# ベイズ統計の基礎

水産資源学におけるベイズ統計の
応用ワークショップ

2007年8月2-3日，中央水研

遠洋水産研究所　外洋資源部
鯨類管理研究室
岡村　寛

# なぜベイズ法を使うのか？

- 水産資源学･生態学では，必ずしも実験デザインをきちんとコントロールした上でデータが得られるわけではない
- その場合，データとその扱いに様々な問題が起こる

**データ間の相関，個体差によるばらつき，**

**コントロールされない自然変動，**

**小さな標本サイズ，多くの欠測値，**

**直接的なデータはないが間接的な情報ならある，**

**複数のデータセットの統合，複雑な意思決定**

# なぜベイズ法を使うのか？

- そのような場合，パラメータの数は多くなり，モデルは複雑な階層構造を持つものになりがち.
- そんなときベイズ法は大変ありがたく，強力な武器となる.
- ベイズ法のひとつのネックは計算が大変であるということだったが，理論と道具の発展により，敷居がどんどん低くなってきている.

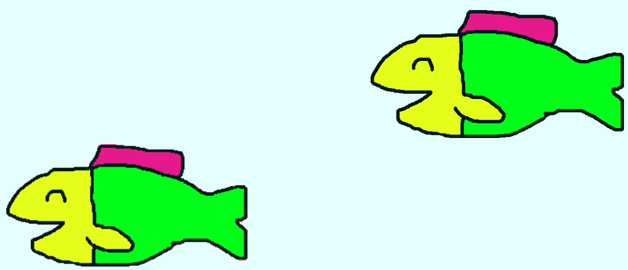
# このトークの内容

- ベイズ法の基本的考え方
- 従来の統計学とベイズ統計学
- 水産資源管理に対するベイズ法の応用例
- 階層ベイズモデル，モデル選択

# ベイズ法の基本的考え方

# 釣り人の漁場選択

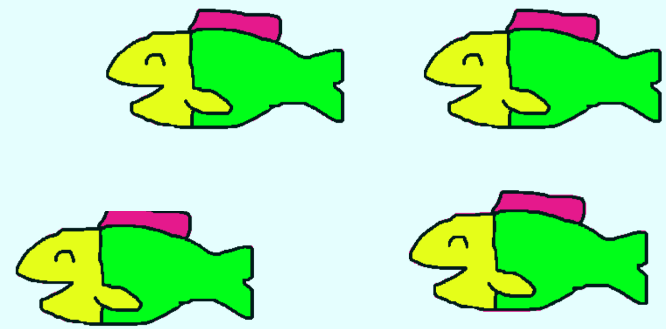
## 漁場A：0.8



**魚が釣れる確率 0.1**

**魚が釣れない確率 0.9**

## 漁場B: 0.2



**魚が釣れる確率 0.6**

**魚が釣れない確率 0.4**

# 確率記号で

- 漁場Aを選んだとき釣れる確率（条件付確率）
  P(釣れる|A) = 0.1
- 漁場Bを選んだとき釣れる確率（条件付確率）
  P(釣れる|B) = 0.6
- ある人が漁場Aを選んで，魚が釣れる確率
  P(A, 釣れる) = P(A)×P(釣れる|A)
  = 0.8×0.1 = 0.08
  P(釣れる|A) = P(A, 釣れる)/P(A)

# 条件付確率

- Xが起こったという条件のもとで，Yが起こる確率
  $P(Y|X) = P(X, Y)/P(X)$
- Yが起こったという条件のもとで，Xが起こる確率
  $P(X|Y) = P(X, Y)/P(Y)$

- $P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$

# 結果から原因を推測する

**$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X)$**

X: 原因, Y: 結果とせよ.

P(X|Y) = P(原因|結果)
= P(Y|X)P(X)/P(Y) = P(結果|原因)P(原因)/P(結果)
結果（データ）から原因（パラメータ, 仮説）の推測へ

**サンプル** ⟶ **自然現象**

# ベイズの公式

**$P(X|Y) = P(Y|X)P(X)/P(Y)$**

- $P(X|Y)$ : データYが与えられた時のパラメータXの確率分布（事後分布）
- $P(Y|X)$ : パラメータXが与えられた時のデータYが得られる確率（尤度）
- $P(X)$ : パラメータXのデータを得る前の確率分布（事前分布）
- $P(Y)$ : 可能なすべてのパラメータの値のもとでデータYが得られる確率分布　$P(Y) = \Sigma_X P(Y|X)P(X)$

# ベイズの公式

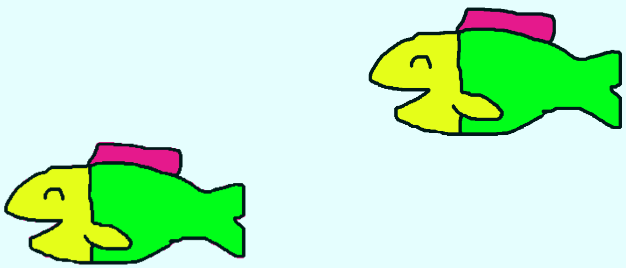
$$P(X|Y) = P(X,Y)/\Sigma_X P(X,Y)$$
$$= P(Y|X)P(X)/\Sigma_X P(Y|X)P(X)$$
$$= P(Y|X)P(X)/P(Y)$$

Xが興味の対象であり，Yは定数であることに注意すると，右辺の定数項P(Y)を無視して，

$P(X|Y) \propto P(Y|X)P(X)$（$\propto$：比例）

# 釣り人の漁場選択

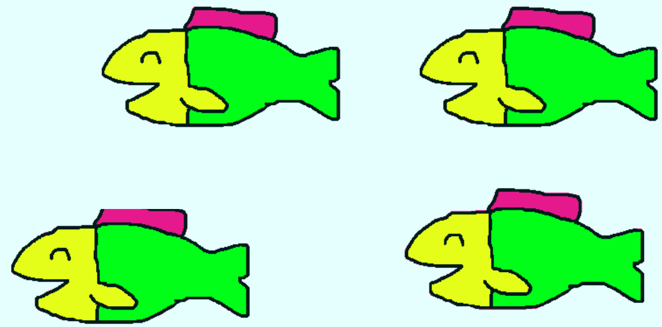
## 漁場A：0.8



**魚が釣れる確率 0.1**

**魚が釣れない確率 0.9**

## 漁場B: 0.2



**魚が釣れる確率 0.6**

**魚が釣れない確率 0.4**

# 例題

- 男性(37)が魚釣りに出かけて，魚を釣って帰ってきた．漁場は先のA，Bどちらかである．あなたは男性(37)がどちらの漁場に行ったと推理しますか？

**P(A|釣れた)？**

**P(B|釣れた)？**

**P(A|釣れた) {>, =, <} P(B|釣れた)**

# 解答

ベイズの公式より

P(A|釣れた) = P(釣れた|A)P(A)/P(釣れた)

P(釣れた|A) = 0.1, P(A) = 0.8

P(釣れた)

= P(釣れた|A)P(A)+P(釣れた|B)P(B)

= 0.1×0.8 + 0.6×0.2 = 0.08 = 0.2

# 解答

P(A|釣れた) = 0.1×0.8 / 0.2 = 0.4

同様に,

P(B|釣れた) = 0.6×0.2 / 0.2 = 0.6

故に,

P(B|釣れた) = 0.6 > 0.4 = P(A|釣れた)

男性（37）はBに行ったと推理するのが正しい！

# 従来の統計学とベイズ統計学

# 従来の統計学

真実はひとつである．パラメータθは決まっている．

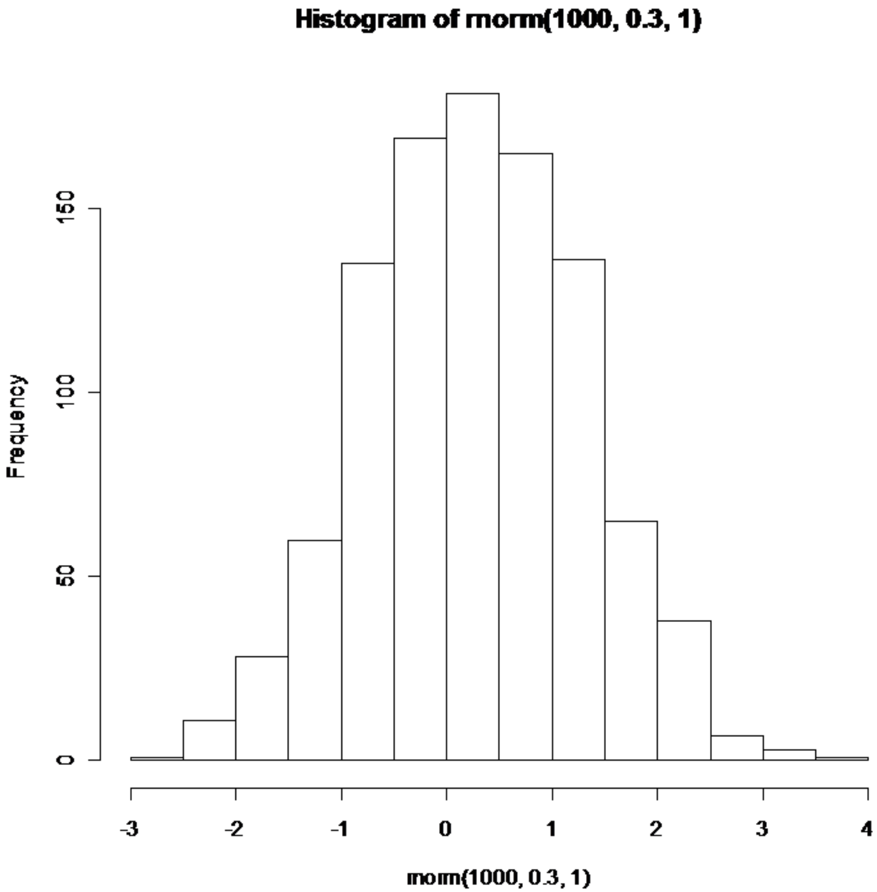
データはある真実の確率分布

$$x \sim P(X|\theta)$$

から得られる．

そのように考えた時，$P(\theta|X)$には意味がない．

# 従来の統計学




左図はN(0.3,1)からのランダムサンプルである.

しかし，もとの正規分布の平均θ=0.3は不変であって，確率的に変わるものではない.

# 従来の統計学

仮説検定

真

データ

| | H0 | H1 |
|---|---|---|
| H0 | 1－α | Type II error：β |
| H1 | Type I error：α | 1－β：検出力 |

$\alpha = \Pr(x \geq c | H_0)$

$\beta = \Pr(x < c | H_1)$

**しかし我々が真に知りたいのは$\Pr(H_0 | x \geq c)$とか$\Pr(H_1 | x < c)$とかではないのか？**

# 従来の統計学

信頼区間の解釈

- 頻度論流

  真の値は(データを得る前に)この区間の中に95%の確率で含まれる. 真の値がaであるかそうでないかは常に決まっている.

- ベイズ流

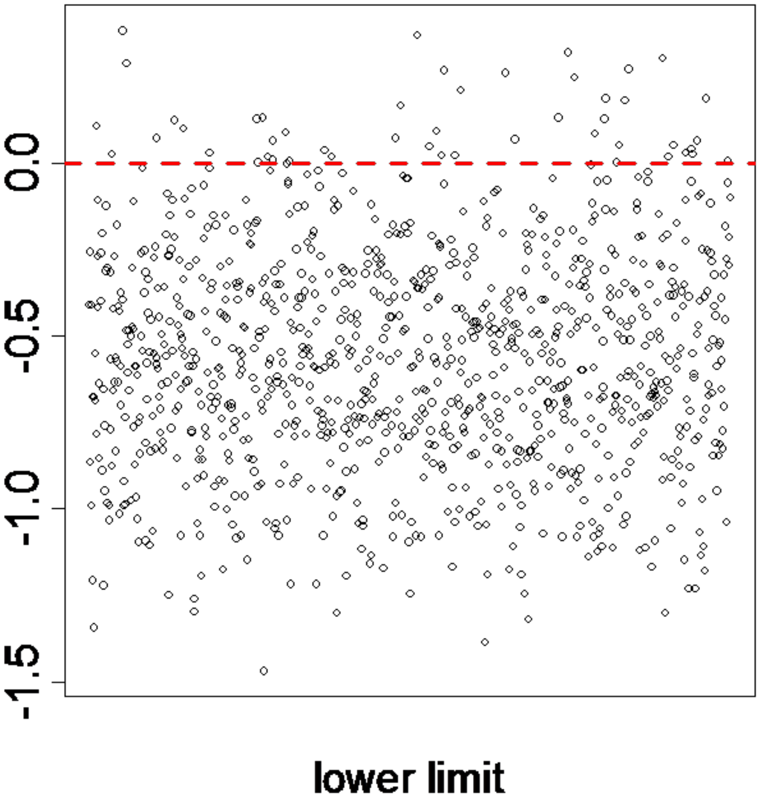
  真の値は(データを得た後に)この区間の中に95%の確率で含まれる. 真の値がaである確率はx%と言える.

# 従来の統計学

- 頻度論では，何回も何回も繰り返しサンプルする場合にどうなるかということを念頭においている．現在手元にあるデータはその一例にすぎない．ベイズ法では，現在手元にあるデータが全てである．
- このことは，シミュレーションをやってみると理解しやすい．

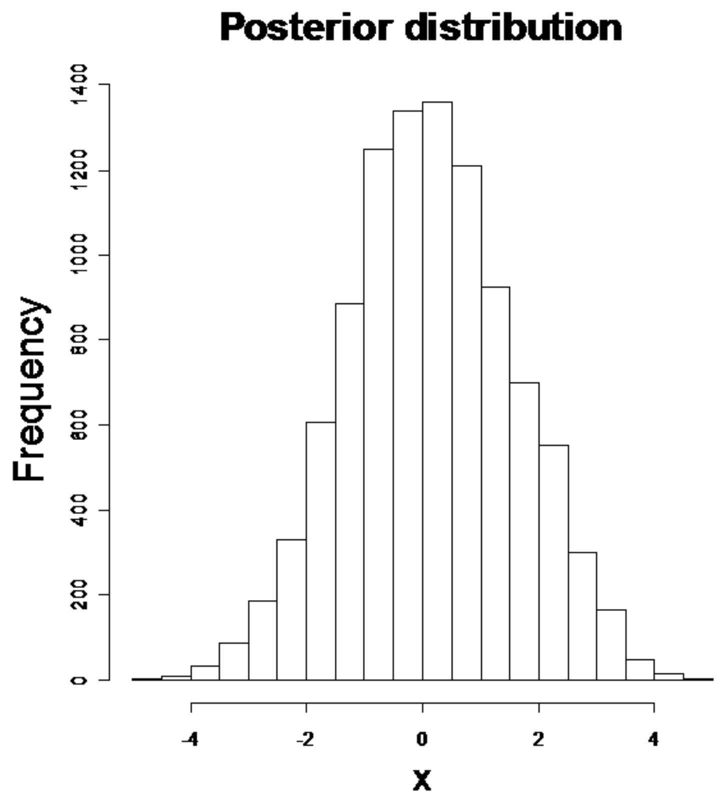
# 頻度論の信頼区間イメージ

```
x <- matrix(NA,nrow=10000,ncol=10)
for (i in 1:10){
  x[,i] <- rnorm(10000)
}
mx <- rowMeans(x)
vx <- apply(x,1,var)
lx <- mx+qt(0.05,9)*sqrt(vx/10)

> sum(mx > 0)/10000
[1] 0.5057
> sum(lx > 0)/10000
[1] 0.0513
```



# ベイズ信頼区間イメージ

```
theta <- seq(-10,10,by=0.001)
x <- rnorm(10)
f1 <- function(x) dnorm(x,theta,1)
post1 <- rowSums(apply(as.matrix(x),1,f1))
n1 <- sample(length(theta),10000,replace=T,prob=post1)

> quantile(theta[n1], probs=c(0.05,0.5,0.95))
     5%      50%      95% 
-2.189    0.111    2.530
```



# 頻度論とベイズ法の考え方

- 頻度論

  パラメータ　固定

  データ　確率変数

- ベイズ法

  パラメータ　確率変数

  データ　固定

未観測のデータがあるときは，どちらでも確率変数として扱われる．その場合，頻度論の方は，ランダム効果モデルと呼ばれるものを使うことになる．

# 従来の統計学

最尤法：データが得られた確率分布P(x|θ)を，データが与えられたときにパラメータθが得られる尤もらしさの指標として見てやり，尤度L(θ|x)とするとき，これを最大にする推定量を最尤推定量と呼ぶ.

例題：コインを3回投げて，結果は表，表，裏だった．表が出る確率は？

答：表が出る確率をpとする．確率は$p^2(1-p)$．これをpの関数と見て，最大化するpはp = 2/3.

# 最尤法

- 正規分布 N(μ,1)

$$P(x\mid\mu)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right)$$

これが最大になるのは$(x-\mu)^2 = 0$のとき，すなわち$\mu = x$が解．これは最小二乗法による解でもある．すなわち，この場合，尤度を最大にすることは，データに最も近いパラメータを選ぶことになる．

# 最尤法 ~ AIC

AIC = -2 log(尤度) + 2×パラメータ数

AICはKullback-Leibler距離（真のデータ発生プロセスとモデルとの近さの指標）と関連

モデルが一緒でパラメータ数も一緒なら，AIC最小モデル=KL距離最小モデルは尤度関数を最大にするものである
AICはモデル選択に使われる（色々と便利）

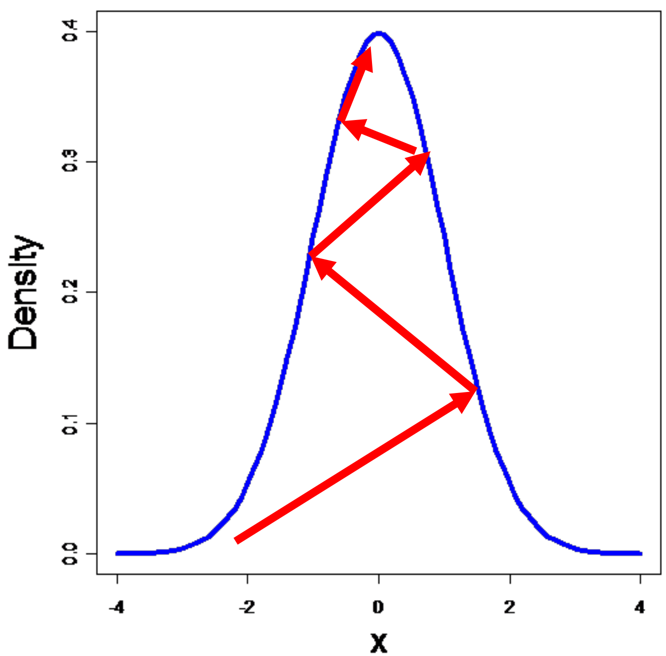
# 最尤法とベイズ法の関係

- $P(\theta|x) \propto P(x|\theta)P(\theta)$

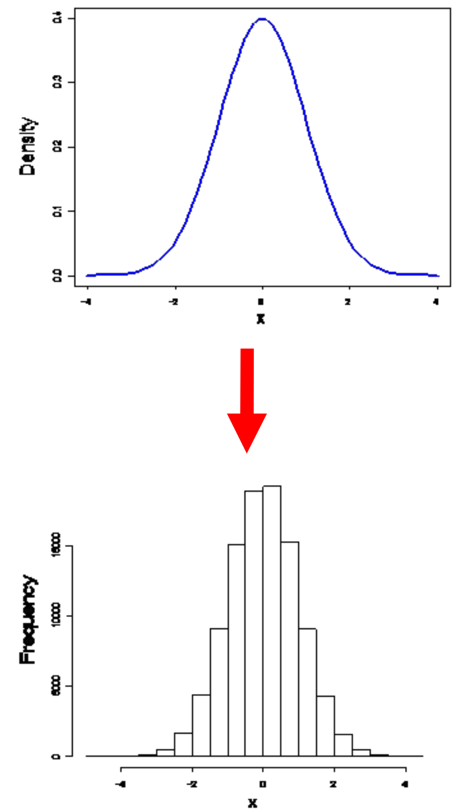
$\theta$は$\theta_1, \ldots, \theta_N$と離散的で，どれが起こる（事前）確率も等しいとする
$P(\theta) \propto 1$ より，
$P(\theta|x) \propto P(x|\theta) \propto L(\theta|x)$
尤度を最大化するパラメータは事後確率を最大化するものに他ならない
この意味では，ベイズ法 > 最尤法

# 最尤法とベイズ法

## 最尤法




## ベイズ法



# ベイズ法の利点

- 間接的なデータ（他種の情報，専門家の意見など）を事前分布の形で公式に取り込める
- 事後予測分布を使って，予測や外挿の問題を自然に扱える
- 複雑な階層モデルの扱いが最尤法より楽ちん
- 非専門家にも理解しやすい解釈
- 意思決定との密接な関係

# ベイズ法の欠点

- データの情報が少ない時，事前分布の影響が大きくなる
- どのパラメータに事前分布を与えるかで結果が変わる
- 一般に計算が大変である
- しばしば多少のプログラミングスキルを必要とする

# 水産資源管理に対するベイズ法の応用例

# 個体数トレンド

$N_t = N_0 \exp(-r\ t) \exp(u_t)$,   $u_t \sim N(0, s^2)$

$\log(N_t) = \log(N_0) - r\ t + u_t$

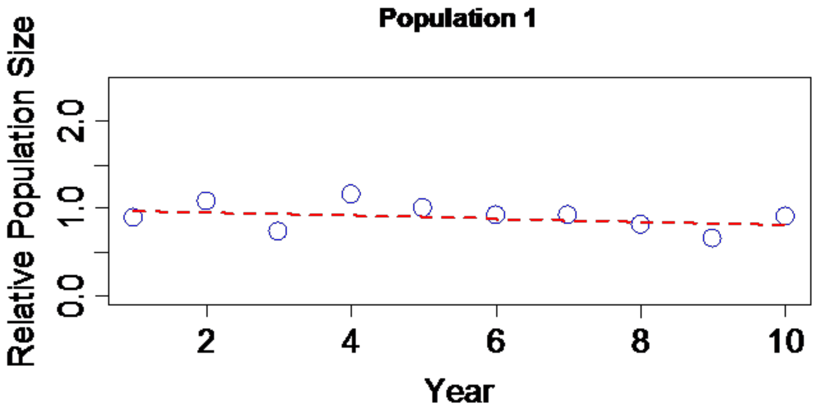
$\log(D_t) = -r\ t + u_t$   $(D_t = N_t/N_0)$

sは既知とするとき,

尤度関数  $\log(L) = -\Sigma\{\log(D_t) - ( - r\ t)\}^2/s^2$

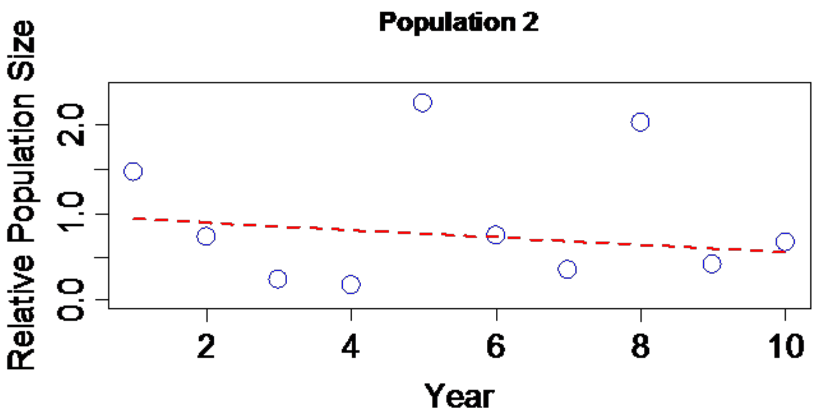
rの推定が興味の対象.

# 2つの個体群トレンドデータ




r = -0.02（年に約2%減）

P-値 0.04 (t-検定)




r = -0.06（年に約6%減）

P-値 0.23 (t-検定)

有意差のあるなしで，Pop. 1は減少している危険な個体群，Pop. 2は減少しているとは言えない安心個体群と評価する科学者をあなたはどう思うか？

# グリッド法による事後確率の計算

$s_1 = 0.2$, $s_2 = 0.9$は既知

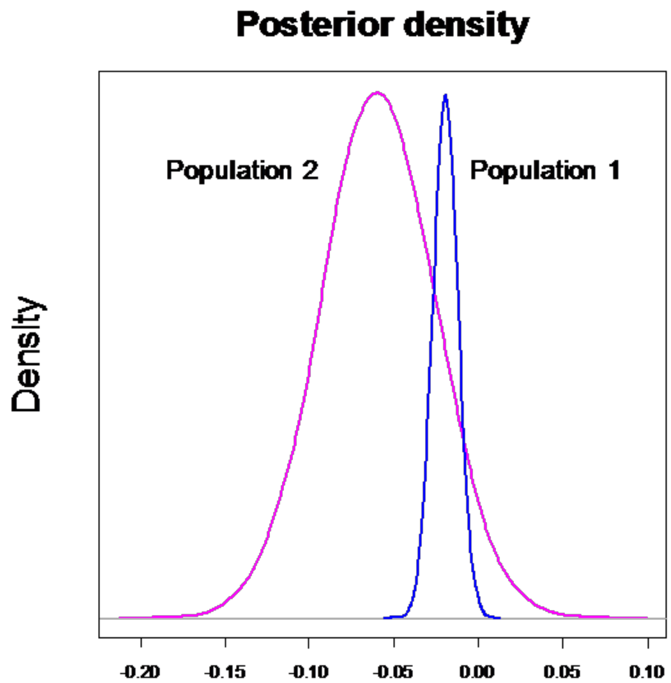
事後サンプルの生成（グリッド法）

1. 減少率rのグリッドデータを作る

   -0.300,-0.299,-0.298,…

2. それぞれのrに対して尤度を計算
3. 計算した尤度の大きさに従ってrをサンプルする　→ rの事後サンプル

# 事後分布



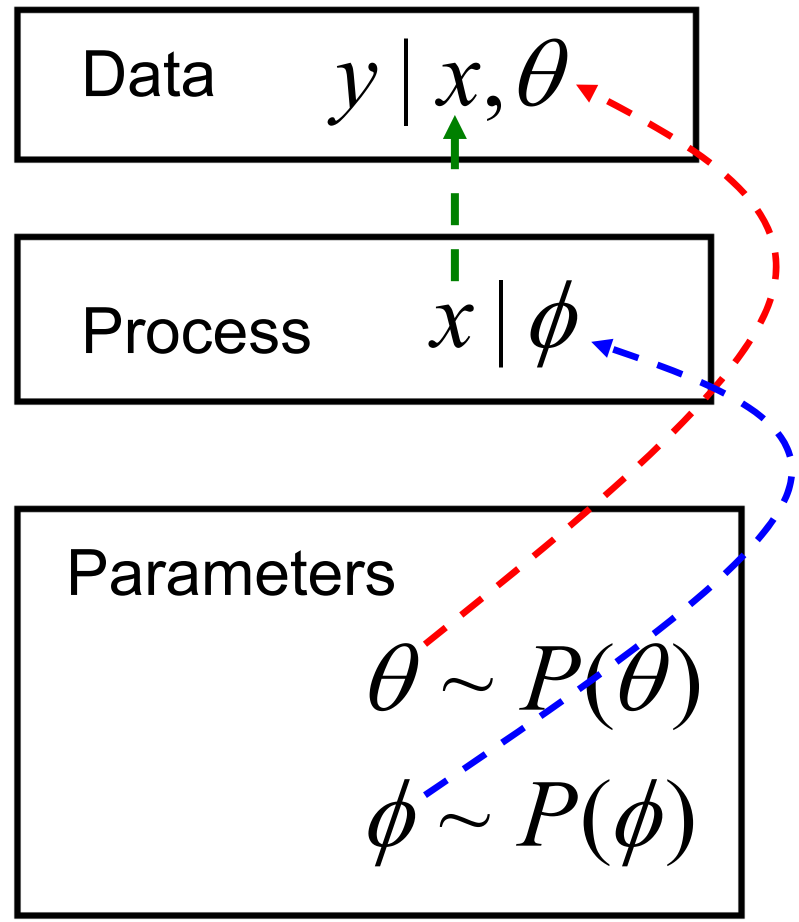

| 事後確率 | r < - 0.05 | - 0.05 ≤ r < 0.0 | r ≥ 0.0 |
|---|---|---|---|
| Population 1 | 0.000 | 0.996 | 0.004 |
| Population 2 | 0.604 | 0.361 | 0.035 |

| Decision | r < - 0.05 | - 0.05 ≤ r < 0.0 | r ≥ 0.0 | E(Loss) for Pop. 1 | E(Loss) for Pop. 2 |
|---|---|---|---|---|---|
| Rapid decline | 0 | 0.5 | 1 | 0.502 | 0.216 |
| Slow decline | 0.5 | 0 | 0.5 | 0.002 | 0.320 |
| No decline | 1 | 0.5 | 0 | 0.498 | 0.784 |

# 階層ベイズモデル，モデル選択

# 階層ベイズモデル

**P(parameters, process|data)**
**∝P(data|process, parameters)**
**×P(process|parameters)**
**×P(parameters)**




**P(θ, φ, x|y)**
**∝P(y|x, θ)P(x|φ)P(θ)P(φ)**

# 階層ベイズモデルの例

状態空間モデル

状態方程式 $B_{t+1} = \{B_t + rB_t(1-B_t/K)-C_t\}\exp(v_t)$

観測方程式 $CPUE_t = q\ B_t \exp(\varepsilon_t)$

パラメータ $r \sim U[0, 0.3]$, $K \sim U[0.1M, 10M]$

$\log(q) \sim U[-5, 5]$

$\varepsilon_t \sim N(0,\sigma^2)$

$v_t \sim N(0, \tau^2)$

$y = CPUE$, $x = B$, $\theta= (q, \sigma^2)$, $\varphi= (r, K, \tau^2)$

# モデル選択：Bayes Factor

$$\mathrm{BF} = \frac{p(y|H_2)}{p(y|H_1)} = \frac{\int p(y,\theta_2|H_2)d\theta_2}{\int p(y,\theta_1|H_1)d\theta_1}$$

事後確率の比 = 事前確率の比×Bayes factor

尤度比検定のベイズ版と考えられる.

周辺事後分布の計算が大変.

# モデル選択: BIC, DIC

- **BIC**

  $BIC = -2\log(L(\theta^*|x))+\log(n)p$
- **DIC**

  $DIC = deviance + pD$

  deviance: $E[-2\log(L(\theta|x))]$,

  pD: $E[-2\log(L(\theta|x))] - [-2\log(L(\theta^*|x))]$

  WinBUGSでアウトプットされる

  事後分布がlog concaveであることを仮定しているので，そうでないときはうまくいかない

BIC，DICともに最小の値をとるものがベストモデル.

# 参考文献


Malakoff, D. 1999. Bayes offers a ‘new’ way to make sense of numbers. Science 286: 1460-1464.

Wade, P. R. 2000. Bayesian methods in conservation biology. Conservation Biology 14: 1308-1316.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8: 2-14.