ベイズ統計について考えるゼミ
～R+WinBUGsを用いて～

担当：　　原口　岳
と
直江先生（小川なう）

なぜ「面倒な」統計を使うのか1

線形モデル→一般化

カウント・2値など、明らかに正規分布しないデータを扱いたい

AICなどの情報量基準に従う統計

現象を説明する要因の候補があまりに多い

（かつ、共線性もしばしば発生する）

複雑すぎる説明式は人間の理解を超えている

非線形モデル

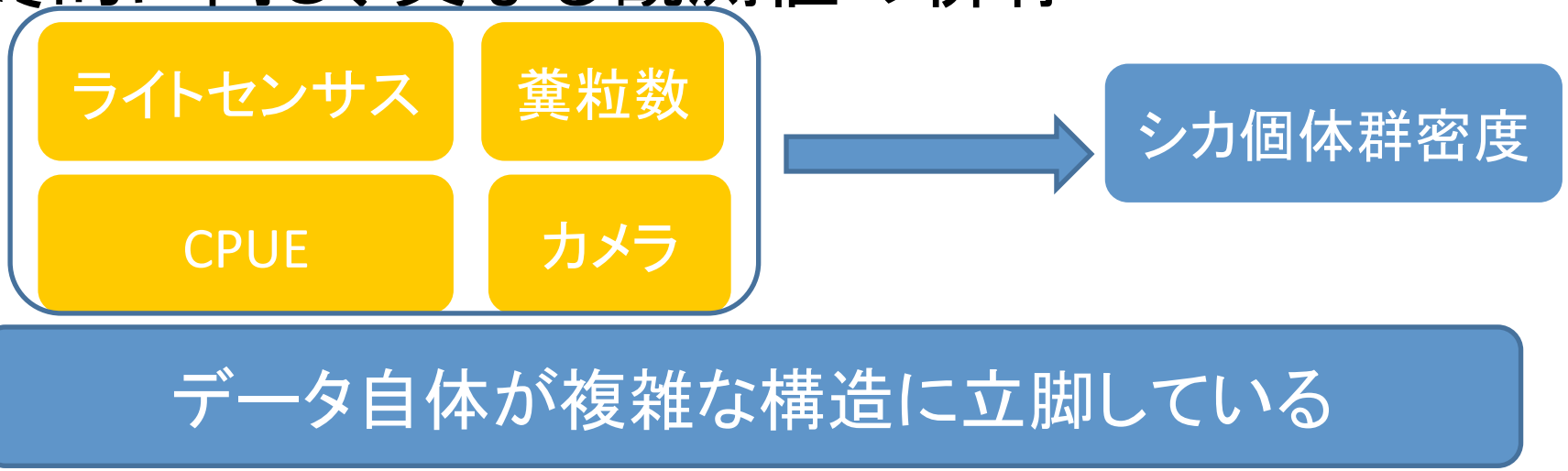
一山型の反応への対応や、時系列変化の理解？

よりデータに対する説明力の高いモデルを作りたい

なぜ「面倒な」統計を使うのか2

知りたいこと　≠　測れる項目

質的に同じ、異なる観測値の併存



ベイズ的アプローチの原理

頻度主義統計学的な問い

母集団（イデア）の中から標本を抽出した、と仮定する

繰返し標本抽出すれば、各標本集団の平均の平均は母平均

→帰無仮説が真である場合にデータを得る確率はどれくらい？

ベイズ

母集団を仮定しない／与えられたデータからは、求めたい値がどのような範囲に分布すると推測されるのか、その「確率分布」を求める

→データが観測された場合に仮説が真である確率分布は？

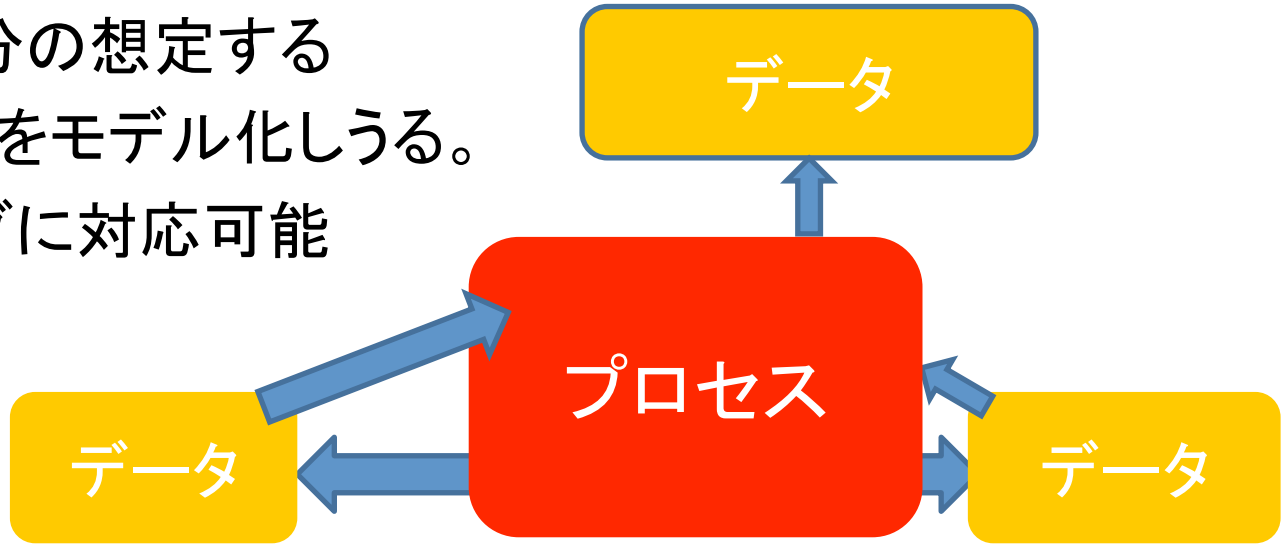
思考プロセスに従った考え方→「私はこう考える」というアイディアを明文化する必要がある

ベイズの良い点・面倒な点：プロセスを仮定してモデリング

パラメータ間の関係は多様である

シカ糞の例で挙げたように、（何らかの理論に基づいて）関係性が既知のパラメータの場合には、それを活用したい（他にも土壌炭素蓄積とか、生物の代謝速度と外部環境とか…）。

ベイズ推定…「ハイパーパラメータ」の与え方次第で、自分の想定する ecological process　をモデル化しうる。
→複雑なモデリングに対応可能



では、どれだけ面倒なのか？

カベハシリの存在率（%）を16年記録

割り算値、時系列データですが…

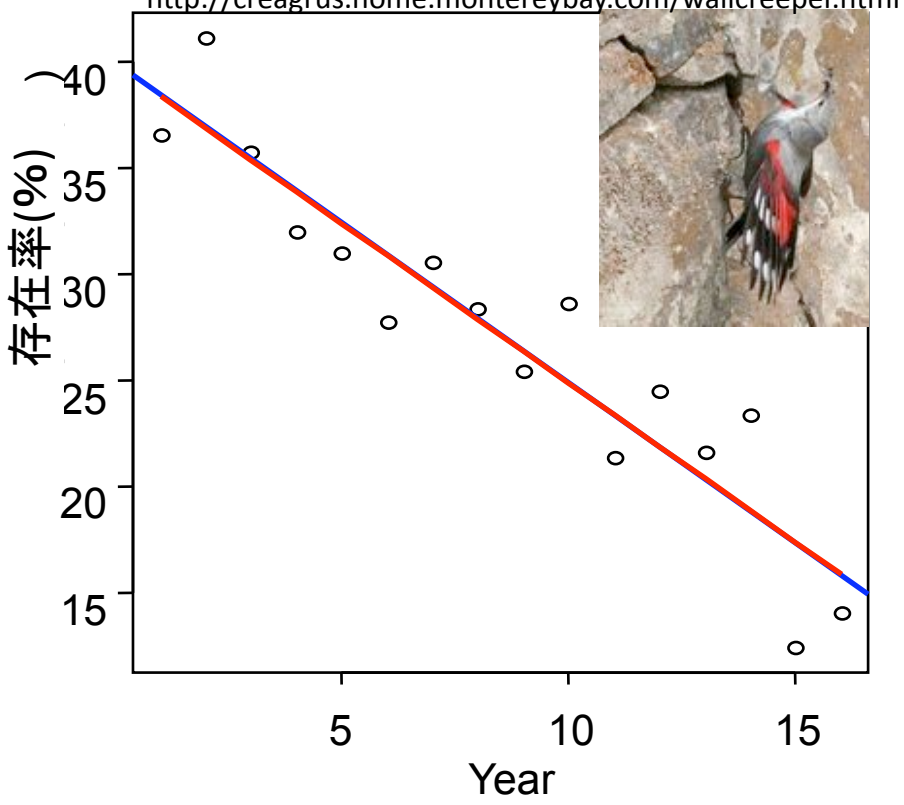
だんだん見られる頻度が減ってきたような気がする

→存在率yと年xの関係を線形回帰してみる(Kery, 2010)

http://creagrus.home.montereybay.com/wallcreeper.html

glmで線形回帰

lm (y ~ x)・・・だけ！



R+ winBUGsによるベイズ推定/骨子

```
model {
alpha ~ dnorm(0,0.001); beta ~ dnorm(0,0.001); sigma ~ dunif(0, 100)
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau); mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)}
",fill=TRUE)
inits <- function(){ list(alpha=rnorm(1), beta=rnorm(1), sigma = rlnorm(1))}
params <- c("alpha","beta",  "sigma")
nc = 3  ;  ni=11000  ;  nb=1000  ;  nt=1
out <- bugs(data = win.data, inits = inits, parameters = params, model =
"linreg.txt",n.thin = nt, n.chains = nc, n.burnin = nb, n.iter = ni, debug = TRUE,
working.directory=getwd(),bugs.directory="D:/Users/TFHaraguchi/
Documents/bayes/WinBUGS14")
```

意味的にはlm(y~x)と一緒だが…？？

原理説明：式を理解するために

```
model {
alpha ~ dnorm(0,0.001); beta ~ dnorm(0,0.001); sigma ~ dunif(0, 100)
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau); mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)}
",fill=TRUE)
inits <- function(){ list(alpha=rnorm(1), beta=rnorm(1), sigma = rlnorm(1))}
params <- c("alpha","beta",  "sigma")
nc = 3  ;  ni=11000  ;  nb=1000  ;  nt=1
out <- bugs(data = win.data, inits = inits, parameters = params, model =
"linreg.txt",n.thin = nt, n.chains = nc, n.burnin = nb, n.iter = ni, debug = TRUE,
working.directory=getwd(),bugs.directory="D:/Users/TFHaraguchi/
Documents/bayes/WinBUGS14")
```

なにゆえ繰返し構文？？

ベイズの定理

事後確率∝尤度×事前確率

$P(H|D) \propto P(D|H)P(H)$

あるデータが得られた条件で、仮定が支持される確率は、

その仮説のもとでデータを得る確率と、仮説そのものの確率の積である（データ＝fixの前提のもとでの条件付確率の式変形）

仮定…ある母数（確率分布）に従う

各確率は離散的ではないことが多いので、確率→確率密度

事後分布∝尤度関数×事前分布

事後確率密度は尤度関数×P(H)

ベイズの定理

事後分布∝尤度関数×事前分布

$P(H|D) \propto P(D|H)P(H)$

あるデータが得られた条件で、仮定が支持される確率は、
その仮説のもとでデータを得る確率と、仮説そのものの確率の積である（データ＝fixの前提のもとでの条件付確率の式変形）
仮定…ある母数（確率分布）に従う

各確率は離散的ではないので、確率→確率密度と読み換える

事後確率密度は尤度関数×P(H)

確率密度関数と尤度関数

e.g. [正規] 確率密度関数(Probability function)

$P(x)=(2\pi v)^{-\frac{1}{2}}\times e^{-\frac{(x-\mu)^{\wedge}2}{2v}}$

尤度関数(likelifood function)

→ベイズの定理から「階層的に」求められる

$L(\mu, v) = \prod_{i=1}^{n}(2\pi v)^{-\frac{1}{2}} e^{-\frac{(yi-\mu)^{\wedge}2}{2v}}$

観測値は所与のものなので、iでフィックスし、確率密度関数をμとvについての関数と見たものが尤度関数。(尤度関数を最大化するμとvを見つけるのが、最尤推定（MLE）です)

同様に、事前分布・事後分布も階層的に与えられる

尤度関数と確率密度関数は裏表

ベイズの定理の拡張

事後分布∝尤度関数×事前分布

更に、あるデータセット(i=1~n-1)を得た時の事後確率はn個目のデータセットを得た時の事前確率となる。

（例：コイントスでn-1回中s回表が出たとき、n回目のトスで表が出て、s/nだけ表が出る確率）

尤度原理→可換性・同一性

同一性…$P(y|\Theta_a) = P(y|\Theta_b)$ ならば $\Theta a=\Theta b$

可換性… 得られるデータの順番は確率に影響を与えない

初期の事前確率さえ与えれば、データを与えた時の事後確率を求めることが出来る

ベイズで推定しているものの実態

- P(Θ|D) ∝ P(D|Θ)P(Θ)

ベイズ推定で出力されるのは、常に「確率分布」である
（どこまで行っても「真値」の存在を仮定しない）
事前分布の形状を決定するパラメータ（ハイパー～）を置く
つまり、p(Θ|λ)= $\prod_{i=1}^{n} p(\theta i|\lambda)$
事後分布はP(Θ,λ|y)=ℓ(y|Θ,λ)p(Θ|λ)p(λ)
＊式下線部が事前分布p(Θ)をλの条件付き確率に変形したもの
（階層事前分布）
ハイパーパラメータはいくらでも作れる→つまり、階層的なデータの構造を、そのまま解析に組み込める

ハイパー～で、データ間の複雑な構造を表現

原理に従って式を読むと…

```
model {
alpha ~ dnorm(0,0.001); beta ~ dnorm(0,0.001); sigma ~ dunif(0, 100)
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau); mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)}
",fill=TRUE)
```

1. Y[i]は平均mu[i], 分散1/tauの正規分布

2. 平均値muはx[i]についての関数であり、傾きbeta, 切片alphaの線形回帰で表せる

3. 1個目のデータから推定したalpha,beta,tau分布は2個目のデータからalpha,beta,tauを推定するための事前分布となり、これを繰り返してn個のデータからパラメータの分布を推定する

実用上の問題点は…

事後分布∝尤度関数×事前分布

次々データを足すという発想で、事前分布から事後分布が求められる／一番最初の事前分布は？

→最初の事前分布を(自分で)決定する

ベイズはどんな分布型にも対応できるの？

→色々な分布型があるが、「自然な共役分布」であることが解析的にベイズ推定が可能な条件／モデルが複雑だと自然な共役分布にはならない。／シュミレーション(MCMCなど)の世界へ

事前分布の決め方とMCMCの活用が、
ベイズ推定を実用化せしめた重要な要素

再び線形回帰のモデル式

```
model {
alpha ~ dnorm(0,0.001); beta ~ dnorm(0,0.001); sigma ~ dunif(0, 100)
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau); mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)}
",fill=TRUE)
inits <- function(){ list(alpha=rnorm(1), beta=rnorm(1), sigma = rlnorm(1))}
params <- c("alpha","beta",  "sigma")
nc = 3  ;  ni=11000  ;  nb=1000  ;  nt=1
out <- bugs(data = win.data, inits = inits, parameters = params, model =
"linreg.txt",n.thin = nt, n.chains = nc, n.burnin = nb, n.iter = ni, debug = TRUE,
working.directory=getwd(),bugs.directory="D:/Users/TFHaraguchi/
Documents/bayes/WinBUGS14")
```

傾きと切片, σの確率分布を指定するようだ

事前分布の決定

最初の事前分布の与え方は事後分布に影響

→恣意的な事前分布を与えることは出来ない

提案1:無情報事前分布を与える

一般にdunif()や、N(0,10^3), Ga(0.01,10^3)など、極めて薄く広い分布を与えている

注：winbugsではN(a,b)をdnorm(a,b^{-1})と書く

（標準偏差の逆数τで式を書く）

提案2:先行研究から得られている情報を使う

例：同位体ソースミキシングモデルで、捕食者のエサ推定の事前分布として胃内容物からの推定結果を使う

根拠がないなら無情報事前分布を。

MCMCに関わる要素

```
model {
alpha ~ dnorm(0,0.001); beta ~ dnorm(0,0.001); sigma ~ dunif(0, 100)
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau); mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)}
",fill=TRUE)
inits <- function(){ list(alpha=rnorm(1), beta=rnorm(1), sigma = rlnorm(1))}
params <- c("alpha","beta",  "sigma")
nc = 3  ;  ni=11000  ;  nb=1000  ;  nt=1
out <- bugs(data = win.data, inits = inits, parameters = params, model =
"linreg.txt",n.thin = nt, n.chains = nc, n.burnin = nb, n.iter = ni, debug = TRUE,
working.directory=getwd(),bugs.directory="D:/Users/TFHaraguchi/
Documents/bayes/WinBUGS14")
```

alpha, beta, sigmaの初期値？／
なんか色々決めとかなきゃいけないらしい

MCMC（マルコフ連鎖もんてかるろ）法

参考になるサイト：http://tombo.sub.jp/doc/esj55/MCMC.pdf

各パラメータは確率分布として返されるが、「分布」のままでは解析的に解けるケースは稀。

マルコフ過程：未来に影響するのは現在だけである

確率過程（p(t+1)はp(t)のみによって決まる）

WinBUGs…Bayesian inference using Gibbs sampler

ギブズサンプラーというMCMCの一種を使って、（一部MH法）事前分布に従う標本抽出をやってくれる／パラメーター一つだけについて値をちょっと動かしてみる試行を何度も繰り返す

∴歩幅の変わるランダムウォーク：

分布の山の上では足踏み、山のすそ野では大股歩き

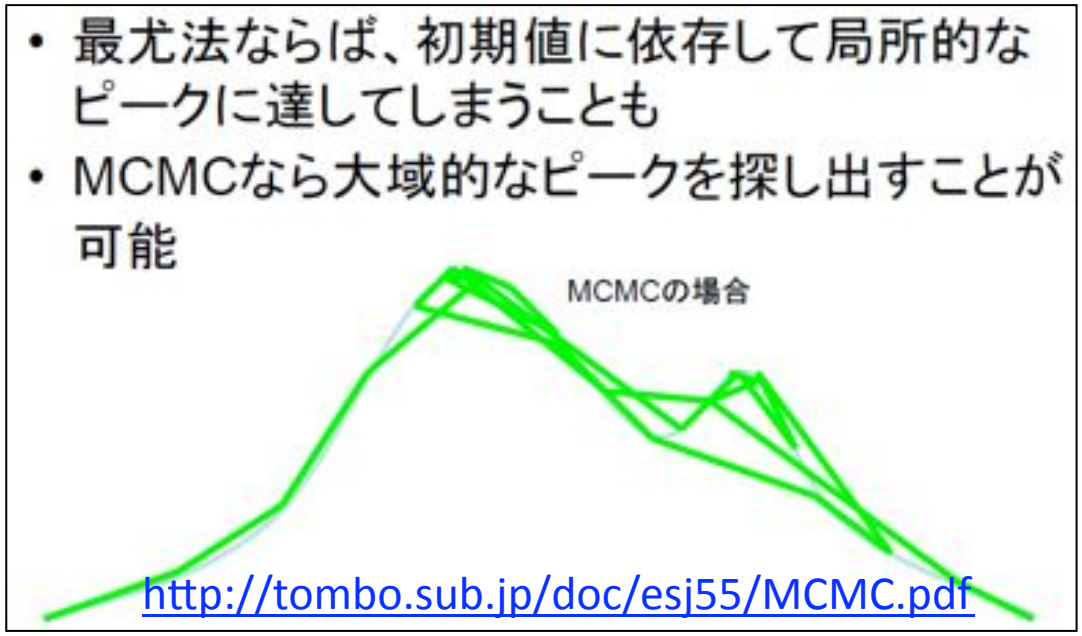
全分布型を再現するための、確率分布に従う標本抽出方法

初期値の決定

初期値とは…標本抽出過程で「山のどこからスタートするか」なんでもよい

∵初期値の影響が無いくらい標本抽出を多回行うため

というか、
初期値の影響を受けている状態では、標本抽出回数が不充分



分布型を再現するために、標本抽出は多数回おこなう

MCMCに関わる要素

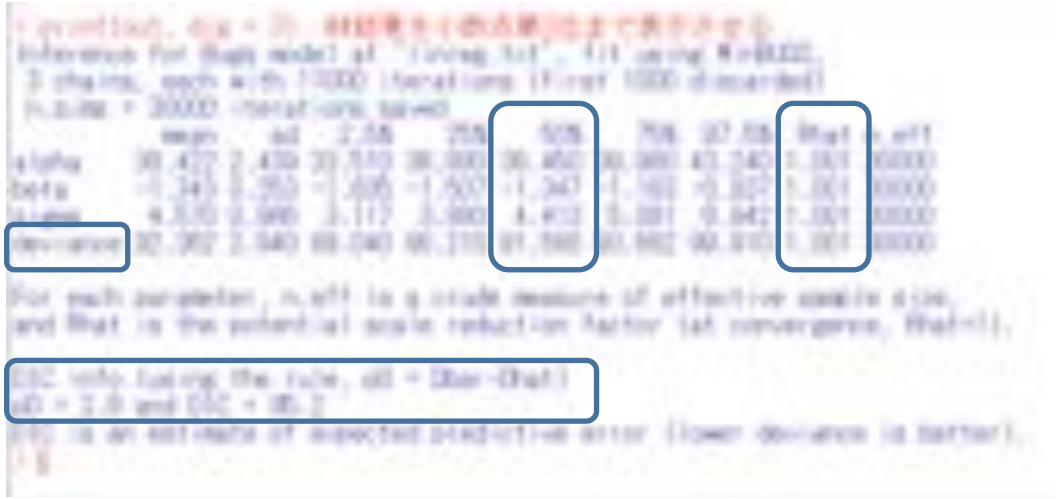
alpha, beta, sigmaの初期値
＝なんでもよい(論理的にありえない数値以外)
＊「推定結果を知りたい項目」を指定する必要あり

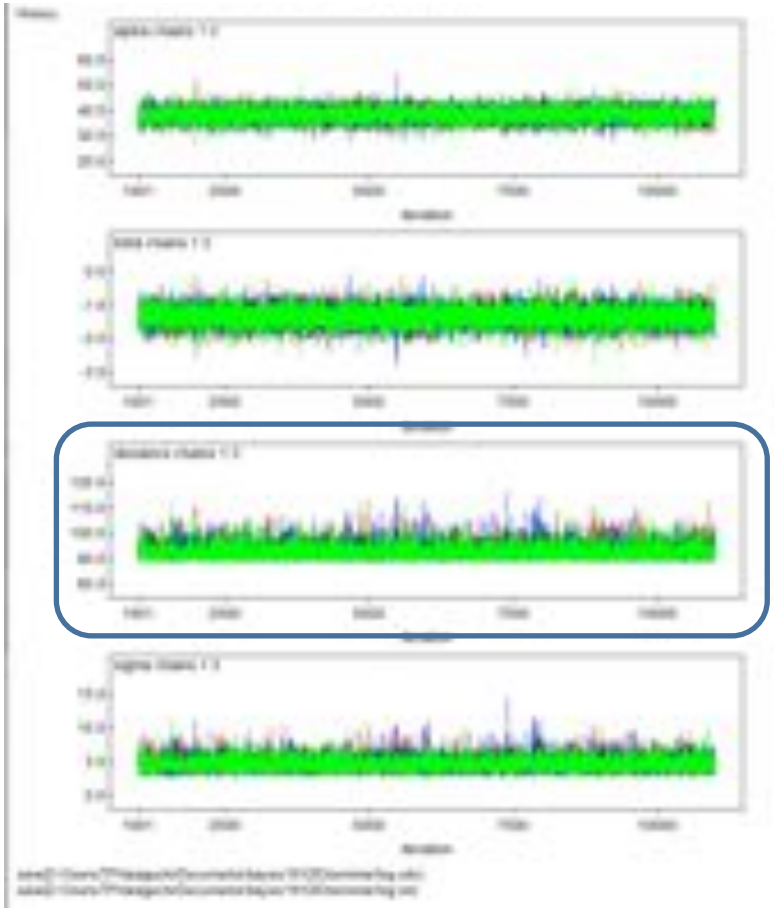
```
,inl=TRUE)
inits <- function(){ list(alpha=rnorm(1), beta=rnorm(1), sigma = rlnorm(1))}
params <- c("alpha","beta",  "sigma")
nc = 3  ;  ni=11000  ;  nb=1000  ;  nt=1
out <- bugs(data = win.data, inits = inits, parameters = params, model =
```

No. of iterations = 何回MCMCを実行するか

なんか、まだ指定している項目がある
→結果のverificationのための項目
ひとまず、実際に動かしてみよう。

出力される結果





median (＝値の推定値)

2.5%～97.5%(＝　95%信頼区間)

devianceとはなんぞ？

DICが95.2で、低いほど良いよ、とは？

標本抽出の結果が3本書かれている

(黄緑・赤・青)

収束する、ってどういうことか？

MCMCで得た値が適切なのか判断する基準

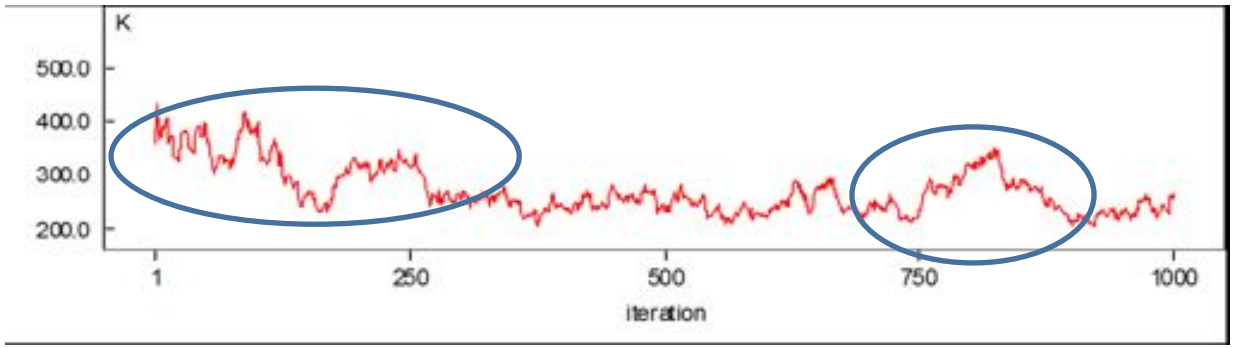
前提：多回シュミレーションを行うと、特定の平衡状態に達する

収束していない値は信頼できない

→iterationを増やせば（論理上）いつか収束する、とされている

But　収束しづらい例も（多々）ある　e.g. パラメータ間に相関

- 初期値の影響を排除する[burn in]
- 推定値の自己相関を間引き再抽出で薄める[thinning]



http://cse.fra.affrc.go.jp/okamura/bayes/mcmc.pdf

モデルの評価（収束）

まず、推定値が収束しているかどうか確かめる

1. 初期値の影響がない

　　初期値を変えて再計算しても予測値は同じ

2. Stochasticな変動が無い

　　同じ試行を繰り返せば、同じ予測値を返す

→[no.chain]を2以上にして与える初期値を変えておくと、chain間の予測値の差を評価することで判定できる

これらを数値化したもの：Rハット値 (Gelman Rubin統計量)

sqrt(f(bet. chain var.(B)／with. chain var.(W),m,n))

…nはburn in後の標本数、mはchain数であり、B／W比(0だと良収束)を、標本数について補正したもの　　1に近いほど収束

モデルの評価（変数選択）

情報量基準に従ってパラメータ選択が可能

AICの改良版が色々ある c.f.データが多いほどパラメータが落ちにくくなる効果を補正

最も標準的な使用例…DIC（=ABIC）によるモデル評価

尤度　パラメータ数　データ数

$$AIC_m = -2l_m(\theta_m \mid X) + 2k_m$$

$$CAIC_m = -2l_m(\theta_m \mid X) + k_m(\log N + 1)$$

$$BIC(MDL)_m = -2l_m(\theta_m \mid X) + k_m(\log N)$$

$$DIC = -2\log L(\sigma_2, d)$$

周辺対数尤度

6.2.2 Gelman and Rubin (1992)

モデルの評価（変数選択）

DIC（=ABIC）によるモデル評価

推定する関数fについて、その滑らかさを決めるH. param.を置く

データの誤差εiが正規分布N(0,σ²)に従うとすると、σ²に関する尤度関数（周辺尤度）が最大のものが最良のモデルである

従って、モデルとしての当てはまりが良いほどDICは小さいと言える　$DIC = -2\log L(\sigma_2, d)$

DIC＝Deviance Information Criterion

∴解析結果には自分で設定しなくても必ずDevianceというパラメータが生成される

推定すべき関数 $f(x)$ を等間隔な離散点 x_j 上の値として $f_j = f(x_j)$ と表現し、f を式 7 のよ

線形回帰式(これで最後！)

model {

params <- c("alpha","beta", "sigma")

```
nc = 3 ; ni=11000 ; nb=1000 ; nt=1
```

```
out <- bugs(data = win.data, inits = inits, parameters = params, model =
"linreg.txt",n.thin = nt, n.chains = nc, n.burnin = nb, n.iter = ni, debug = TRUE,
working.directory=getwd(),bugs.directory="D:/Users/TFHaraguchi/
Documents/bayes/WinBUGS14")
```

1chainあたり標本抽出を11000サイクル繰返
最初の1000サンプルを捨てる
得られたパラメータ推定の結果から1標本
おきに取り出した値のセットを使う

今まで与えた命令通りに実行

線形回帰の結果

•カベハシリの存在率（%）を16年記録・・・割り算値、時系列データですが、
存在率yと年xの関係を線形回帰してみる(Kery, 2010)

•最尤推定

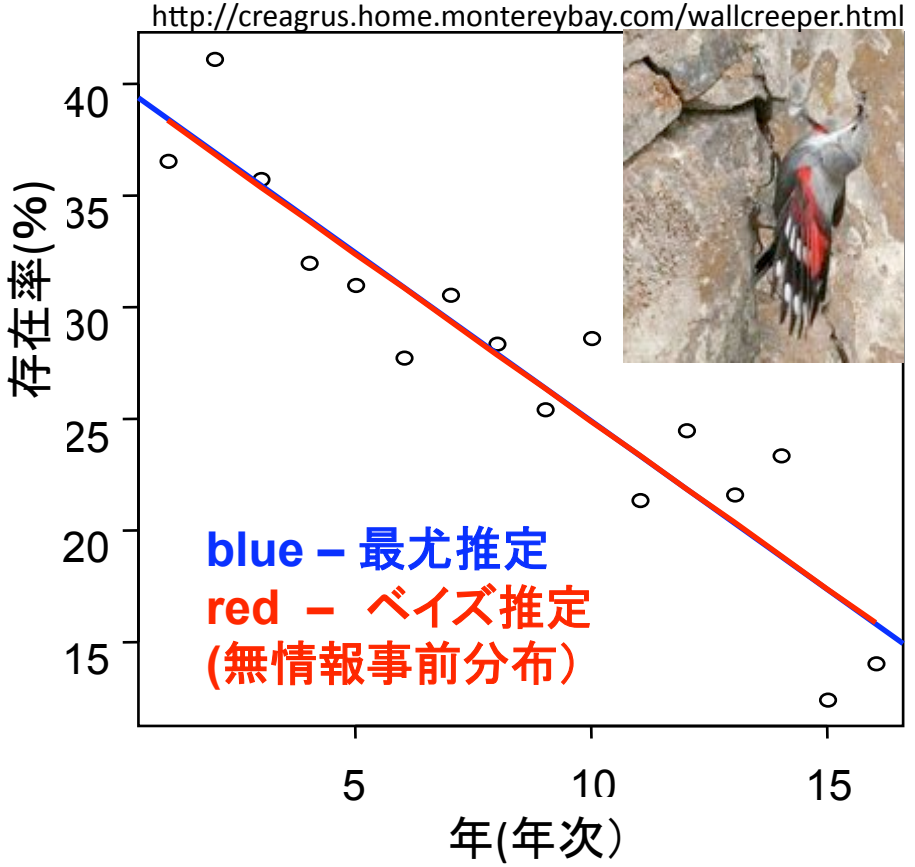
lm (y ~ x)・・・だけ！

•ベイズ推定

基本的なモデル構造

```
model {
for (i in 1:n) { y[i] ~ dnorm(mu[i], tau)
                 mu[i] <- alpha + beta*x[i]}
tau <- 1/ (sigma * sigma)   }
```

http://creagrus.home.montereybay.com/wallcreeper.html



無情報事前分布の場合、最尤推定もベイズ推定も推定値自体はほぼ同じ！

実例2:オックスフォードボートレース

http://www.theboatrace.org/article/introduction/pastresults

1836年から続くcambridge vs Oxford対抗戦

・昨年出場した学生の一部は今年も出場する

・強いチームには良い人材が集まりやすい

・勝つ確率の密度関数pの結果として実際の勝敗が決まる

→時間軸方向に自己相関する説明変数q[i]を設定する

本来p~qだが、ベルヌーイ試行なので、リンク関数はロジット

Log(p/1-p)=qより、 p = exp(q)/1+exp(q)

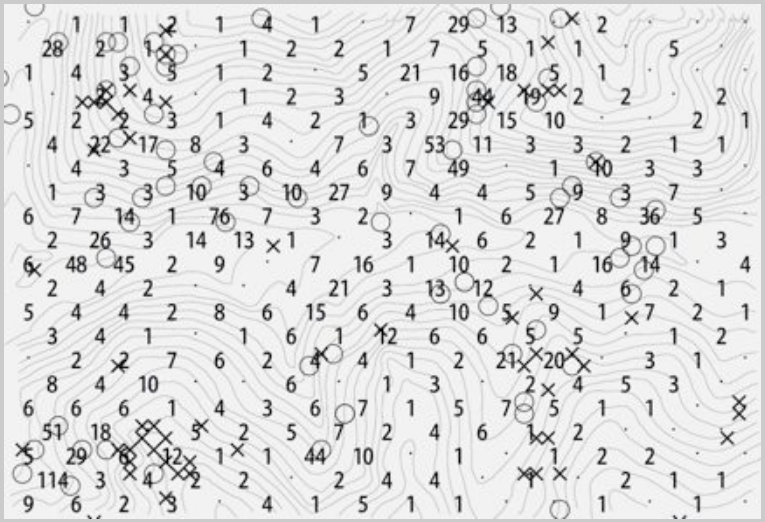
```
model{
for(t in 1:N){p[t]<-exp(q[t])/(1+exp(q[t])); y[t]~dbern(p[t])}
q[1]~dnorm(0,0.001)
for(t in 2:N){q[t]~dnorm(q[t-1],tau)}
tau~dgamma(0.001,0.001);Isigma<-log(1/tau)}
```

実例3：鳥が運ぶ種子の空間分布

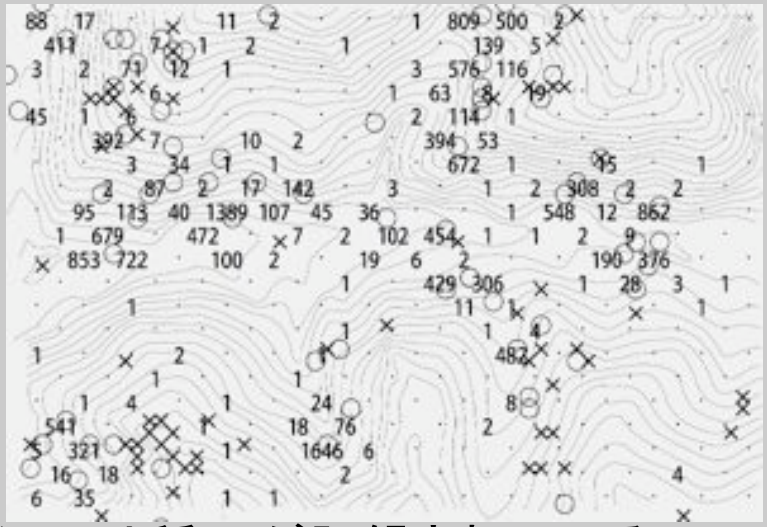
- 樹種Aの鳥散布種子の分布を、樹種A・B・Cの結実木の分布で説明したい

結実木下に鳥が種を運ぶと予測：*Aの鳥散布種子数～Aの自然落下種子数＋Bの自然落下種子数＋Cの自然落下種子数*

樹種Aの鳥散布種子の空間分布



樹種Aの自然落下種子の空間分布



どちらも種子が多い場所の周辺では種子が記録されている
＝目的変数も説明変数も空間自己相関している

- ある場所の種子数は近隣の場所の種子数の影響を受ける
→空間自己相関を考慮したモデル

実例3 : 鳥が運ぶ種子の空間分布

なぜベイズ推定を用いたモデルなのか？

1. 目的変数の空間自己相関を扱えるモデルは？・・・SGLMM, GAM, GEE, SEVM, AR, SAR, CAR, GLS, SDM, GWR, MGWR等
   ＋ベイズ推定のモデル
2. 目的変数がポアソン分布である・・・SGLMM, GAM, GEE, SEVM, AR＋ベイズ推定のモデル
3. 説明変数の空間自己相関も扱いたい
   ・・・ この時点で現実的なのはベイズ推定のモデルのみ
4. さらに複雑なモデルにしたい（e.g., 時間自己相関,樹種D,E,F,G,H・・・）

→モデルが複雑になるほど、柔軟なモデリングが出来る
ベイズ推定を用いるしかなさそう

注意など:

新しい解析手法である

ベイズを使うからには、相応の「ベイズでなくてはならない」理由づけが要るだろう

相応の計算機資源を消費する

1回の解析に数時間かかることも珍しくない (e.g., 富田ら 2009では3日間）

解析結果が必ず収束すると言い切れるものではない

妥当なプロセスモデルを提示することが大切

論文を読む上で留意すべきこと

参考文献

説明が丁寧「ベイズ統計データ解析R＆WinBUGS」 古谷知之著　朝倉書店

WinBUGS only/実例が生態学的テーマに即している「生態学のためのベイズ法」 Michael A. McCarthy著　野間口眞太郎訳　共立出版

考え方「道具としてのベイズ統計」 涌井良幸著　日本実業出版社

「http://www.okada.jp.org/RWiki/?R%A4%C7%A5%D9%A5%A4%A5%BA%C5%FD%B7%D7%B3%D8 （Rjp.wikiのベイズページ）」

モデル式の実例

「http://www.rhasumi.net/wiki/wiki.cgi?page=Bayes」

リンク集「http://hosho.ees.hokudai.ac.jp/~kubo/ce/BayesianMcmc.html」

Tips「http://www7.atwiki.jp/hayatoiijima/pages/40.html」

謝辞

久保田先生・松田先生（統計ゼミ主催＆講師@琉球大学, 2010 Aug）

参考文献(続き)

とりあえず動かしてみる：RとWinBUGSを用いて、t検定からGLMM+αまでを最尤推定とベイズ推定で算出、比較している。Tipsも。

「Intoduction to WinBUGS for Ecologists」 Marc Kery, Academic Press, 2010. http://137.227.242.23/software/kerybook/で、WinBUGS単体での使用法を解説した章（4章）を公開しており、ベイズをお手軽に体験できる

もっと専門的：詳細なベイズ周りの説明とWinBUGSを用いた様々な例。

「Bayesian Inference with Ecological Applications」 Link and Barker, Academic Press, 2010.

Macでもやりたい：JAGS (Just Another Gibbs Sampler)を用いる。
http://web.sfc.keio.ac.jp/~maunz/wiki/index.php?ESTRELA%CF%A2%BA%DC
に日本語での簡単な紹介と実例あり(第81～84回）。