第5回 DSRT
医薬品の臨床試験および製造販売後調査における ベイズ統計学の活用について

議題２：ベイズ流の統計手法と 頻度論の統計手法 【ディスカッションテーマ】

背景

- 現在… 医薬品開発で利用される統計的推測手法
  - ベイズ流 (Bayes factor, 事後確率, 信用区間)：限られた方法・状況 (用量探索試験など)
  - 頻度論 (仮説検定, 信頼区間)：幅広く用いられる
- ベイズ統計学に基づく統計手法の特性や特徴は, 頻度論の手法に比べ, 十分理解されているとは言い難い
- 理論, シミュレーション等を通じ, 両手法の特性や特徴について議論し, 医薬品開発の文脈におけるベイズ流の統計手法の現状についての理解を深めたい

ディスカッションテーマ

- ベイズ流・頻度論の統計手法について議論
  - 二値判断：Bayes factorと仮説検定
  - 推定：パラメータの事後分布の信用区間と信頼区間
- 「ベイズ流 vs 頻度論を議論の目的とはしない」

- 各手法の特徴・特性は？
- 各手法はどのような状況で使える？ベイズ流の統計手法があまり使われていない理由は？

二値判断

Bayes Factorと仮説検定

二値判断の統計手法

- 試験の結果から「有効・無効」,「 あり・なし」 などを判断
- さまざまな手法があるが, 本テーマでは以下を扱う
  - ベイズ流：Bayes factor (BF)
    - ASAの声明 (2016) でも言及
  - 頻度論：仮説検定

Bayes factor (BF)

- 事後オッズ比を用いた仮説の評価 (Jeffreys, 1961)

$$\Omega = \frac{\Pr(H_0 \mid data)}{\Pr(H_1 \mid data)} = \frac{f(data \mid H_0)}{f(data \mid H_1)} \frac{\Pr(H_0)}{\Pr(H_1)}$$

  - 事前オッズ：$\Pr(H_0)/\Pr(H_1)$
  - 周辺尤度：$f(data \mid H) = \int_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathrm{H}}} f_H(data \mid \boldsymbol{\theta}) p_H(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$
- Bayes Factorは周辺尤度, $f(data \mid H_k)$, の比 (≠尤度比検定)

$$BF_{01} = \frac{f(data \mid H_0)}{f(data \mid H_1)} \quad \left(= \Omega \frac{\Pr(H_1)}{\Pr(H_0)} \right)$$

  - 多くのベイズ統計学の教科書に記載

BFの判断基準の例

- Jeffreys (1961)
  - $BF_{01} \geq 3$ or $BF_{01} \leq 1/3$ ：”some evidence”
  - $BF_{01} \geq 10$ or $BF_{01} \leq 1/10$ ：”strong evidence”
  - $BF_{01} \geq 30$ or $BF_{01} \leq 1/30$ ：”very strong evidence”
- Kass & Raftery (1995)
  - $1 < BF_{01} \leq 3$ ：”not worth more than a bare mention”
  - $3 < BF_{01} \leq 20$ ：”positive”
  - $20 < BF_{01} \leq 150$ ：”strong”
  - $BF_{01} > 150$ ：”very strong”

二標本平均の二値判断手法

- 仮説$H_0: \delta = \mu_1 - \mu_2 = 0, H_1: \delta \neq 0$の評価
  - 各群1&2から$n_1 = n_2$個の標本が得られた時の, 群間の母平均に差があるかどうかを検討
- 検討する手法
  - 頻度論：二標本*t*検定 (詳細は省略)
  - ベイズ流 (BF)：BFを用いたベイズ流*t*-Test (Gönen et al., 2005)

BFを用いたベイズ流t-Test (Gönen et al., 2005)

- パラメータ： $\delta = \mu_1 - \mu_2, \mu = 0.5(\mu_1 + \mu_2), \sigma^2$
- 事前分布：$\frac{\delta}{\sigma} \sim N(\lambda, \tau^2), \mathrm{Pr}(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ (nuisance)

H_0の周辺尤度

$$BF_{01} = \frac{T_\nu(t_* \mid 0)}{T_\nu\left(t_*/\sqrt{1+m\tau^2} \mid m^{1/2}\lambda/\sqrt{1+m\tau^2}\right)/\sqrt{1+m\tau^2}}$$

H_1の周辺尤度

- $T_\nu(\cdot \mid c)$：非心度c, 自由度νの非心t分布の密度関数, t_*：標本のt統計量, $m = (n_1^{-1} + n_2^{-1})^{-1}$：effective sample size, $\nu = n_1 + n_2 - 2$, 事前オッズ比は1を仮定

シミュレーション

- いくつかの条件下で, 両手法の特徴・特性を理解したい

- 目的
  - BFと仮説検定の両手法の特徴を検討する
  - あくまで比較は目的としない

※特徴を見るためにあえて極端な事前分布を設定

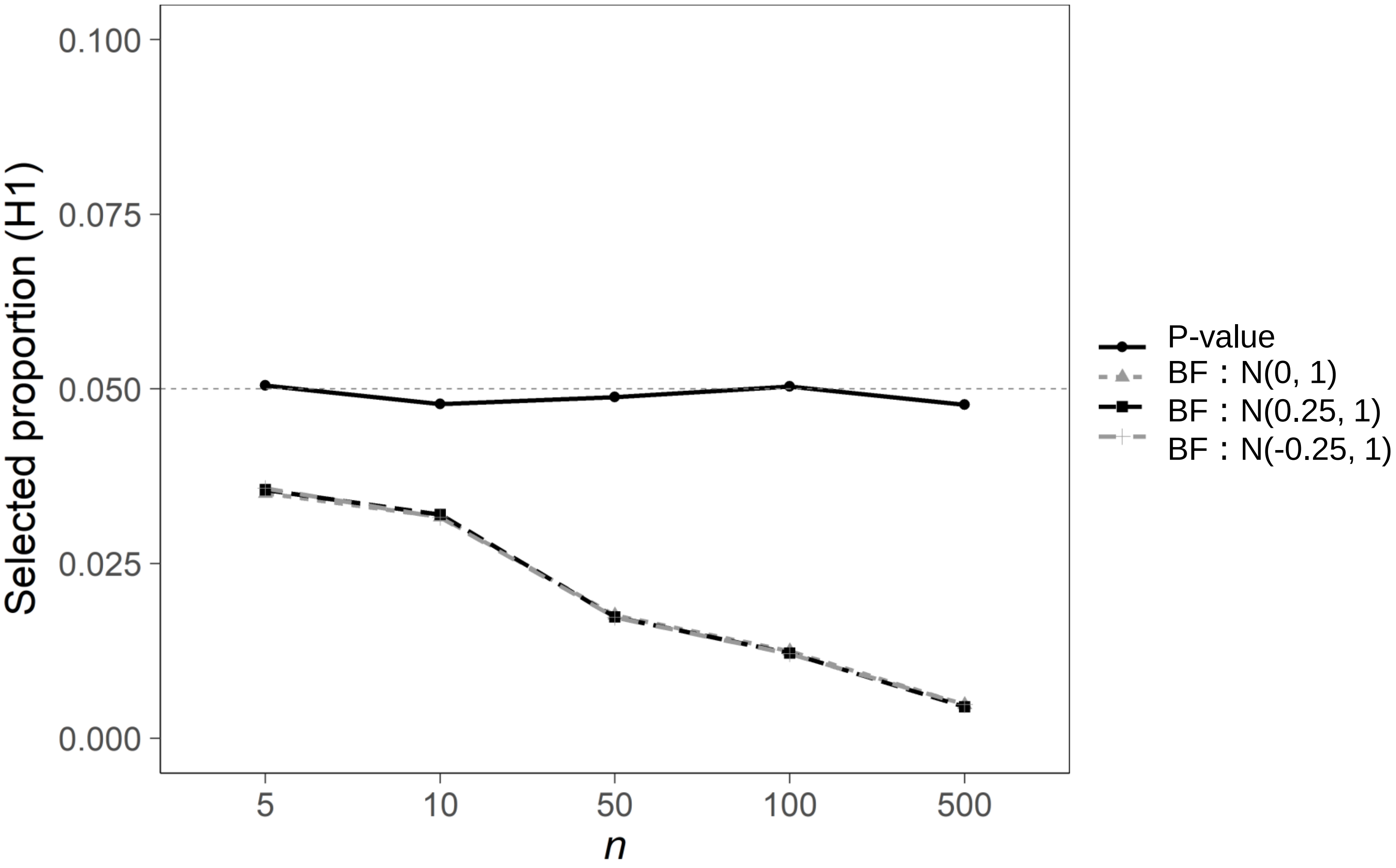
シミュレーションデータの生成条件

- 標本を正規乱数で生成 (25,000セット)
  - $n_1 = n_2 = 5, 10, 50, 100, 500$
- シナリオ1：$N(0, 1)$
  - 群間差がない (帰無仮説が正しいとき)
  - $\delta = \mu_1 - \mu_2 = 0$
- シナリオ2：$N(-0.25, 1)$
  - 群間差がある (対立仮説が正しいとき)
  - $\delta = \mu_1 - \mu_2 = -0.25$

シミュレーションで検討する手法 (1)

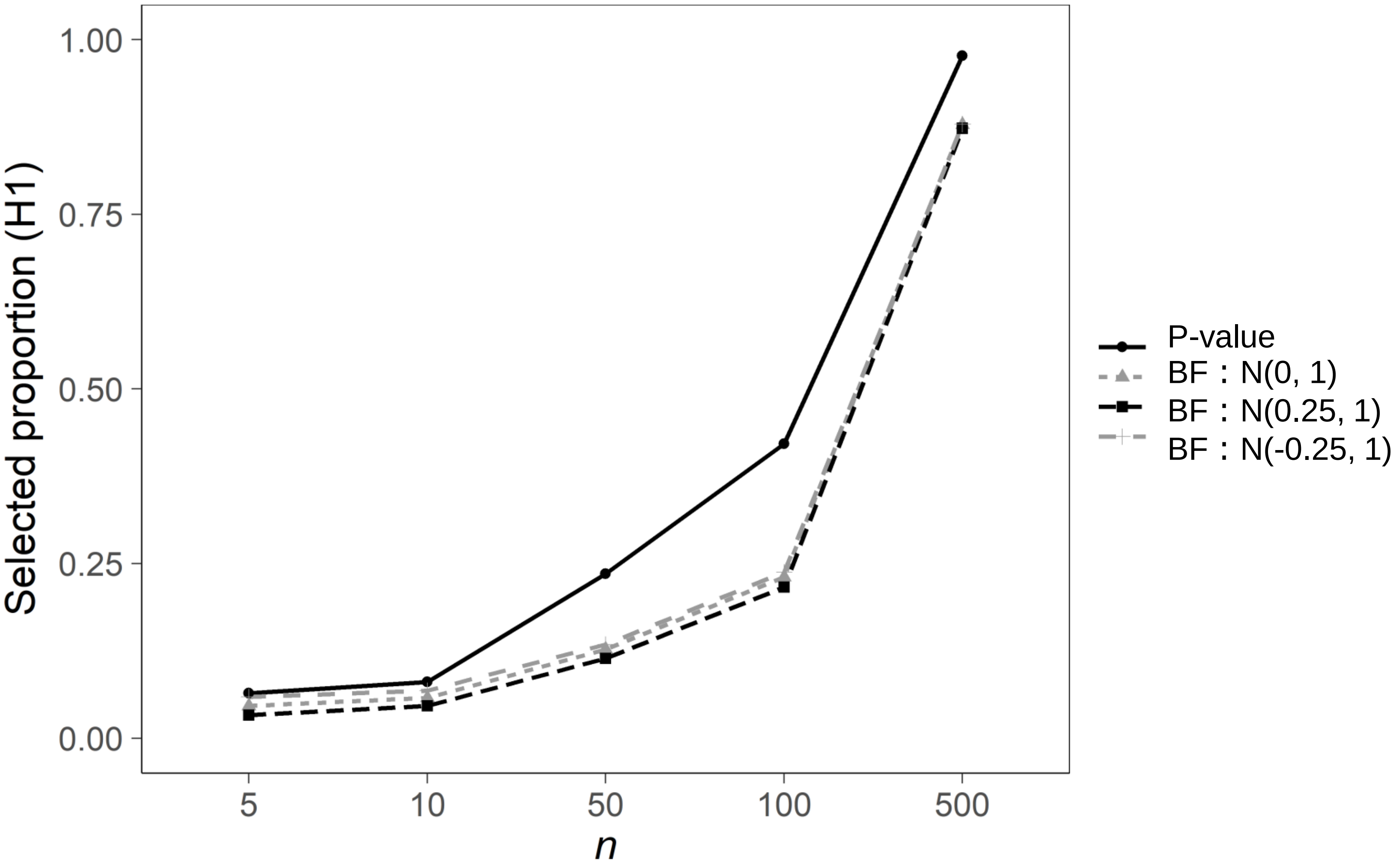
- 二標本t検定（基準：$p < 0.05$）
- BF (基準：BF ≦ 1/3)
  - 情報が少ない事前分布：$\delta/\sigma \sim N(0,1)$
  - 情報が少ない対立仮説からずれた事前分布：$\delta/\sigma \sim N(0.25,1)$
  - 情報が少ない対立仮説に近い事前分布：$\delta/\sigma \sim N(-0.25,1)$
- 性能指標：H_1を選択する割合

結果：シナリオ1



0.100
0.075
0.050
0.025
0.000
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 1)
BF：N(-0.25, 1)

結果：シナリオ2



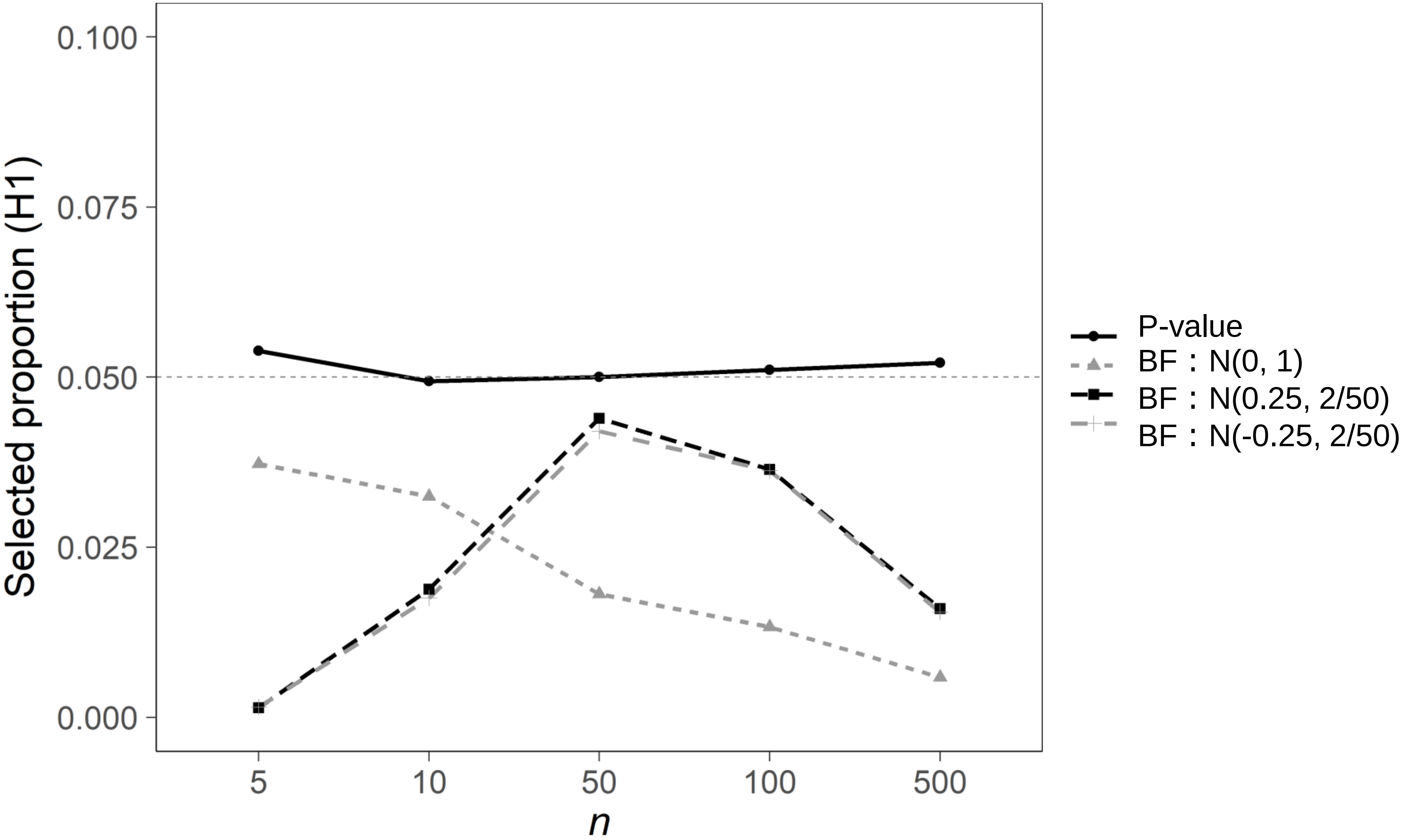
1.00
0.75
0.50
0.25
0.00
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 1)
BF：N(-0.25, 1)

シミュレーションで検討する手法 (2)

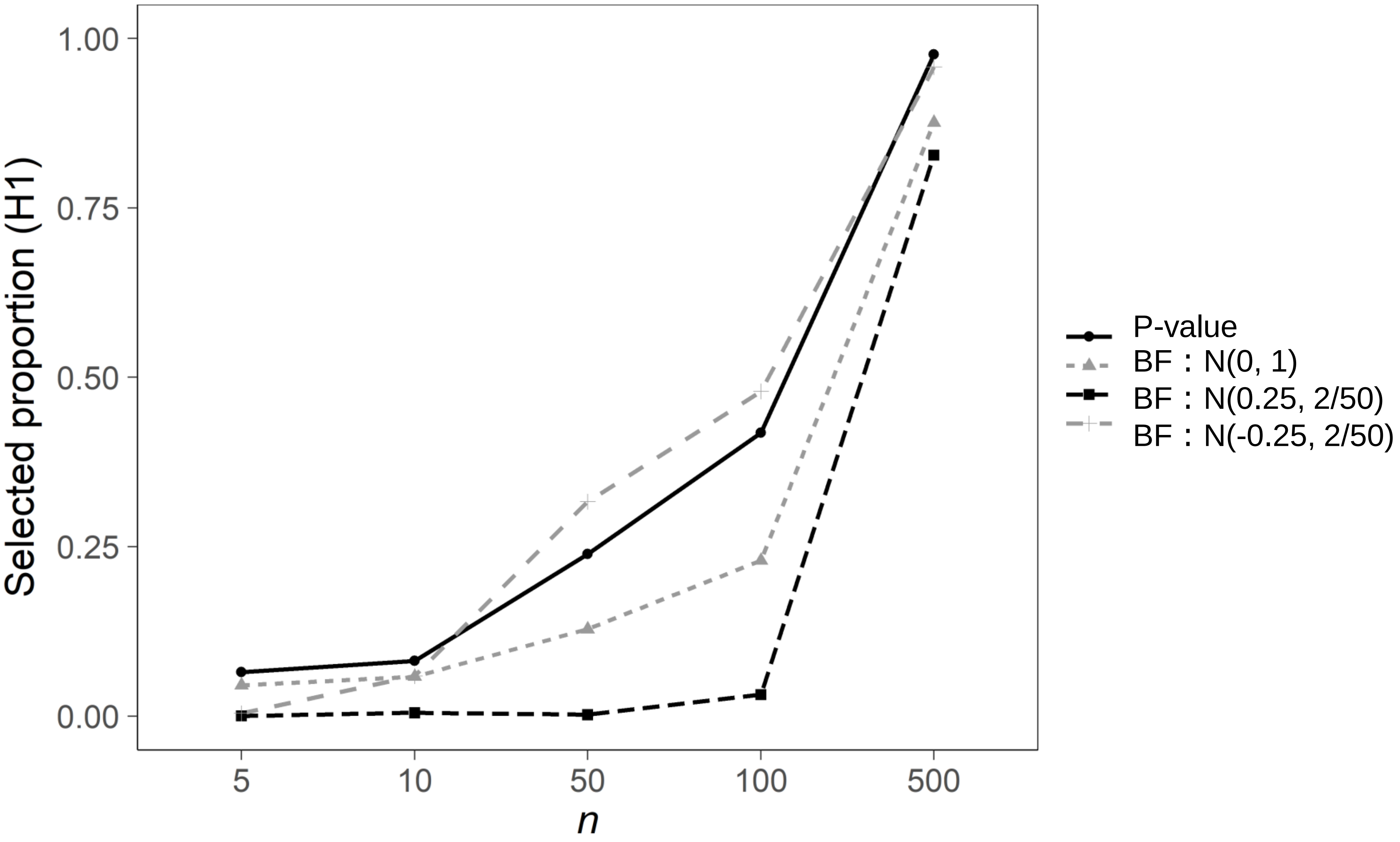
- 二標本t検定（基準：$p < 0.05$）
- BF (基準：BF ≦ 1/3)
  - 情報が少ない事前分布：$\delta/\sigma \sim N(0,1)$
  - 対立仮説からずれた事前分布：$\delta/\sigma \sim N(0.25,2/50)$
  - 対立仮説に近い事前分布：$\delta/\sigma \sim N(-0.25,2/50)$

- 性能指標：H_1を選択する割合

結果：シナリオ1


0.100
0.075
0.050
0.025
0.000
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 2/50)
BF：N(-0.25, 2/50)

結果：シナリオ2



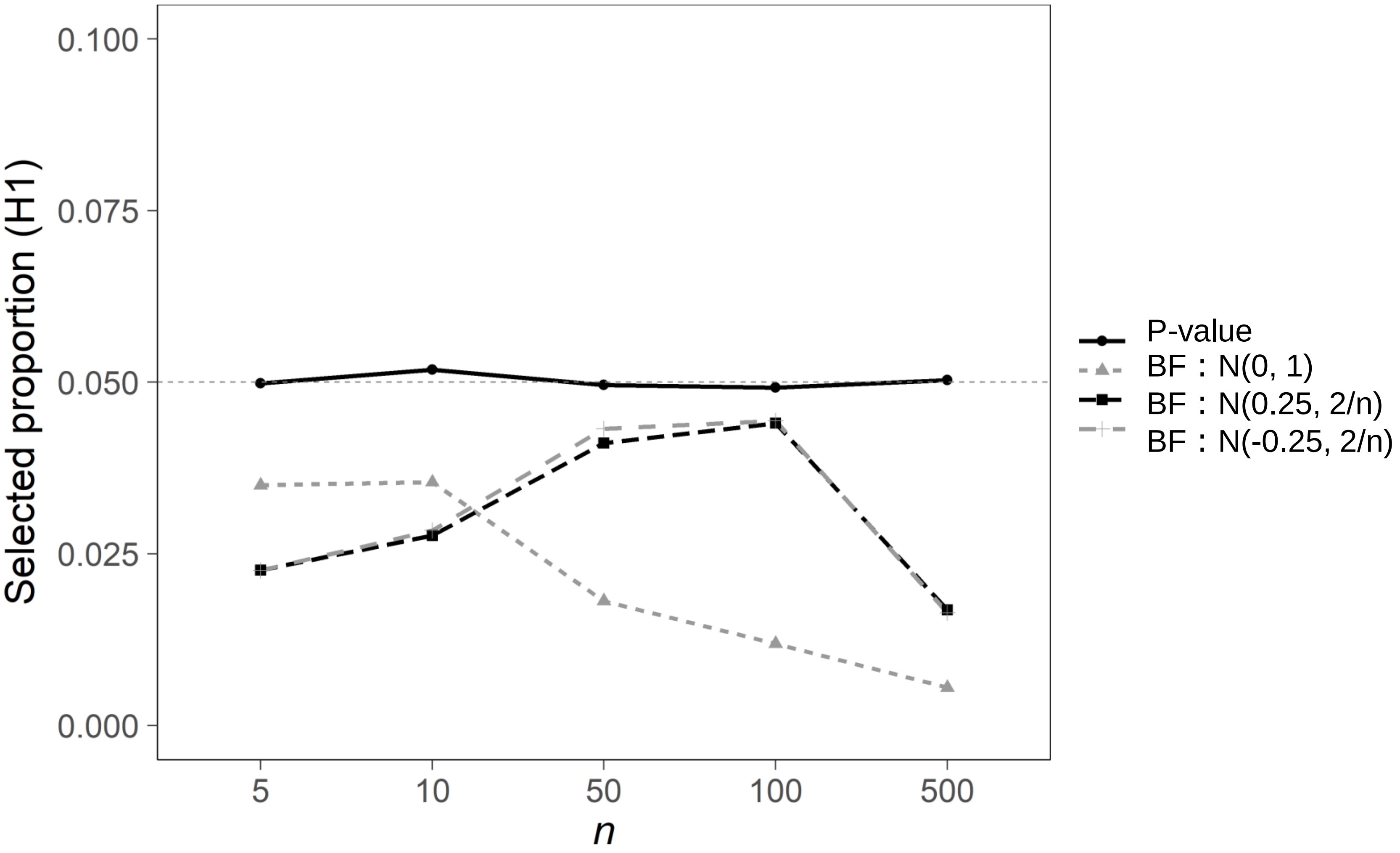
1.00
0.75
0.50
0.25
0.00
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 2/50)
BF：N(-0.25, 2/50)

シミュレーションで検討する手法 (3)

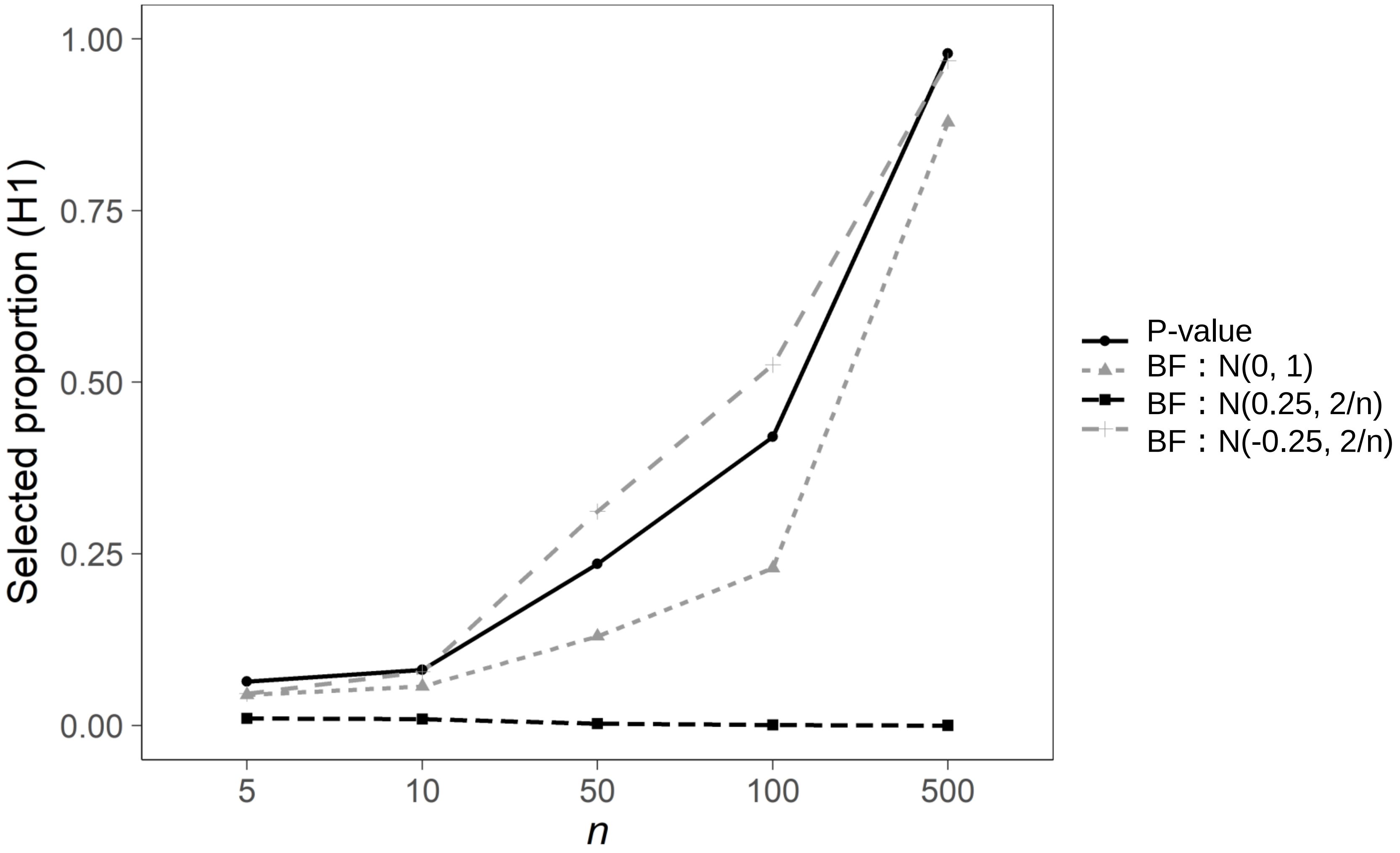
- 二標本t検定（基準：$p < 0.05$）
- BF (基準：BF ≦ 1/3)
  - 情報が少ない事前分布：$\delta/\sigma \sim N(0,1)$
  - 対立仮説からずれた事前分布：$\delta/\sigma \sim N(0.25,2/n_2)$
  - 対立仮説に近い事前分布：$\delta/\sigma \sim N(-0.25,2/n_2)$

- 性能指標：H_1を選択する割合

結果：シナリオ1


Selected proportion (H1)
0.100
0.075
0.050
0.025
0.000
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 2/n)
BF：N(-0.25, 2/n)

結果：シナリオ2



Selected proportion (H1)
1.00
0.75
0.50
0.25
0.00
5
10
50
100
500
n
P-value
BF：N(0, 1)
BF：N(0.25, 2/n)
BF：N(-0.25, 2/n)

補足

- 他の条件下でのシミュレーションは付録に示した
- BFの閾値を変化させた場合については, 基本的には選択割合が上下に変動
    - 基準を厳しくする ⇒ 第一種の過誤に相当する割合が下がる
    - 基準を緩くする ⇒ 第一種の過誤に相当する割合が上がる

ディスカッション前のまとめ：各手法の特徴・特性について気づいた点などを挙げてください

- 仮説検定

・BF

ディスカッション

- なぜBFはあまり使われていないのか？なぜ頻度論の方法をよく使うのか？
- BFを利用するとしたら, どういう情報が必要か？
- BFの使用を考えた際に何が気になるか？何が障壁になるか？
- BFについて理解できない点は？
- BFは仮説検定の代替として使えそうか？どのようなとき？
- BFはPhase IIIの検証試験等で使えそうか？
- BFを使う際にはどのような点に注意が必要か？
- BFの閾値の設定は？

理論的結果の補足

- BF
  - Model selection consistency (Casella et al., 2009)
    - $n \to \infty$で確率1で真のモデルを選択
    - 例えば, H_0が正しいときには$n \to \infty$においてH_1が選択される確率が0に収束
  - BICの差は対数BFの近似 (Kass & Raftery, 1995)
- 仮説検定：αエラーはnによらない
- 両者はかなり異なる性質を持つ

推定

信用区間と信頼区間

区間推定の統計手法

- 試験の結果から, 特定のパラメータに関する推測を行う

- さまざまな手法があるが, 本テーマでは以下を扱う
  - ベイズ流：パラメータの事後分布の信用区間
  - 頻度論：信頼区間

二標本平均の差の推定手法

- $\delta = \mu_1 - \mu_2$の区間推定問題
- 頻度論
  - 最尤法ベースの正規近似に基づく両側95%信頼区間
- ベイズ流
  - 正規分布モデルベース
  - 事前分布：$\mu_j \sim N(\mu_{j0}, \tau_{j0}^2), \sigma^2 \sim \mathrm{Inv-gamma}(0,0)$
  - 群ごとにμ | $\mathbf{y}$の事後分布から1000回サンプリングして差を求め, 2.5%点と97.5%点を求める

シミュレーションデータの生成条件

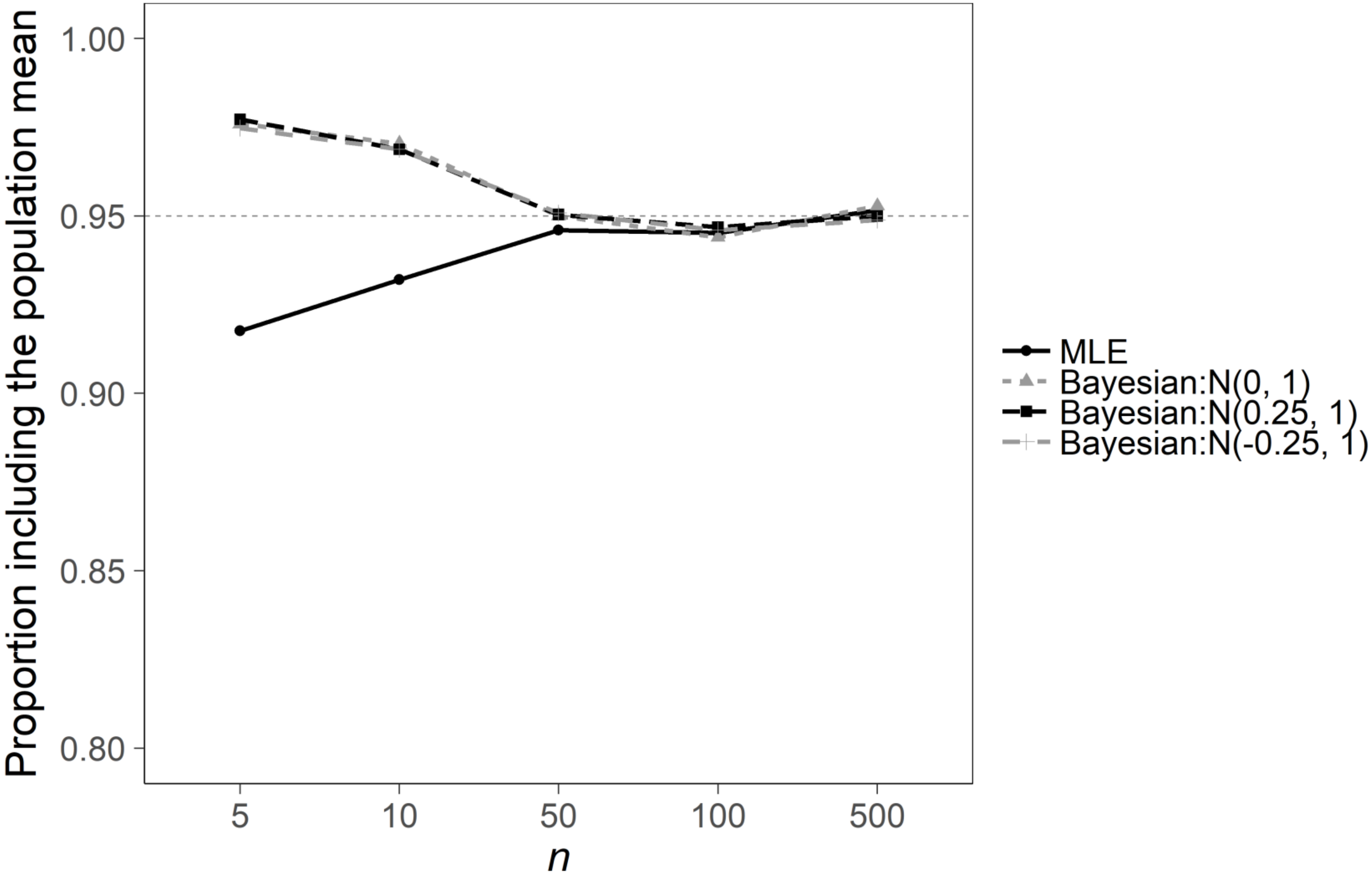
- 標本を正規乱数で生成 (2,500セット)
  - $n_1 = n_2 = 5, 10, 50, 100, 500$
  - シナリオ1：群間差がない場合, $\delta = \mu_1 - \mu_2 = 0$
    - 群1：$N(0, 1)$
    - 群2：$N(0, 1)$
  - シナリオ2：群間差がない場合, $\delta = \mu_1 - \mu_2 = -0.25$
    - 群1：$N(0, 1)$
    - 群2：$N(0.25, 1)$

シミュレーションで検討する手法 (1)

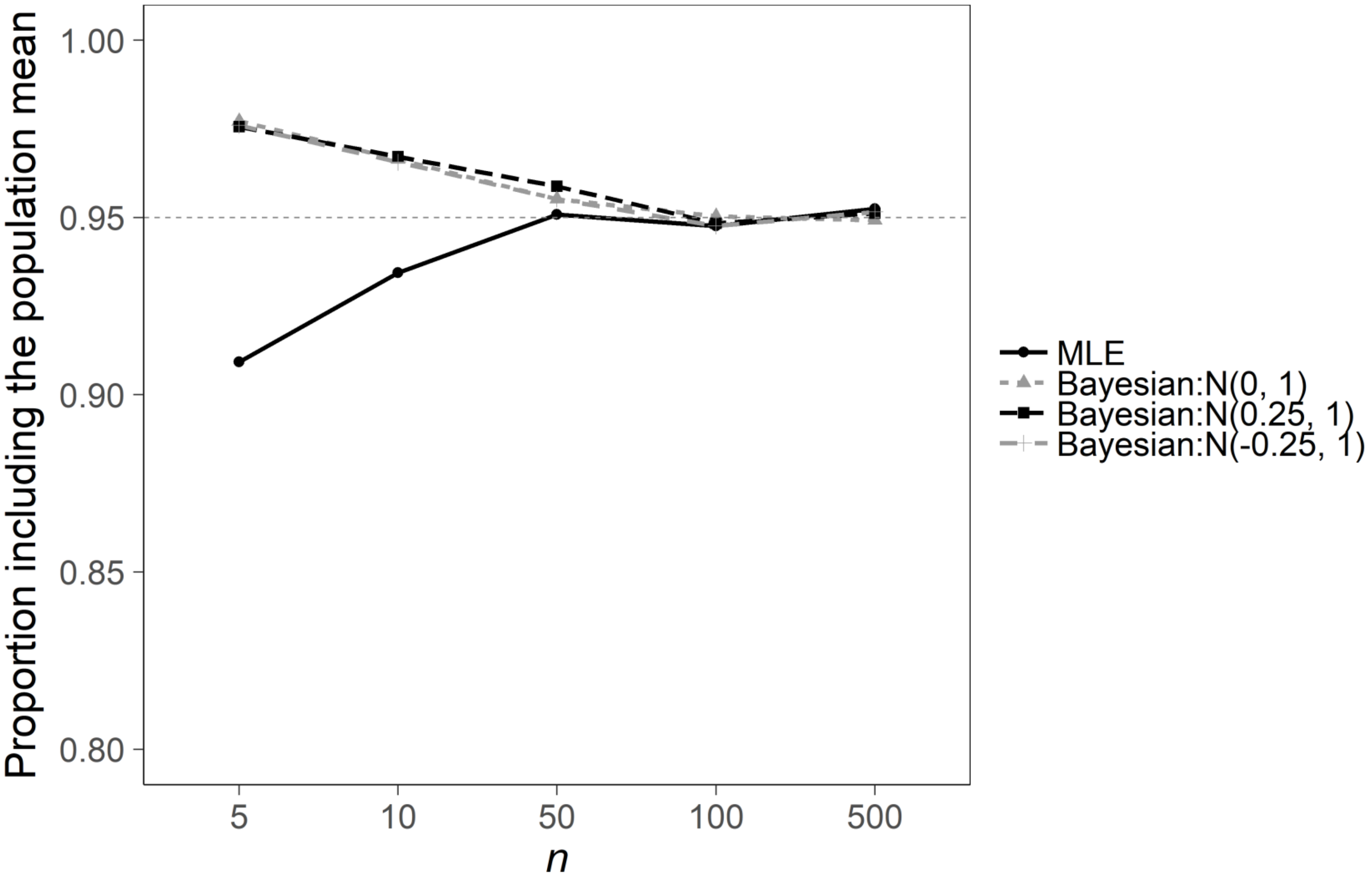
- 最尤法ベースの正規近似に基づく両側95%信頼区間
- ベイズ信用区間 (事前分布の情報量が少ない場合)：$\mu_1 \sim N(0,1)$は共通
  - 平均0の事前分布：$\mu_2 \sim N(0,1)$
  - 平均0.25の事前分布：$\mu_2 \sim N(0.25,1)$
  - 平均-0.25の事前分布：$\mu_2 \sim N(-0.25,1)$

  ※特徴を見るためにあえて極端な事前分布を設定
- 評価指標：各区間が群間差の母集団平均を含む割合

結果：シナリオ1


Proportion including the population mean
1.00
0.95
0.90
0.85
0.80
5
10
50
100
500
n
MLE
Bayesian:N(0, 1)
Bayesian:N(0.25, 1)
Bayesian:N(-0.25, 1)

結果：シナリオ2



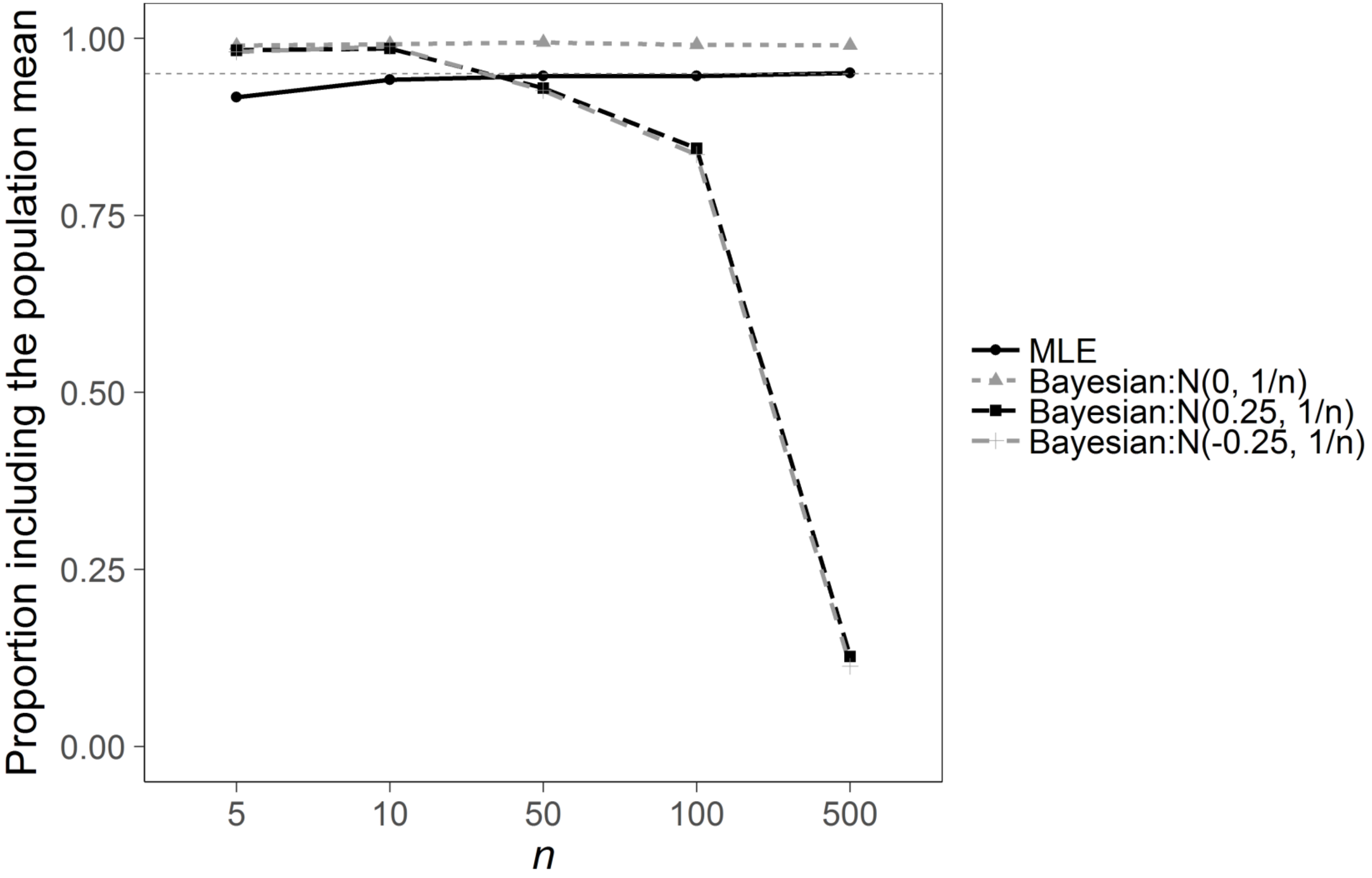
Proportion including the population mean
1.00
0.95
0.90
0.85
0.80
5
10
50
100
500
n
MLE
Bayesian:N(0, 1)
Bayesian:N(0.25, 1)
Bayesian:N(-0.25, 1)

シミュレーションで検討する手法 (2)

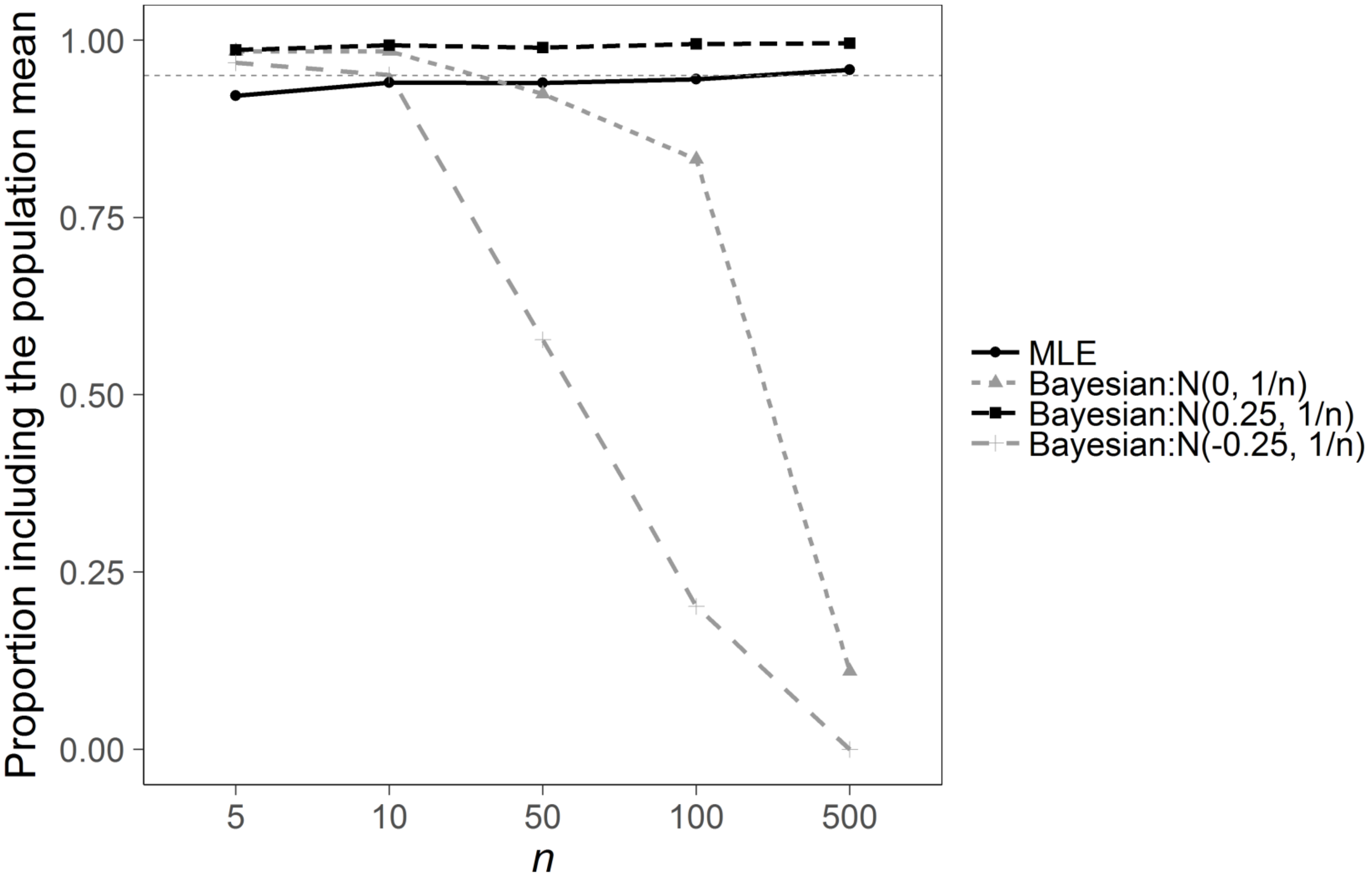
- 最尤法ベースの正規近似に基づく両側95%信頼区間
- ベイズ信用区間 (事前分布の情報量が多い場合)：$\mu_1 \sim N(0,1/n_1)$は共通
  - 平均0の事前分布：$\mu_2 \sim N(0,1/n_2)$
  - 平均0.25の事前分布：$\mu_2 \sim N(0.25,1/n_2)$
  - 平均-0.25の事前分布：$\mu_2 \sim N(-0.25,1/n_2)$

  ※特徴を見るためにあえて極端な事前分布を設定
- 評価指標：各区間が群間差の母集団平均を含む割合

結果：シナリオ1


Proportion including the population mean
1.00
0.75
0.50
0.25
0.00
5
10
50
100
500
n
MLE
Bayesian:N(0, 1/n)
Bayesian:N(0.25, 1/n)
Bayesian:N(-0.25, 1/n)

結果：シナリオ2

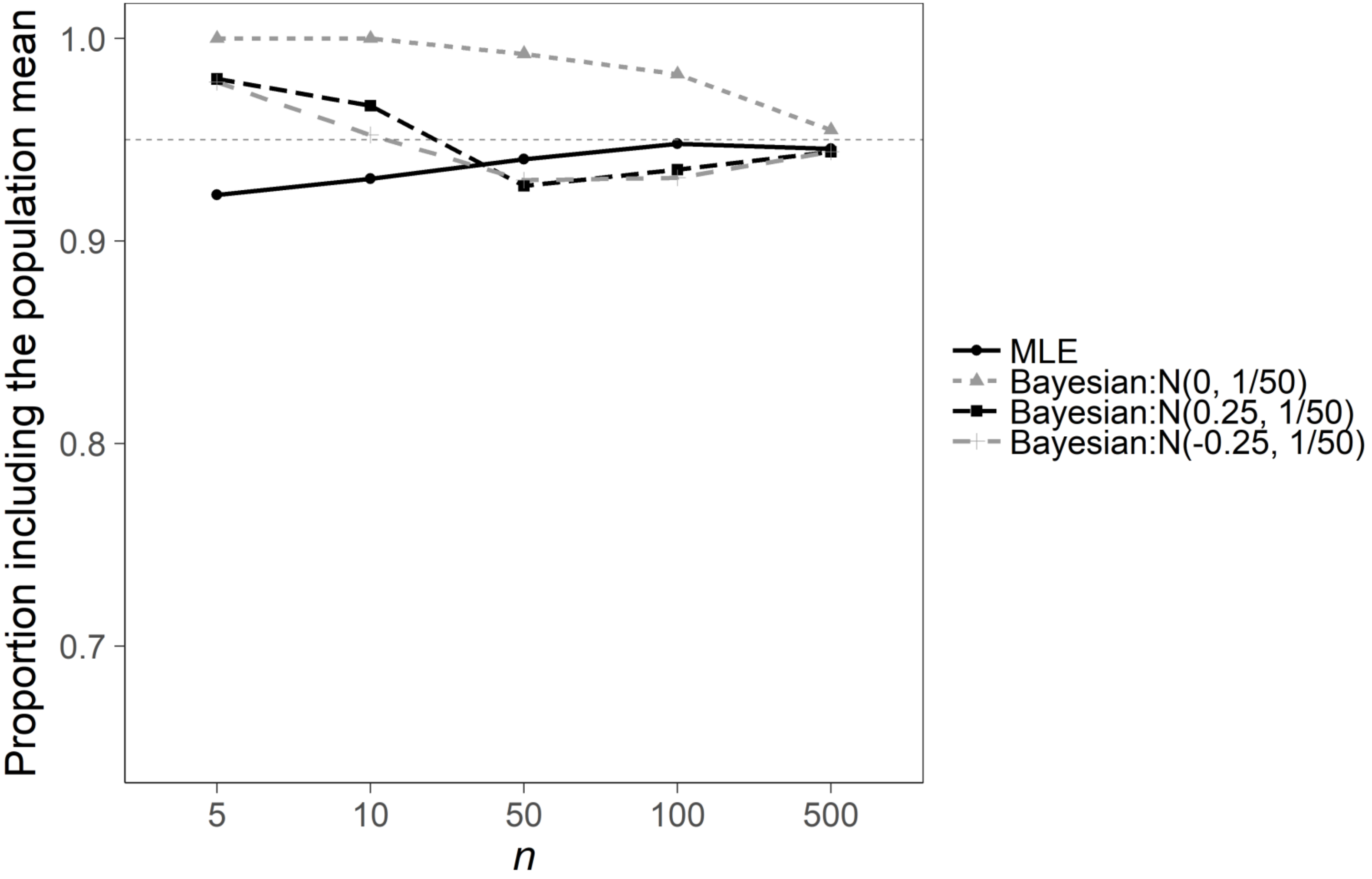

Proportion including the population mean
1.00
0.75
0.50
0.25
0.00
5
10
50
100
500
n
MLE
Bayesian:N(0, 1/n)
Bayesian:N(0.25, 1/n)
Bayesian:N(-0.25, 1/n)

シミュレーションで検討する手法 (3)

- 最尤法ベースの正規近似に基づく両側95%信頼区間
- ベイズ信用区間 (事前分布の情報量が多い場合)：$\mu_1 \sim N(0,1/50)$は共通
  - 平均0の事前分布：$\mu_2 \sim N(0,1/50)$
  - 平均0.25の事前分布：$\mu_2 \sim N(0.25,1/50)$
  - 平均-0.25の事前分布：$\mu_2 \sim N(-0.25,1/50)$

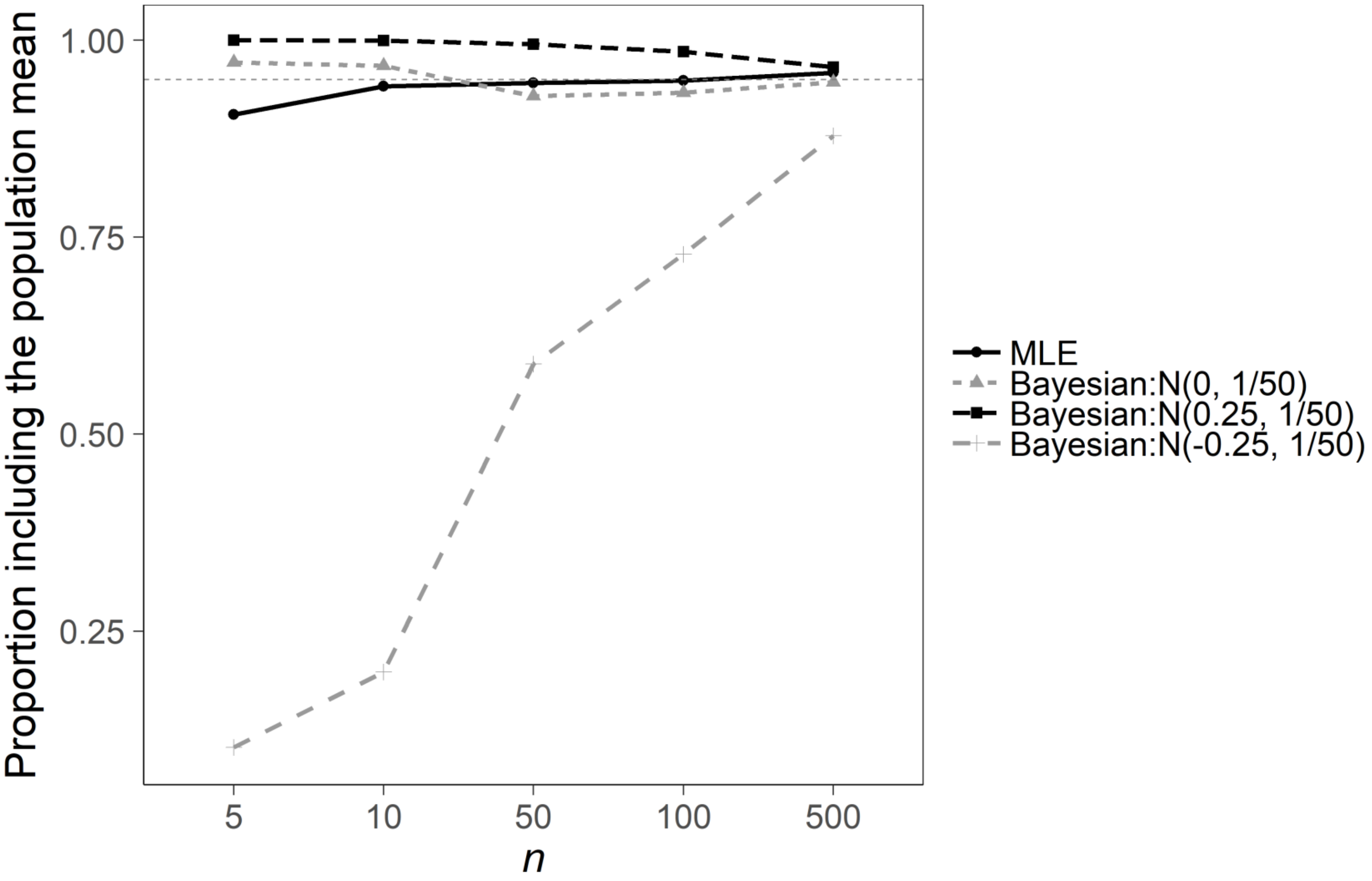
  ※特徴を見るためにあえて極端な事前分布を設定
- 評価指標：各区間が群間差の母集団平均を含む割合

結果：シナリオ1



Proportion including the population mean
1.0
0.9
0.8
0.7
5
10
50
100
500
n
MLE
Bayesian:N(0, 1/50)
Bayesian:N(0.25, 1/50)
Bayesian:N(-0.25, 1/50)

結果：シナリオ2



Proportion including the population mean
1.00
0.75
0.50
0.25
5
10
50
100
500
n
MLE
Bayesian:N(0, 1/50)
Bayesian:N(0.25, 1/50)
Bayesian:N(-0.25, 1/50)

ディスカッション前のまとめ：各手法の特徴・特性について気づいた点などを挙げてください

- 頻度論

・ベイズ流

ディスカッション

- ベイズ流の手法を利用した経験はありますか？
- なぜベイズ流の手法はあまり使われていないのか？なぜ頻度論の方法をよく使うのか？
- 信用区間を利用するとしたら, どういう情報が必要か？
- 信用区間の使用を考えた際に何が気になるか？
- 信用区間について理解できない点は？
- 信用区間はPhase IIIの検証試験等で使えそうか？
- 信用区間を使う際にはどのような点に注意が必要か？

補足

- 参考のため, 別の手法を用いた場合の追加シミュレーション結果を付録に示した
  - 頻度論：最尤法ではなく, t分布に基づく信頼区間
  - ベイズ流：Rouder et al. (2009) の無情報事前分布を用いた方法

理論的結果の補足：漸近理論 (正則モデル)

- Bernstein–von Mises theorem (e.g., Van der Vaart, 1998; Nickl, 2013)
  - $n \to \infty$の粗い近似：(不正確だが) ざっくり説明すると, 適当な条件の下で, パラメータの事後分布が平均がMLEで分散が$n^{-1}i(\theta_0)$の正規分布に収束
- nが大きくなると, 両者の結果が近づいてゆく
- 有限標本の時, 両者は異なる傾向を持ちうる

理論的結果の補足：ベイズ推測

- 事後分布そのもの$\mathrm{Pr}(\boldsymbol{\theta} \mid \mathbf{y})$や事後分布の期待値$\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\mathrm{Pr}(\boldsymbol{\theta} \mid \mathbf{y})\,\mathrm{d}\boldsymbol{\theta}$はもちろん重要
- 予測分布$\mathrm{Pr}(Y \mid \mathbf{y}) = \int \mathrm{Pr}(Y \mid \boldsymbol{\theta})\,\mathrm{Pr}(\boldsymbol{\theta} \mid \mathbf{y})\,\mathrm{d}\boldsymbol{\theta}$へつなげやすい点も結構重要
- 事後分布が正規分布で近似できる場合, nが大きければ$\hat{\boldsymbol{\theta}}$と最尤推定量は同じ分布に近づいてゆく
- 近似できない場合には大きく違う挙動になる (渡辺, 2006, 2012)

参考文献

- Wasserstein RL, Lazar NA. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician* 2016; **70**(2): 129–133.
- Casella G, Girón FJ, Martínez ML, Moreno E. Consistency of Bayesian procedures for variable selection. *The Annals of Statistics* 2009; **37**(3): 1207–1228.
- Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association* 1995; **90**(430): 773–795.
- Gönen M, Johnson WO, Lu Y, Westfall PH. The Bayesian two-sample t test. *The American Statistician* 2005; **59**(3): 252-257.
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review* 2009; 16: 225–237.
- Van der Vaart AW. *Asymptotic Statistics*. Cambridge University Press, 1998.
- Nickl R. *Statistical Theory*. 2013. Available at: http://www.statslab.cam.ac.uk/~nickl/Site/__files/stat2013.pdf
- 渡辺澄夫．代数幾何と学習理論．森北出版株式会社，2006．
- 渡辺澄夫．ベイズ統計の理論と方法．コロナ社，2012．

あとがき

- 当日は70分間でスライドの説明とディスカッションを行ったため，非常に限られた内容になっています．不正確・不十分な部分が多々あると思われますが，ご容赦下さい
- シミュレーションに用いたプログラムなどの情報を公開しています
  - http://nshi.jp/

付録

ベイズ推定 (Bayesian Inference)

- パラメータθ_0を持つ確率変数$\mathbf{Y}$の母集団分布を, 確率変数θがθ_0に等しいことを与えたもとでの確率変数$\mathbf{Y}$の条件付き分布とみなす：$f(\mathbf{Y} \mid \theta = \theta_0)$
- データは分布$f(\mathbf{Y} \mid \theta = \theta_0)$からのサンプリングとみなす
- 事前分布はランダムなパラメータの分布：$\Pr(\theta)$
- 事後分布はランダムなパラメータの観測値の条件付き分布：$\Pr(\theta \mid \mathbf{y})$
- このモデルのもとで, 事後分布はベイズの定理を使って計算可能である

Bernstein–von Mises theorem

- 適切な正則条件 (Theorem 3 of Nickl, 2013; (不正確だが) 正則モデル) のもとで, 確率変数$Y_1, \ldots, Y_n$が独立同一分布$f(\theta_0)$にしたがい, $\hat{\theta}_n$を標本に基づくMLEとする．事前分布はパラメータ真値の近傍で連続かつ正でLebesgue-densityをもつとする．このとき，事後分布$\Pr(\theta \mid \mathbf{y})$ の分布関数$\Pi(B \mid \mathbf{y}) = \int_B \Pr(\theta \mid \mathbf{y})\, d\theta$は ($B$はパラメータ空間のBorel部分集合) 以下を満たす;

$$\sup_{B \in \mathcal{B}} \left| \Pi(B \mid \mathbf{y}) - N\left(B \mid \hat{\theta}_n, n^{-1} i(\theta_0)\right) \right| \to^{P_{\theta_0}} 0 \text{ as } n \to \infty.$$

付録

二値判断の追加シミュレーション

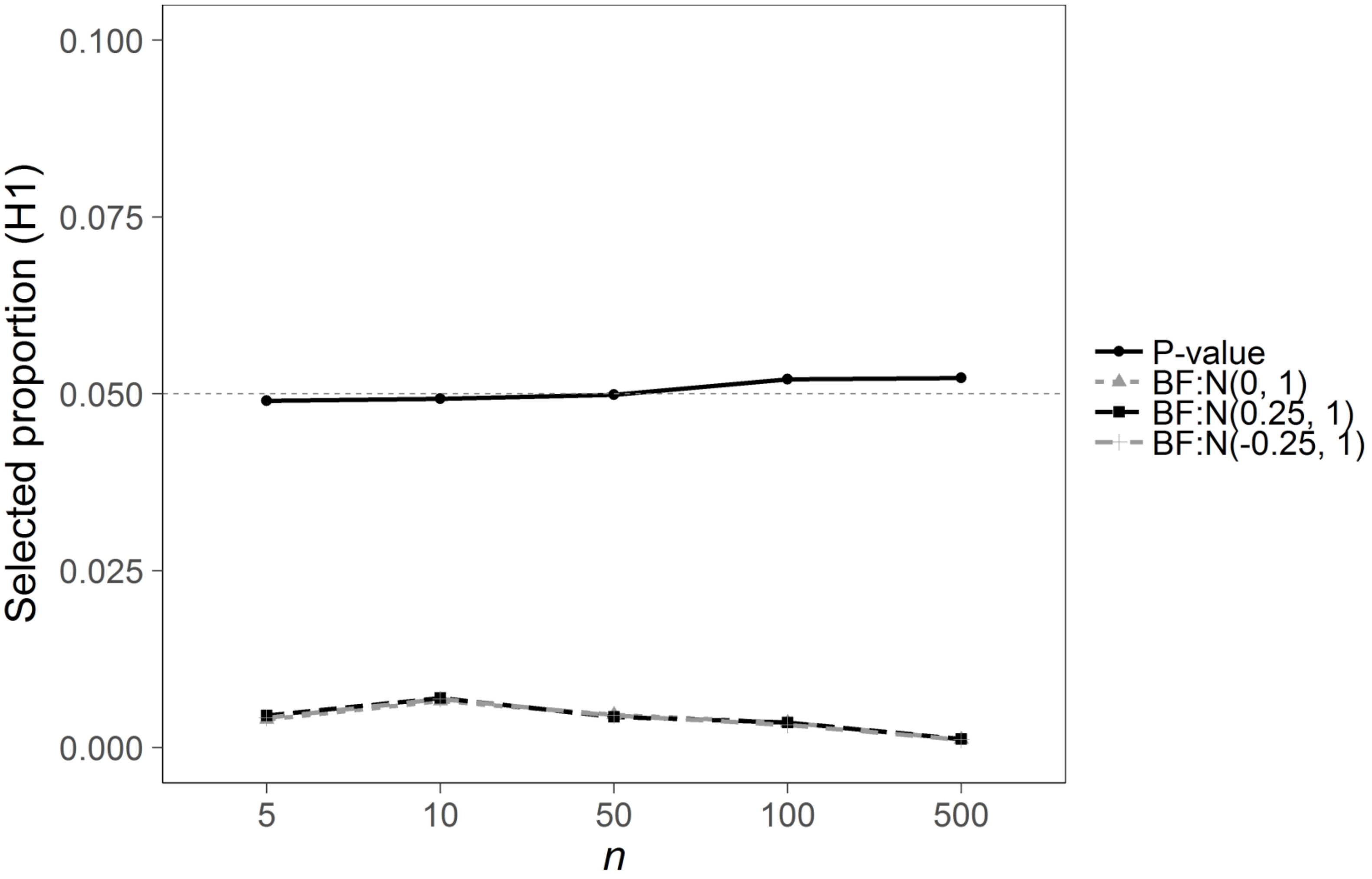
シミュレーション結果の要約：検定

- シナリオ1：帰無仮説が正しい
  - 仮説検定はnによらず対立仮説採択確率が一定（α）
  - BFはnが増加すると対立仮説採択確率が低下
- シナリオ2：対立仮説が正しい
  - 仮説検定は特筆する点はないがnとともに増大
  - BFは誤った方向の事前分布を与えると対立仮説採択確率が低く, 無情報は中間, 正しい方向の事前分布を与えると対立仮説採択確率が高い

追加シミュレーション (1)

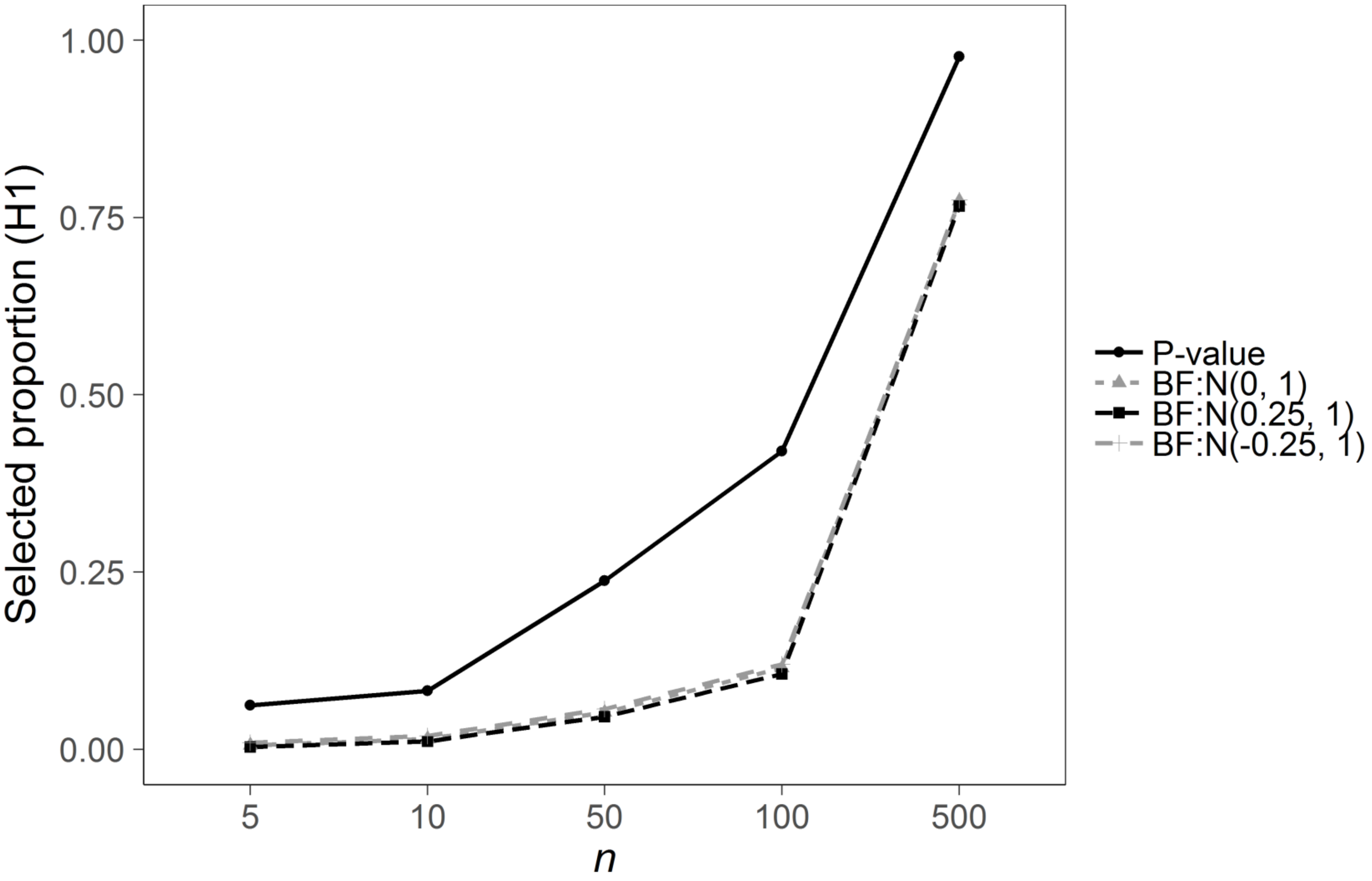
- 二標本t検定（基準：$p < 0.05$）
- BF (基準：BF ≦ 1/10)
  - 情報が少ない事前分布：$\delta/\sigma \sim N(0,1)$
  - 情報が少ない対立仮説からずれた事前分布：$\delta/\sigma \sim N(0.25,1)$
  - 情報が少ない対立仮説に近い事前分布：$\delta/\sigma \sim N(-0.25,1)$
- 性能指標：H_1を選択する割合

結果：シナリオ1



0.100
0.075
0.050
0.025
0.000
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF:N(0, 1)
BF:N(0.25, 1)
BF:N(-0.25, 1)

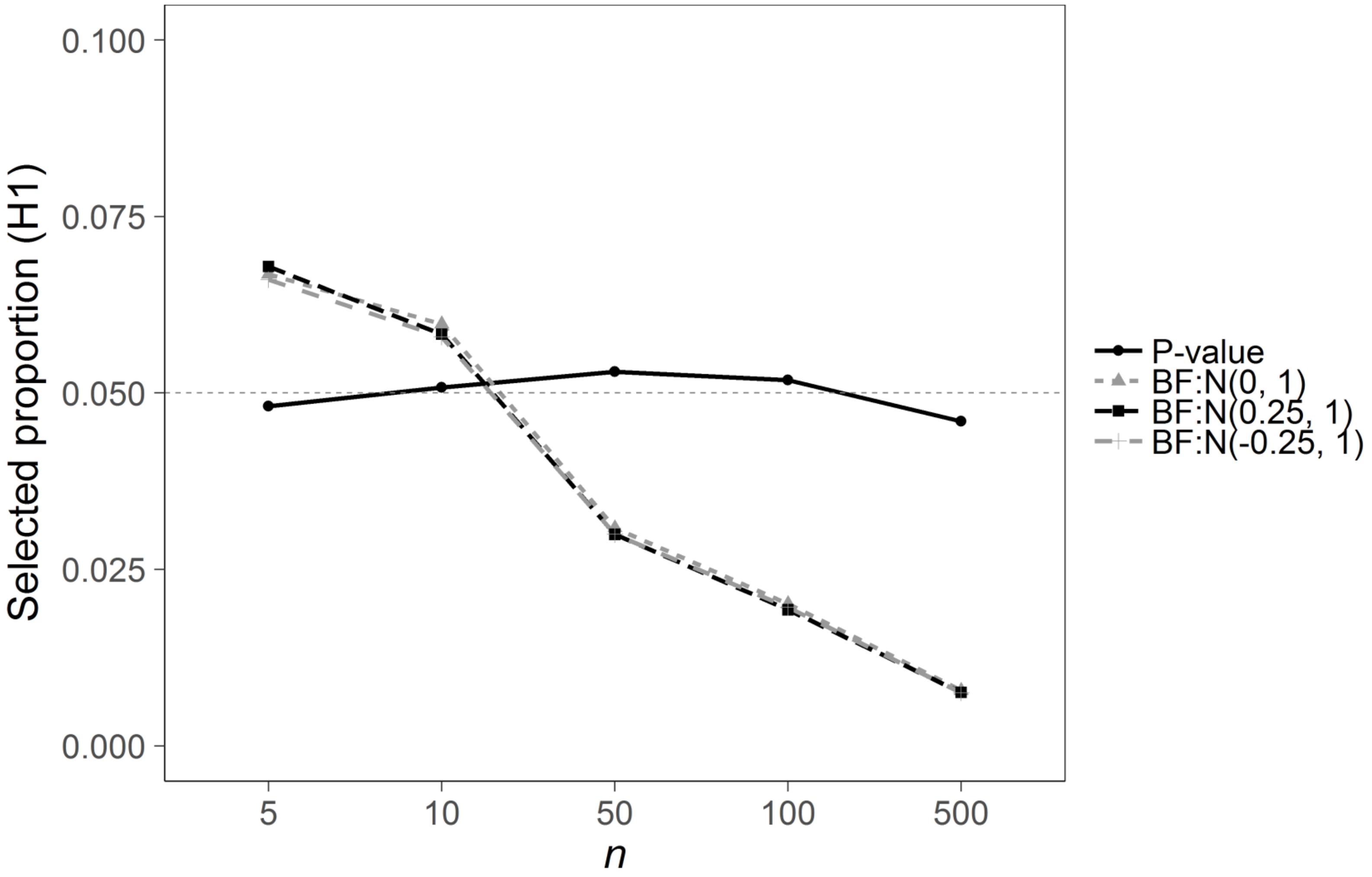
結果：シナリオ2


Selected proportion (H1)
1.00
0.75
0.50
0.25
0.00
5
10
50
100
500
n
P-value
BF:N(0, 1)
BF:N(0.25, 1)
BF:N(-0.25, 1)

追加シミュレーション (2)

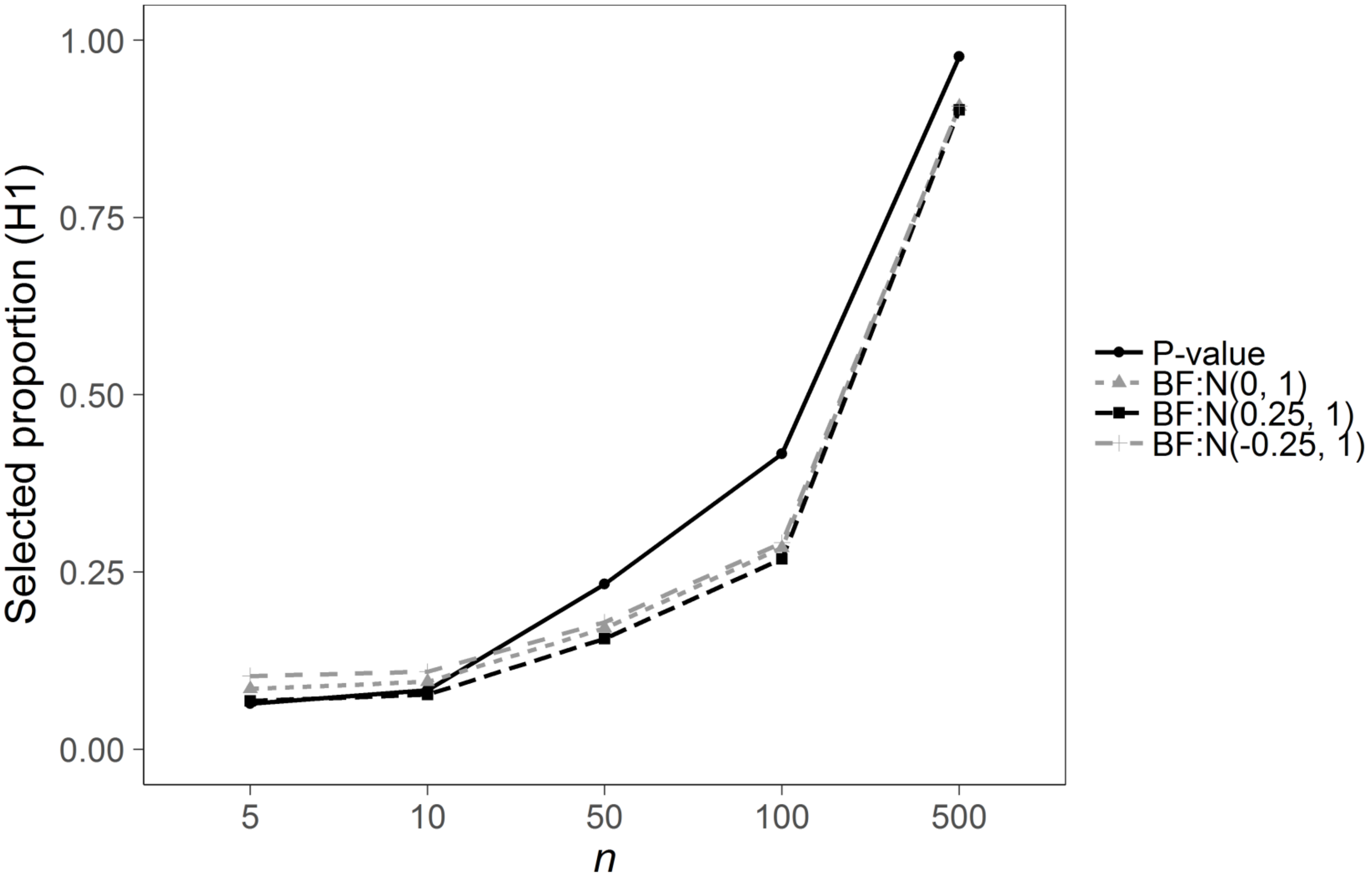
- 二標本t検定（基準：$p < 0.05$）
- BF (基準：BF ≦ 1/2)
    - 情報が少ない事前分布：$\delta/\sigma \sim N(0,1)$
    - 情報が少ない対立仮説からずれた事前分布：$\delta/\sigma \sim N(0.25,1)$
    - 情報が少ない対立仮説に近い事前分布：$\delta/\sigma \sim N(-0.25,1)$
- 性能指標：H_1を選択する割合

結果：シナリオ1



0.100
0.075
0.050
0.025
0.000
Selected proportion (H1)
5
10
50
100
500
n
P-value
BF:N(0, 1)
BF:N(0.25, 1)
BF:N(-0.25, 1)

結果：シナリオ2



Selected proportion (H1)
1.00
0.75
0.50
0.25
0.00
5
10
50
100
500
n
P-value
BF:N(0, 1)
BF:N(0.25, 1)
BF:N(-0.25, 1)

付録

区間推定の追加シミュレーション

t分布に基づく信頼区間を用いた場合