{ i.i.d. } N(\theta_p,\sigma_p^2)
\end{aligned}
$$

Some inferential goals:

- confidence interval for the mean of each county:

$$\Pr(\theta_j \in C_j(\mathbf{y})|\boldsymbol{\theta},\boldsymbol{\sigma}^2) = 1-\alpha.$$

- prediction interval for a new observation:

$$\Pr(\tilde{y}_j \in A_j(\mathbf{y})|\boldsymbol{\theta},\boldsymbol{\sigma}^2) = 1-\alpha.$$

Model and inferential goals

$$
\begin{aligned}
y_{1,1}, \ldots, y_{n_1,1} &\sim \text{ i.i.d. } N(\theta_1, \sigma_1^2) \\
y_{1,2}, \ldots, y_{n_2,2} &\sim \text{ i.i.d. } N(\theta_2, \sigma_2^2) \\
&\vdots \qquad\qquad \vdots \\
y_{1,p}, \ldots, y_{n_p,p} &\sim \text{ i.i.d. } N(\theta_p, \sigma_p^2)
\end{aligned}
$$

Some inferential goals:

- confidence interval for the mean of each county:

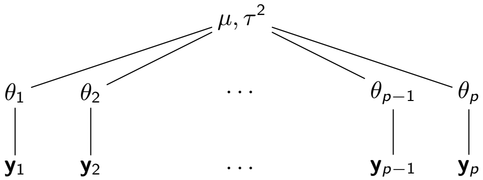
$$\Pr(\theta_j \in C_j(\mathbf{y}) | \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = 1 - \alpha.$$

- prediction interval for a new observation:

$$\Pr(\tilde{y}_j \in A_j(\mathbf{y}) | \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = 1 - \alpha.$$

Indirect methods via hierarchical models

- Share information across groups with a *linking model*.
- linking model + sampling model = *hierarchical model*.



$$
\begin{aligned}
y_{1,j}, \ldots, y_{n_j,j} | \boldsymbol{\theta} &\sim \text{i.i.d. } N(\theta_j, \sigma_j^2) \\
\theta_1, \ldots, \theta_p &\sim \text{i.i.d. } N(\mu, \tau^2)
\end{aligned}
$$

Indirect methods via hierarchical models

$$\theta_j|\{\mathbf{y},\sigma_j^2,\mu,\tau^2\} \sim N(\hat\theta_j, 1/(n_j/\sigma_j^2 + 1/\tau^2))$$

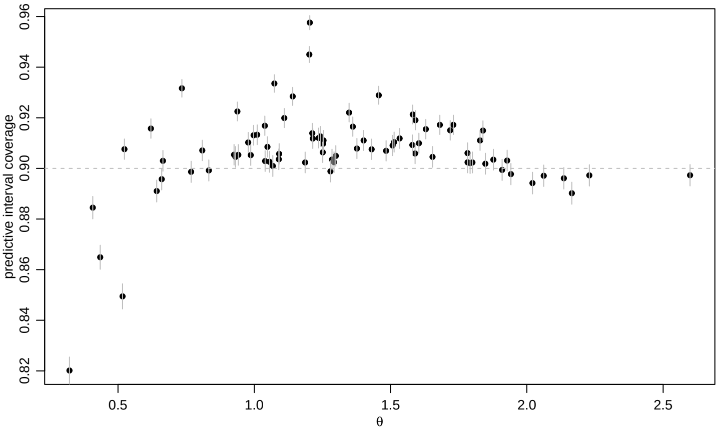
$$\hat\theta_j \approx \frac{n_j/s_j^2}{n_j/\hat s_j^2 + 1/\hat\tau^2}\bar y_j + \frac{1/\hat\tau^2}{n_j/s_j^2 + 1/\hat\tau^2}\hat\mu$$

$$C_j = \hat\theta_j \pm t_{1-\alpha/2}/\sqrt{s_j^2/n_j + 1/\hat\tau^2}$$

$$A_j = \hat\theta_j \pm t_{1-\alpha/2}\sqrt{s_j^2 + 1/(n_j/s_j^2 + 1/\hat\tau^2)}$$

- information is shared via $\hat\mu$, $\hat\tau^2$
- narrower than direct methods
- coverage varies across groups/θ_j's

Nonconstant coverage: Radon data



$$\Pr(\theta_j \in C(\hat{\theta}_j)) \approx 1 - \alpha$$

$$\Pr(\theta_j \in C(\hat{\theta}_j)|\boldsymbol{\theta}) \quad \text{depends on } \theta_j.$$

Adaptive FAB for multigroup inference (Yu and Hoff 2019)

For each group $j = 1, \ldots, p$:

1. Obtain $\hat{\mu}$, $\hat{\tau}^2$, $\hat{\sigma}^2$ using data from groups other than j;
2. Obtain $\hat{w}_j(\theta) = g^{-1}(2\hat{\sigma}(\theta - \hat{\mu})/\hat{\tau}^2)$;
3. Construct $C_{\hat{w}_j}(\bar{y}_j)$.

- Exact $1 - \alpha$ coverage *for each group*, even if hierarchical model is wrong.
- Improved precision *on average across groups*.

Adaptive FAB for multigroup inference (Yu and Hoff 2019)

For each group $j = 1, \ldots, p$:

1. Obtain $\hat{\mu}$, $\hat{\tau}^2$, $\hat{\sigma}^2$ using data from groups other than j;
2. Obtain $\hat{w}_j(\theta) = g^{-1}(2\hat{\sigma}(\theta - \hat{\mu})/\hat{\tau}^2)$;
3. Construct $C_{\hat{w}_j}(\bar{y}_j)$.

- Exact $1 - \alpha$ coverage *for each group*, even if hierarchical model is wrong.
- Improved precision *on average across groups*.

Spatial small area inference (Burris and Hoff 2019)

Spatial small area inference (Burris and Hoff 2019)

Sampling model: $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$ independently across groups.

Linking Model: $\theta_j = \boldsymbol{\beta}^\top \mathbf{x}_j + e_j$, $\text{Cov}[\boldsymbol{\theta}] = \Sigma$ (spatial FH model).

Direct interval: $\bar{y}_j \pm \hat{\sigma}_j t_{1-\alpha/2}$

AFAB interval: For each area $j = 1, \ldots, p$

1. using areas other than j, obtain estimates of $\boldsymbol{\theta}_{-j}$, β and Σ;
2. obtain "prior" distribution for θ_j from estimates and working model;
3. compute optimal w-function and construct FAB interval for θ_j.

- Both intervals have $1 - \alpha$ area-specific coverage, under random sampling within each area. The linking model need not be correct.
- FAB intervals make use of information from neighboring areas and known area-level characteristics (surficial radium).

Spatial small area inference (Burris and Hoff 2019)

Sampling model: $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$ independently across groups.

Linking Model: $\theta_j = \boldsymbol{\beta}^\top \mathbf{x}_j + e_j$, $\mathrm{Cov}[\boldsymbol{\theta}] = \Sigma$ (spatial FH model).

Direct interval: $\bar{y}_j \pm \hat{\sigma}_j t_{1-\alpha/2}$

AFAB interval: For each area $j = 1, \ldots, p$

1. using areas other than j, obtain estimates of $\boldsymbol{\theta}_{-j}$, $\boldsymbol{\beta}$ and Σ;
2. obtain "prior" distribution for θ_j from estimates and working model;
3. compute optimal w-function and construct FAB interval for θ_j.

- Both intervals have $1-\alpha$ area-specific coverage, under random sampling within each area. The linking model need not be correct.
- FAB intervals make use of information from neighboring areas and known area-level characteristics (surficial radium).

Spatial small area inference (Burris and Hoff 2019)

Sampling model: $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$ independently across groups.

Linking Model: $\theta_j = \boldsymbol{\beta}^\top \mathbf{x}_j + e_j$, $\text{Cov}[\boldsymbol{\theta}] = \Sigma$ (spatial FH model).

Direct interval: $\bar{y}_j \pm \hat{\sigma}_j t_{1-\alpha/2}$

AFAB interval: For each area $j = 1, \ldots, p$

1. using areas other than j, obtain estimates of $\boldsymbol{\theta}_{-j}$, $\boldsymbol{\beta}$ and Σ;
2. obtain "prior" distribution for θ_j from estimates and working model;
3. compute optimal w-function and construct FAB interval for θ_j.

- Both intervals have $1 - \alpha$ area-specific coverage, under random sampling within each area. The linking model need not be correct.
- FAB intervals make use of information from neighboring areas and known area-level characteristics (surficial radium).

Spatial small area inference (Burris and Hoff 2019)

Sampling model: $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$ independently across groups.

Linking Model: $\theta_j = \boldsymbol{\beta}^\top \mathbf{x}_j + e_j$, $\text{Cov}[\boldsymbol{\theta}] = \Sigma$ (spatial FH model).

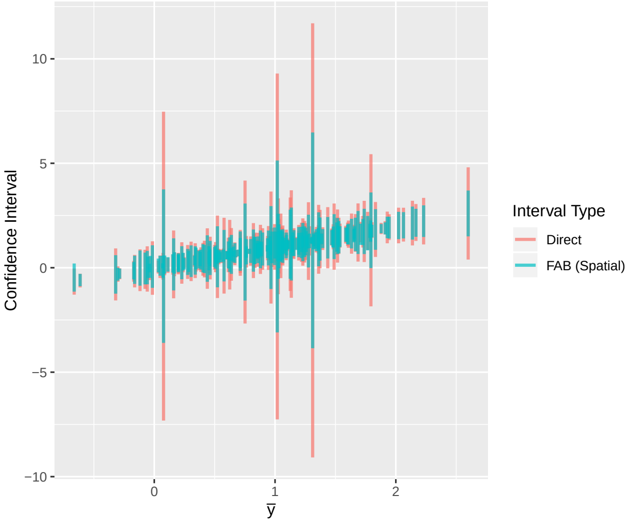
Direct interval: $\bar{y}_j \pm \hat{\sigma}_j t_{1-\alpha/2}$

AFAB interval: For each area $j = 1, \ldots, p$

1. using areas other than j, obtain estimates of $\boldsymbol{\theta}_{-j}$, $\boldsymbol{\beta}$ and Σ;
2. obtain "prior" distribution for θ_j from estimates and working model;
3. compute optimal w-function and construct FAB interval for θ_j.

- Both intervals have $1 - \alpha$ area-specific coverage, under random sampling within each area. The linking model need not be correct.
- FAB intervals make use of information from neighboring areas and known area-level characteristics (surficial radium).

Interval comparisons



By sharing information, FAB intervals can improve on across-group performance, even if the linking model is wrong.

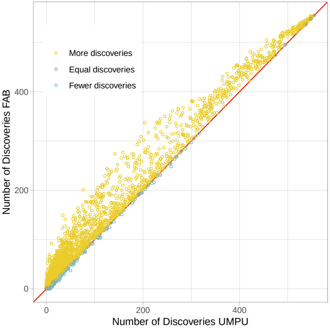
FAB testing

Testing problem:

$$H : Y \sim P_{\theta_0} \quad \text{versus}$$
$$K : Y \sim P_{\theta}, \ \theta \in \Theta$$

Bayes-optimal frequentist test: MP level-α test of

$$H : Y \sim P_{\theta_0} \text{versus}$$
$$K_{\pi} : Y \sim \int P_{\theta}\, \pi(d\theta).$$



- Bayes-optimal frequentist tests and p-values (Hoff [2022])
- Tensor models for genomics data (Bryan and Hoff [2023])
- Multivariate populations (McCormack and Hoff [2023]).

FAB prediction

$$\Pr(Y \in C(X)|\theta) = 1 - \alpha \; \forall \theta \in \Theta.$$

Conceptual steps:

1. Characterize all $1-\alpha$ frequentist prediction procedures;
2. Compute the prior expected volume of each;
3. Use the procedure with minimum prior expected volume.

Some results:

- Bayes optimal conformal procedure uses $p(Y_{n+1}|Y_1, \ldots, Y_n)$ as conformity score (Hoff [2023]).
- Multigroup conformal inference for continuous and categorical data (Bersson and Hoff [2023,2024]).

FAB prediction

$$\Pr(Y \in C(X)|\theta) = 1 - \alpha \ \forall \theta \in \Theta.$$

Conceptual steps:

1. Characterize all $1 - \alpha$ frequentist prediction procedures;
2. Compute the prior expected volume of each;
3. Use the procedure with minimum prior expected volume.

Some results:

- Bayes optimal conformal procedure uses $p(Y_{n+1}|Y_1, \ldots, Y_n)$ as conformity score (Hoff [2023]).
- Multigroup conformal inference for continuous and categorical data (Bersson and Hoff [2023,2024]).

Summary

Group-level inference motivates group-level error rate guarantees.

Error rate guarantees do not preclude inclusion of prior or indirect information.

FAB inference

- maintains group-level frequentist error rates;
- improves performance on-average across groups;
- rates maintained even if the prior/linking model is wrong.

Summary

Group-level inference motivates group-level error rate guarantees.

Error rate guarantees do not preclude inclusion of prior or indirect information.

FAB inference

- maintains group-level frequentist error rates;
- improves performance on-average across groups;
- rates maintained even if the prior/linking model is wrong.

Summary

Group-level inference motivates group-level error rate guarantees.

Error rate guarantees do not preclude inclusion of prior or indirect information.

FAB inference

- maintains group-level frequentist error rates;
- improves performance on-average across groups;
- rates maintained even if the prior/linking model is wrong.