

Ⅱ．分担研究報告

課題 3．総合的な摂取量評価における推定値の
精緻化および信頼性の向上に関する研究

研究分担者 鈴木 美成

令和3年度厚生労働行政推進調査事業費補助金 食品の安全確保推進研究事業

食品や環境からの農薬等の摂取量の推計と国際標準を導入するための研究

研究分担報告書

総合的な摂取量評価における推定値の精緻化および信頼性の向上に関する研究

研究分担者

国立医薬品食品衛生研究所食品部 鈴木美成

研究要旨

有害物質の摂取量推定値は、健康リスクの管理を目的とする規格値策定等の行政施策の検討、及び効果検証のための科学的根拠となる。本研究では、国民が食品を介して摂取する農薬の量をより精緻に推定するために、1) 不検出例を含むデータに対するベイズ推定 (BE)法の有用性、2) 真値に0を含むデータにおける統計妥当性の高い推定法に関する数値シミュレーション、3) 二次元モンテカルロシミュレーションを用いた摂取量分布の推定に関して検討を行った。

1日農薬摂取量の理論平均値に関する事後予測分布とLOQの0.2倍の値を代入した推定値およびMLE法と比較した。MLE法による1日農薬摂取量の平均値は全てBE法による平均値の事後予測分布の四分位範囲内に収まっていたが、BE法とMLE法で推定した平均値の相対的な差には次に示す傾向が認められた。すべての食品群で不検出となったサンプルの割合が高い場合、推定した分布の σ/μ 比が大きい場合、用いたデータのサンプルサイズが小さい場合に、BE法とMLE法による推定値の相対的な差が大きくなる傾向が認められた。これらの結果より、BE法とMLE法の差異が大きくなった要因は、歪みの大きい確率密度分布であること、サンプルサイズが小さいこと、および下限値が0となるサンプルの割合が大きかったことの影響が大きいと考えられた。MLE法は、有効なサンプルサイズが少なくなるほど、また分布の歪みが大きくなるほど、推定の精度が低くなることが報告されている。この点を踏まえると、BE法による推定結果の方がMLE法による推定結果よりも妥当である可能性が示唆

された。

また、NDの問題と関連して、非天然型農薬は、使用履歴が無ければ、検体中の残留農薬含有量が0である可能性もありうる。そこで、真値に0を含むデータの推定に関して、ゼロ過剰対数正規 (ZILN) 分布を仮定してベイズ推定法と最尤推定法による推定値の真度・精度・妥当性について数値シミュレーションを用いて検討した。数値シミュレーションの結果、MLE法よりもBE法の方が、推定の真度・精度・妥当性が良好であった。さらに、この傾向は事前分布の設定に寄らず、無情報事前分布を用いた場合でもBE法の方がMLE法よりも良好な真度・精度・妥当性を示した。以上のことから、ZILN分布に従う不検出例を含むデータの解析においても、BE法が有効であることが示された。

上記の結果を基にゼロ過剰モデルも考慮して、各食品群中の農薬濃度分布をBE法で推定し、2D-MCSによる残留農薬摂取量分布を推定した。ADIを超過する確率は、アセフェートの0.001%が最大であったことから、検討したいずれの農薬も健康リスクは小さいと示唆された。さらに、2D-MCSから推定した結果と、摂取量のデータからBE法で推定した結果を比較したところ、農薬摂取量の平均値の事後予測分布が良い一致を示した。以上のことから、いずれの推定方法も妥当であると考えられた。

以上の検討により、食品に含まれる残留農薬を対象とした統計的推定においてBE法は有効であることが示された。

A. 目的

食品を介した農薬の摂取量評価は、残留農薬基準値といった健康リスクの管理を目的とする規格値策定等の行政施策の検討、及び効果検証のための科学的根拠となる。

食品中の農薬濃度については打ち切り (censoring) 問題が生じる。打ち切りは、ある値よりも大きい、小さい、またはその両方の値を非表示にする。切り捨てられた (truncated) データとは異なるのは、打ち切られたデータポイントの数がわかっている点である。より具体的に残留農薬

分析の場合について言及すると、検出限界 (LOD) あるいは定量下限値 (LOQ) 未満のデータは、妥当な数値を割りてることは困難となる。とくに、本研究課題の様に複数の研究機関が参加した調査の場合には、ある機関のLOQが他の機関のLOQよりも高い場合にNDとなった場合には影響が大きくなる。

これまでの厚生労働省が主体となって行ってきた農薬摂取量評価において、LOD未満・およびLOD以上LOQ未満のデータは、NDとして扱い、NDにはLOQの0.2倍した値を代入し平均的な摂取量

を評価してきた。しかしながら、代入法は問題のある方法として認識されて来ており、限定された場合にのみ使用を推奨されるようになってきている。代入法に代わり、最尤推定法などの代替法の使用が推奨されて来ている。しかしながら、サンプル数が少ない場合は、検出値と検出された割合のみを報告すべきとの提言もある。

このような場合に、ベイズ推定が有効となる可能性がある。ベイズ推定は尤度分布にこれまでの知識を用いて重み付けしたものと解釈できる。適切な事前分布を用いることで、最尤推定よりも安定した推定値を得ることができると期待されている。そこで、2019 年から 2021 年度の間で得られた農薬摂取量のデータに対して、ベイズ法による推定を行い、代入法および最尤推定 (MLE) 法との相違について評価した。

ND の問題と関連して、非天然型農薬は、使用履歴が無ければ、検体中の残留農薬含有量が 0 である可能性もありうる。近年になって、真値に 0 を含むようなデータの解析に関する報告がなされてきた。Friesen ら (10.1038/s41370-021-00331-7) は大気中 Pb 濃度の経年データにゼロ過剰対数正規分布 (ZILN) を適用した時系列解析に適用している。Lavoue と Burstyn (10.1093/annweh/wxaa086) は、LN 分布・ZILN 分布・切断 LN 分布を仮定した数値シミュレーションと大気中化学物質への適用例を報告している。Lavoue と Burstyn は仮定したモデルの相違に注目した検討であり、推定法の差異については検討していない。また、真値が 0 である割合も最大で 60% であり、より広い範囲での検討が必要である。そこで、本研究では真値が 0 とな

る割合および不検出割合について幅広い組合せで、ベイズ推定 (BE) 法と MLE 法による推定結果の真度・精度・妥当性について評価を行った。また、BE 法については事前分布の影響についても検討した。さらに、TD 試料中の農薬濃度について適用を試み、モンテカルロシミュレーション (MCS) を用いた農薬の摂取量推定を行った。

B. 理論

1 尤度と最尤推定法

ここで、解析の基本となる尤度について基本的な数式を示す。 N 個の観測値 Y がパラメータ θ を持つ確率密度関数 $f(Y|\theta)$ に従うと仮定したときの尤度 (L) は、以下の式で算出することができる。

$$L(\theta|Y) = \prod_{i=1}^N f(Y_i|\theta) \quad \dots \text{Eq. 1}$$

打ち切りのあるデータの場合、定量できたレコードには確率密度関数 $f(Y|\theta)$ から確率密度を、打ち切りとなったレコードには累積確率関数 $F(L, U|\theta) = P(L \leq Y \leq U|\theta)$ から累積確率を使用し、尤度関数次のように求めることが出来る。

$$L(\theta|Y) = \prod_{i=1}^{N_{\text{obs}}} f(Y_i|\theta) \cdot \prod_{j=1}^{N_{\text{cen}}} F(L_j, U_j|\theta) \quad \dots \text{Eq. 2}$$

ここで、 N_{obs} は定量値が得られたサンプル数、 N_{cen} は定量値未満のサンプル数、 L_j はサンプル j の下限値、 U_j は上限値を示す。

MLE 法は、尤度が最大となるパラメータを探索する。

2 ベイズ推定

尤度を基にした計算方法として、MLE法とBE法による方法がある。MLE法では尤度が最大となるパラメーターを計算する方法である。つまり、得られたデータの同時確率が最大となるパラメーターを推定する方法である。

BE法の場合も尤度を基にパラメーターを探索するという点は同じだが、最尤法と大きく異なる点が2つある。1つ目の相違は、探索する際に事前分布の情報が必要であることで、2つ目の相違は点ではなく分布を得る点である。BE法で得られる事後予測分布 $f_{\text{post}}(\theta|Y)$ は、ベイズの定理を用いて以下のようにあらわすことができる。

$$f_{\text{post}}(\theta|Y) = \frac{L(Y|\theta)f_{\text{pri}}(\theta)}{f_{\text{obs}}(Y)} \dots \text{Eq. 3,}$$

ここで、 $f_{\text{pri}}(\theta)$ は事前分布を、 $f_{\text{obs}}(Y)$ は観測値 Y の確率密度分布を示す。観測値 Y は観測後には変化しないため、定数とみなせるので、

$$f_{\text{post}}(\theta|Y) \propto L(Y|\theta)f_{\text{pri}}(\theta) \dots \text{Eq. 4}$$

と簡略化することができる。つまり、ベイズ推定は尤度を事前知識に基づいて重み付したものであると解釈できる。

一般的にMLE法は、サンプルサイズが小さい場合（例えば50未満）には、使用を勧められていない。原因の一つとして、少ないサンプルサイズでは標本の偏りによる影響が大きいことが挙げられる。BE法は、少ないサンプルサイズの場合に生じる標本の偏りを補正するために、事前分布を利用してMLE法を補正しているとも解釈できる。

3 検討する確率密度分布

本研究においては、1日農薬摂取量が従う確率密度分布に、対数正規分布、ガンマ分布、ワイブル分布を仮定した。

対数正規分布は、形状パラメーター e^μ 、尺度パラメーター e^σ の2つのパラメーターで特徴づけられ、その確率密度関数は以下の式で表される。

$$f(Y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \frac{1}{Y} \exp\left[-\frac{1}{2}\left(\frac{\log Y - \mu}{\sigma}\right)^2\right] \dots \text{Eq. 5}$$

形状パラメーター e^μ は幾何平均値、尺度パラメーター e^σ は幾何標準偏差に相当する。

ガンマ分布は形状パラメーター α 、比率パラメーター β の2つのパラメーターで特徴づけられ、その確率密度関数は以下の式で表される。

$$f(Y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} Y^{\alpha-1} e^{-\beta Y} \dots \text{Eq. 6}$$

ここで、 $\Gamma(\alpha)$ はガンマ関数である。ガンマ分布は、主に信頼性工学における電子部品の寿命分布や通信工学におけるトラフィックの待ち時間分布に応用される。また所得分布にも応用される。

ワイブル分布は、形状パラメーター m 、尺度パラメーター η の2つのパラメーターで特徴づけられ、その確率密度関数は以下の式で表される。

$$f(Y|m, \eta) = \frac{m}{\eta} \left(\frac{Y}{\eta}\right)^{m-1} \exp\left[-\left(\frac{Y}{\eta}\right)^m\right] \dots \text{Eq. 7}$$

ワイブル分布は、物体の強度を統計的に記述するために提案された確率密度分布であり、時間に対する劣化現象や寿命

を統計的に記述するためにも利用される。

4 ゼロ過剰モデル

ゼロ過剰 (zero-inflated) モデルは混合分布の1種であり、 ω の確率で 0、 $(1-\omega)$ の確率である確率密度分布に従うモデルである。つまり、農薬が含まれているかどうかはベルヌーイ分布に従い、含まれていた場合の残留農薬濃度がある確率密度分布に従うと解釈できる。残留農薬濃度は 0 以上の連続変数であるので、対数正規分布・ガンマ分布・ワイブル分布と組み合わせ、ゼロ過剰対数正規 (zero-inflated lognormal, ZILN) 分布、ゼロ過剰ガンマ (zero-inflated gamma, ZIG) 分布、およびゼロ過剰ワイブル (zero-inflated Weibull, ZIW) 分布への適用を検討した。

C. 試料と方法

1 残留農薬摂取量の推定

データには、2019 年から 2021 年に本研究班で行ったトータルダイエツトスタデー調査による、残留農薬摂取量の下限值および上限値を用いた。ただし、下限値がすべて 0 となった農薬及び、サンプル数が 2 未満の農薬は除いた。

解析には R (3.4.0) を用いた。ベイズモデルの計算は rstan パッケージ (2.16.2) を用いた。

作成したベイズモデルの選択は、下記の式で示した広く使える情報量規準 (widely applicable information criterion (WAIC)) を用いて WAIC が最も低いモデルを採用した。

$$WAIC = L_{WAIC} + p_{WAIC}$$

ここで、 L_{WAIC} はベイズ学習損失を示す対数尤度の平均値で、 p_{WAIC} は推定有効パラメ

ーター数を示す事後予測分布の揺らぎを示す。 L_{WAIC} と p_{WAIC} は以下のように定義される。

$$L_{WAIC} = -\frac{1}{n} \sum_{i=1}^n \log E(L_i)$$
$$p_{WAIC} = \frac{1}{n} \sum_{i=1}^n V(\log L_i).$$

ここで、 n はサンプル数、 L_i は事後予測分布から得られた各データポイントに対する尤度を、 $E(X)$ は変数 X の平均値、 $V(X)$ は変数 X の分散を示す。

最尤推定は Excel のソルバー機能を使用して、一般化簡約勾配法により最適解を算出した。

2 数値シミュレーション

以下の④～⑥に従って、MLE 法および BE 法で分布パラメーターの推定を行った。

- ① サンプルサイズ=200、幾何平均値=1、幾何標準偏差=3 の乱数を生成した。
- ② 真値が 0 のレコードとなる割合 ω (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) を決め、200 のサンプルから ω の割合で 0、 $(1-\omega)$ の割合で 1 を生成し、①と掛け合わせることで ZILN 分布に従う乱数を得た。
- ③ 不検出割合 φ (0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95) を決めた。ただし、 $\varphi > \omega$ を満たす。
- ④ 不検出とした最大値—検出された最小値間の一様分布から、不検出となる基準 (reporting limit, RL) を決めた。不検出となった測定値の下限値を 0, 上限値を RL とした。
- ⑤ 無情報、弱情報、強情報事前分布を用いた BE 法および MLE 法を用いて、分布パラメーター等を推定した。その際には Eq.2 で示した尤度関数を用いた。ただし、 L の値には 0 を、 U の値には RL

を用いた。

⑥ ①～⑤を 1000 回 (N_{sim}) 繰返す。

BE 法においては、3 つの事前分布を用いた；無情報事前分布、弱情報事前分布 [gsd \sim cauchy(3,1), mean_est \sim cauchy(mean^{MB}, mean^{MB})], 特定情報事前分布 [$\omega \sim N(\omega, \omega/10)$, gm $\sim N(1,0.1)$, gsd $\sim N(3, 0.3)$]。ここで、mean^{MB}は不検出となった観測値に LOQ の半値を代入して算出した平均値である。

推定した分布パラメータおよび期待値等について、以下の項目で評価した。

$$\text{真度: } rMSE(\hat{\theta}) = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{N_{sim}}}$$

精度: RSD($\hat{\theta}$)

妥当性: 平均値の 95% 推定区間に母平均が含まれる確率

平均値の 95% 推定区間は、BE 法では事後予測分布の 95% 信用区間を用い、MLE 法ではブートストラップ法 (500 回) を用いて計算した平均値の 95% 区間を用いた。

解析には R (4.1.0) を用いた。ベイズモデルの計算は cmdstanr パッケージ (0.4.0) を用いた。

3 二次元モンテカルロシミュレーション

TD 試料中の残留農薬濃度データには、2019 年から 2021 年に本研究班で行った TD 試料中の残留農薬濃度の下限値および上限値を用いた。各食品群の喫食量には、2014 年から 2016 年の国民健康・栄養調査のデータを用いた。

解析には R (3.4.0) を用いた。ベイズ推定には rstan パッケージ (2.16.2) を用いて、2000 個のモンテカルロサンプルを事後予測分布として得た。得られた事後予測分布を分布パラメータとする乱数を 50 個発生させ、計 10 万 (2000 \times 50) 個の

乱数を残留農薬濃度及び喫食量について生成した。

D. 結果と考察

1 確率モデルの設定

当初の計画では、各食品群について農薬濃度分布を推定し、喫食量の分布と掛け合わせることで、モンテカルロシミュレーションを用いた確率論的摂取量評価を行うことを想定していた。

しかしながら、ほぼすべての農薬において、すべての地域で農薬検出されない食品群が確認された。このように、全て不検出の場合は下限値-上限値間が最も狭い範囲内に確率密度分布が収まるように最適化される結果、パラメータが発散する、あるいは不適切な解が得られてしまう。つまり、14 食品群すべてで農薬濃度分布を推定することは現実的ではないと考えられた。そのため、農薬濃度の分布ではなく、14 食品群の総和である農薬摂取量の下限値及び上限値を用いて推定を行うこととした。

2 不検出レコードがあるデータの統計的解析手法について

化学分析で得られたデータが検出されたかどうか判断するためには、信号強度がある閾値 [検出限界値 (Limit of detection (LOD), 定量下限値 (Limit of Quantification (LOQ))] 以上である必要がある。分析化学においては、得られた信号は以下の 3 つのパターンに分けることが出来る。

- 1) 定量するに十分な信号強度が得られた。
- 2) 定量するに十分な信号強度が得られてはいないが、検出したとみなせる

信号強度が得られた。

- 3) 検出したとみなせる信号強度は得られなかった。

データに非定量値が存在すると、信頼できる値を割り当てることができないため、統計上の計算に困難を生じる。LOD や LOQ など、特定の基準より低い値が欠損したデータは、左打ち切りデータとして知られる。2006年に、EPAのガイダンスは、検出されなかった15%未満のデータで、0、LOD/2、またはLODを代入する計算方法をサポートしていた。しかし、2015年にEPAは左打ち切りデータの割合が5%未満で、データがわずかに歪んでいる場合にのみLOD/2による置換を使用するようにガイドラインを改訂した。代入法に代わる方法として、Kaplan-Meier (KM) 法、robust regression on order statistics (rROS) 法、最尤推定推定 (MLE)、ベイズ推定 (BE) 法などの利用が提案されている。

左打ち切りデータは2)と3)に分類することが出来る。この2つのパターンを分けた上で解析することで情報が細分化でき、より妥当な解析結果を得ることが出来ると考えられる。RのEnvStatパッケージには、KM, ROS, MLE法だけでなくその他多くの推定法を指定することができるが、2)と3)を区別して指定することには対応していない。

一方で、BE法にはBUGS, JAGS, Stanなどの確率的言語プログラムを利用した報告がなされている。商用の統計解析ソフトのようにグラフィックインターフェース環境が整っているわけではないため、利用者が明示的にコードを書く必要があり導入の敷居は高いが、利用者が様々な確率密度関数/累積確率を指定できるた

め自由度が高い。

3 ベイズ法を用いた農薬摂取量の推定

対数正規分布を仮定した場合のStanコードをFig. 1に示す。コード6つのブロックからなり、1~6行目はdataブロックで、解析に使用するデータを宣言している。2行目にはサンプル数、3行目には下限値、4行目には上限値をしている。5行目のN_newは2次元モンテカルロシミュレーション用に生成させる乱数の個数を指定している。

8~12行目はtransformed dataブロックで、データを変換して別のデータを生成している。11行目では、下限値と上限値の平均から、この分布の推定平均値($\hat{\mu}^{RL/2}$)を算出している。

14~17行目はparametersブロックで、推定したいパラメーターを宣言している。このモデルでは、対数正規分布を仮定しているため、幾何平均値(e^{μ})と幾何標準偏差(e^{σ})を推定するようにここで宣言している。GSDは1以上の実数であるため、GSDの取りうる範囲について制限している。

19~25行目はTransformed parametersブロックで、parametersブロックで宣言したパラメーターから、理論平均値および理論標準偏差を計算している。

27~44行目がmodelブロックで、そのなかでも29, 30行目は事前分布に相当する。形状パラメーターである幾何標準偏差の事前分布には、以下の確率密度関数で表せるコーシー分布を用いた。

$$f(Y|y_0, \gamma) = \frac{1}{\pi} \frac{\gamma}{(Y - y_0)^2 + \gamma^2}$$

対数正規分布を仮定した場合には $(y_0, \gamma) = (3, 1)$ を用い、ガンマ分布及びワイブル分布には $(y_0, \gamma) = (1, 1)$ をコーシー分布のパラメーターに用いた。

パラメーターから変換した理論平均値の事前分布には、下限値と上限値の平均値から求めた平均値 $\hat{\mu}^{RL/2}$ を用いて、 $N(\hat{\mu}^{RL/2}, \hat{\mu}^{RL/2})$ の正規分布を採用した。

33~44 行目がデータに対するモデル部分に相当する。34 行目では定量できたデータであるのかどうかを識別し、その場合は確率密度の対数値を計算し足しこむことを 35 行目で行っている。定量できなかった場合は (36 行目)、さらに LOD 未満のデータかどうかを 37 行目で識別し、38 行目で 0~下限値までの累積確率の対数値を計算し足しこんでいる。2 つのパターンのどちらでもない場合 (29 行目) は、下限値~上限値までの累積確率の対数値を計算し足しこむ (40 行目)。この足し込み ("target +=") を 23 行目で 1~N まで繰り返す操作を行うことで、尤度が高くなるパラメーターを探索・推定している。

46~63 行目の generated quantity ブロックでは推定したパラメーターを利用して種々の計算を行う。ここでは、WAIC を計算するための各サンプルの対数確率密度あるいは対数累積確率 (log_lik) を 50~60 行目において計算している。62 行目では、二次元モンテカルロシミュレーション用の乱数を生成している。

対数正規分布・ガンマ分布・ワイブル分布を仮定した BE を行い、WAIC を比較して WAIC が低い確率密度分布を採用した。また、その確率密度分布の元に得られた計算結果を Table 4 に示す。

BE 法で得られる推定結果は、事後予測分布として得られる。事後予測分布から推定値としては、3 つの点推定値が頻繁に用いられる; 事後予測平均値 (expected a posteriori, EAP), 事後予測確率最大値 (maximum a posteriori, MAP), 事後予測中央値 (median of posterior predictive distribution, MED)。MED は計算が簡便でかつ分布の歪みに比較的頑健であるため、本研究では MED を代表値として用いた。しかしながら、どの点推定値を採用するのかについては、シミュレーション研究および継続的な評価によって検討すべきであろう。

各農薬のパラメーターおよび摂取量の理論平均値 ($\hat{\mu}^{BE}$) と理論標準偏差 ($\hat{\sigma}^{BE}$) の事後予測分布を Fig. 3 に示す。Table 1 にはベイズ推定の結果として事後予測中央値 (MED) と四分位範囲を表記した。BE 法に用いることのできるサンプル数は 2~10 であったが、サンプル数が小さい場合であっても事後予測分布の広がりが大きくなるといった傾向は無く、比較的妥当な推定が行えたと判断できた。また、昨年度の推定結果と比較して、全体的に事後予測分布の範囲が狭くなった。これはサンプル数が増えたことによって推定の精度が増加したことによるものであると考えられた。

BE 法では、対数正規分布が 12 農薬で、ガンマ分布が 4 農薬で、ワイブル分布が 4 農薬で最も適していると判断された。

対数正規分布の形状パラメーターは平均して 3.14 ± 1.20 、ガンマ分布の形状パラメーターは 1.63、ワイブル分布の形状パラメーターは平均して 0.97 ± 0.51 であ

った。これらの値は、設定した事前分布との乖離は小さかった。この結果から、ベイズ推定に用いた形状パラメーターの事前分布は妥当であったと判断できた。

4 最尤推定法を用いた農薬摂取量の推定

Table 2 には、最尤推定法による農薬摂取量の確率密度分布の推定結果および理論平均値 ($\hat{\mu}^{\text{MLE}}$) と理論標準偏差 ($\hat{\sigma}^{\text{MLE}}$) の値を表記した。MLE 法では、対数正規分布が9農薬で、ガンマ分布が8農薬で、ワイブル分布が3農薬で最も適していると判断された。15農薬では、BE法とMLE法で最適な確率密度分布が一致していた。

5 推定法間の比較

BE法とMLE法で異なる確率密度分布と推定された農薬が散見された。MLE法でガンマ分布が最適と判断されたもののうち、フルアジホップブチル・フルアジホップブチル代謝物・フルベンジアミドは、BE法ではワイブル分布が最適であると判断された。

1日農薬摂取量の理論平均値に関する事後予測分布とLOQの0.2倍の値を代入した推定値およびMLE法と比較した結果をFig. 4に示す。NDにLOQの0.2倍の値を代入して推定した値よりもBE法による推定値の方が低かったのは、アセフェート・メタミドホスであった。MLE法よりもBE法による推定値の方が低かったのは、アセフェート・クロチアニジン・クロルピリホス・チアクロプリド・ノバルロン・ピフェントリン・ヘキサジノン・メタミドホスであった。インドキサカルブ・ピフェントリン・プロプロフェジン・フルベン

ジアミド代謝物のBE法による平均値の事後予測分布の四分位範囲内には、0.2LOQを代入して推定した値が含まれていなかった。MLE法による1日農薬摂取量の平均値は、全てBE法による平均値の事後予測分布の四分位範囲内に収まっていた。

MLE法とBE法の差異に与える要因を解析するために、BE法とMLE法で推定した平均値の相対的な差 $[2(\hat{\mu}^{\text{BE}} - \hat{\mu}^{\text{MLE}}) / (\hat{\mu}^{\text{BE}} + \hat{\mu}^{\text{MLE}})]$ について解析を行った (Fig. 5)。

すべての食品群で不検出となったサンプルの割合が高い場合、推定した分布の σ/μ 比が大きい場合、用いたデータのサンプルサイズが小さい場合に、BE法とMLE法による推定値の相対的な差が大きくなる傾向が認められた。

$L=0$ となるサンプルの割合は、有効なサンプルサイズに相当すると考えられる。また、本研究で仮定した確率密度分布の場合、 σ/μ が大きくなると歪度が大きくなることを意味する。これらの結果より、BE法とMLE法の差異が大きくなった要因は、歪みの大きい確率密度分布であること、サンプルサイズが小さいこと、および $L=0$ となるサンプルの割合が大きかったことの影響が大きいと考えられた。MLE法は、有効なサンプルサイズが少なくなるほど、また分布の歪みが大きくなるほど、推定の精度が低くなることが報告されている。この点を踏まえると、MLE法が不得意なデータにおいて、事前分布を用いることでBE法がMLE法の弱点をカバーしているとも捉えることができる。つまり、BE法による推定結果の方がMLE法による推定結果よりも妥当である可能性が示唆され

た。

しかしながら、真の値は不明であるため、代入法、MLE法、BE法のどちらの推定値が真値に近いかは判断できない。真の値との差(真度)、平均値の推定区間内に真の値が含まれているかどうか(妥当性)、異なるデータセットにおいても同様の解析結果を返すことができるか(繰返し性能)について、乱数を繰返し生成するシミュレーションを行い評価する必要があるだろう。

6 真値に 0 を含むデータの推定 (数値シミュレーション)

BE法の点推定値として事後分布の中央値(MED)を用いて、真度(Fig. 1)・精度(Fig. 2)・妥当性(Fig. 3)を評価した。MLE法は、 ω よりも φ による影響を強く受け、 φ が大きくなるほど、推定の真度および精度が悪くなる傾向があった。一方で、BE法は φ よりも ω による影響の方が強い傾向にあった。また、MLE法よりもBE法の方が、推定の真度・精度・妥当性が良好であった。さらに、この傾向は事前分布の設定に寄らず、無情報事前分布を用いた場合でもBE法の方がMLE法よりも良好な真度・精度・妥当性を示した。MLEとBE(NIP)を比較すると、 ω と gm の真度・精度はBE法が広い φ と ω の組合せでMLE法よりも良い推定結果となる傾向を示したが、 gsd に関してはそれほど大きな差異は認められなかった。この結果は、事前分布に形状パラメーターである gsd の情報を追加することで推定結果が向上する可能性を示唆するものと考えられた。実際に、弱事前分布として、 gsd に関する事前分布を用いることで、推定の真度・精

度・妥当性が向上することが示された。以上のことから、ZILN分布に従う不検出例を含むデータの解析においても、BE法が有効であることが示された。また、事前分布の利用は推定の真度・精度・妥当性を向上させるが、通常の利用では真値は不明であるため、弱情報事前分布の使用を検討するのが現実的であろう。事前分布に利用できるデータベースを整備することが有効であると考えられる。

7 二次元モンテカルロシミュレーションによる農薬摂取量分布の推定

ゼロ過剰モデルも考慮に入れ、TD試料中の残留農薬濃度を、BE法を用いて推定した。解析の対象としたのは、14群全ての農薬が検出され、デーサンプルサイズが6以上のものとした。ここでは、サンプルサイズの最も大きかったチアメトキサムの結果をTable 4に示す。14群中13群でゼロ過剰モデルが最適なモデルであると推定され、ZILN分布は11の食品群で、ゼロ過剰ワイブル分布が2つの食品群で最適な分布であった。検出率が最も高かった8群でもゼロ過剰モデルが採用されたことから、ゼロ過剰モデルが適切かどうかは検出率だけで決定されるものではないと示唆された。

農薬濃度分布の推定の不確かさを考慮するために、BE法で得られた2000個の事後予測サンプルから50個の乱数を生成し、計10万個の乱数を生成した。別途国民健康・栄養調査のデータを解析して得られた体重当たりの喫食量分布(Table 5)の乱数と掛け合わせ、摂取量分布を推定した。チアメトキサムの場合、5, 25, 50, 75, 96パーセンタイル値は、0.0005, 0.0024,

0.0060, 0.0148, 0.0586 $\mu\text{g}/\text{kg}/\text{day}$ であり、ADI (18 $\mu\text{g}/\text{kg}/\text{day}$) を超過する確率は0%であった (Table 6)。検討した農薬についての推定摂取量の平均値や各パーセントイル値等を Table 7 に示す。ADI を超過する確率は、アセフェートの0.001%が最大であったことから、検討したいずれの農薬も健康リスクは小さいと示唆された。

農薬摂取量の平均値について、2D-MCS から推定した結果と、摂取量のデータから BE 法で推定した結果を比較した (Fig. 9)。異なるアプローチで算出した平均値の分布が良い一致を示したことから、いずれの推定方法も妥当であると考えられた。

E 結論

食品を介して摂取する農薬の量をより精緻に推定するために、不検出 (ND) となった例を含むデータにおける統計妥当性の高い推定法に関して検討を行った。農薬摂取量は、1) 定量下限値 (LOQ) 以上, 2) 検出限界値 (LOD) 以上 LOQ 未満, 3) LOD 未満の三パターンに分け、定量値, 下限値, 上限値に喫食量を乗じて1日摂取量の下限値と上限値を算出した。これらの値を用いて、BE 法を用いて推定したところ、相対標準偏差の推定値が大きくなるほど、また下限値が0のデータ (全ての食品群で不検出だった地域) が多いほど、サンプルサイズが小さいほど、BE 法と MLE 法との推定値の相対的な差が大きくなる傾向にあった。このようなデータは MLE 法の妥当性が低くなることから、BE 法による推定結果の方が MLE 法による推定結果よりも妥当である可能性が示唆された。

また、農薬は使用履歴がない場合には真値に0を含むデータもありうる。そこで、ZILN 分布を対象とした数値シミュレーションを行った結果、BE 法は MLE 法と比較して真度・精度・妥当性に優れた推定結果を得ることができた。

以上の結果を基にゼロ過剰モデルも考慮して、各食品群中の農薬濃度分布を BE 法で推定し、2D-MCS による残留農薬摂取量分布を推定した。ADI を超過する確率は、アセフェートの0.001%が最大であったことから、検討したいずれの農薬も健康リスクは小さいと示唆された。さらに、農薬摂取量の平均値について、2D-MCS から推定した結果と、摂取量のデータから BE 法で推定した結果を比較したところ、平均値の分布が良い一致を示した。以上のことから、いずれの推定方法も妥当であると考えられた。

F. 研究発表

1. 論文発表

なし

2. 学会発表

- 1) 鈴木美成, 青柳光敏, 戸田英汰, 伊藤功一, 福光徹, 萩尾真人, 林孝子, 新宅沙織, 井原紗弥香, 川崎恭寛, 中島安基江, 佐藤環, 飛石和太, 堀就英, 穠山浩; ベイズ法を用いた食品を介した残留農薬摂取量の推定の試み, 第29回環境化学討論会 2021年6月.
- 2) 鈴木美成, 青柳光敏, 戸田英汰, 伊藤功一, 福光徹, 萩尾真人, 林孝子, 新宅沙織, 井原紗弥香, 川崎恭寛, 中島安基江, 佐藤環, 飛石和太, 堀就英,

穂山浩; 不検出値を含むデータを用いたベイズ推定による残留農薬摂取量の精緻化の試み, 第 58 回全国衛生化学技術協議会年会 2021 年 12 月.

H. 知的財産権の 出願, 登録状況
なし

I. 健康危機情報
なし

Table 1 2019–2021 年に調査した農薬 1 日摂取量 ($\mu\text{g}/\text{person}/\text{day}$) のベイズ推定による結果*

	確率密度分布	Parameter1	Parameter2	平均値	標準偏差
アセタミプリド	対数正規分布	1.14 (0.99–1.31)	2.24 (2.01–2.53)	1.59 (1.35–1.88)	1.51 (1.11–2.12)
アセタミプリド代謝物	ワイブル分布	2.08 (1.72–2.45)	0.151 (0.134–0.174)	0.135 (0.119–0.156)	0.067 (0.056–0.083)
アセフェート	ガンマ分布	0.527 (0.394–0.703)	2.46 (1.68–3.51)	0.321 (0.234–0.434)	0.512 (0.349–0.798)
インドキサカルブ	対数正規分布	0.005 (0.003–0.007)	2.88 (2.31–3.47)	0.008 (0.006–0.011)	0.011 (0.007–0.019)
クロチアニジン	ガンマ分布	1.68 (1.36–2.08)	2.65 (2.07–3.32)	0.649 (0.566–0.738)	0.489 (0.421–0.582)
クロルピリホス	ガンマ分布	2.06 (1.61–2.65)	12.9 (9.92–17.1)	0.161 (0.141–0.183)	0.110 (0.094–0.132)
チアクロプリド	対数正規分布	0.110 (0.086–0.142)	4.49 (3.94–5.22)	0.354 (0.263–0.477)	1.04 (0.648–1.68)
チアメトキサム	ワイブル分布	0.787 (0.690–0.886)	0.698 (0.562–0.865)	0.820 (0.665–0.126)	1.04 (0.816–1.37)
ニテンピラム	対数正規分布	0.165 (0.119–0.219)	2.91 (2.49–3.38)	0.288 (0.217–0.383)	0.409 (0.270–0.656)
ノバルロン	対数正規分布	0.036 (0.029–0.045)	4.07 (3.62–4.68)	0.101 (0.076–0.133)	0.245 (0.166–0.381)
ビフェントリン	ワイブル分布	2.36 (1.89–2.97)	0.213 (0.187–0.244)	0.191 (0.167–0.220)	0.084 (0.067–0.108)
ピリダベン	ワイブル分布	0.609 (0.497–0.750)	0.051 (0.033–0.070)	0.076 (0.057–0.105)	0.129 (0.086–0.207)
フェントロチオン	対数正規分布	0.112 (0.079–0.148)	3.39 (2.92–3.99)	0.240 (0.176–0.322)	0.442 (0.289–0.719)
ブプロフェジン	対数正規分布	0.068 (0.053–0.087)	2.84 (2.45–3.25)	0.117 (0.092–0.151)	0.162 (0.107–0.249)
フルアジホップブチル	対数正規分布	0.010 (0.006–0.014)	3.86 (3.26–4.73)	0.025 (0.017–0.037)	0.059 (0.035–0.103)
フルアジホップブチル代謝物	対数正規分布	0.024 (0.016–0.036)	3.44 (2.95–4.15)	0.054 (0.037–0.080)	0.105 (0.062–0.189)
フルベンジアミド	対数正規分布	0.250 (0.198–0.309)	2.91 (2.57–3.32)	0.435 (0.349–0.561)	0.635 (0.443–0.937)
ヘキサジノン	対数正規分布	0.010 (0.007–0.013)	3.88 (3.32–4.62)	0.026 (0.019–0.035)	0.058 (0.037–0.096)
ボスカリド	対数正規分布	1.22 (1.08–1.36)	2.05 (1.85–2.29)	1.58 (1.40–1.82)	1.29 (1.00–1.71)
メタミドホス	ガンマ分布	0.227 (0.158–0.312)	4.82 (3.18–7.21)	0.049 (0.034–0.067)	0.098 (0.070–0.141)

* Parameter 1,2 は、対数正規分布を仮定した場合には、それぞれパラメーター e^μ と e^σ を、ガンマ分布の場合はパラメーター β および α を、ワイブル分布の場合はパラメーター η および m を示す。全ての推定値は事後予測分布の中央値と四分位範囲を示す。

Table 2 2019–2021 年に調査した農薬 1 日摂取量 ($\mu\text{g}/\text{person}/\text{day}$) の最尤推定による結果

	確率密度分布	Parameter1	Parameter2	平均値	標準偏差
アセタミプリド	対数正規分布	1.11	1.95	1.39	1.62
アセタミプリド代謝物	ワイブル分布	2.26	0.143	0.126	0.059
アセフェート	ガンマ分布	0.370	1.69	0.242	0.398
インドキサカルブ	ワイブル分布	1.71	0.007	0.007	0.004
クロチアニジン	ガンマ分布	1.69	2.54	0.665	0.512
クロルピリホス	ガンマ分布	2.23	13.5	0.165	0.111
チアクロプリド	対数正規分布	0.100	5.27	0.399	2.03
チアメトキサム	ワイブル分布	0.755	0.617	0.731	0.982
ニテンピラム	対数正規分布	0.157	2.33	0.226	0.377
ノバルロン	対数正規分布	0.033	4.62	0.106	0.466
ビフェントリン	対数正規分布	0.183	1.35	0.192	0.076
ピリダベン	ガンマ分布	0.233	3.91	0.060	0.123
フェニトロチオン	対数正規分布	0.088	3.63	0.202	0.660
ブプロフェジン	対数正規分布	0.067	2.28	0.093	0.149
フルアジホップブチル	ガンマ分布	0.328	13.6	0.024	0.042
フルアジホップブチル代謝物	ガンマ分布	0.335	8.68	0.039	0.067
フルベンジアミド	ガンマ分布	1.50	4.26	0.353	0.288
ヘキサジノン	対数正規分布	0.007	4.88	0.026	0.121
ボスカリド	対数正規分布	1.197	1.83	1.43	1.45
メタミドホス	ガンマ分布	0.151	2.38	0.064	0.163

* Parameter 1,2 は、対数正規分布を仮定した場合には、それぞれパラメーター e^{μ} と e^{σ} を、ガンマ分布の場合はパラメーター β および α を、ワイブル分布の場合はパラメーター η および m を示す。全ての推定値は事後予測分布の中央値と四分位範囲を示す。

Table 3 TD 試料中のチアマトキサム濃度分布の推定結果 (事後予測分布の平均値 ± 標準偏差)

Food group	Detection ratio	Distribution ^a	Distribution parameter 1 ^b	Distribution parameter 2 ^c	Distribution parameter 3 ^d	Estimated mean	Estimated SD
FG01	0.188	ZILN	0.267 ± 0.110	0.081 ± 0.058	3.66 ± 1.40	0.043 ± 0.028	0.248 ± 0.509
FG02	0.188	ZIW	0.370 ± 0.135	3.26 ± 1.59	0.022 ± 0.007	0.007 ± 0.004	0.011 ± 0.005
FG03	0.125	ZILN	0.237 ± 0.118	0.047 ± 0.033	3.20 ± 1.11	0.019 ± 0.013	0.095 ± 0.128
FG04	0.125	ZILN	0.171 ± 0.080	0.286 ± 0.196	3.31 ± 1.07	0.088 ± 0.063	0.506 ± 0.603
FG05	0.063	ZILN	0.190 ± 0.135	0.036 ± 0.041	3.49 ± 1.67	0.011 ± 0.011	0.097 ± 0.249
FG06	0.500	Weibull	0.273 ± 0.046	0.231 ± 0.222		3.55 ± 3.34	32.1 ± 52.1
FG07	0.563	ZIW	0.601 ± 0.123	0.549 ± 0.131	2.00 ± 1.24	2.19 ± 1.31	6.97 ± 7.08
FG08	0.875	ZILN	0.845 ± 0.085	0.836 ± 0.242	3.33 ± 0.59	1.46 ± 0.50	3.11 ± 1.77
FG09	0.250	ZILN	0.340 ± 0.120	0.094 ± 0.050	3.44 ± 0.99	0.064 ± 0.035	0.271 ± 0.387
FG10	0.063	ZILN	0.174 ± 0.116	0.044 ± 0.049	3.45 ± 1.38	0.013 ± 0.012	0.105 ± 0.261
FG11	0.063	ZILN	0.180 ± 0.122	0.035 ± 0.040	3.58 ± 1.64	0.011 ± 0.010	0.099 ± 0.240
FG12	0.063	ZILN	0.203 ± 0.136	0.027 ± 0.033	3.68 ± 1.96	0.010 ± 0.010	0.111 ± 0.534
FG13	0.500	ZILN	0.558 ± 0.129	0.128 ± 0.052	3.21 ± 0.71	0.136 ± 0.059	0.367 ± 0.278
FG14	0.125	ZILN	0.238 ± 0.118	0.025 ± 0.019	3.15 ± 1.15	0.010 ± 0.009	0.053 ± 0.088

a: ZILN: ゼロ過剰対数正規, ZIW: ゼロ過剰ワイブル

b: Weibull の形状パラメーター m 、ZILN 分布と ZIW 分布の $1-\omega$ を示す。

c: Weibull の位置パラメーター η 、ZILN 分布の e^μ 、ZIW 分布の m 、を示す。

d: ZILN 分布の e^σ 、ZIW 分布の η を示す。

Table 4 体重当たりの喫食量に関する確率密度分布のパラメーターに関するベイズ推定の結果

Food Group	Distribution	Parameter 1 ^a	Parameter 2 ^b	Parameter 3 ^c	Estimated theoretical mean	Estimated theoretical SD ^d
1	ZIG	0.965 ± 0.001	3.13 ± 0.03	0.488 ± 0.005	6.19 ± 0.02	3.75 ± 0.02
2	Tweedie	3.26 ± 0.02	1.059 ± 0.008	1.553 ± 0.005	3.26 ± 0.02	2.58 ± 0.01
3	ZILN	0.869 ± 0.002	0.321 ± 0.003	4.55 ± 0.03	0.88 ± 0.01	2.84 ± 0.07
4	Tweedie	0.901 ± 0.002	1.49 ± 0.01	6.63 ± 0.06	0.203 ± 0.001	0.187 ± 0.001
5	ZIG	0.759 ± 0.003	1.19 ± 0.01	0.750 ± 0.008	1.204 ± 0.009	1.44 ± 0.01
6	ZIG	0.659 ± 0.003	1.19 ± 0.01	0.356 ± 0.004	2.20 ± 0.02	2.94 ± 0.02
7	ZIG	0.940 ± 0.001	1.19 ± 0.01	0.574 ± 0.006	1.95 ± 0.01	1.91 ± 0.01
8	Tweedie	0.9937 ± 0.0005	2.14 ± 0.02	0.534 ± 0.005	3.98 ± 0.02	2.75 ± 0.02
9	Tweedie	11.54 ± 0.06	2.14 ± 0.02	1.527 ± 0.003	11.54 ± 0.06	9.46 ± 0.06
10	Tweedie	1.309 ± 0.009	1.26 ± 0.01	1.476 ± 0.003	1.309 ± 0.009	1.369 ± 0.009
11	ZIG	0.963 ± 0.001	2.08 ± 0.02	0.815 ± 0.008	2.46 ± 0.01	1.80 ± 0.01
12	Tweedie	0.746 ± 0.003	0.799 ± 0.007	0.196 ± 0.002	3.04 ± 0.03	4.32 ± 0.04
13	ZILN	0.9990 ± 0.0002	1.204 ± 0.006	2.299 ± 0.008	1.70 ± 0.01	1.70 ± 0.02

a: パラメーター1は頻度パラメーターで、Tweedie分布の θ 、ZILN分布とZIG分布の p を示す。

b: パラメーター2は位置あるいは割合パラメーターで、Tweedie分布の μ 、ZILN分布の e^μ 、ZIG分布の β を示す。

c: パラメーター3は形状パラメーターで、Tweedie分布の ϕ 、ZILNの e^σ 、ZIG分布の α を示す。

d: standard deviation

Table 5 ベイズ法と二次元モンテカルロ法を併用して推定した食品を介したチアメトキサムの1日摂取量 ($\mu\text{g}/\text{kg}/\text{day}$) 分布の推定結果

Food Group	Mean	Standard deviation	5 th percentile	25 th percentile	50 th percentile	75 th percentile	95 th percentile	Contribution ratio (%)
1	0.00028	0.00245	0.00000	0.00000	0.00000	0.00001	0.00119	4.0 ± 13.0
2	0.00002	0.00005	0.00000	0.00000	0.00000	0.00003	0.00012	1.2 ± 5.8
3	0.00002	0.00029	0.00000	0.00000	0.00000	0.00003	0.00012	0.5 ± 4.0
4	0.00002	0.00021	0.00000	0.00000	0.00000	0.00000	0.00007	0.6 ± 4.5
5	0.00001	0.00016	0.00000	0.00000	0.00000	0.00000	0.00003	0.4 ± 3.7
6	0.00725	0.09014	0.00000	0.00000	0.00000	0.00032	0.0157	12.5 ± 26.1
7	0.0042	0.0222	0.00000	0.00000	0.00002	0.00152	0.0188	18.9 ± 30.3
8	0.0058	0.0171	0.00000	0.00041	0.00182	0.00545	0.0229	45.7 ± 36.9
9	0.00073	0.00553	0.00000	0.00000	0.00000	0.0002	0.0033	7.7 ± 19.0
10	0.00002	0.00026	0.00000	0.00000	0.00000	0.00000	0.00005	0.5 ± 4.0
11	0.00003	0.00036	0.00000	0.00000	0.00000	0.00000	0.00008	0.7 ± 4.9
12	0.00003	0.00089	0.00000	0.00000	0.00000	0.00000	0.00006	0.6 ± 4.8
13	0.00023	0.00089	0.00000	0.00000	0.00002	0.00017	0.00098	5.4 ± 14.2
14	0.00005	0.00030	0.00000	0.00000	0.00000	0.00000	0.00021	1.4 ± 6.7
Total	0.01864	0.09453	0.0005	0.0024	0.0060	0.0148	0.0586	

Table 5 二次元モンテカルロシミュレーションを用いた残留農薬の1日摂取量 (µg/kg/day) 分布

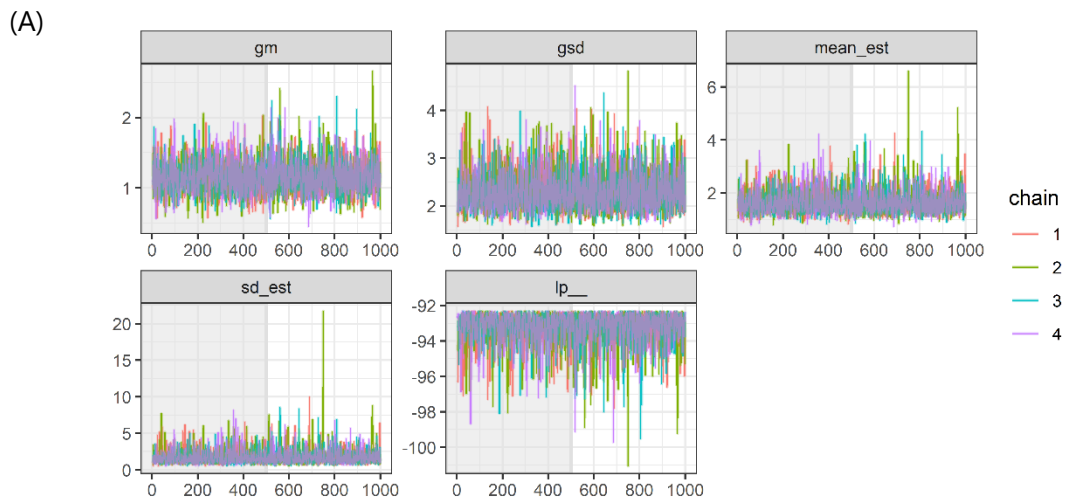
Pesticide	Mean	Standard deviation	5 th percentile	25 th percentile	50 th percentile	75 th percentile	95 th percentile	Probability of exceeding ADI (%)
Acephate	0.0049	0.0182	0.0002	0.0006	0.0016	0.0043	0.0181	0.001
Acetamiprid	0.032	0.093	0.001	0.006	0.017	0.038	0.105	0
Buprofedine	0.0027	0.0132	0.0003	0.0007	0.0011	0.0019	0.0079	0
Flubendiamide	0.0084	0.0141	0.0005	0.0023	0.0048	0.0093	0.0268	0
Hexazinone	0.00059	0.00171	0.00002	0.00009	0.00024	0.00058	0.00206	0
Methamidphos	0.00111	0.00417	0.00003	0.00012	0.00027	0.00071	0.00437	0
Thiacloprid	0.0091	0.0428	0.0000	0.0002	0.0009	0.0042	0.0394	0
Thiamethoxam	0.0186	0.0945	0.0005	0.0024	0.0060	0.0148	0.0586	0

```

1. data {
2.   int<lower=1> N;
3.   vector<lower=0>[N] Y_L;
4.   vector<lower=0>[N] Y_U;
5.   int<lower=1> N_new;
6. }
7.
8. transformed data {
9.   real mean_RL2;
10.
11.   mean_RL2 = mean((Y_L + Y_U)/2);
12. }
13.
14. parameters {
15.   real<lower=0> gm;
16.   real<lower=1> gsd;
17. }
18.
19. transformed parameters {
20.   real mean_est;
21.   real sd_est;
22.
23.   mean_est = exp( log(gm) + 0.5*log(gsd)^2 );
24.   sd_est = mean_est * sqrt( exp( log(gsd)^2 ) - 1 );
25. }
26.
27. model {
28.   // prior distribution
29.   gsd ~ cauchy(1, 1);
30.   target += normal_lpdf(mean_est | mean_RL2, mean_RL2);
31.
32.   // model
33.   for (i in 1:N) {
34.     if ( Y_L[i]==Y_U[i] ) {
35.       target += lognormal_lpdf( Y_U[i] | log(gm), log(gsd) );
36.     } else {
37.       if ( Y_L[i]==0 ) {
38.         target += lognormal_lcdf( Y_U[i] | log(gm), log(gsd) );
39.       } else {
40.         target += log( exp(lognormal_lcdf( Y_U[i] | log(gm), log(gsd) )) - exp(lognormal_lcdf( Y_L[i] | log(gm),
41. log(gsd) )) );
42.       }
43.     }
44.   }
45.
46. generated quantities {
47.   real log_lik[N];
48.   real<lower=0> Y_new[N_new];
49.
50.   for (i in 1:N) {
51.     if ( Y_L[i]==Y_U[i] ) {
52.       log_lik[i] = lognormal_lpdf( Y_U[i] | log(gm), log(gsd) );
53.     } else {
54.       if ( Y_L[i]==0 ) {
55.         log_lik[i] = lognormal_lcdf( Y_U[i] | log(gm), log(gsd) );
56.       } else {
57.         log_lik[i] = log( exp(lognormal_lcdf( Y_U[i] | log(gm), log(gsd) )) - exp(lognormal_lcdf( Y_L[i] | log(gm),
58. log(gsd) )) );
59.       }
60.     }
61.
62.     for (i in 1:N_new) Y_new[i] = lognormal_rng(log(gm), log(gsd));
63.   }
64. }

```

Fig. 1 対数正規分布を仮定した農薬摂取量に対するベイズモデルの Stan コード



(B) Convergence indices for MCMC parameters (iterations=1000, warm-up=500, chain=4, thinning=1)

Parameter	Mean	SD	\hat{R}	n_{eff}/N	mcse/sd
gm	1.17	0.25	0.999	0.474	0.0301
gsd	2.31	0.41	0.999	0.627	0.0306
mu_est	1.68	0.48	1.000	0.491	0.0300
sd_est	1.79	1.11	1.000	0.581	0.0292
lp_	-93.27	0.98	1.004	0.396	0.0371

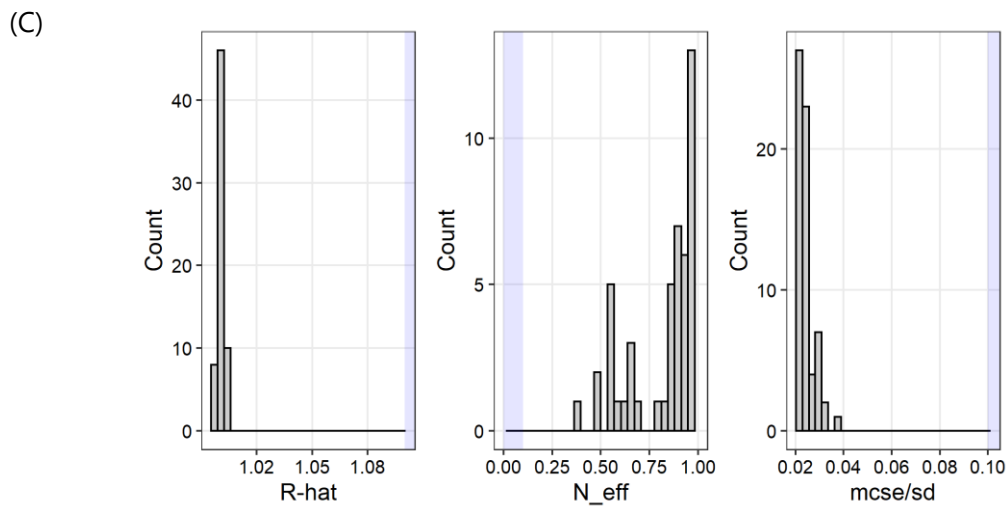


Fig. 2 アセタミプリドの1日摂取量推定におけるベイズ推定によるパラメーターの収束判断 (A) トレースプロット, (B) 推定したパラメーター等の記述統計量, (C) 収束判断指標 (\hat{R} , n_{eff}/N , mcse/sd) のヒストグラム

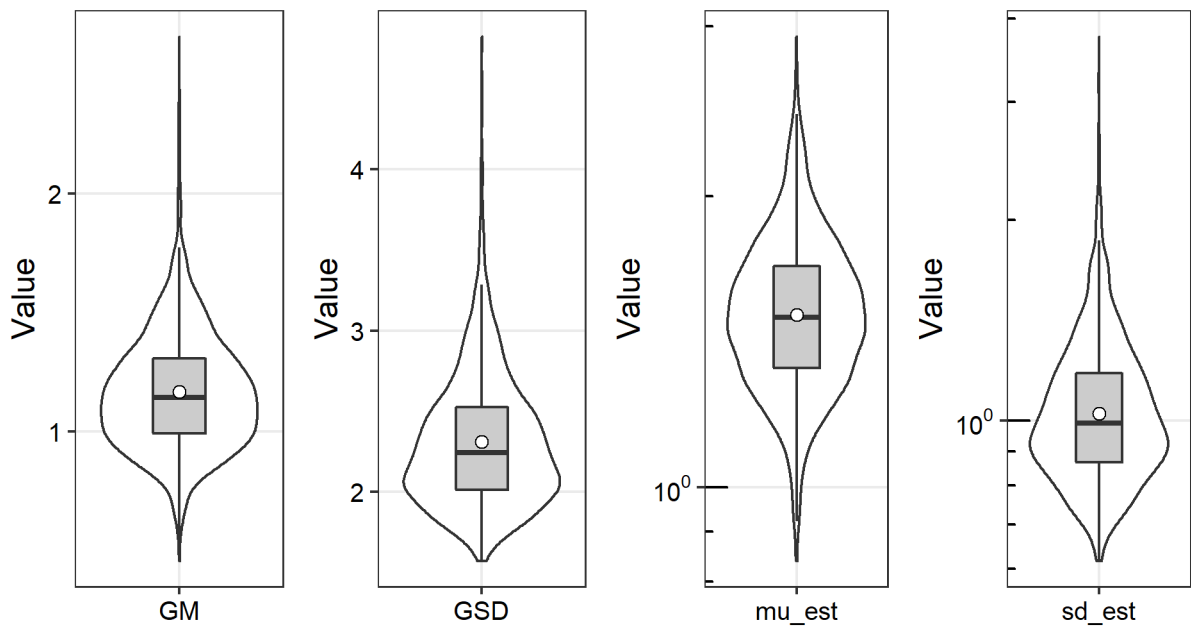


Fig. 3-1 ベイズ推定によるアセタミプリドの1日摂取量推定値に関わる事後予測分布

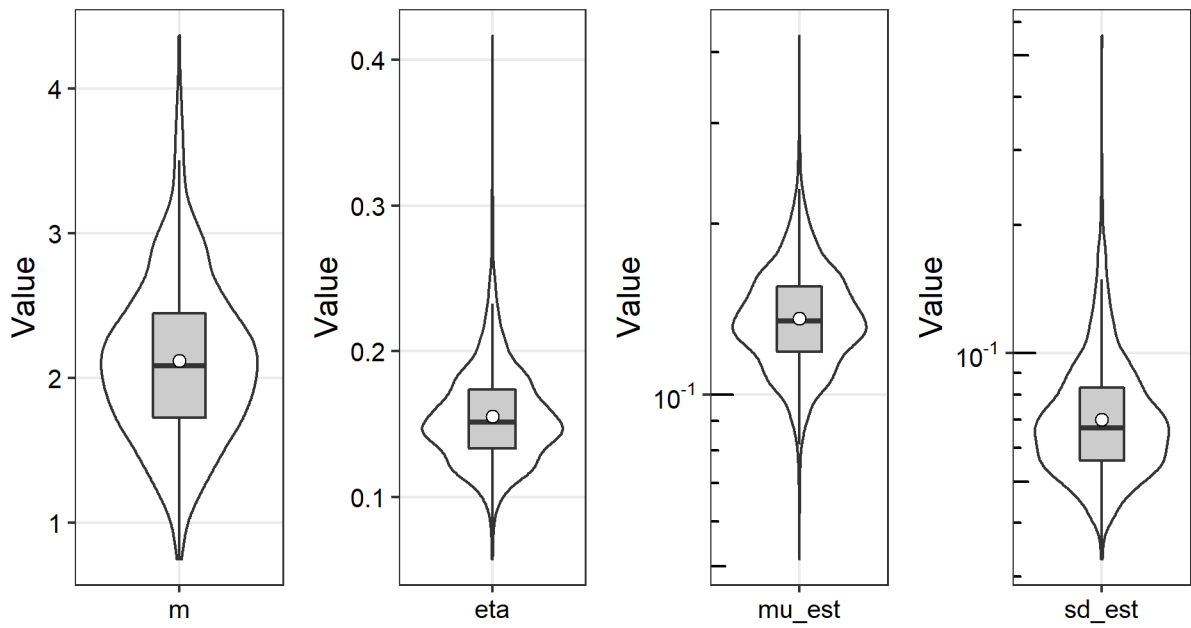


Fig. 3-2 ベイズ推定によるアセタミプリド代謝物の1日摂取量推定値に関わる事後予測分布

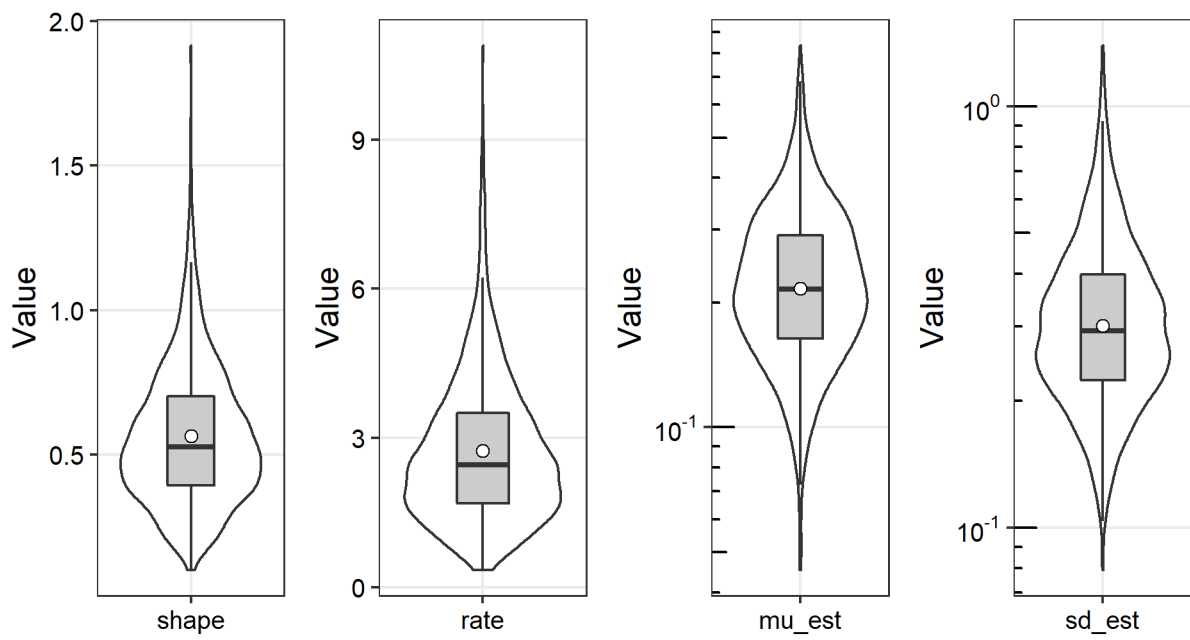


Fig. 3-3 ベイズ推定によるアセフェートの1日摂取量推定値に関わる事後予測分布

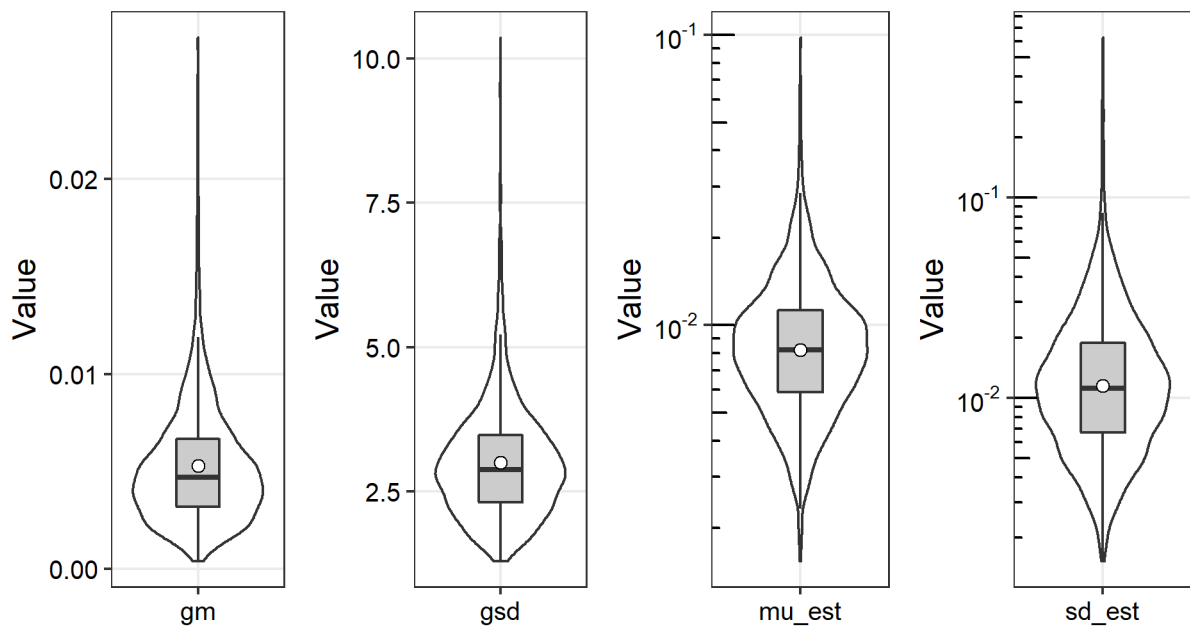


Fig. 3-4 ベイズ推定によるインドキサルブの1日摂取量推定値に関わる事後予測分布

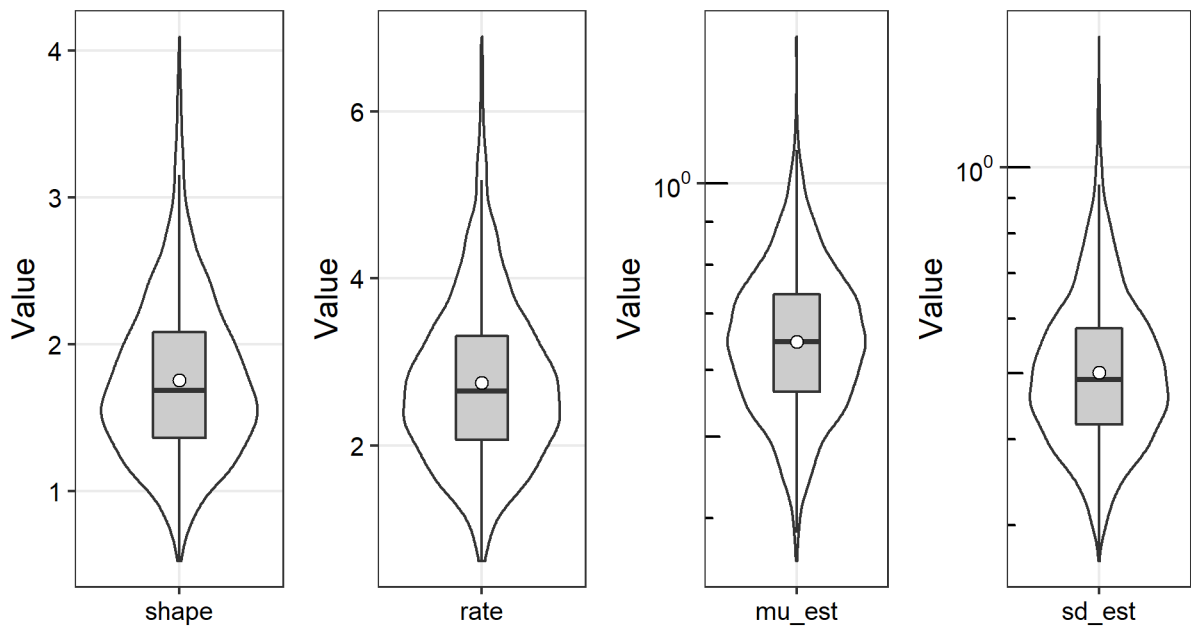


Fig. 3-5 ベイズ推定によるクロチアニジンの1日摂取量推定値に関わる事後予測分布

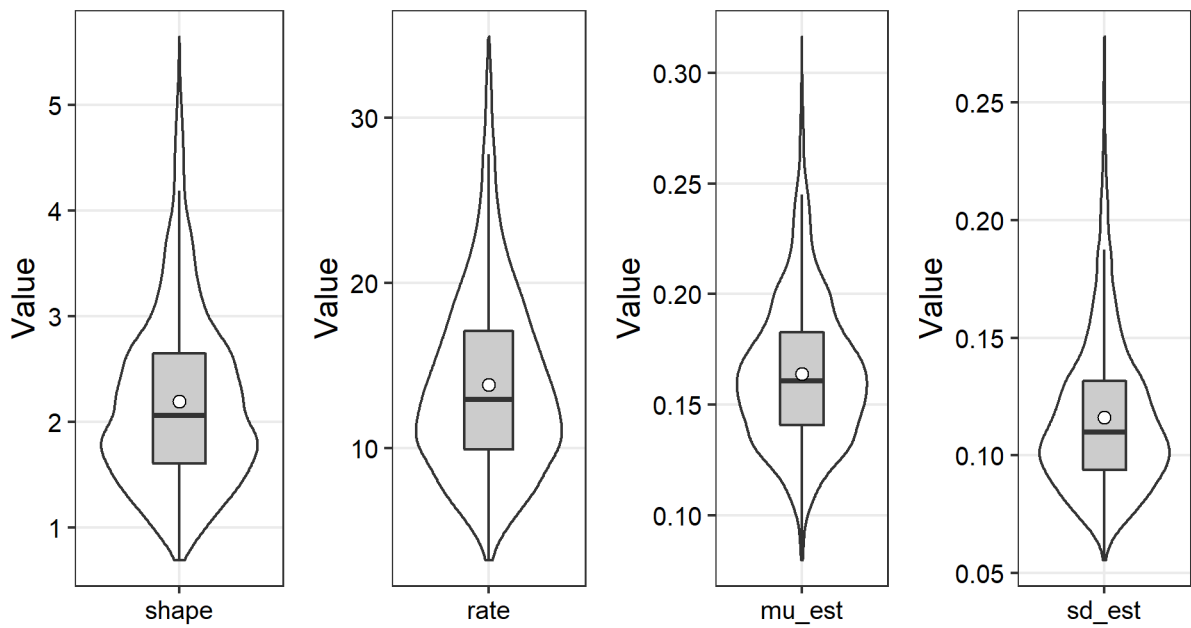


Fig. 3-6 ベイズ推定によるクロルピリホスの1日摂取量推定値に関わる事後予測分布

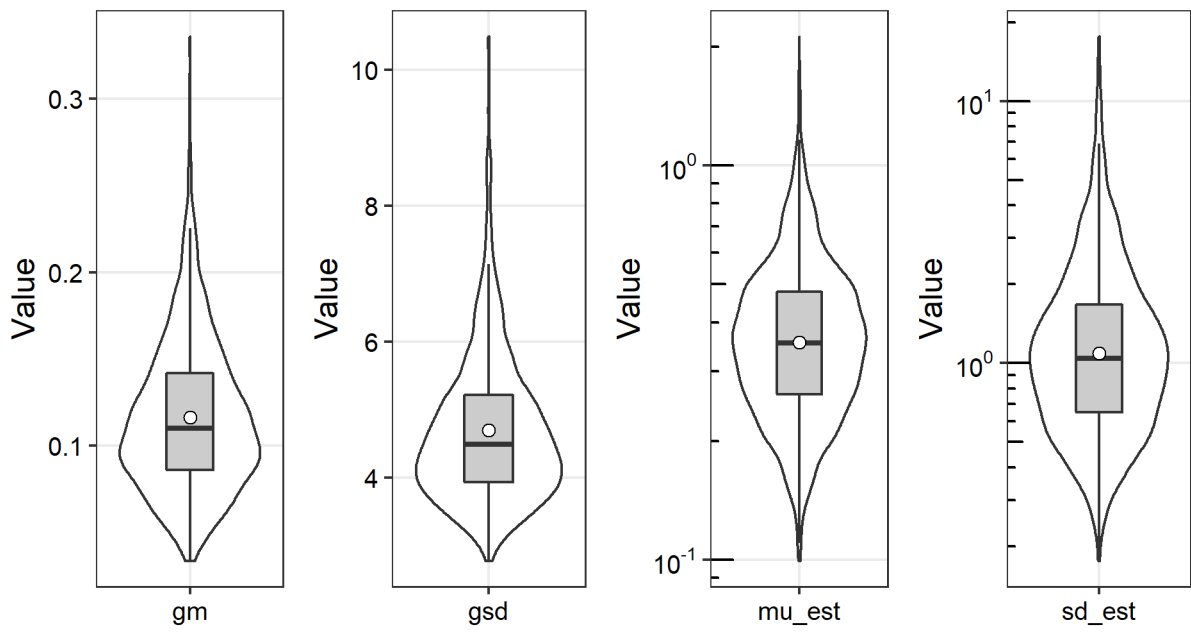


Fig. 3-7 ベイズ推定によるチアクロプリドの1日摂取量推定値に関わる事後予測分布

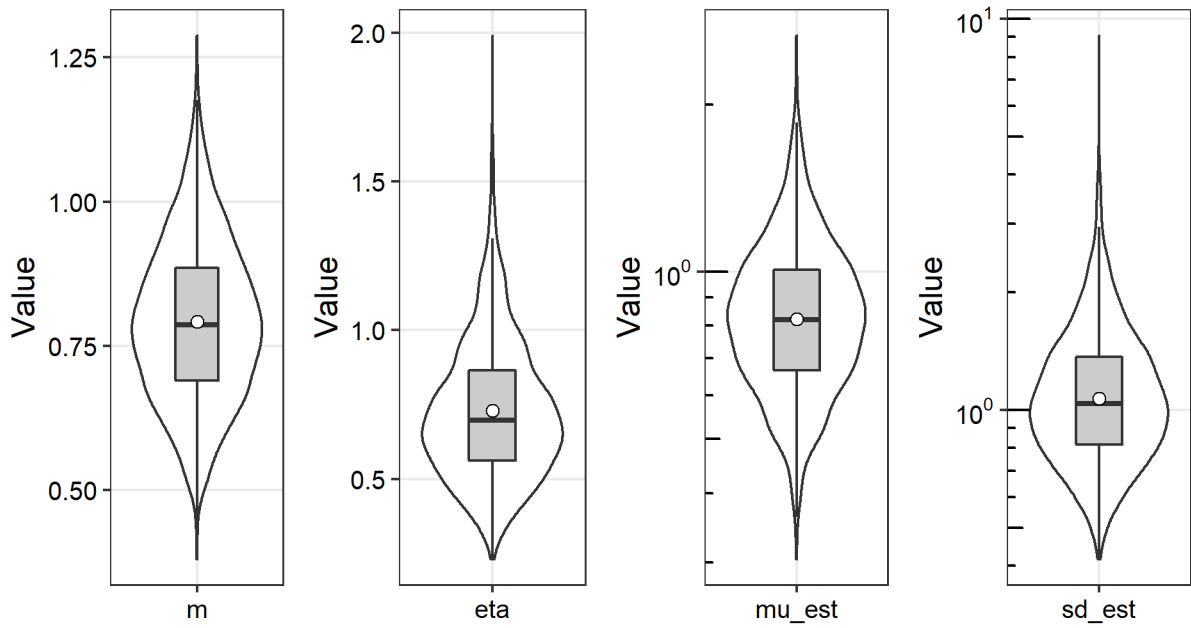


Fig. 3-8 ベイズ推定によるチアメトキサムの1日摂取量推定値に関わる事後予測分布

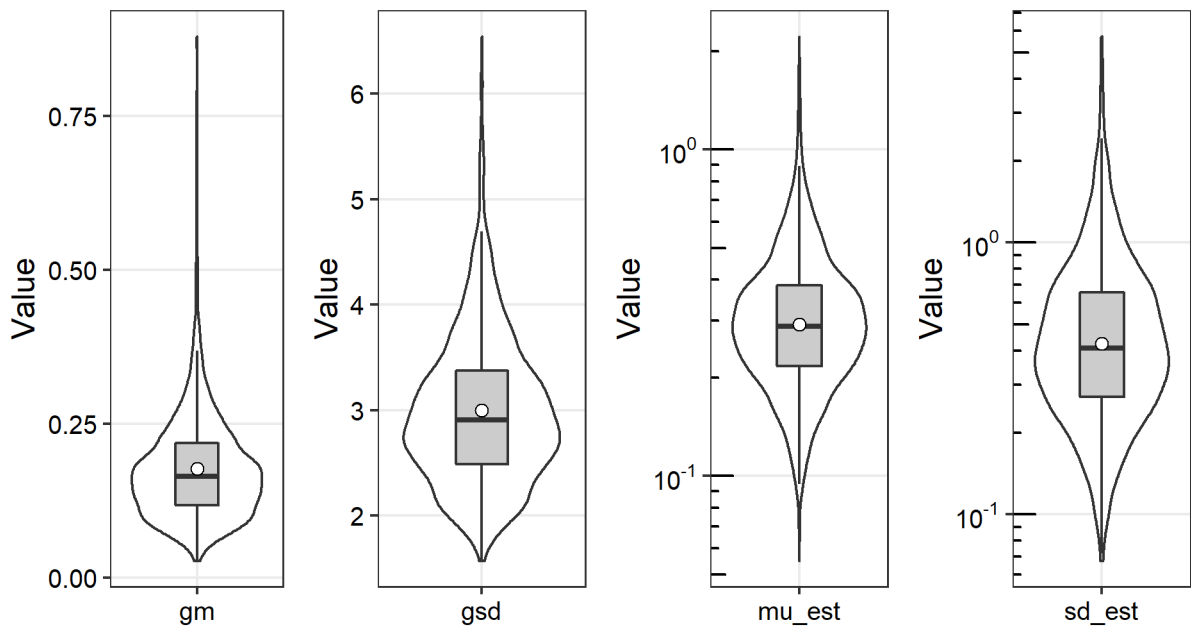


Fig. 3-9 ベイズ推定によるニテンピラムの1日摂取量推定値に関わる事後予測分布

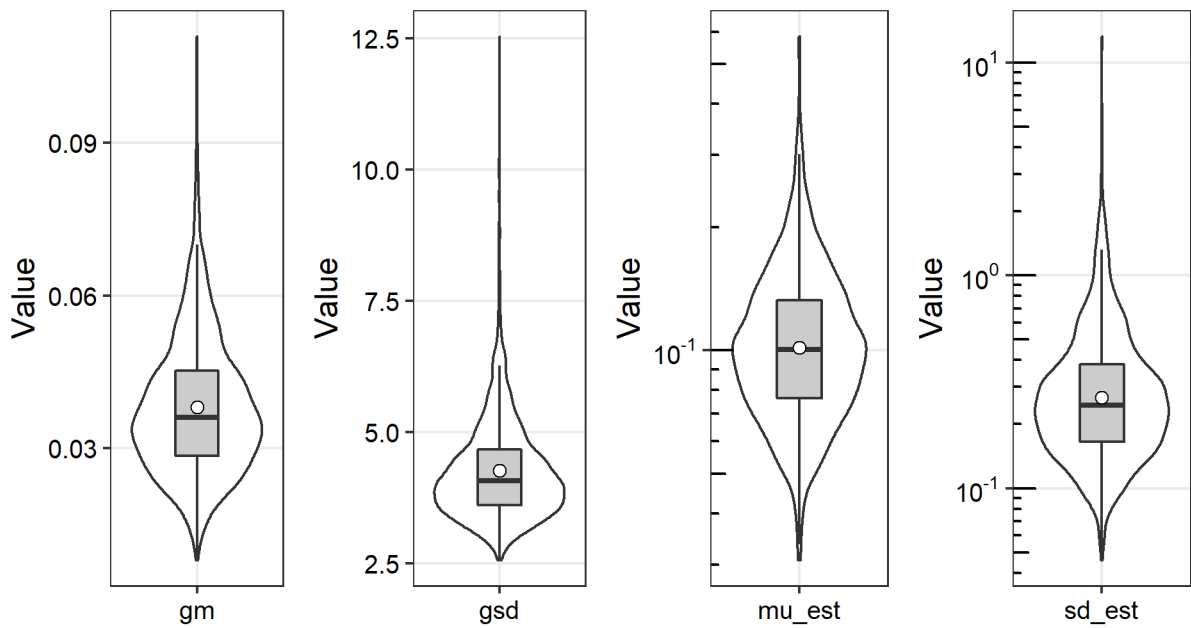


Fig. 3-10 ベイズ推定によるノバルロンの1日摂取量推定値に関わる事後予測分布

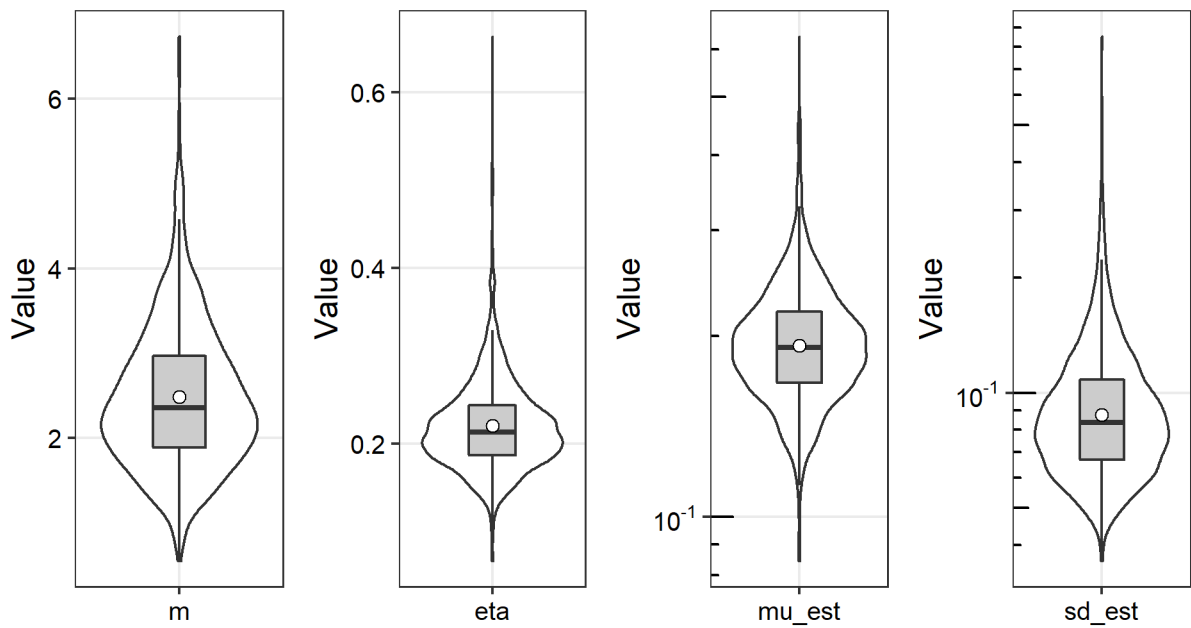


Fig. 3-11 ベイズ推定によるピフェントリンの1日摂取量推定値に関わる事後予測分布

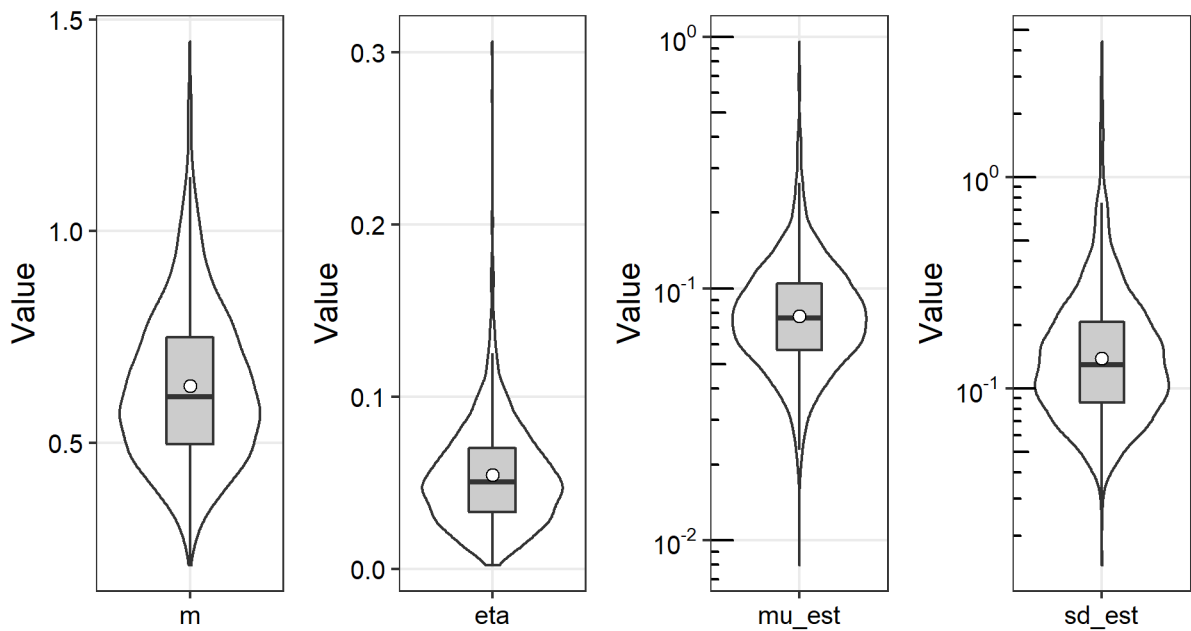


Fig. 3-12 ベイズ推定によるピリダベン1日摂取量推定値に関わる事後予測分布

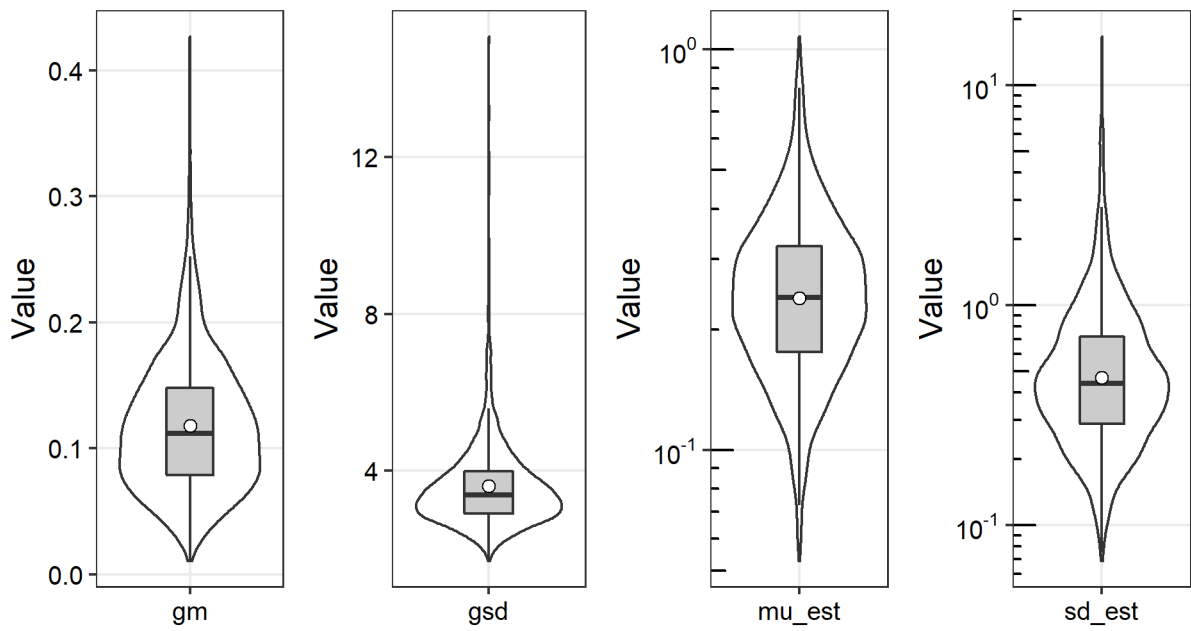


Fig. 3-13 ベイズ推定によるフェニトロチオンの1日摂取量推定値に関わる事後予測分布

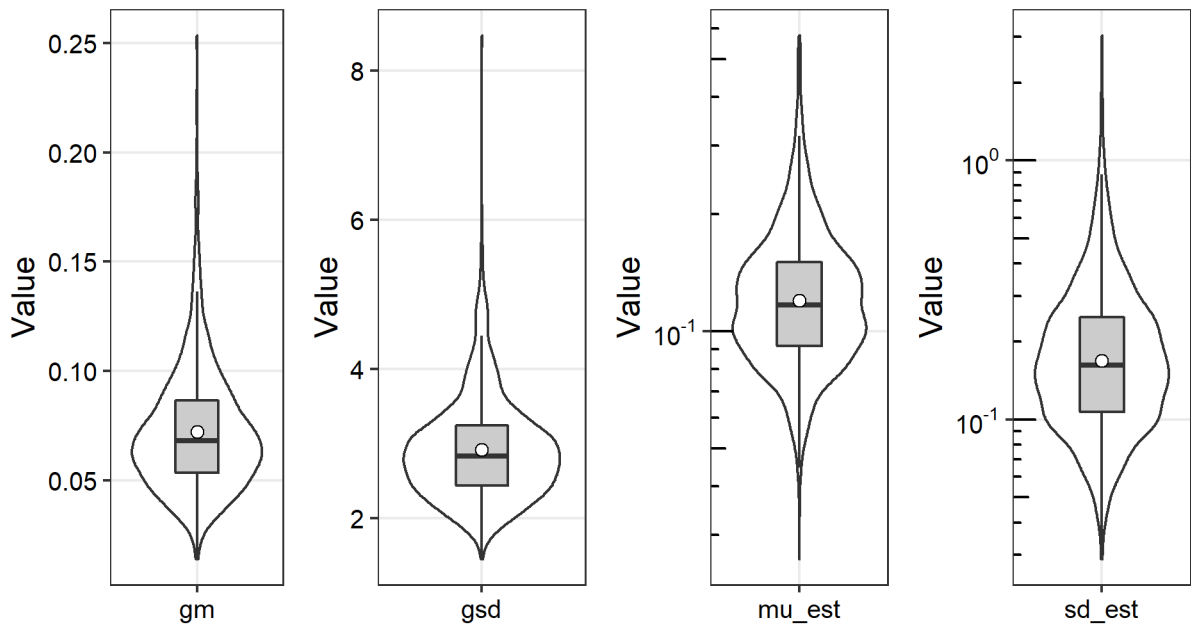


Fig. 3-14 ベイズ推定によるプロフェジンの1日摂取量推定値に関わる事後予測分布

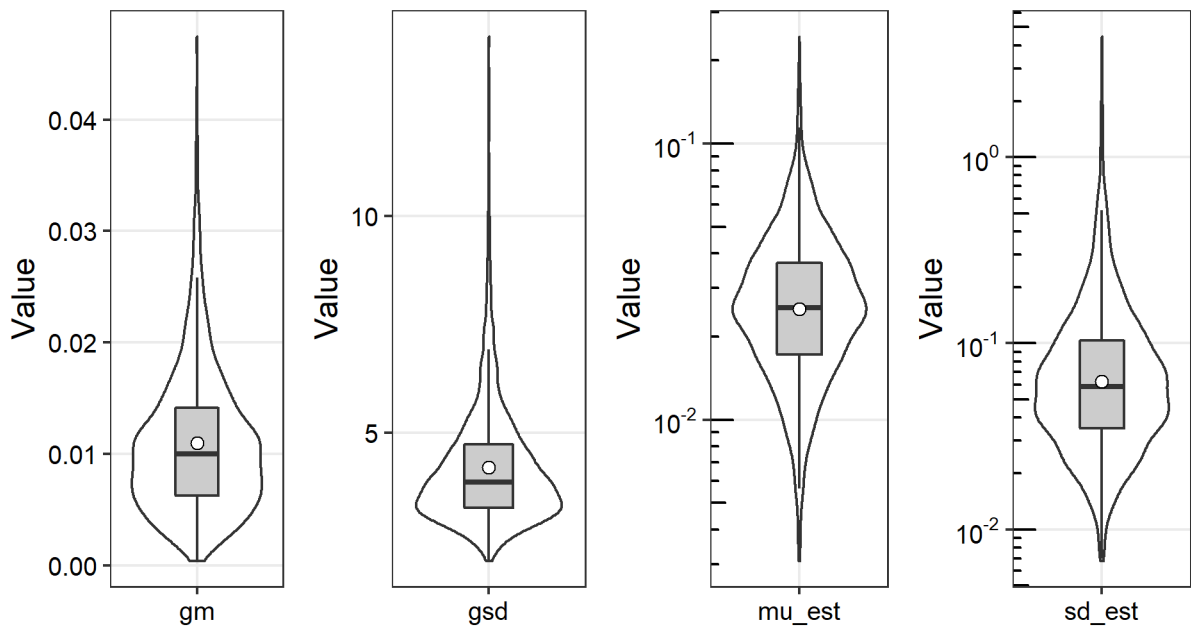


Fig. 3-15 ベイズ推定によるフルアジホップブチルの1日摂取量推定値に関わる事後予測分布

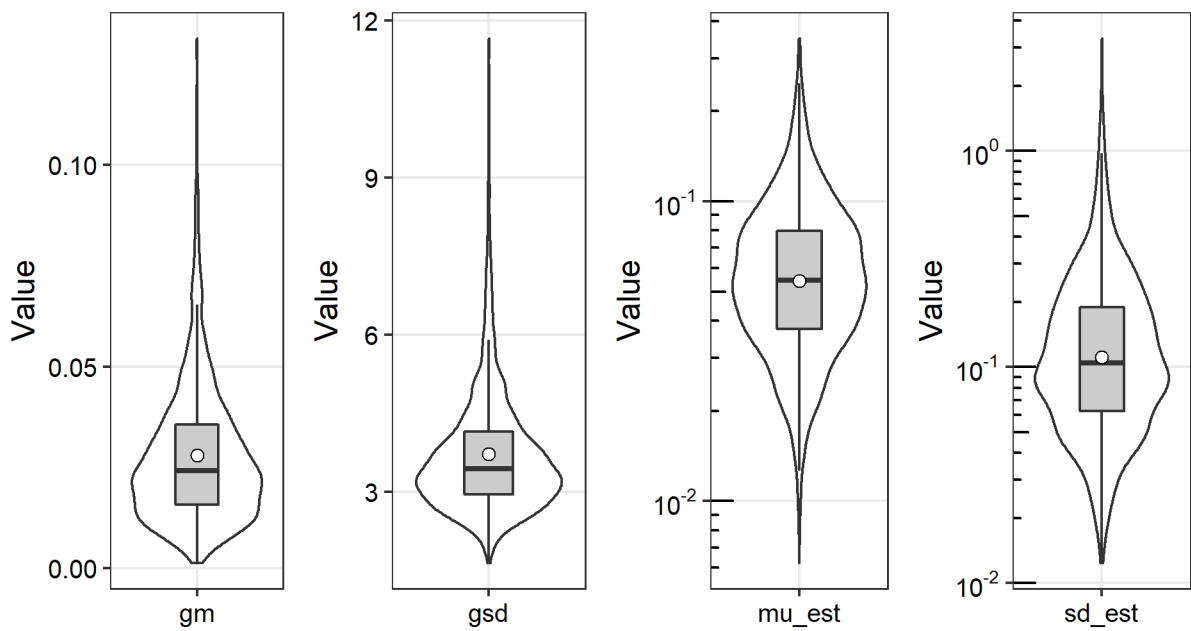


Fig. 3-16 ベイズ推定によるフルアジホップブチル代謝物の1日摂取量推定値に関わる事後予測分布

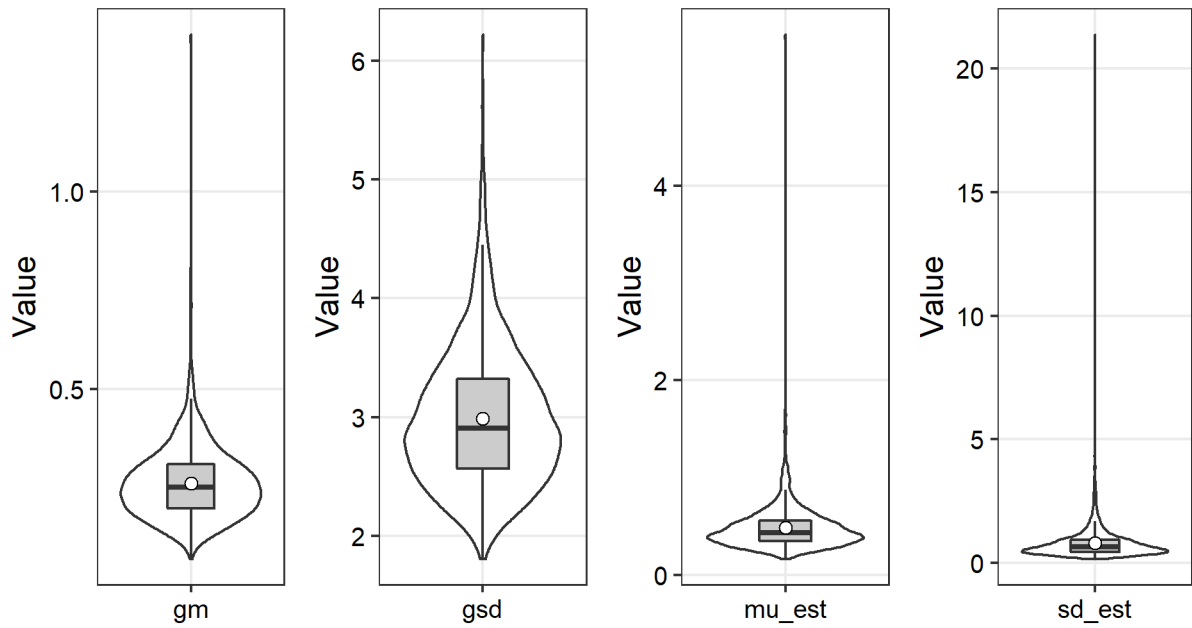


Fig. 3-17 バイズ推定によるフルベンジアミドの1日摂取量推定値に関わる事後予測分布

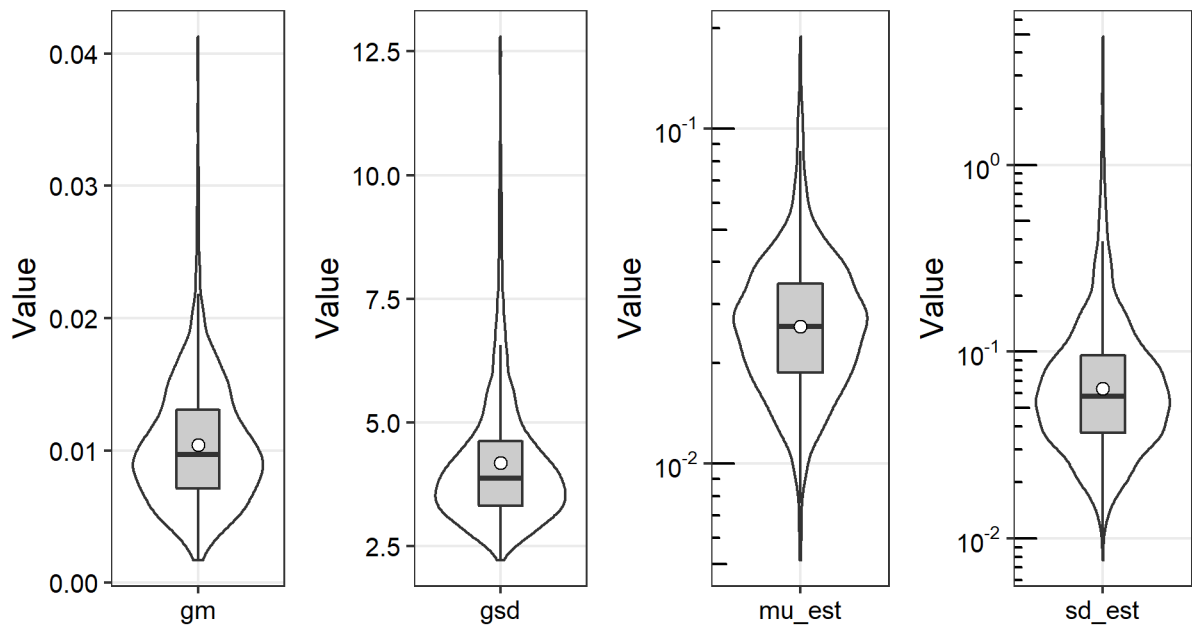


Fig. 3-18 バイズ推定によるヘキサジノンの1日摂取量推定値に関わる事後予測分布

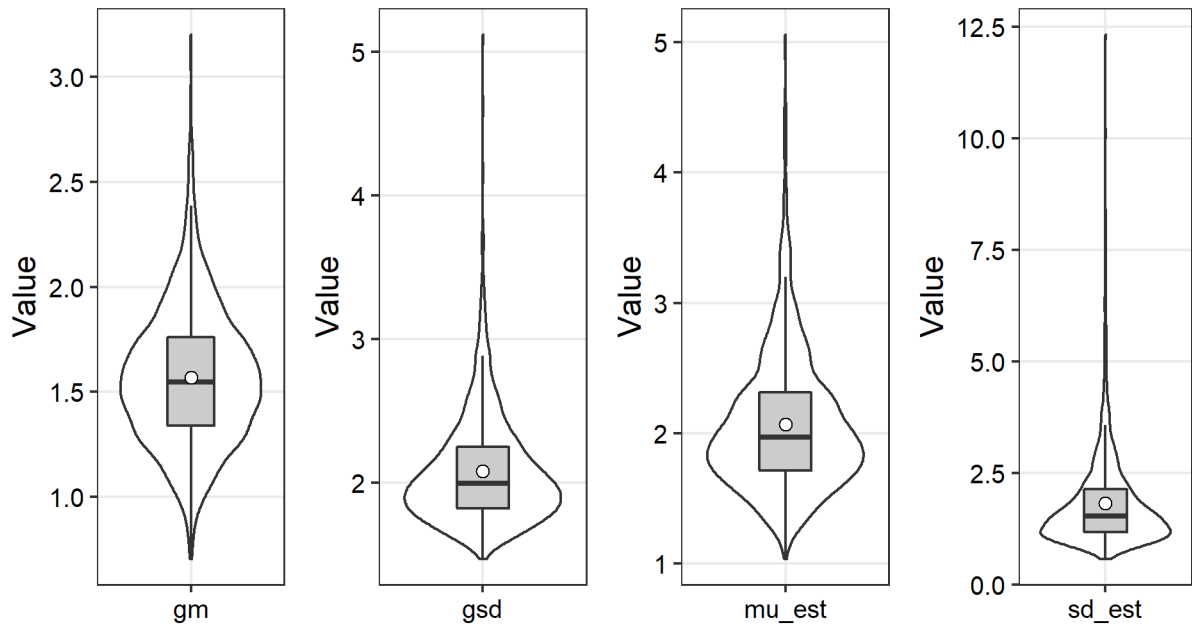


Fig. 3-19 バイズ推定によるボスカリドの1日摂取量推定値に関わる事後予測分布

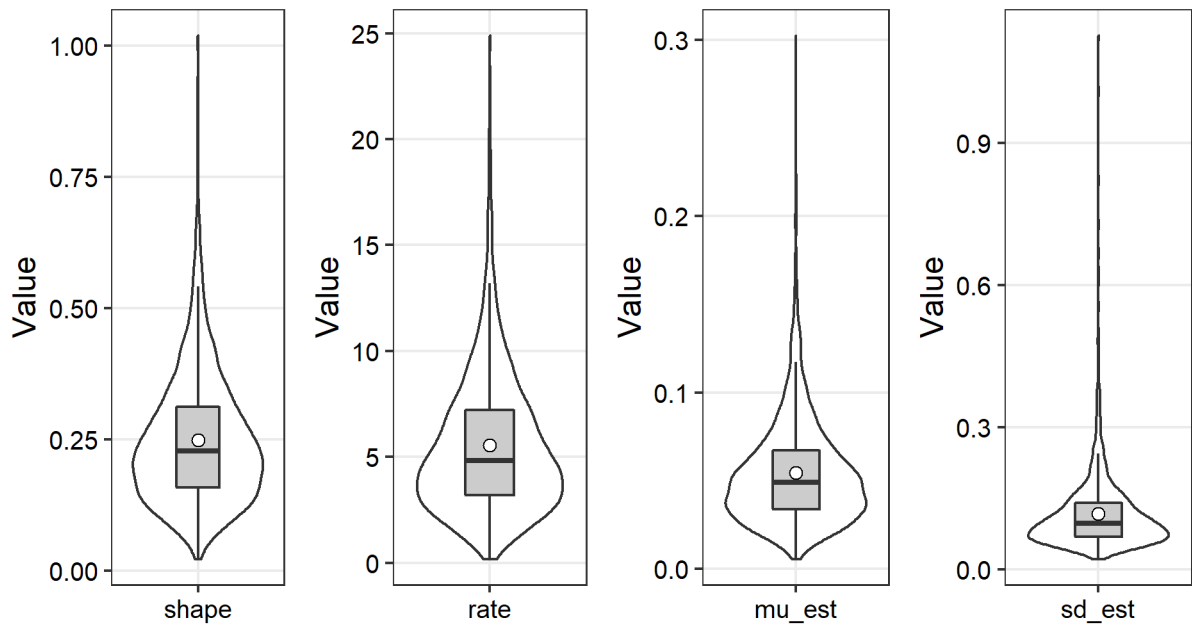


Fig. 3-20 バイズ推定によるメタミドホスの1日摂取量推定値に関わる事後予測分布

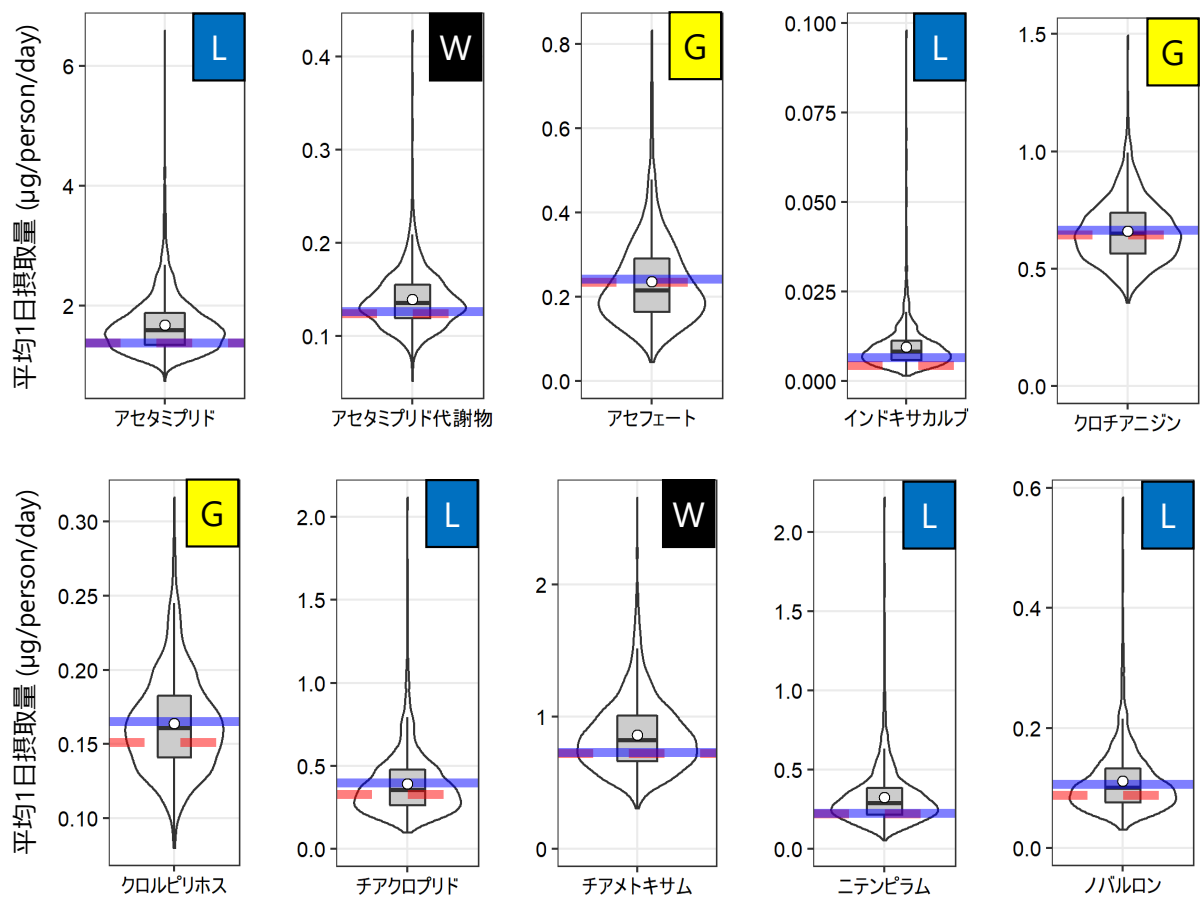


Fig. 4 不検出例を含むデータにおける平均農薬 1 日摂取量の推定

事後予測分布の確率密度分布 (バイオリン), 平均値 (白丸), 黒い横棒 (中央値), 四分位範囲 (灰色箱). 赤破線は RL/5 法による推定値, 青横線は最尤推定値, G はガンマ分布, L は対数正規分布, W はワイブル分布を示す.

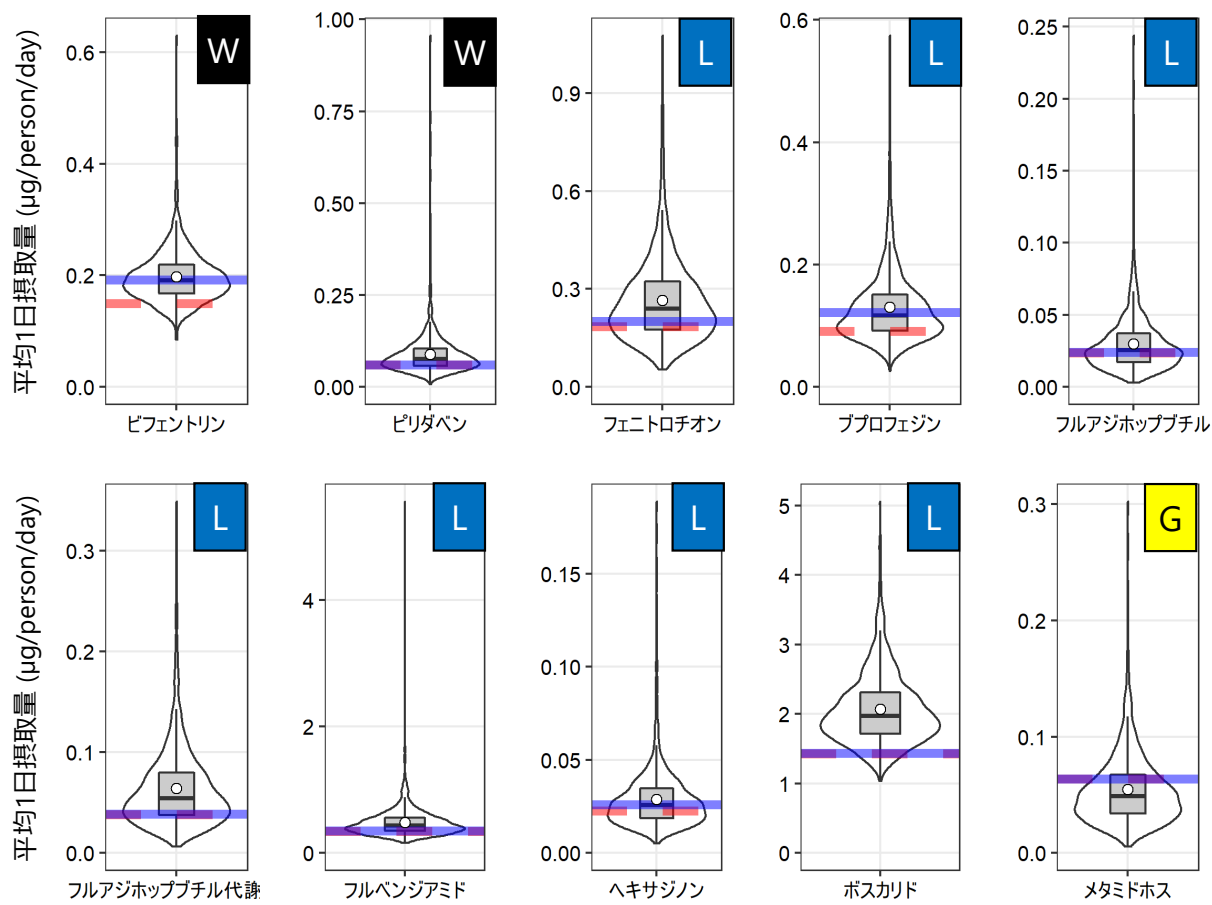


Fig. 4 続き

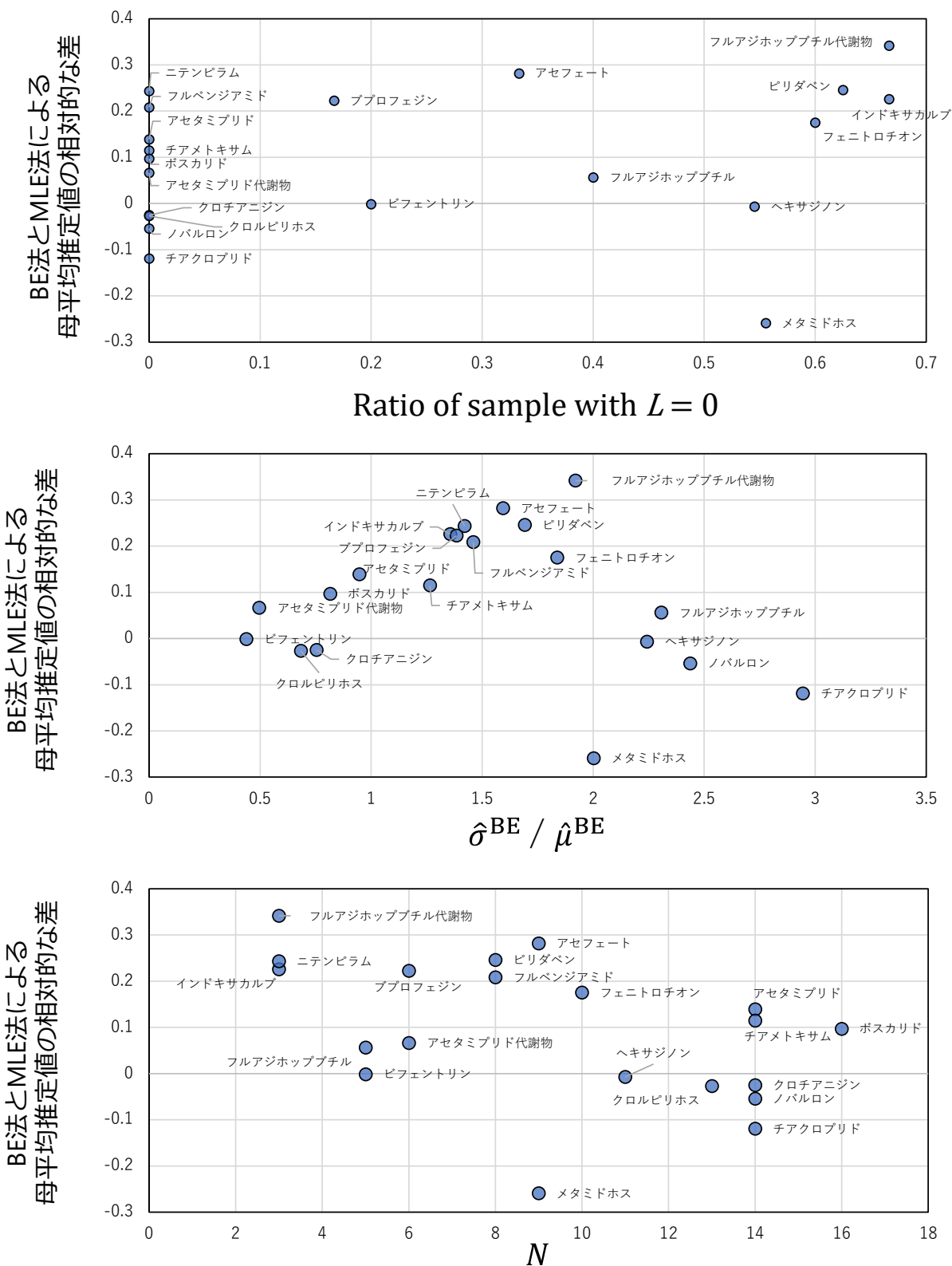


Fig. 5 ベイズ法と最尤推定法の推定結果の際に影響を与える要因

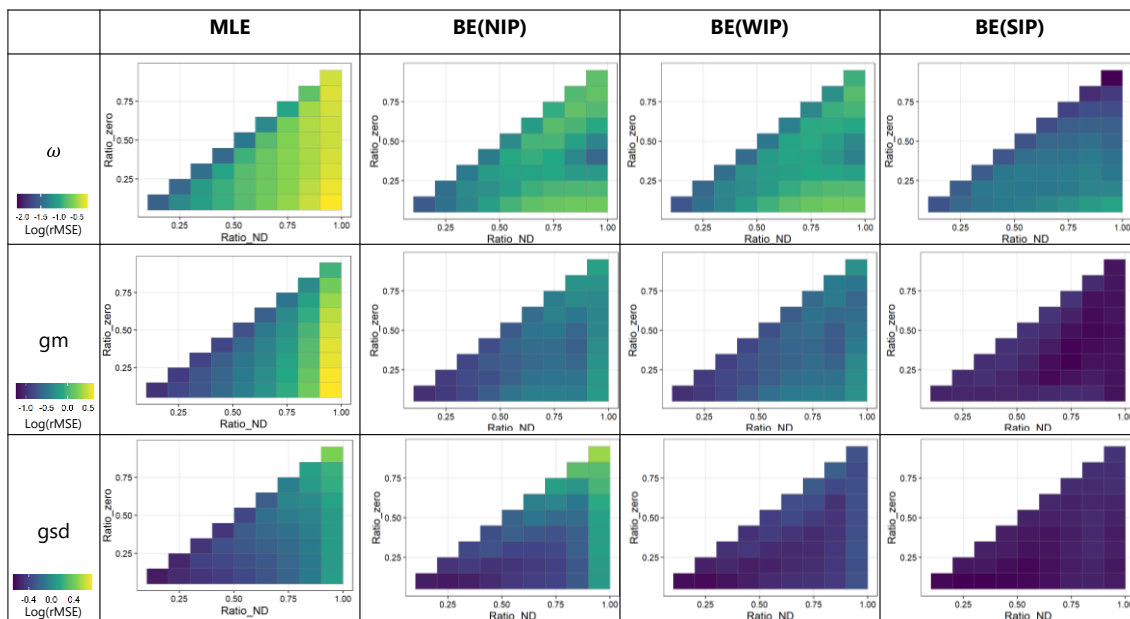


Fig. 6 ゼロ過剰対数正規分布に対する推定方法間の真度の比較

MLE: maximum likelihood estimation, BE: Bayesian estimation, NIP: no-informative prior, WIP: weak informative prior, SPI: specific informative prior.

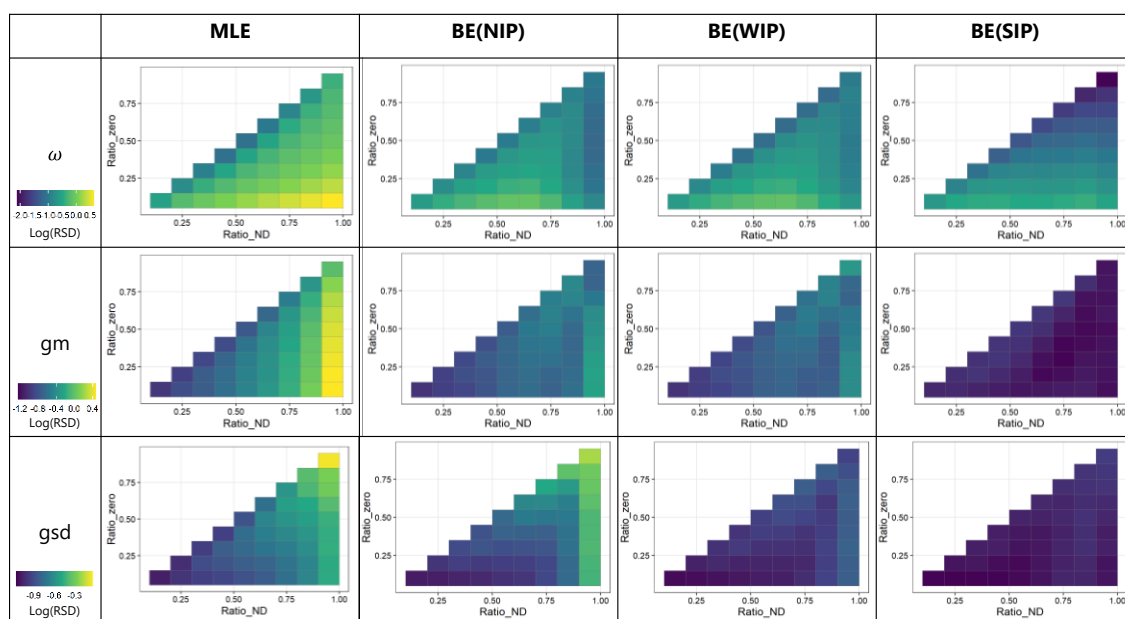


Fig. 7 ゼロ過剰対数正規分布に対する推定方法間の精度の比較

MLE: maximum likelihood estimation, BE: Bayesian estimation, NIP: no-informative prior, WIP: weak informative prior, SPI: specific informative prior.

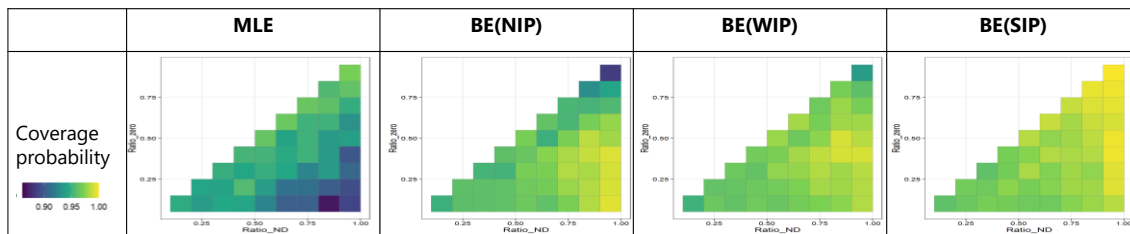


Fig. 8 ゼロ過剰対数正規分布の期待値に対する推定方法間の妥当性の比較

MLE: maximum likelihood estimation, BE: Bayesian estimation, NIP: no-informative prior, WIP: weak informative prior, SPI: specific informative prior

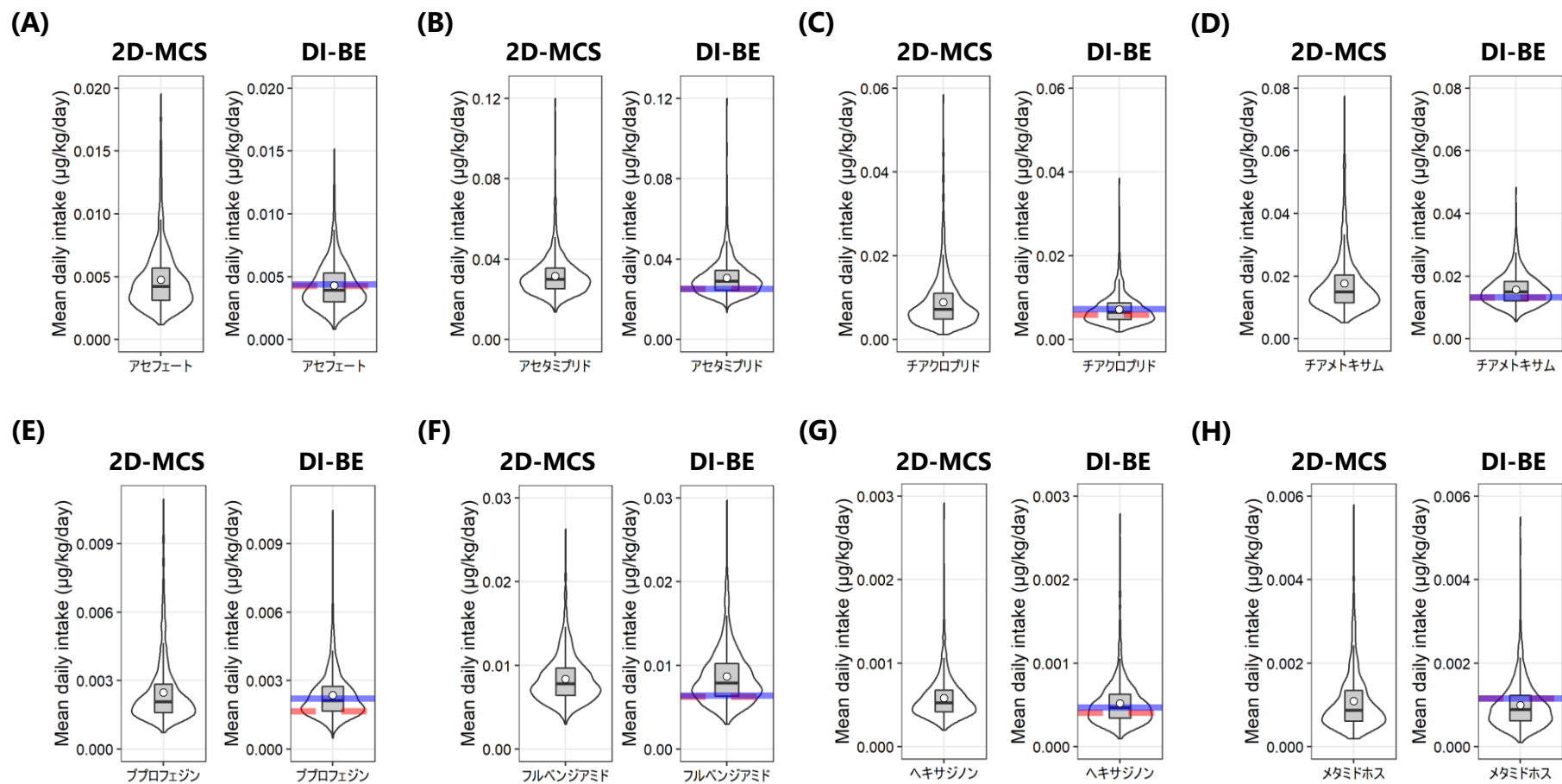


Fig. 9 平均農薬 1 日摂取量の推定値の推定方法間の違い

(A) アセフェート, (B) アセタミプリド, (C) チアクロプリド, (D) チアメトキサム, (E) プロフェジン, (F) フルベンジアミド, (G) ヘキサジノン, (H) メタミドホス。2D-MCS は濃度および喫食量ともに分布を推定した結果、DI-BE は農薬摂取量のデータをベイズ推定した結果、事後予測分布の確率密度分布 (バイオリン), 平均値 (白丸), 黒い横棒 (中央値), 四分位範囲 (灰色箱)。赤破線は RL/5 法による推定値, 青横線は最尤推定値