経験ベイズ入門

粕谷英一（九大・理・生物）

階層的データの解析　　　個体差　局所的環境

階層ベイズ　　　経験ベイズ

生態学でのデータ解析への適用

経験ベイズ

empirical Bayes ときにEB

事前分布についてデータから推定

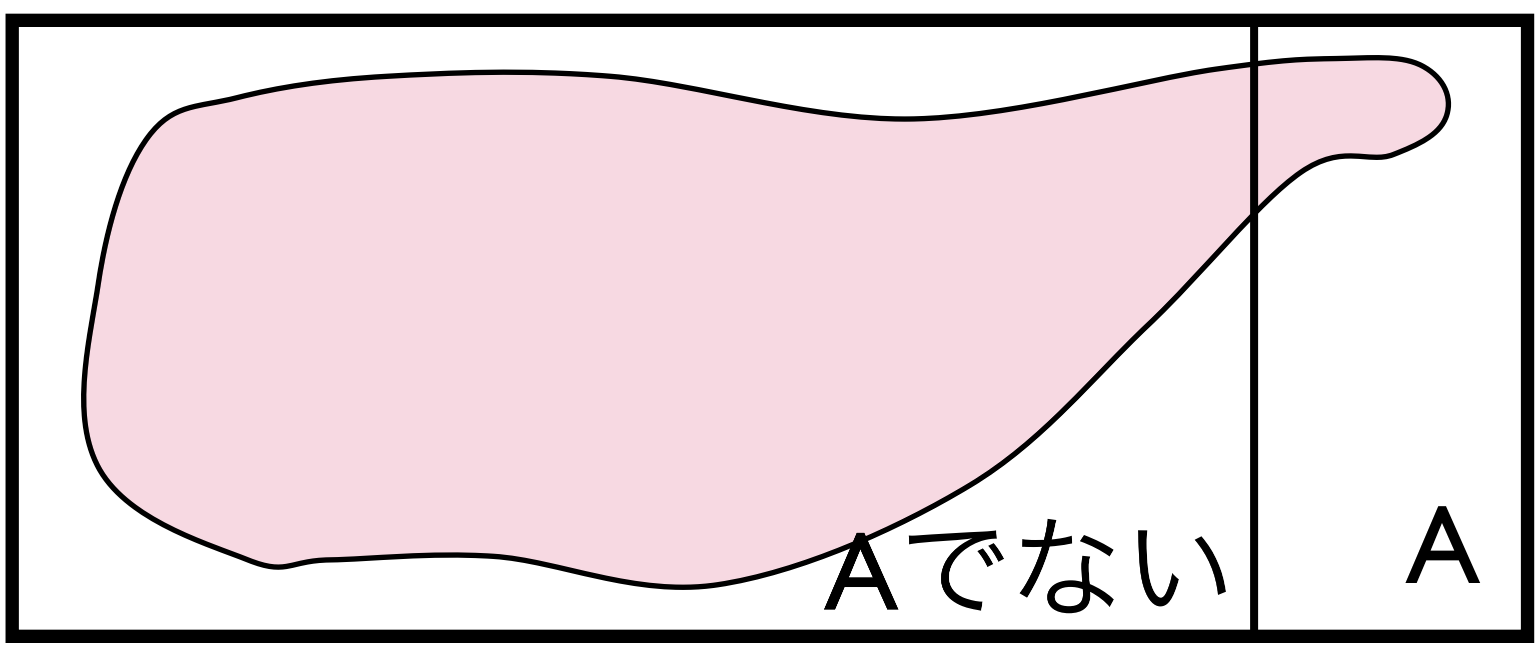
目次

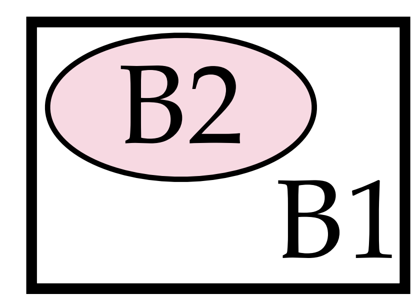
ベイズの定理

条件付き確率P(A|B)

条件AでBの確率

$$P(Bj \mid A) = \frac{P(A \mid Bj) \cdot P(Bj)}{\Sigma P(A \mid Bi) \cdot P(Bi)}$$



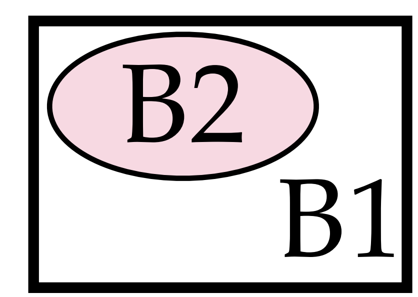


ベイズの定理

条件付き確率P(A|B)

条件AでBの確率

$$P(Bj \mid A) = \frac{P(A \mid Bj) \cdot P(Bj)}{\Sigma P(A \mid Bi) \cdot P(Bi)}$$



２０００万人の中に犯人が１人

遺留された犯人の体組織

体組織検査：誤って同一人としてしまう確率=1/300万
（陽性）

犯人の体組織とX氏の体組織特徴が

検査で一致

Xが犯人である：確率0.13

Xが犯人でない：確率0.87

$$P（Bj \mid A） = \frac{P（A \mid Bj）P（Bj）}{\sum_j P（A \mid Bj）P（Bj）}$$

A:陽性　B1:犯人　B2:犯人でない

P（B1 | A）

犯人でなくて、陽性になる確率
＝（２０００万ー１）／２０００万×（１／３００万）
≒１／３００万

犯人であって、陽性になる確率
＝１／２０００万×１
＝１／２０００万

事前（prior）

Xが犯人である：確率1/2000万

データ

犯人の体組織とX氏の体組織特徴が

検査で一致

事後（posterior）

Xが犯人である：確率0.13

確率モデル　データ

パラメーター

事前分布（prior）

パラメーター

$\boldsymbol{\theta}$　**パラメーター**

y　**データ**

$$p(\theta\,|y)=\frac{p(y|\,\theta\,)\;p(\,\theta\,)}{\int p(y|\,\theta\,)\;p(\,\theta\,)}$$

$$p(\theta\,|y)\propto p(y|\,\theta\,)\;p(\,\theta\,)$$

$$p(\theta\,|y)\propto l(\,\theta\,|y)\;p(\,\theta\,)$$

l **尤度**

事後分布←（データ）・（事前分布）

データの分布ー正規分布 分散がわかっている場合

事前分布→正規分布
事後分布→正規分布

事前分布 正規分布（平均μ、分散σ^2）

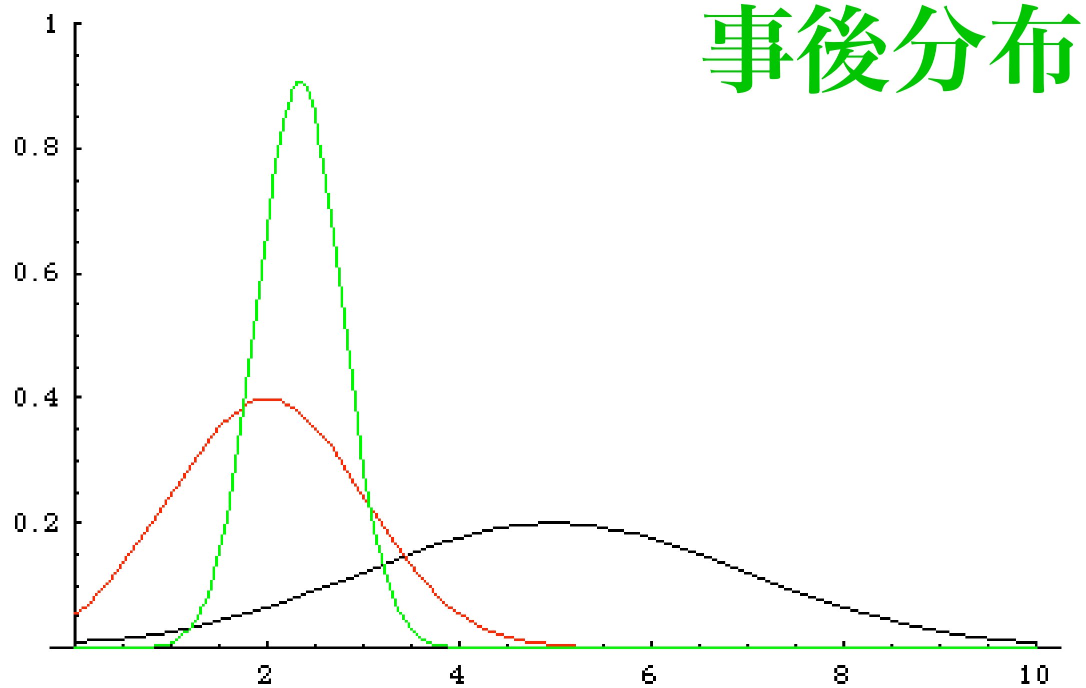
データ　平均$\bar{y}$、分散ϕ^2、標本数n

事後分布 正規分布

平均 μと$\bar{y}$の加重平均

重み($1/\sigma^2$)　重み(n/ϕ^2)

分散　$1/\{(n/\phi^2)+(1/\sigma^2)\}$



事後分布 正規分布（平均μ、分散）

事前分布

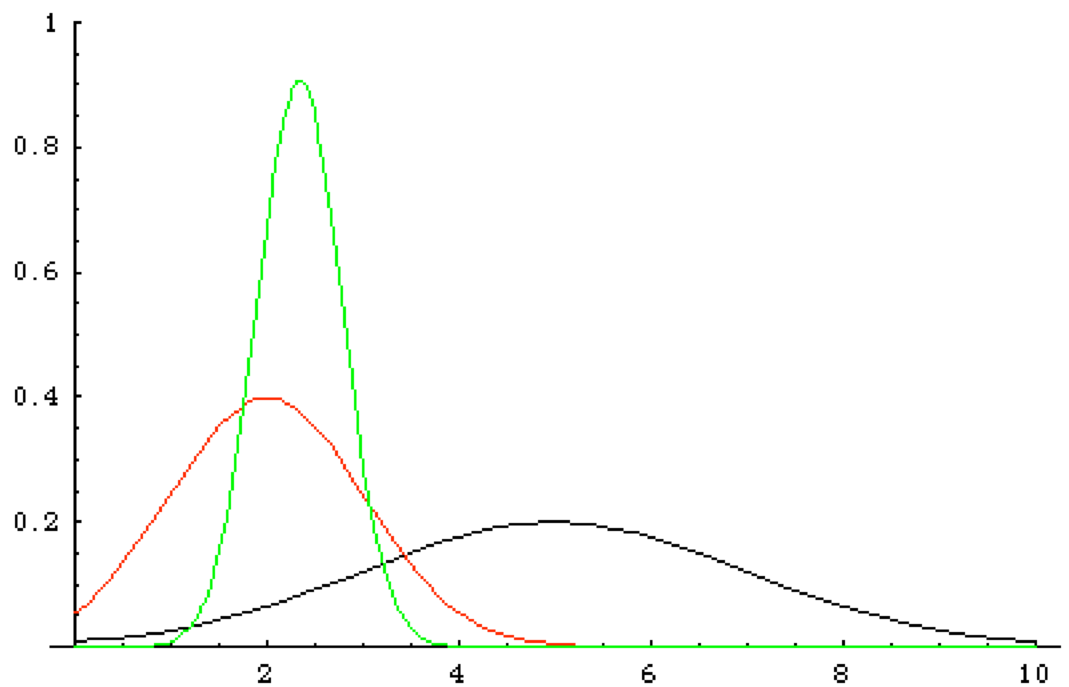
データ

事後分布

点推定値

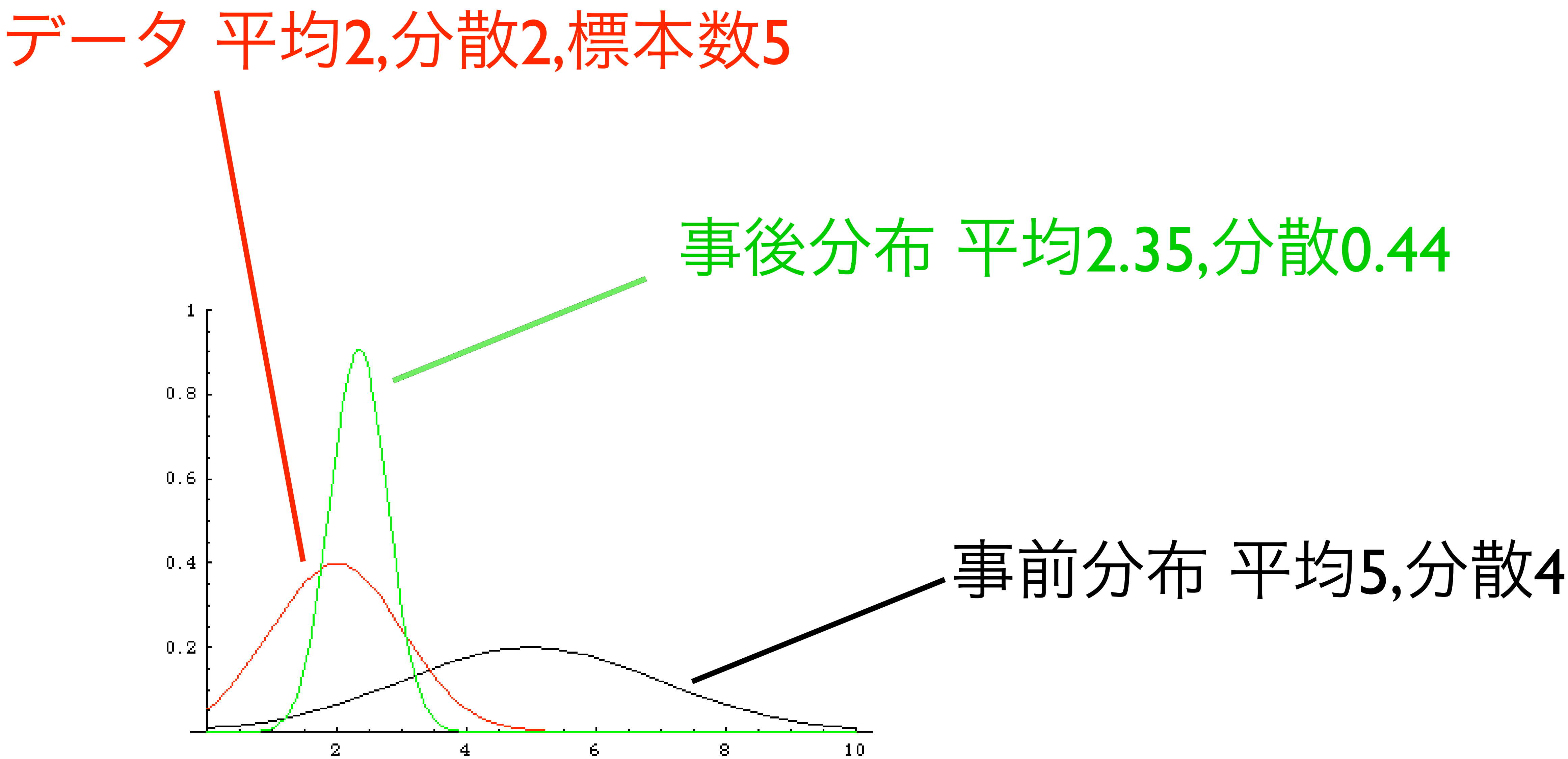
信頼区間

・
・
・



事後分布 正規分布（平均μ、分散）

データの分布ー正規分布



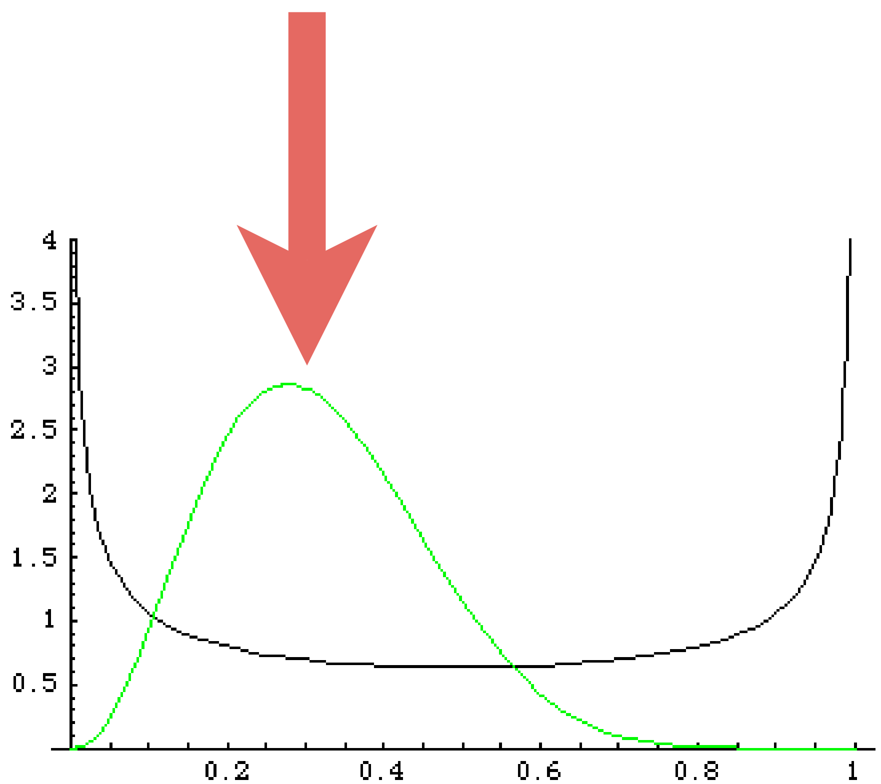
二項分布

事前分布→ベータ分布
事後分布→ベータ分布

ベータ分布　パラメーターα　β

平均 $\dfrac{\alpha}{(\alpha+\beta)}$

分散$(\alpha\beta)/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$



$\alpha=0.5, \beta=0.5$

事前分布　α　β

データ 成功x回,失敗(n-x)回,標本数n

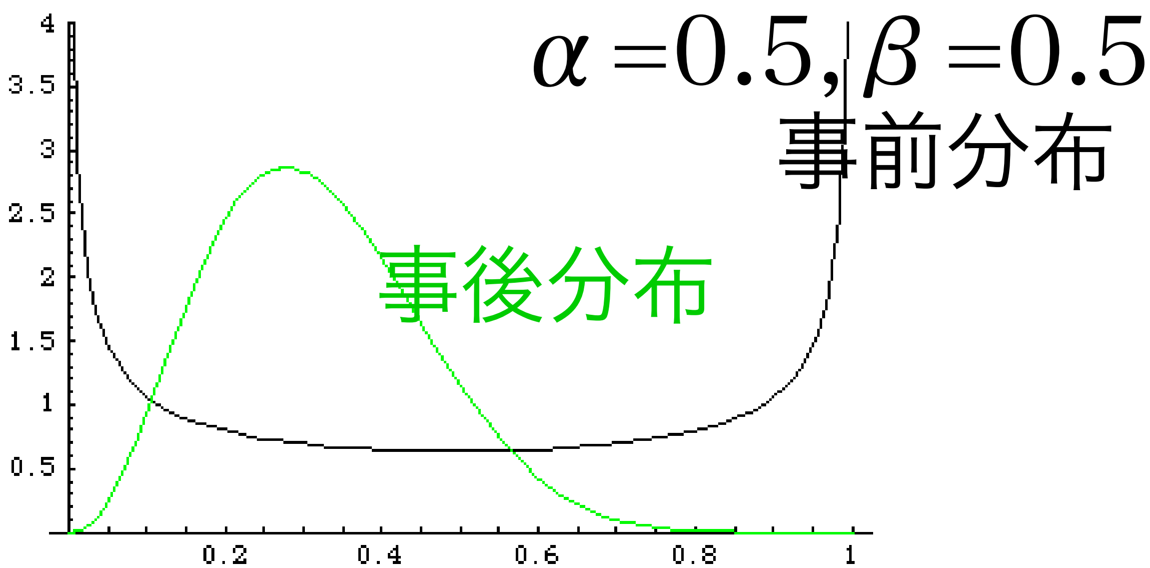
事後分布　$(\alpha+x)$　$(\beta+n-x)$

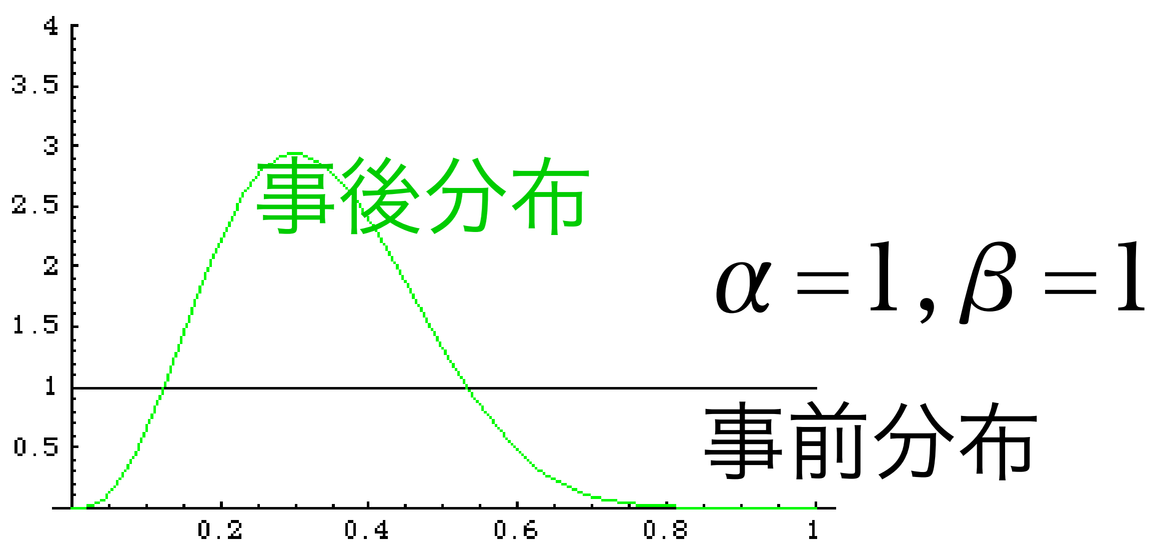
二項分布

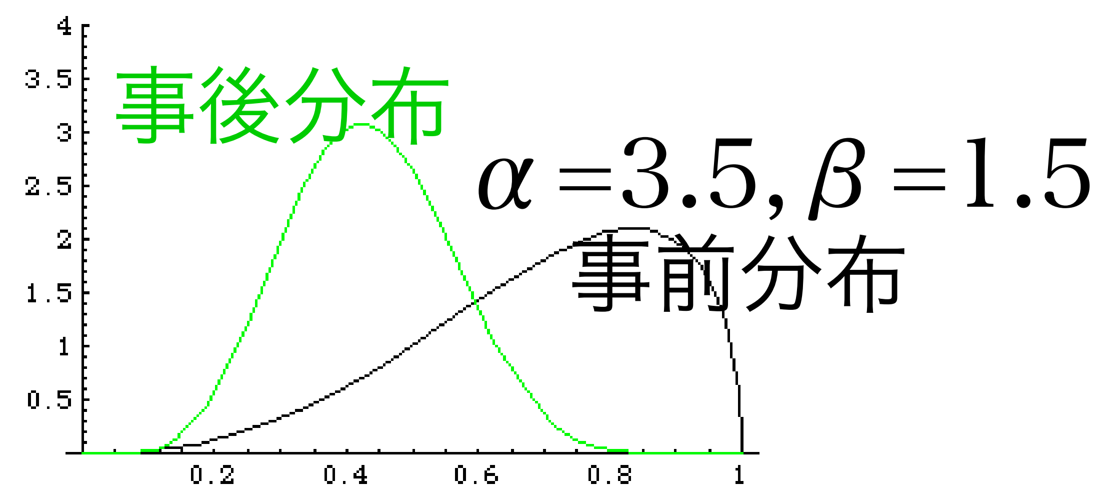
事前分布→ベータ分布

事後分布→ベータ分布

データ 成功3回,失敗7回,標本数10







経験ベイズ

empirical Bayes
ときにEB

事前分布についてデータから推定

パラメトリック経験ベイズ

経験ベイズ　歴史

von Mises(1943)

Robbins(1955)ー経験ベイズという名前の登場

ノンパラメトリック経験ベイズ

J.Neyman(1962)

『（統計的決定理論における）ブレイクスルー』

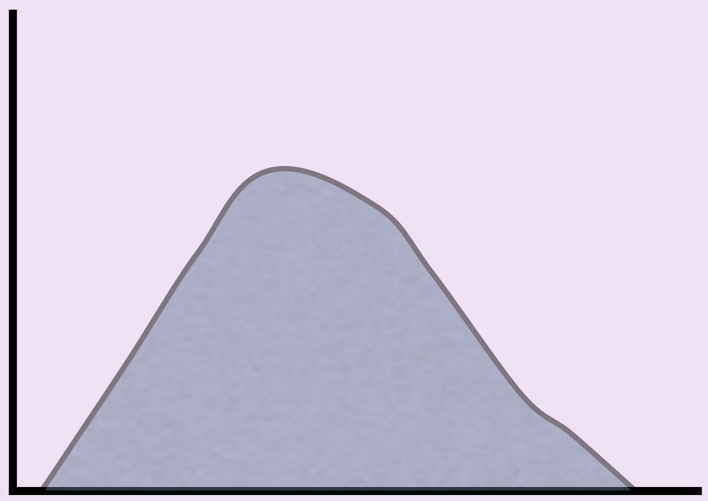
Efron&Morris(1972)　Morris(1983)

パラメトリック経験ベイズ

i.i.d

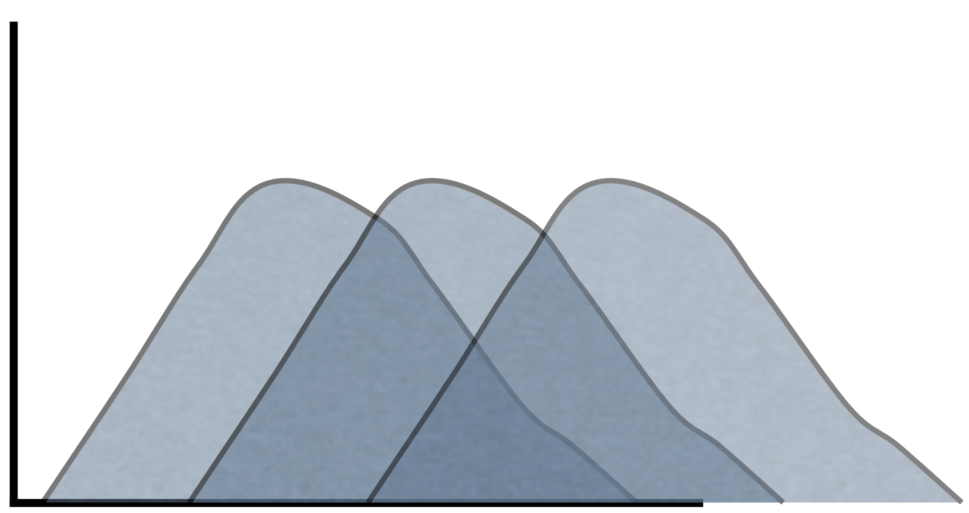
同一分布から独立に

independent　identical distribution



単純な確率モデル

しばしば現実の悪い近似



確率モデル　データ

パラメーター

事前分布（prior）

パラメーター

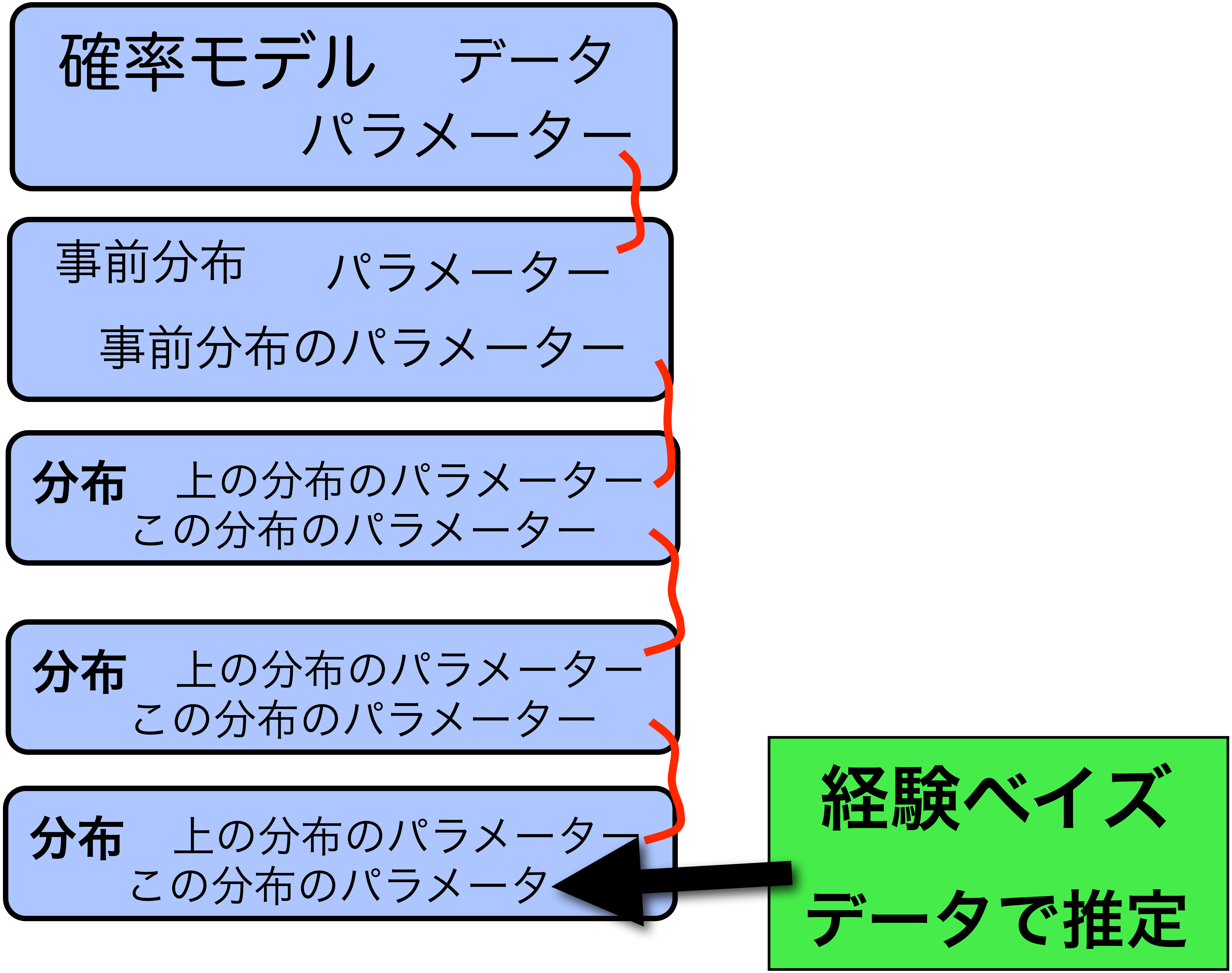
事前分布のパラメーター

超パラメーター

(hyperparameter)

超事前分布（hyperprior）

事前分布のパラメーター


確率モデル　データ
パラメーター
事前分布　パラメーター
事前分布のパラメーター
分布　上の分布のパラメーター
この分布のパラメーター
分布　上の分布のパラメーター
この分布のパラメーター
分布　上の分布のパラメーター
この分布のパラメータ
経験ベイズ
データで推定

経験ベイズ　単純な場合

ここで扱う最も複雑な場合

確率モデル　データ

パラメーター

事前分布（prior）

パラメーター

事前分布のパラメーター

事前分布のパラメーター

(超パラメーター)

データから推定

経験ベイズのしていること

$\boldsymbol{\theta}$　**パラメーター**

y　**データ**

$$p(\theta\,|y) = \frac{p(y|\,\theta\,)\;p(\,\theta\,)}{\int p(y|\,\theta\,)\;p(\,\theta\,)}$$

周辺尤度つまり分母の最大化

正規分布＆等分散の例 一元配置分散分析のような場合

確率モデル　データ Yi　正規分布

パラメーター　**母平均** θi　**母分散** σ^2

事前分布（prior）　パラメーター θi　正規分布

事前分布のパラメーター**母平均** μ　**母分散** τ^2

事前分布のパラメーター（超パラメーター）

μ　τ^2

正規分布&等分散の例

確率モデル　データ Yi　正規分布

パラメーター　**母平均 θ i　母分散 σ^2**

経験ベイズ推定　全体の平均　データ

$$\theta_i \qquad \frac{(k-3)\sigma^2}{(Y_i\text{の平方和})} \qquad 1-\frac{(k-3)\sigma^2}{(Y_i\text{の平方和})}$$

重み付け

事前分布のパラメーター　μ　τ^2

データから推定

θiの経験ベイズ推定値

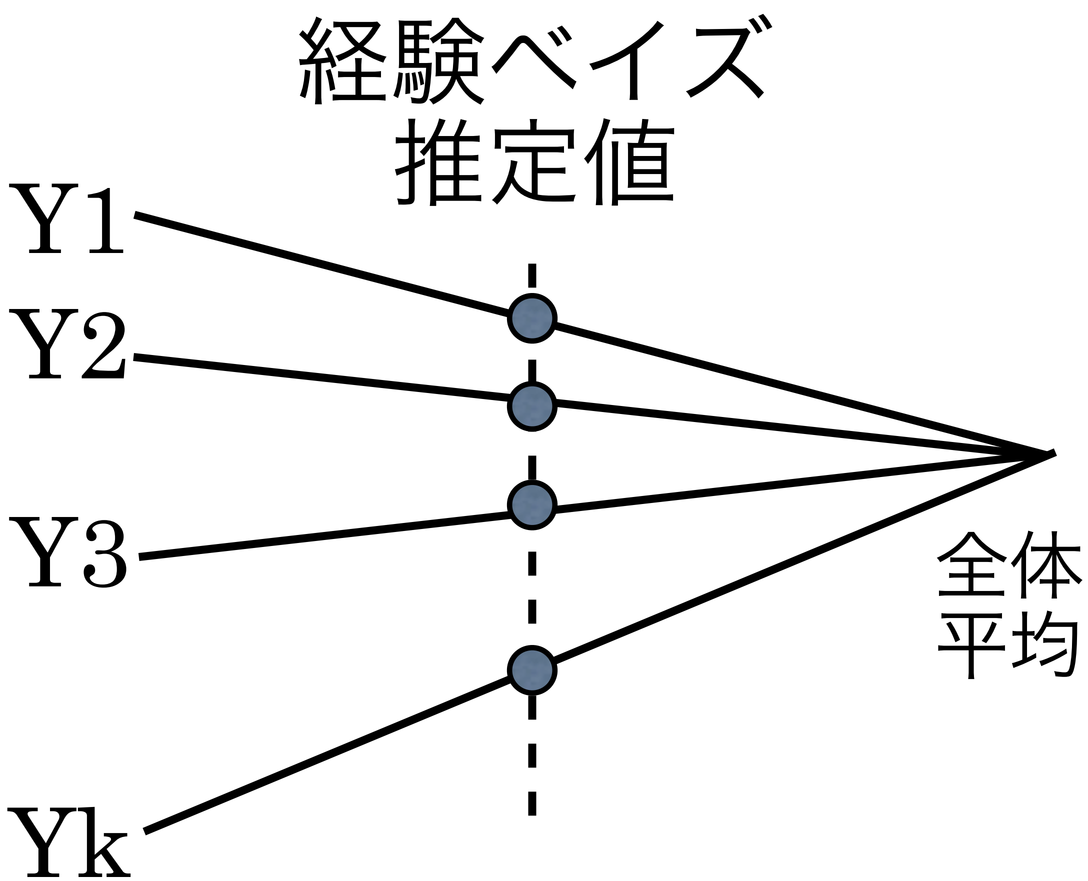
重み付け

全体の平均

$$\frac{(k-3)\,\sigma^2}{(Y_i\text{の平方和})}$$

データ

$$1-\frac{(k-3)\,\sigma^2}{(Y_i\text{の平方和})}$$



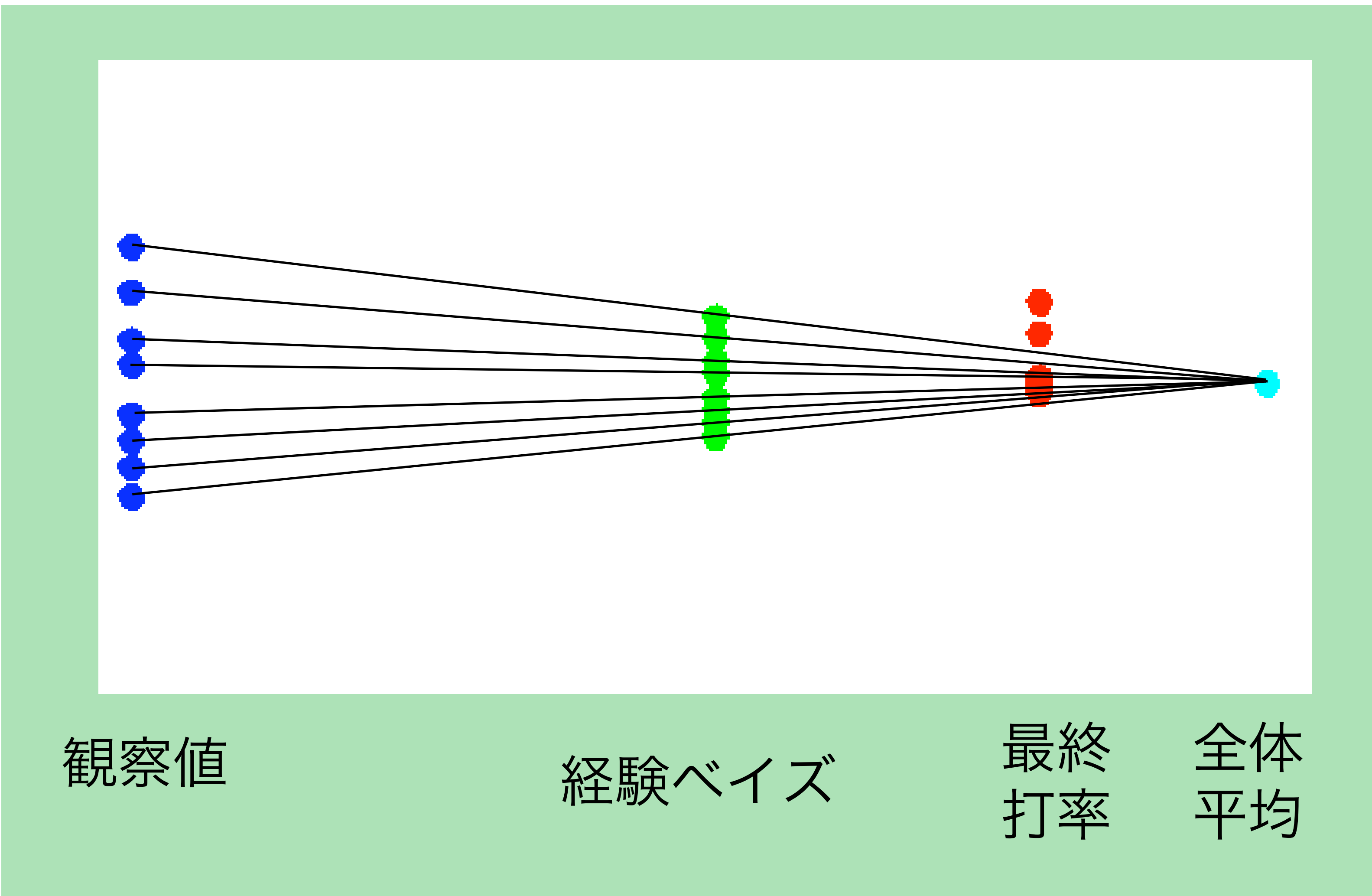
平均二乗誤差がそれぞれの平均よりも小さい

Stein推定量と同じ
（James-Stein推定量）

野球（アメリカMLB）の打者の安打率

Efron&Morris(1983)を
簡単にしたもの

二項比率のデータに正規分布を
あてはめているので無理はある



1970年最初の45打席

最高打率の選手

最低打率の選手

ランダムに選んだ6人

シーズン最終打率と比較

二項分布の場合　標識再捕獲（marking-recapture）

再捕獲率がばらつくとき

二項分布　確率一定

ベータ分布　確率が分布を持ちばらつく

Lloyd&Yip(1991) Chen(2001) Huggins(2002)など

パラメーター α 　β 　　平均 $\dfrac{\alpha}{(\alpha+\beta)}$

ベータ二項分布（負の超幾何分布）型の

overdispersionを組み込んだモデルと同じ状況

ベータ二項分布（負の超幾何分布）型の
overdispersionのモデル

確率が一定でなく分布（ベータ分布）を持つ

確率に相関がある　　など

二項分布（確率一定）の自然な拡張

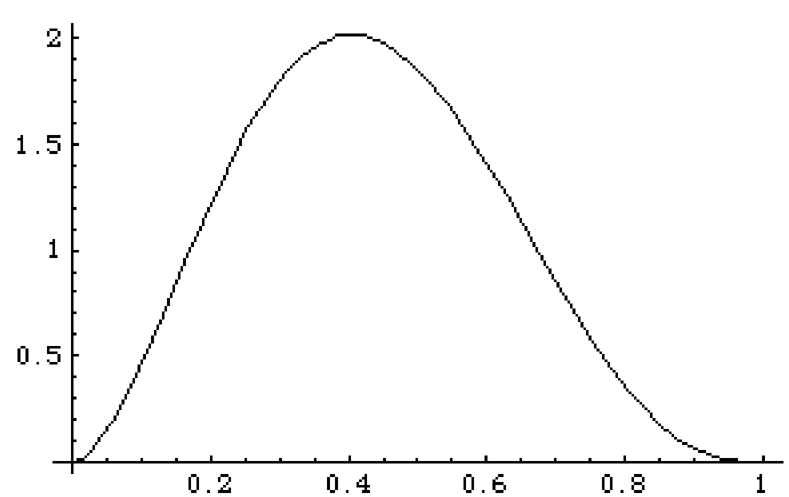
二項分布の場合　　個体間の再捕獲率のばらつき

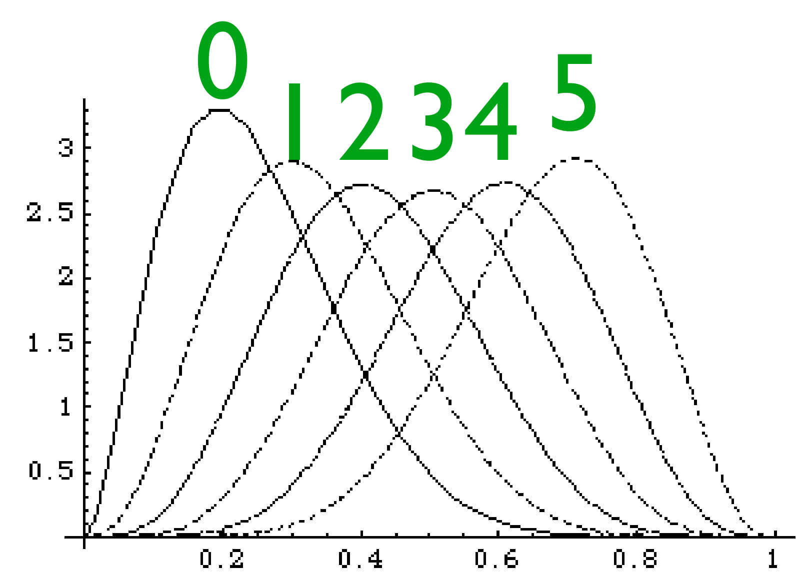
deermice110頭（Hoffmann, from Otisら1978）を一部改変した仮想的なデータ

ベータ分布のパラメーターα　β　データから推定

αの推定値=2.9

βの推定値=3.8





Huggins(2002)に同様な分析の例がある

経験ベイズ

・生態学向き　　疫学の適用例が参考になる

・多くの段階がある階層

・空間的な広がりー近いものの影響を受けやすい

・推定値を使ったデータ解析

役に立った文献

北田修一（2001）栽培漁業と統計モデル分析．　共立出版

岸野洋久（1999）生のデータを料理する．　日本評論社

Carlin&Louis(2000) Bayes andEmpirical Bayes Methods for Data Analysis. 2nd ed. Chapman&Hall

Maritz&Lwin(1989) Empirical Bayes Methods. 2nd ed. Chapman&Hall

Hoggほか（2006）数理統計学ハンドブック．　朝倉書店