

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

令和元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) 確率論的な摂取量推定を行うため、喫食量の確率密度分布の推定、2) 未検出 (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。

1日農薬摂取量にはNDのレコードが含まれる。従来は、NDとなったレコードには定量下限値 (LOQ) の0.2倍を代入して解析を行ってきたが、0.2倍には科学的根拠はない。そこで、より統計的に得られたデータは1) 定量下限値 (LOQ) 以上、2) 検出限界値 (LOD) 以上LOQ未満、3) LOD未満の三つに分け、2)と3)については区間累積確率を用いて、確率密度分布のパラメーターおよび理論平均値と理論標準偏差をベイズ法および最尤推定法を用いて推定した。確率密度分布には、対数正規分布・ガンマ分布・ワイブル分布を仮定した。NDにLOQの0.2倍の値を代入して推定した1日農薬摂取量の平均値と比較すると、BE法による推定値の方が低かったのは、クロチアニジンのみであった。また、BE法によるクロチアニジン、チアクロプリド、ノバルロンの1日農薬摂取量の平均値は、MLE法よりも低かった。また、インドキサカルブ、クローピリホス、ビフェントリンの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

圏内に収まっていた。ベイズ推定と最尤推定の推定値が異なる農薬は、下限値が0のデータの割合が多く、相対標準偏差が大きい傾向が認められた。最尤推定法は、サンプルサイズが小さく、歪度が大きいデータが不得意であることが知られている。この点を踏まえると、本研究の結果は、ベイズ推定の方が最尤推定よりも妥当な推定結果であることを示唆していると考えられた。

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 (truncated) られた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はわかってる点である。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、昨年度および今年度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法との相違について評価した。

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。

$$L(\theta|Y) = \prod_{i=1}^N f(Y_i|\theta)$$

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

て、尤度関数次のように求めることが出来る。

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

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。

最尤推定 (MLE) 法は、尤度が最大となるパラメーターを探索する

2 ベイズ推定

尤度を基にした計算方法として、最尤推定 (MLE) とベイズ推定 (BE) 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。

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

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、

$$f_{\text{post}}(\theta|Y) \propto L(Y|\theta)f_{\text{pri}}(\theta)$$

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け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。

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

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。

$$f(Y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} Y^{\alpha-1} e^{-\beta Y}$$

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。

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

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。

重金属などは拡張元素普存説のように全ての試料には全ての元素が含まれていると考えるのが自然であり、そのため0以上かつ0に漸近しない確率密度分布となる対数正規分布を仮定することが一般的である。一方で、農薬の場合には無農薬の作物もあれば、農薬適用外の食品もあるため、分析化学的に0であることを証明することは難しいものの、統計科学的に0に漸近するような確率密度分布の適用を検討すべきと考えられる。 $\alpha < 1$ のガンマ分布および $m < 1$ ワイブル分布は、確率密度曲線が0に漸近する。無農薬の作物もあれば、農薬適用外の食品もあるため、農薬摂取量は0に漸近する分布となることもありうる。そのため、このような分布になることを想定して、ガンマ分布とワイブル分布についても検討を行った。

B. 試料と方法

1 データ

2019年および2020年に本研究班で行ったトータルダイエツトスタディー調査による、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。

2 方法

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。

C & 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、全て未

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、パラメーターが発散する、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值及び上限値を用いて推定を行うこととした。

2 未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以上 [検出限界値 (Limit of detection (LOD)), 定量下限値 (Limit of Quantification (LOQ))] 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。

MCMC の収束判断に関する例としてアセタミプリドの解析例を Fig. 2 に示す。iterations=1000, warm-up=500, chain=4, thinning=1 の条件で parameter block で宣言した 2 個のパラメーター (GM と GSD) と対数事後確率の合計値 (lp_ (log posterior の略)) が収束した (Fig. 2A)。これらのパラメーターと摂取量の平均値 (mu_est) と標準偏差 (sd_est) について、MCMC サンプルングの結果得られた 2000 個ずつの結果の概要を Fig. 2B に示す。これら 5 つの値以外にも各データの

尤度や予測値などの全ての計算結果において、一般的に収束の目安となる chain 内分散と chain 間分散の割合である \hat{R} が 1.1 以下を満たしていた (Fig. 2C)。また、有効サンプル数の割合や相対的モンテカルロ標準偏差も一般的な基準 (それぞれ 0.1 以上, 0.1 以下) を満たしていた。以上のことから、chain 間での推測値の解離や自己相関の影響は少なく、各 chain は異なる初期値から出発し最終的には同様の値に到着したことを示し、全ての計算値は収束したと考えられた。全ての農薬について収束の判断を確認し、適宜 iteration 数を増やすなど行った。

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 8 農薬で、ガンマ分布が 1 農薬で、ワイブル分布が 11 農薬で最も適していると判断された。一方 MLE 法では、対数正規分布が 7 農薬で、ガンマ分布が 6 農薬で、ワイブル分布が 7 農薬で最も適していると判断された。11 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。

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 ($\hat{\mu}^{MLE}$) と理論標準偏差 ($\hat{\sigma}^{MLE}$) の値を表記した。MLE 法では、対数正規分布が 7 農薬で、ガンマ分布が 6 農薬で、ワイブル分布が 7 農薬で最も適していると判断された。

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、アセフェート・チアメトキサム・ピリダベン・フルアジホップブチル・フルアジホップブチル代謝物・メタミドホスは、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と比較すると、BE 法による推定値の方が低かったのは、クロチアニジンのみであった。MLE 法よりも BE 法による推定値の方が低かったのは、クロチアニジン・チアクロプリド・ノバルロンであった。インドキサカルブ・クロルピリホス・ビフェントリンの 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、 $(\hat{\mu}^{BE} - \hat{\mu}^{MLE})/\hat{\mu}^{BE}$ について解析を行った (Fig. 5)。の傾向が認められた。 $L=0$ 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、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 (正確性)、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出来るか (頑健性) について、乱数を繰り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。

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、未検出 (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、ベイズモデルを用いて推定したところ、クロチアニジン・ブプロフェジン・フルアジホップブチル・フルベンジアミド・メタミドホスの推定 1 日摂取量は、代入法と比較して、BE 法による推定値の方が低かった。一方で、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、BE 法と MLE 法との推定値の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が知られていることから、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。

F. 研究発表

1. 論文発表

1) Suzuki Y, Tanaka N, Akiyama H.; Attempt of Bayesian Estimation from Left-censored Data Using the Markov Chain Monte Carlo Method: Exploring Cr(VI)

Concentrations in Mineral Water Products. Food Safety. 8(4):67-89 (2020).

2. 学会発表

1) 鈴木美成, 穉山浩; トータルダイエツトスタディによる農薬摂取量の推定におけるベイズモデルを用いた精緻化, 日本食品衛生学会第 116 回学術講演会(2020.11-12).

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 2019–2020 年に調査した農薬 1 日摂取量 ($\mu\text{g day}^{-1}$) 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	1.06 (0.89–1.27)	2.29 (2.06–2.61)	1.51 (1.25–1.89)	1.48 (1.07–2.25)
アセタミプリド代謝物	ワイブル分布	0.142 (0.119–0.170)	1.73 (1.33–2.24)	0.129 (0.108–0.156)	0.075 (0.054–0.105)
アセフェート	ワイブル分布	0.116 (0.072–0.171)	4.76 (3.96–5.72)	0.382 (0.227–0.619)	1.21 (0.584–2.52)
インドキサカルブ	ワイブル分布	0.010 (0.007–0.013)	1.47 (1.07–2.11)	0.0095 (0.0071–0.0124)	0.0061 (0.0039–0.0099)
クロチアニジン	ガンマ分布	2.20 (1.66–2.90)	1.65 (1.29–2.08)	0.752 (0.643–0.886)	0.578 (0.476–0.709)
クロルピリホス	ワイブル分布	0.255 (0.219–0.300)	1.58 (1.32–1.90)	0.231 (0.199–0.274)	0.145 (0.117–0.189)
チアクロプリド	対数正規分布	0.151 (0.111–0.201)	4.68 (3.99–5.57)	0.524 (0.381–0.712)	1.65 (1.00–2.84)
チアメトキサム	ワイブル分布	0.960 (0.741–1.23)	0.766 (0.661–0.890)	1.15 (0.911–1.46)	1.49 (1.12–2.06)
ニテンピラム	ワイブル分布	0.398 (0.290–0.519)	1.38 (1.03–1.80)	0.375 (0.287–0.490)	0.263 (0.188–0.401)
ノバルロン	対数正規分布	0.032 (0.024–0.043)	4.06 (3.86–5.70)	0.106 (0.075–0.148)	0.321 (0.183–0.579)
ビフェントリン	対数正規分布	0.221 (0.184–0.257)	1.54 (1.34–1.88)	0.247 (0.212–0.293)	0.107 (0.067–0.185)
ピリダベン	ワイブル分布	0.055 (0.035–0.084)	0.606 (0.491–0.746)	0.091 (0.062–0.126)	0.154 (0.097–0.246)
フェニトロチオン	対数正規分布	0.112 (0.077–0.167)	3.67 (2.98–4.68)	0.284 (0.192–0.415)	0.597 (0.315–1.19)
ブプロフェジン	対数正規分布	0.088 (0.065–0.113)	2.53 (2.15–3.08)	0.140 (0.106–0.185)	0.161 (0.104–0.262)
フルアジホップブチル	ワイブル分布	0.027 (0.018–0.037)	0.695 (0.557–0.874)	0.036 (0.026–0.049)	0.050 (0.033–0.081)
フルアジホップブチル代謝物	ワイブル分布	0.057 (0.038–0.080)	0.964 (0.717–1.24)	0.063 (0.046–0.084)	0.064 (0.042–0.101)
フルベンジアミド	ワイブル分布	0.355 (0.279–0.449)	1.24 (1.00–1.56)	0.338 (0.273–0.434)	0.264 (0.200–0.378)
ヘキサジノン	対数正規分布	0.0062 (0.0039–0.0088)	3.84 (3.05–5.16)	0.016 (0.011–0.023)	0.037 (0.020–0.075)
ボスカリド	対数正規分布	1.55 (1.34–1.76)	2.00 (1.82–2.25)	1.97 (1.71–2.31)	1.53 (1.17–2.13)
メタミドホス	ワイブル分布	0.514 (0.427–0.623)	0.057 (0.035–0.084)	0.114 (0.080–0.158)	0.245 (0.150–0.382)

* 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 2019–2020 年に調査した農薬 1 日摂取量 ($\mu\text{g day}^{-1}$) 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	1.01	2.21	1.38	2.08
アセタミプリド代謝物	ワイブル分布	0.130	2.36	0.116	0.052
アセフェート	ガンマ分布	1.27	0.341	0.268	0.459
インドキサカルブ	ワイブル分布	0.009	2.45	0.008	0.004
クロチアニジン	ワイブル分布	0.860	1.38	0.786	0.577
クロルピリホス	ワイブル分布	0.238	1.79	0.211	0.122
チアクロプリド	対数正規分布	0.139	5.55	0.605	3.27
チアメトキサム	ガンマ分布	0.596	0.596	1.00	1.295
ニテンピラム	ワイブル分布	0.346	1.66	0.309	0.192
ノバルロン	対数正規分布	0.029	5.05	0.107	0.522
ピフェントリン	対数正規分布	0.226	1.25	0.231	0.064
ピリダベン	ガンマ分布	2.58	0.192	0.074	0.170
フェニトロチオン	対数正規分布	0.065	5.20	0.252	1.27
ブプロフェジン	対数正規分布	0.080	2.53	0.122	0.235
フルアジホップブチル	ガンマ分布	12.2	0.347	0.028	0.048
フルアジホップブチル代謝物	ガンマ分布	6.99	0.336	0.048	0.083
フルベンジアミド	ワイブル分布	0.314	1.33	0.289	0.220
ヘキサジノン	ワイブル分布	0.005	0.432	0.015	0.041
ボスカリド	対数正規分布	1.48	1.89	1.81	1.97
メタミドホス	ガンマ分布	2.05	0.182	0.089	0.209

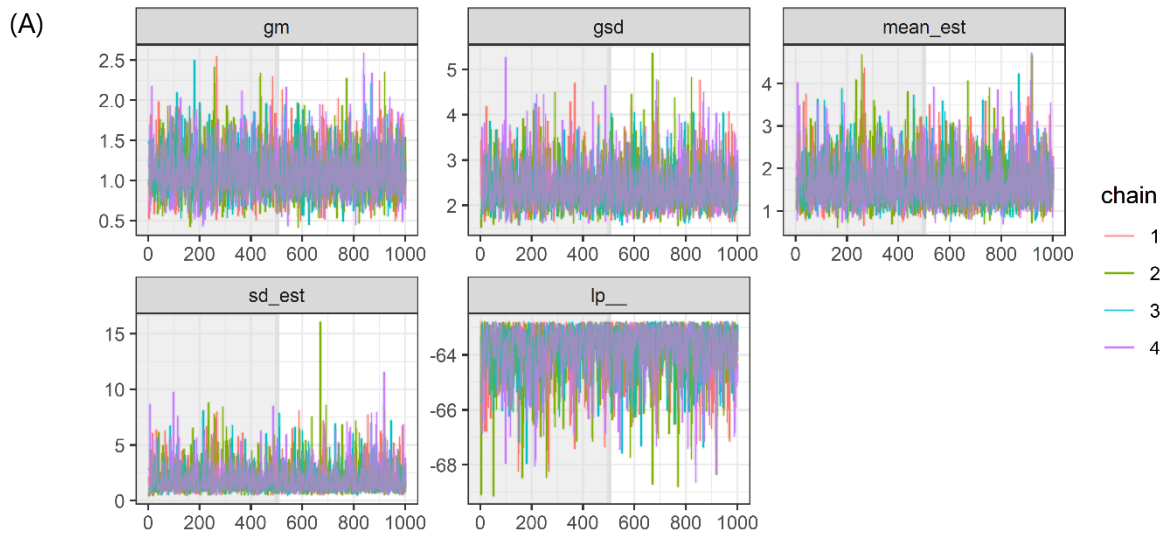
* Parameter 1,2 は、対数正規分布を仮定した場合には、それぞれパラメーター e^{μ} と e^{σ} を、ガンマ分布の場合はパラメーター β および α を、ワイブル分布の場合はパラメーター η および m を示す。全ての推定値は事後予測分布の中央値と四分位範囲を示す。

```

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	1.10	0.30	1.002	0.553	0.0301
gsd	2.39	0.48	1.003	0.534	0.0306
mu_est	1.64	0.55	1.003	0.560	0.0300
sd_est	1.87	1.26	1.004	0.587	0.0292
lp_	-63.79	0.973	1.001	0.364	0.0371

(C)

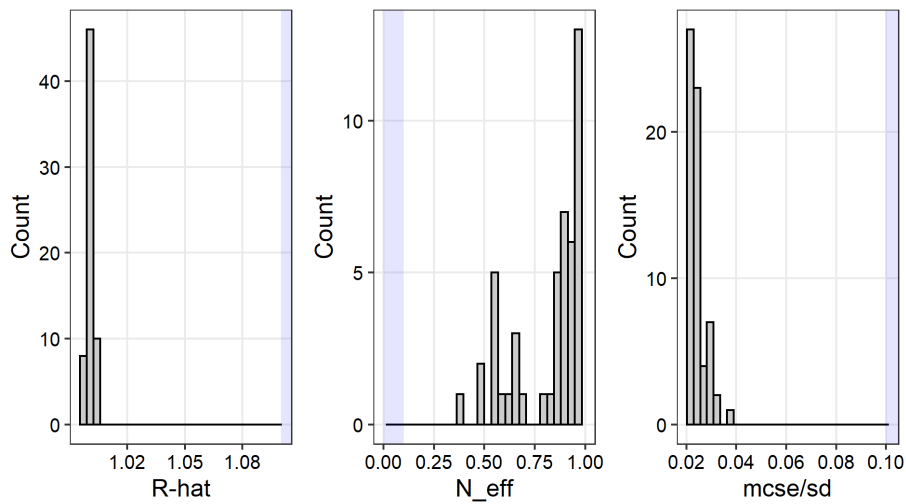


Fig. 2 アセタミプリドの1日摂取量推定におけるベイズ推定によるパラメーターの収束判断 (A) トレースプロット, (B) 推定したパラメーター等の記述統計量, (C) 収束判断指標 (\hat{R} , n_{eff}/N , mcse/sd) のヒストグラム

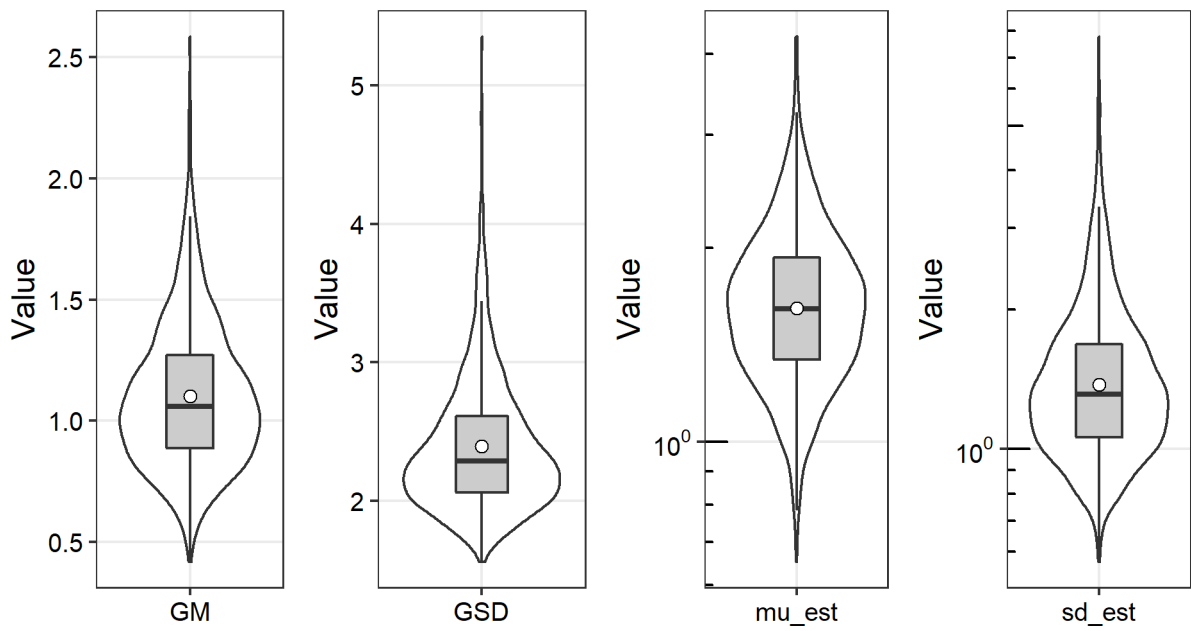


Fig. 3-1 ベイズ推定によるアセタミプリドの1日摂取量推定値に関わる事後予測分布

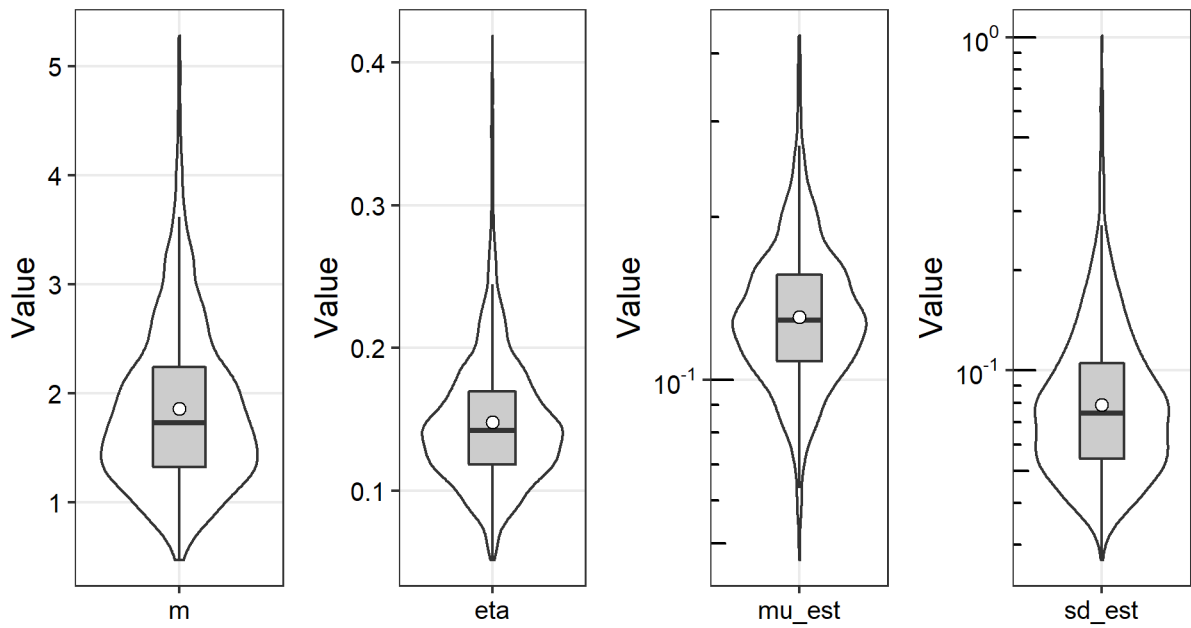


Fig. 3-2 ベイズ推定によるアセタミプリド代謝物の1日摂取量推定値に関わる事後予測分布

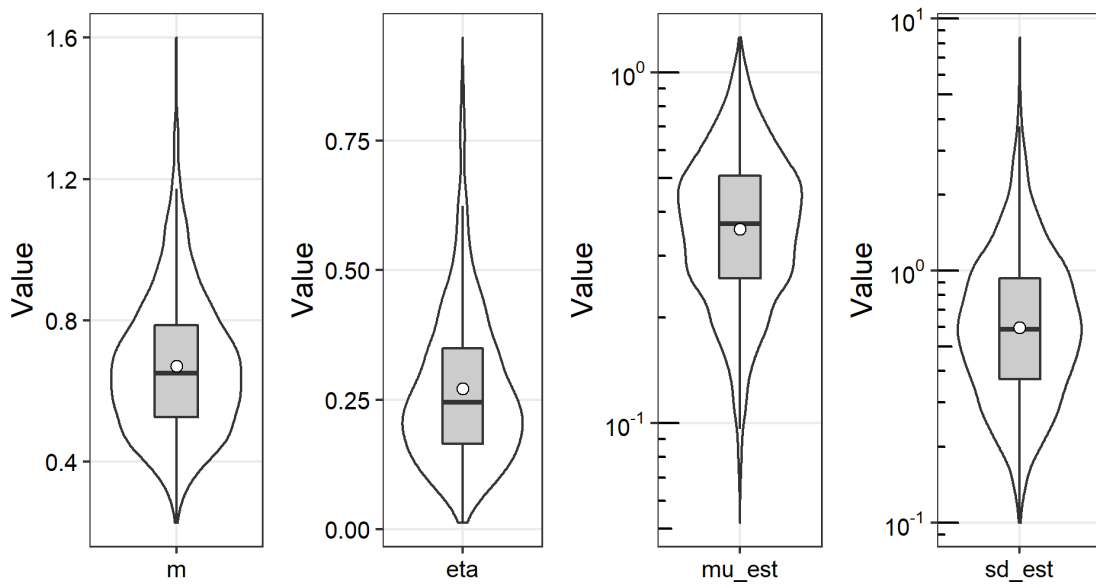


Fig. 3-3 ベイズ推定によるアセフェートの1日摂取量推定値に関わる事後予測分布

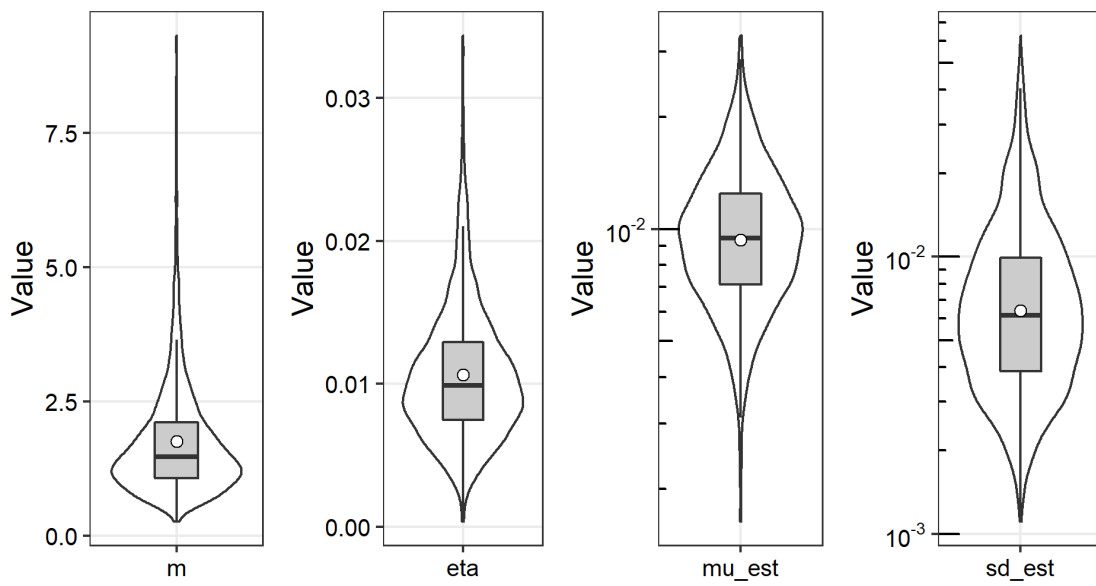


Fig. 3-4 ベイズ推定によるインドキサカルブの1日摂取量推定値に関わる事後予測分布

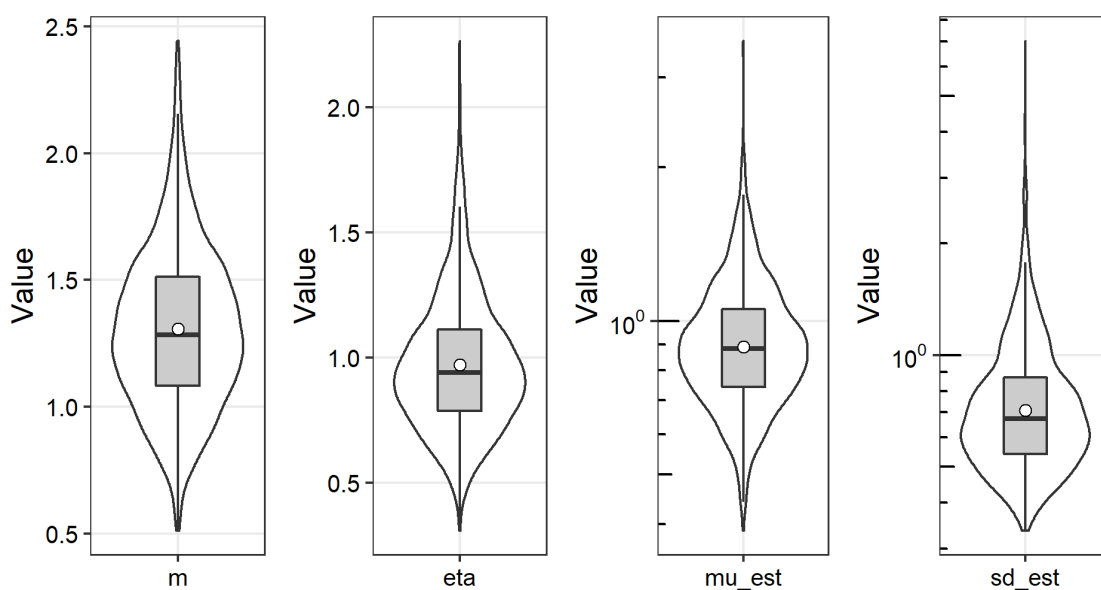


Fig. 3-5 ベイズ推定によるクロチアニジンの1日摂取量推定値に関わる事後予測分布

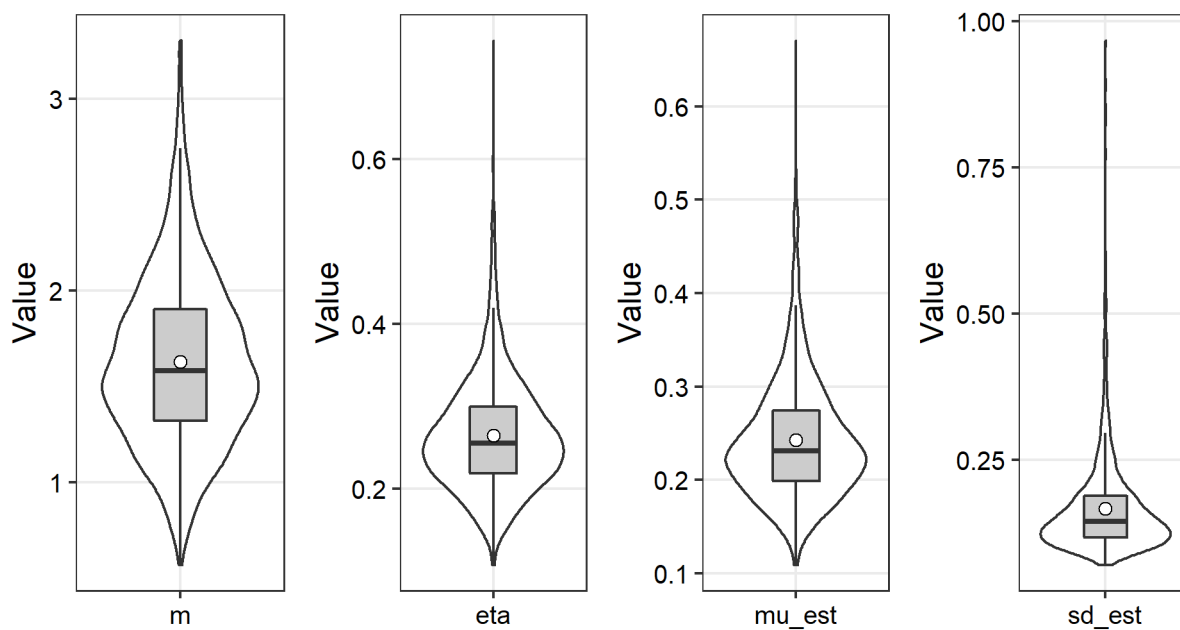


Fig. 3-6 ベイズ推定によるクロルピリホスの1日摂取量推定値に関わる事後予測分布

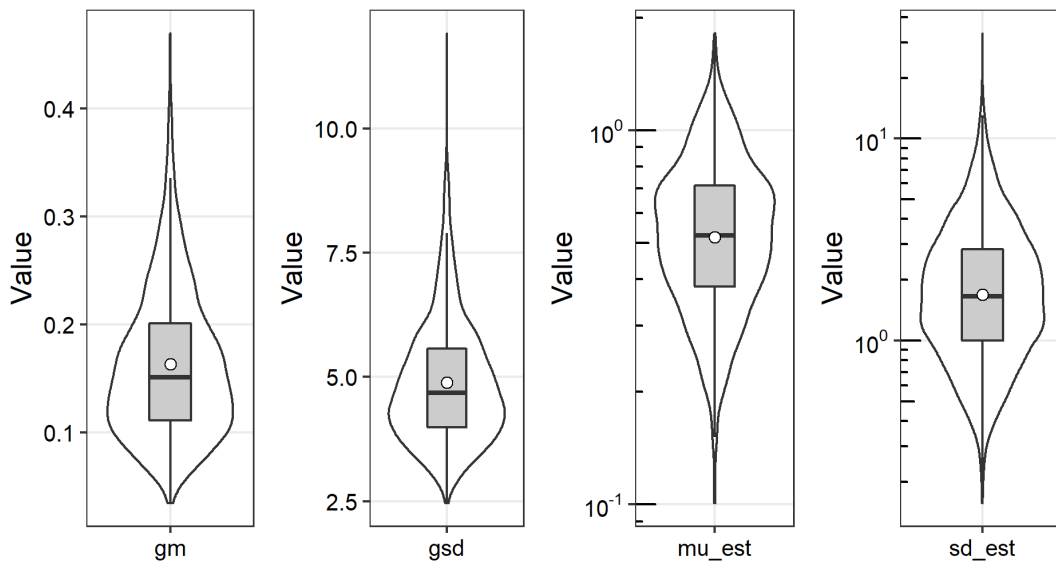


Fig. 3-7 ベイズ推定によるチアクロプリドの1日摂取量推定値に関わる事後予測分布

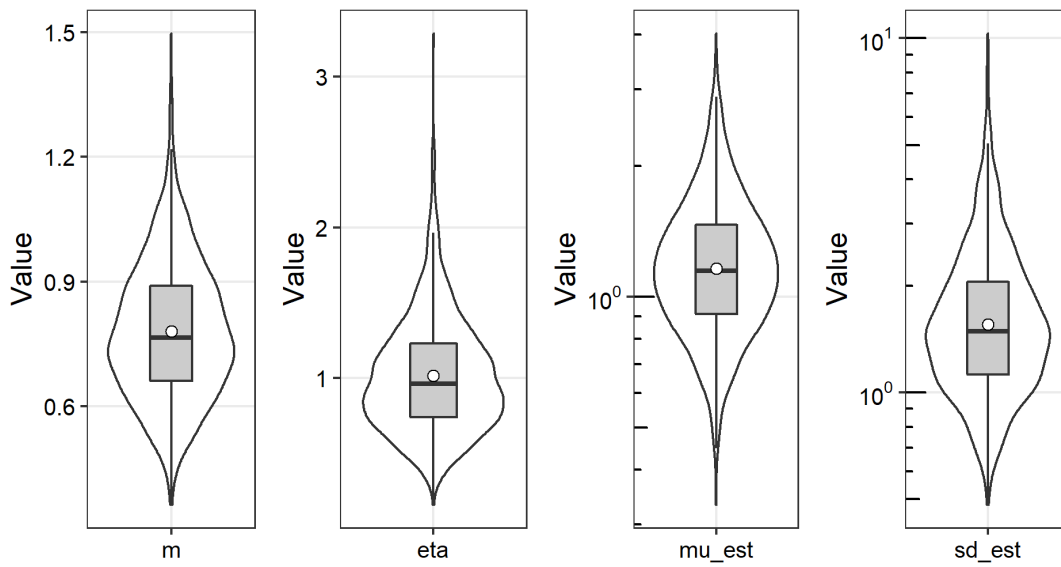


Fig. 3-8 ベイズ推定によるチアメトキサムの1日摂取量推定値に関わる事後予測分布

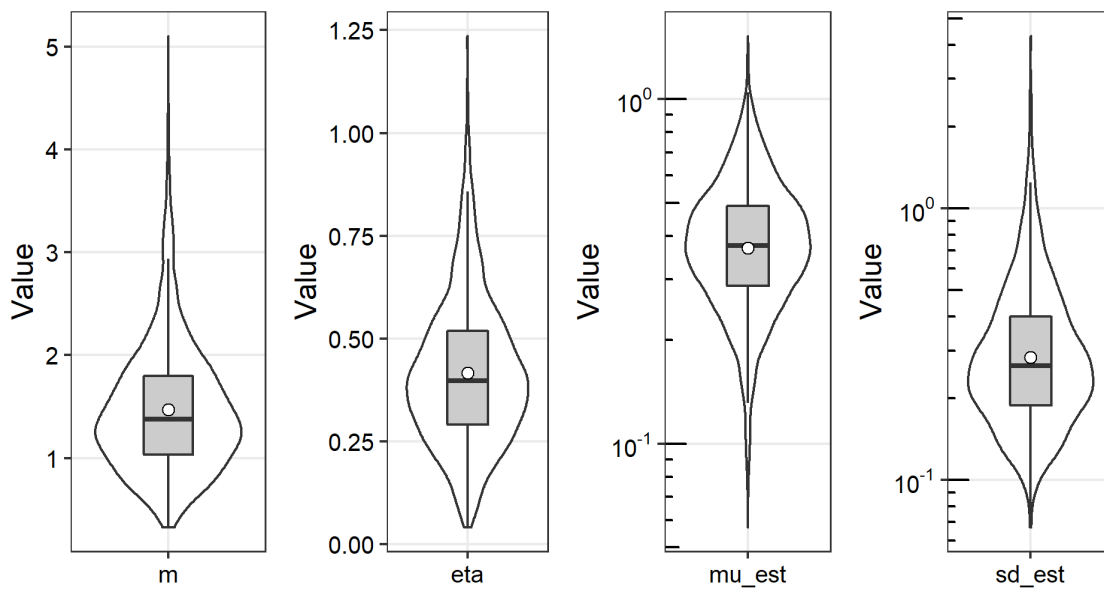


Fig. 3-9 ベイズ推定によるニテンピラムの1日摂取量推定値に関わる事後予測分布

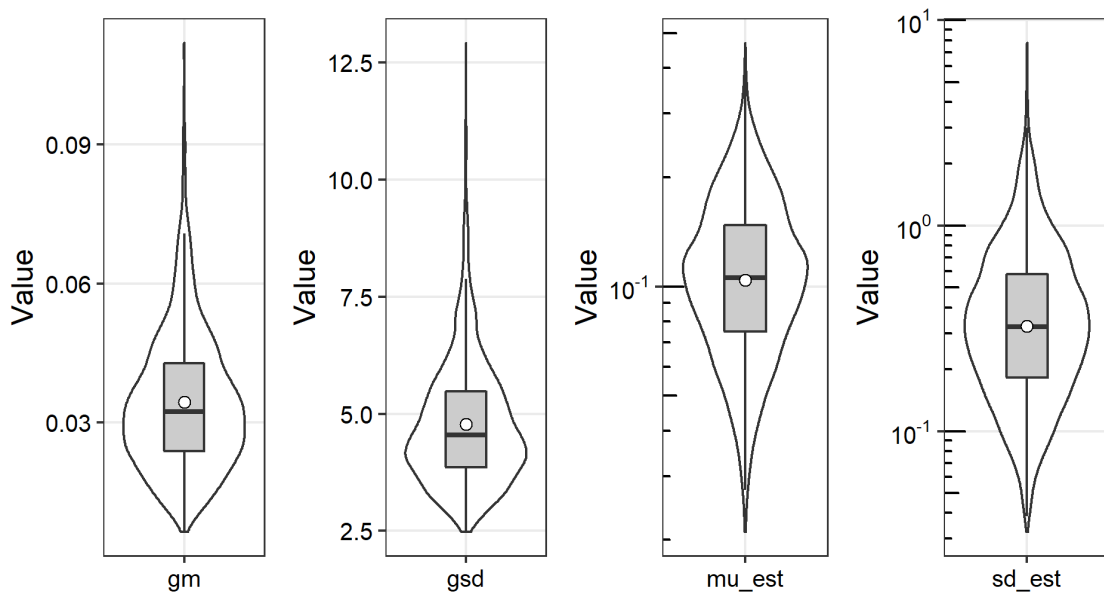


Fig. 3-10 ベイズ推定によるノバルロンの1日摂取量推定値に関わる事後予測分布

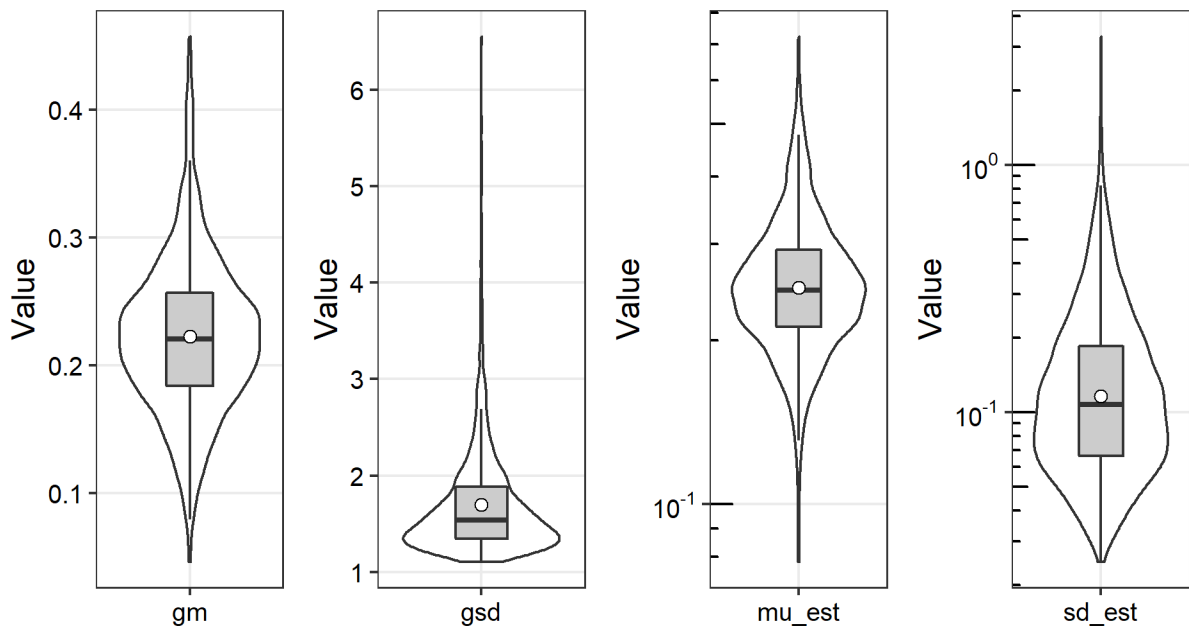


Fig. 3-11 バイズ推定によるピフェントリンの1日摂取量推定値に関わる事後予測分布

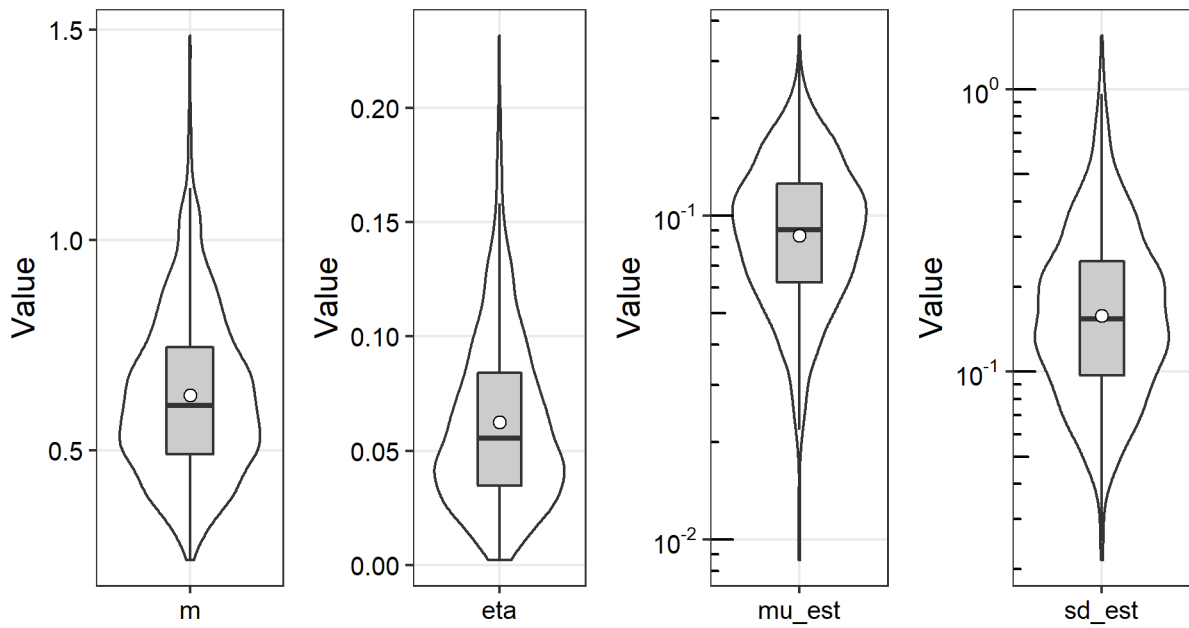


Fig. 3-12 バイズ推定によるピリダベン1日摂取量推定値に関わる事後予測分布

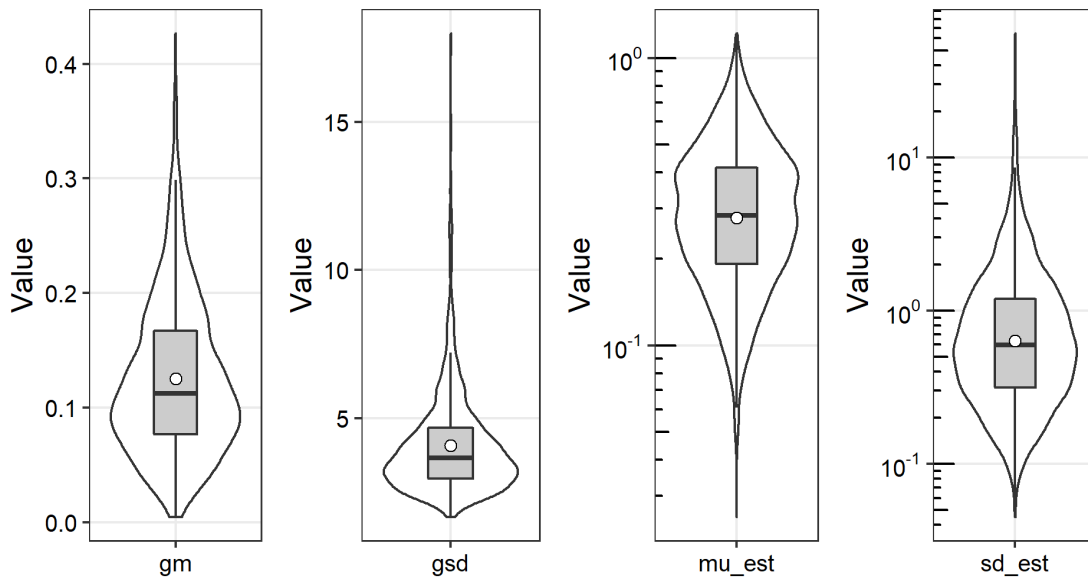


Fig. 3-13 ベイズ推定によるフェントロチオンの1日摂取量推定値に関わる事後予測分布

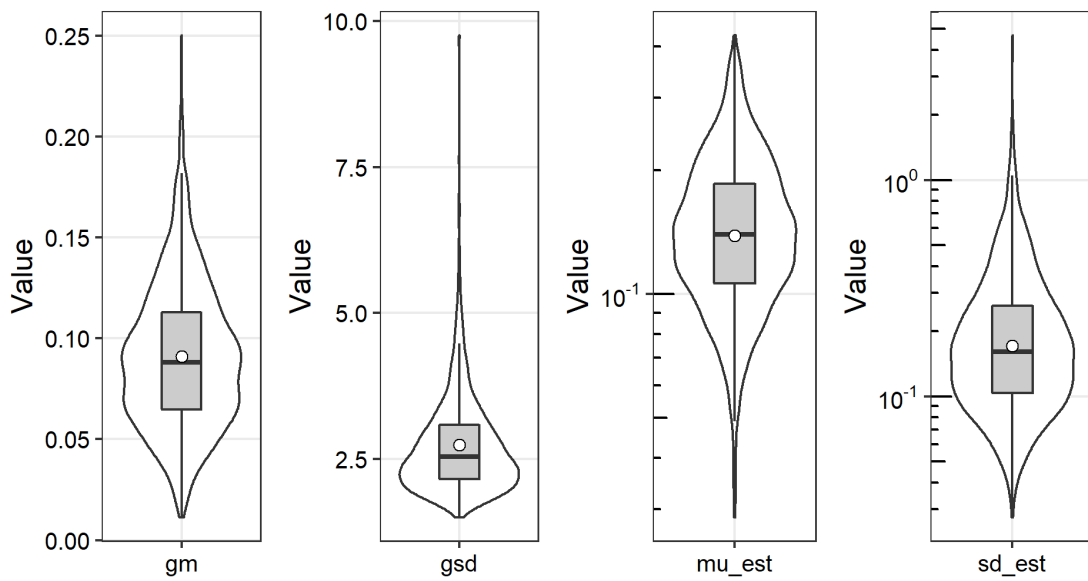


Fig. 3-14 ベイズ推定によるブプロフェジンの1日摂取量推定値に関わる事後予測分布

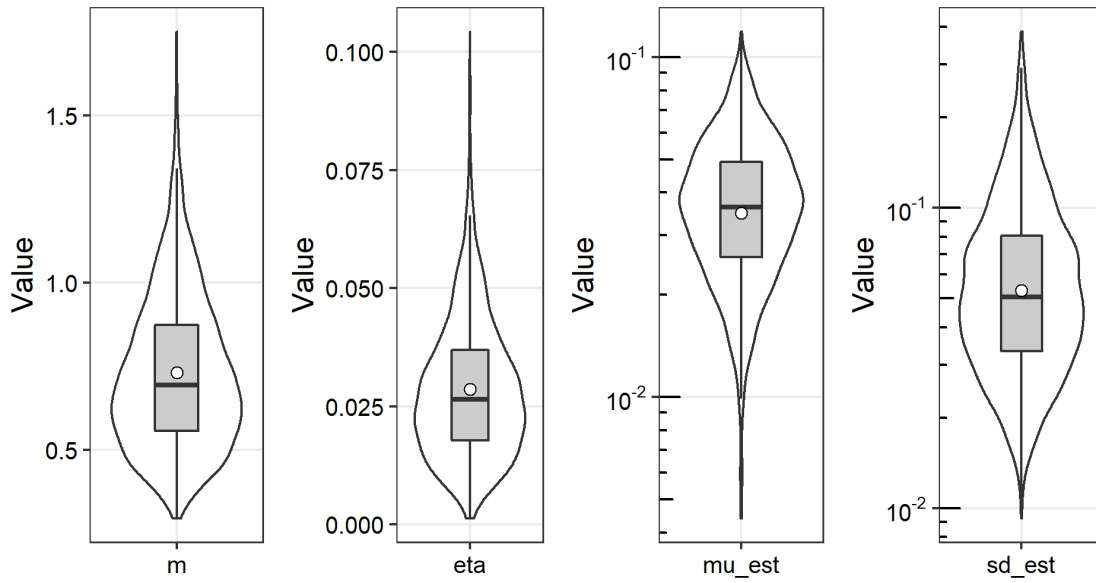


Fig. 3-15 バイズ推定によるフルアジホップブチルの1日摂取量推定値に関わる事後予測分布

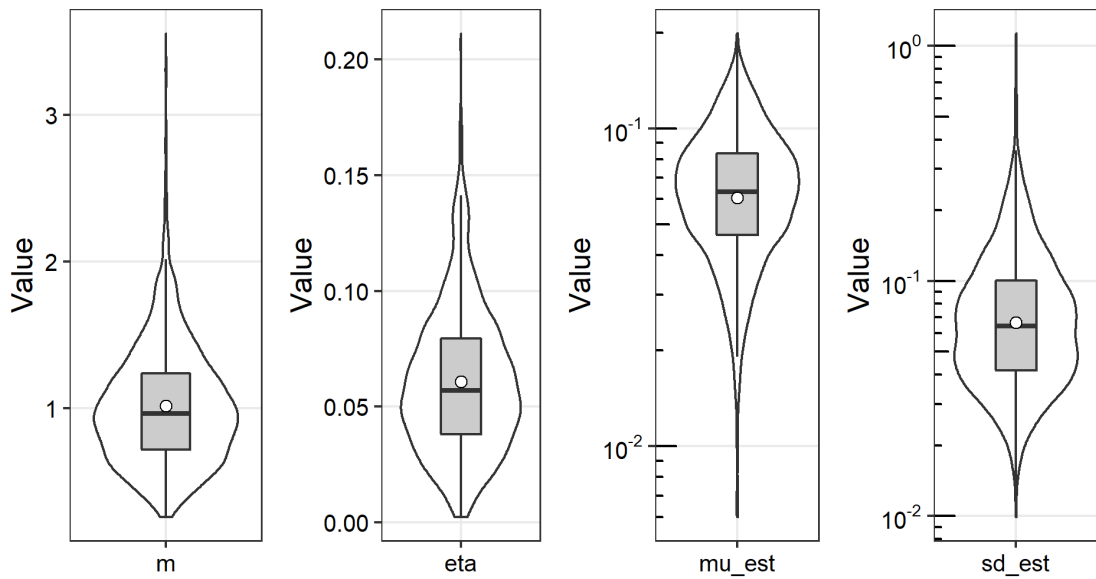


Fig. 3-16 バイズ推定によるフルアジホップブチル代謝物の1日摂取量推定値に関わる事後予測分布

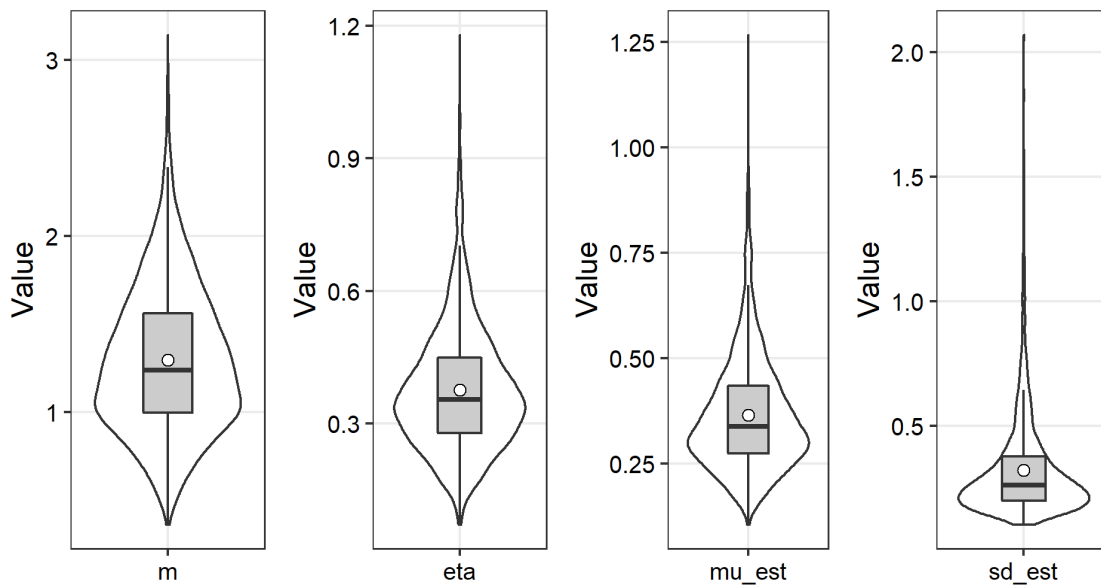


Fig. 3-17 バイズ推定によるフルベンジアミドの1日摂取量推定値に関わる事後予測分布

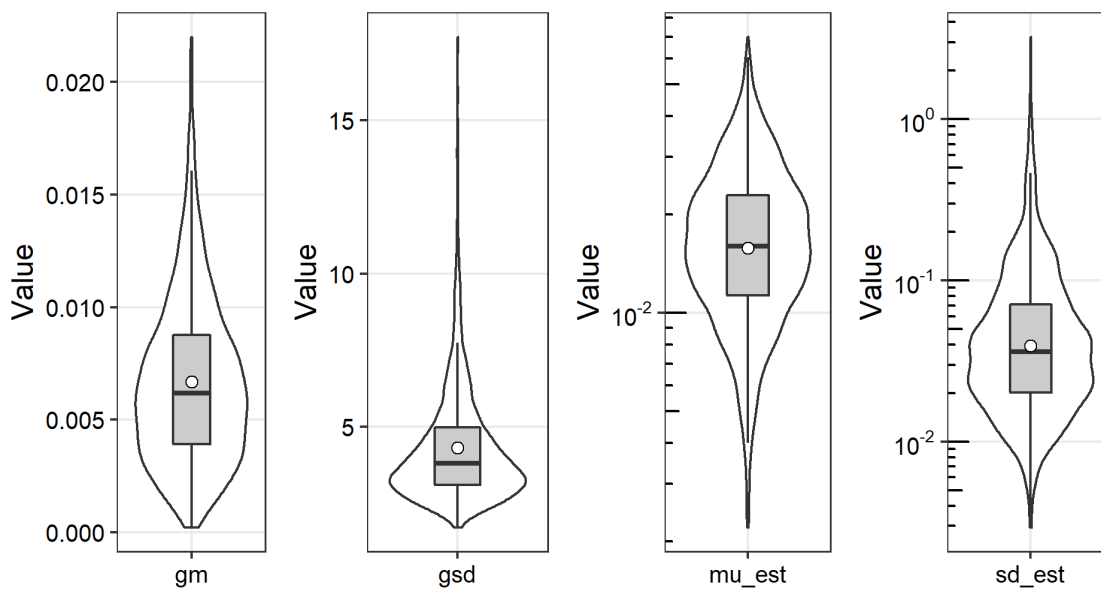


Fig. 3-18 バイズ推定によるヘキサジノンの1日摂取量推定値に関わる事後予測分布

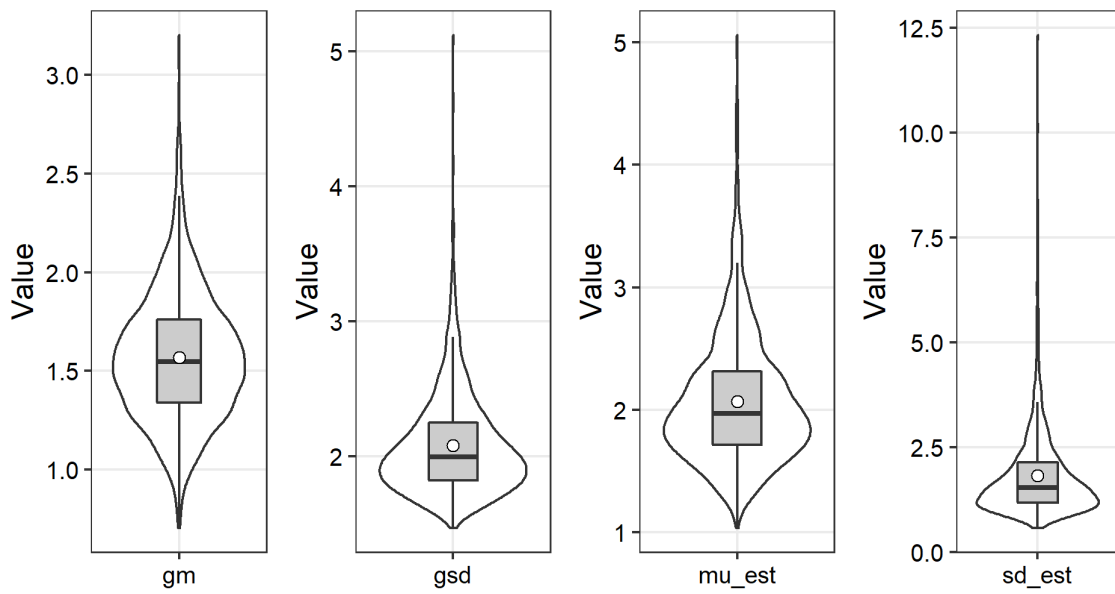


Fig. 3-19 バイズ推定によるボスカリドの1日摂取量推定値に関わる事後予測分布

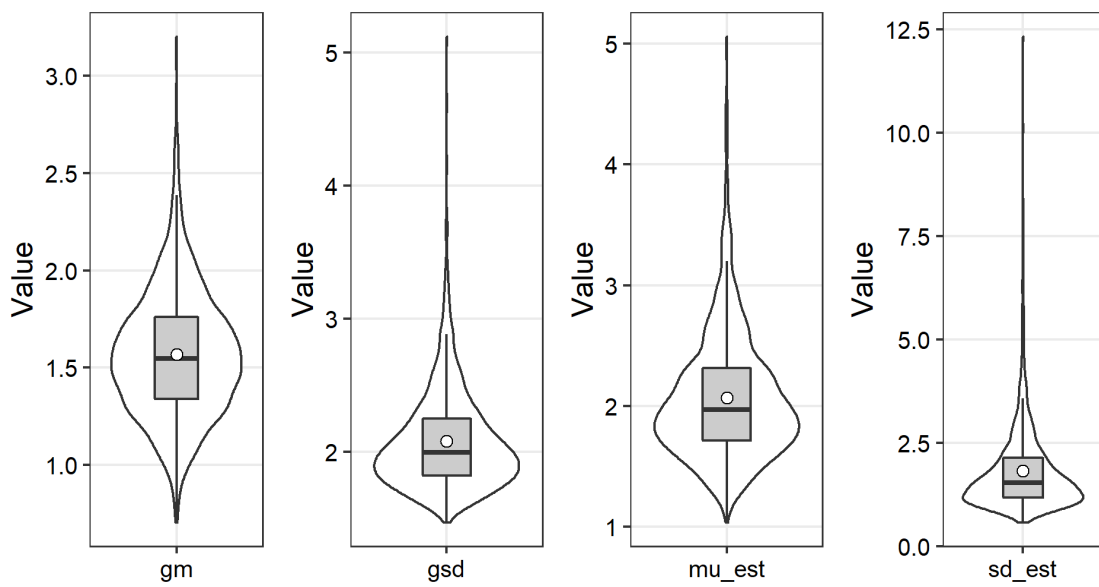


Fig. 3-20 バイズ推定によるメタミドホスの1日摂取量推定値に関わる事後予測分布

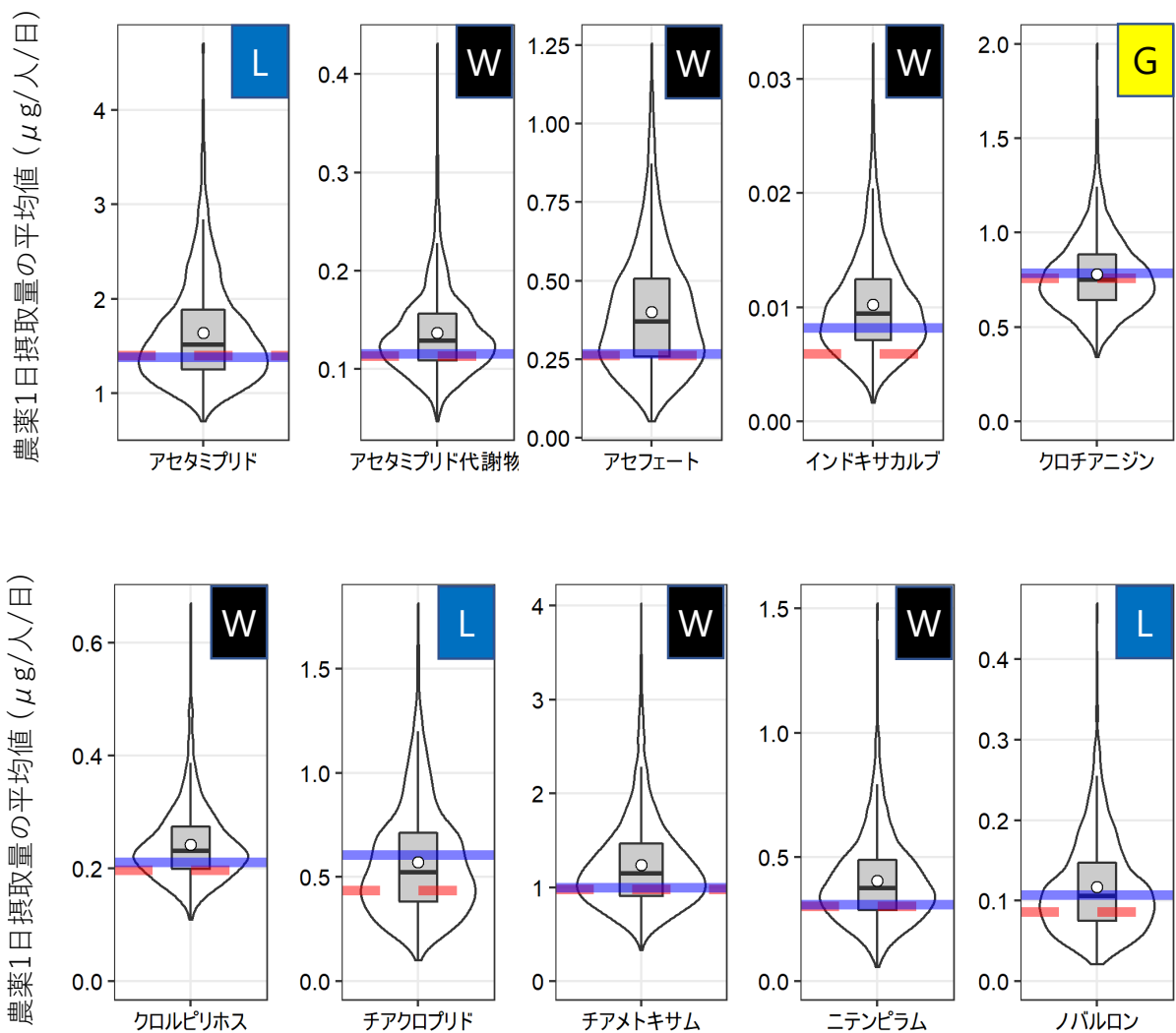


Fig. 4 未検出例を含むデータにおける平均農薬 1 日摂取量の推定

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.

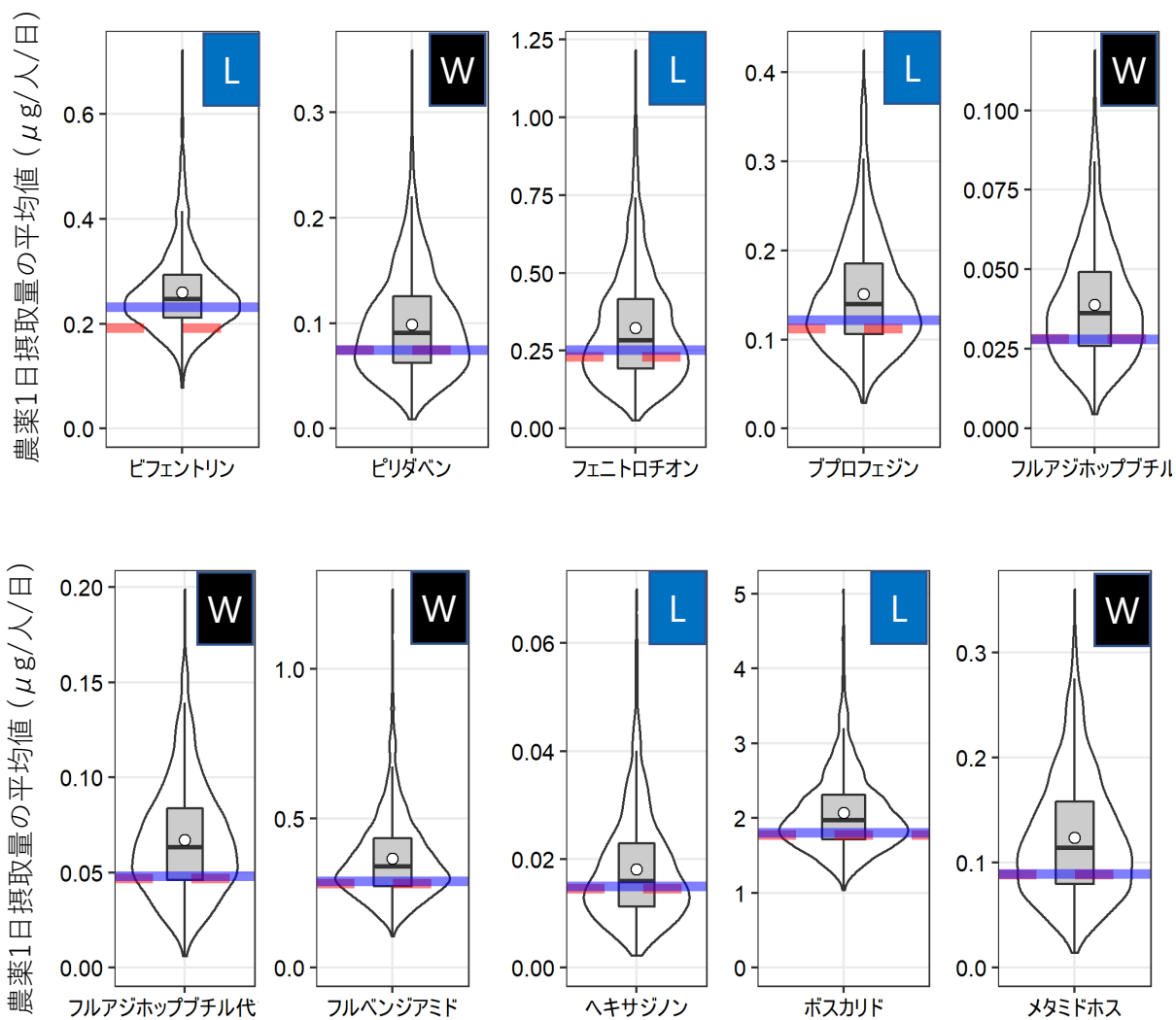


Fig. 4 続き

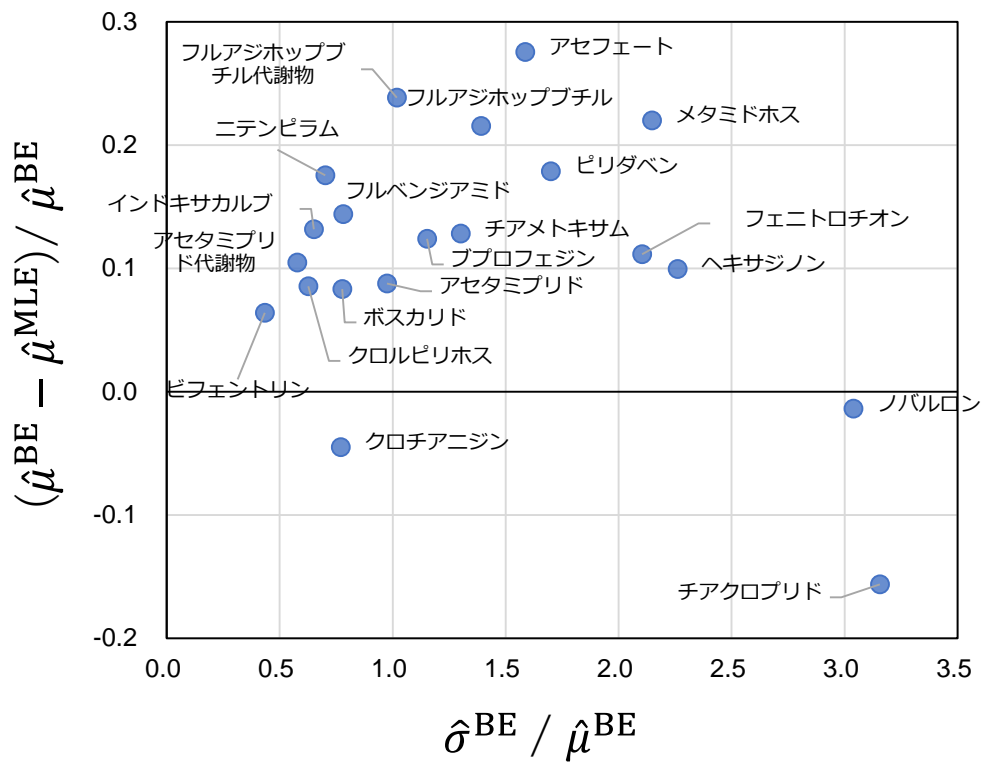
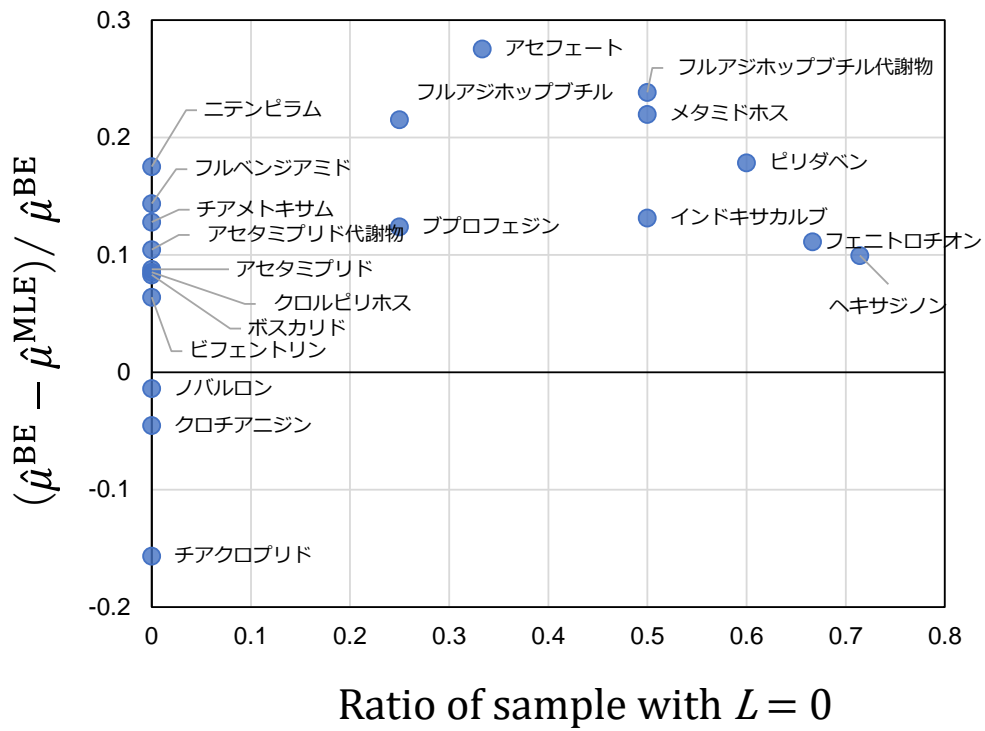


Fig. 5 ベイズ法と最尤推定法の推定結果の際に影響を与える要因