一般化線形モデルと
循環統計モデル

中央水研
岡村　寛

今日のお話

- 循環統計とは？
- 一般化線形モデル（GLM）
- 一般化線形混合モデル（GLMM）
- 循環統計と一般化線形モデル：水産資源データへの応用
- まとめ

循環統計

Circular statistics

単位円周上の確率分布

- 月日（いつ出産するか）
- 方角（海鳥が向かう方向，風向き）
- 経度

ぐるっとまわって同じになるデータ

von Mises分布

二変量正規分布

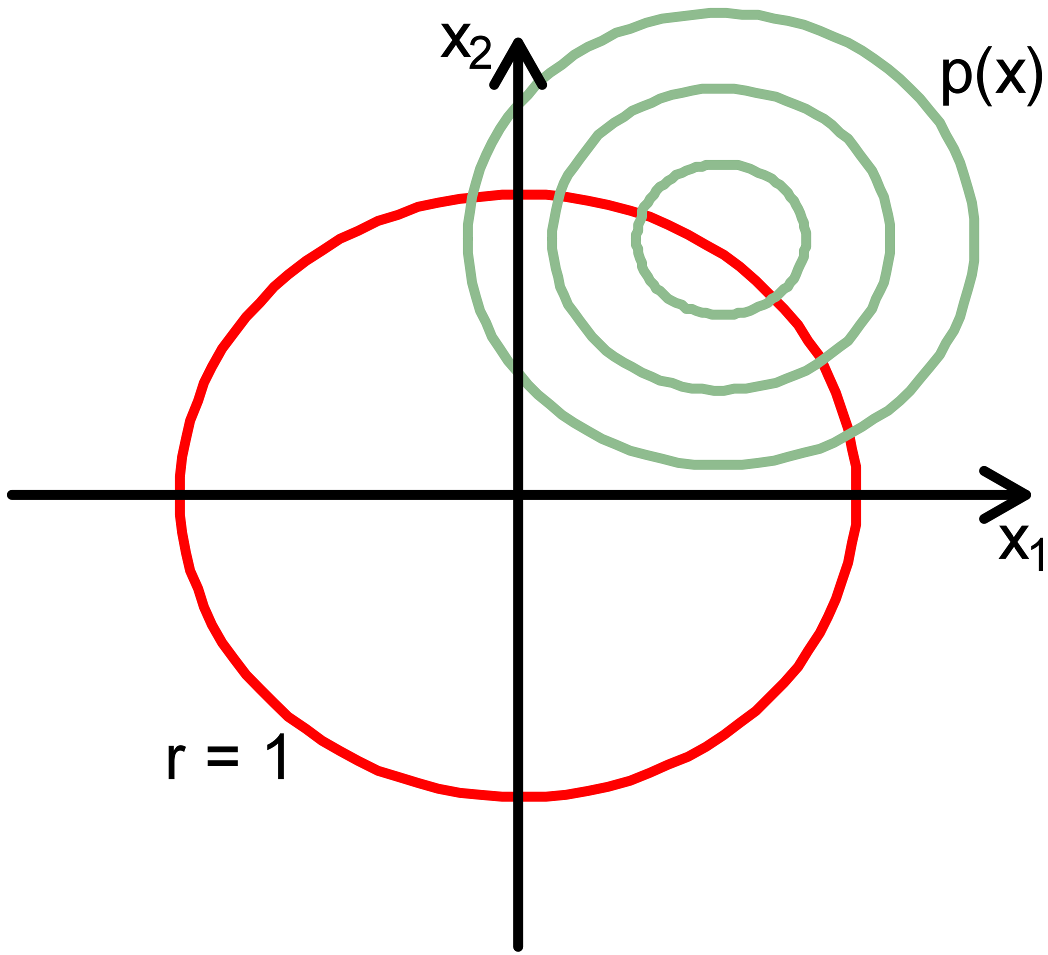
$$f(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2} \right\}$$

x_1 = rcosθ, x_2 = rsinθ, μ_1 = $r_0$cosθ，μ_2 = $r_0$sinθとおいて，r = 1とすると，

　f(θ) ~ exp[m cos(θ ー θ_0)]

θ_0は分布の平均，mは集中度パラメータ（精度）

von Mises分布



Wrapped Cauchy分布

X：実数上の確率変数，g(X)：Xの確率密度関数

θ ≡ X [mod 2π]とすると，

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2k\pi)$$

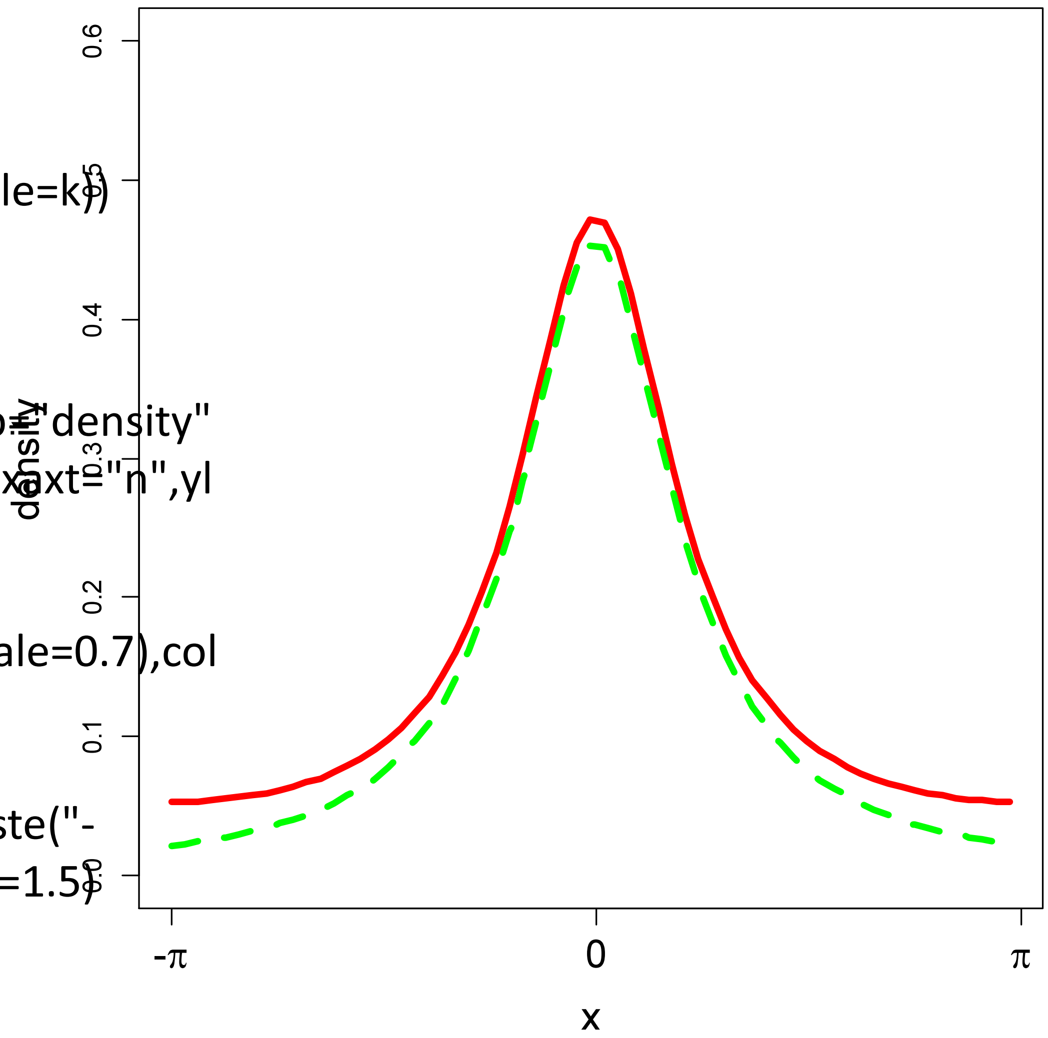
として確率分布を構成できる．実数軸を単位円上に巻き込んだものになっているので，wrapped分布という．g(X)をCauchy分布にしたものがwrapped Cauchy分布．きれいな式になる：

$$f(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho\cos(\theta - \mu)}$$

μ：平均，
ρ：集中度パラメータ

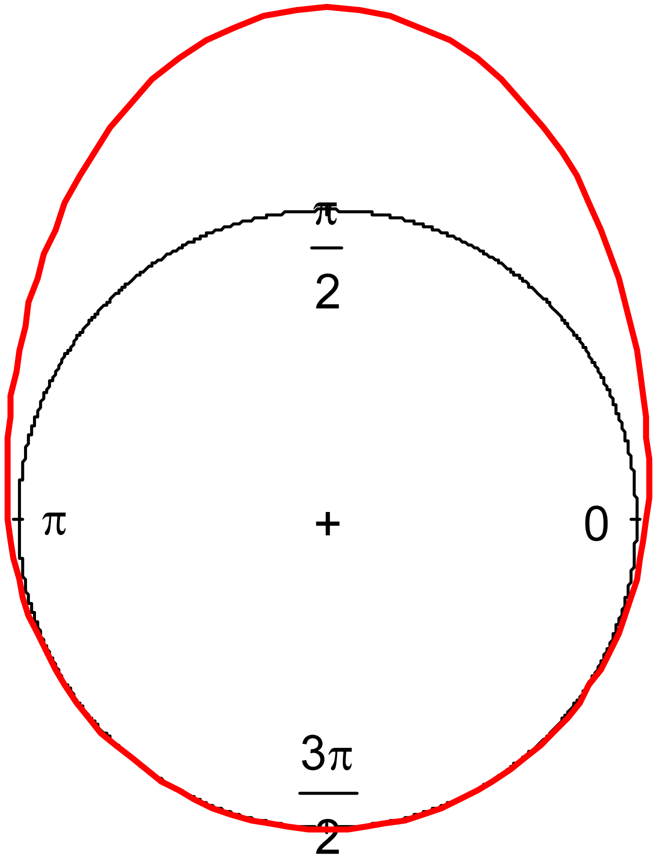
Wrapped Cauchy分布

```
wcauchy <- function(x,mu=0,k=0.7)
sum(dcauchy(x+seq(-
50*pi,50*pi,by=2*pi),location=mu,scale=k))

x <- seq(-pi,pi,by=0.1)

plot(x,sapply(x,wcauchy),xlab="x",ylab="density"
,col="red",type="l",lwd=4,cex.lab=1.5,xaxt="n",yl
im=c(0,0.6))

lines(x,sapply(x,dcauchy,location=0,scale=0.7),col
="green",lwd=4,lty=2)

axis(side=1,c(-pi,0,pi),c(expression(paste("-
",pi,sep="")),0,expression(pi)),cex.axis=1.5)
```

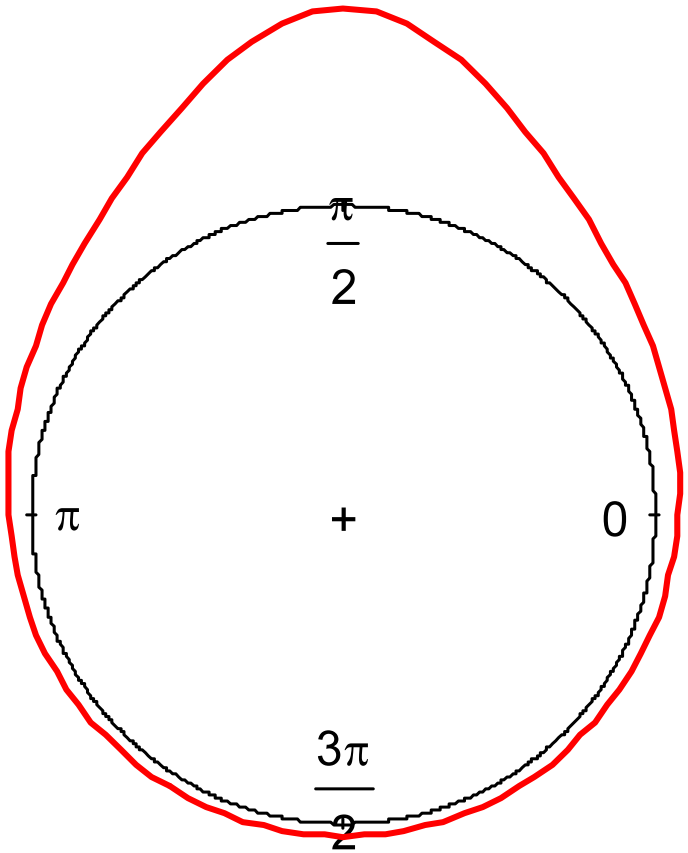


例

von Mises分布：μ=π/2，m=3



wrapped Cauchy分布：μ=π/2，ρ=0.6



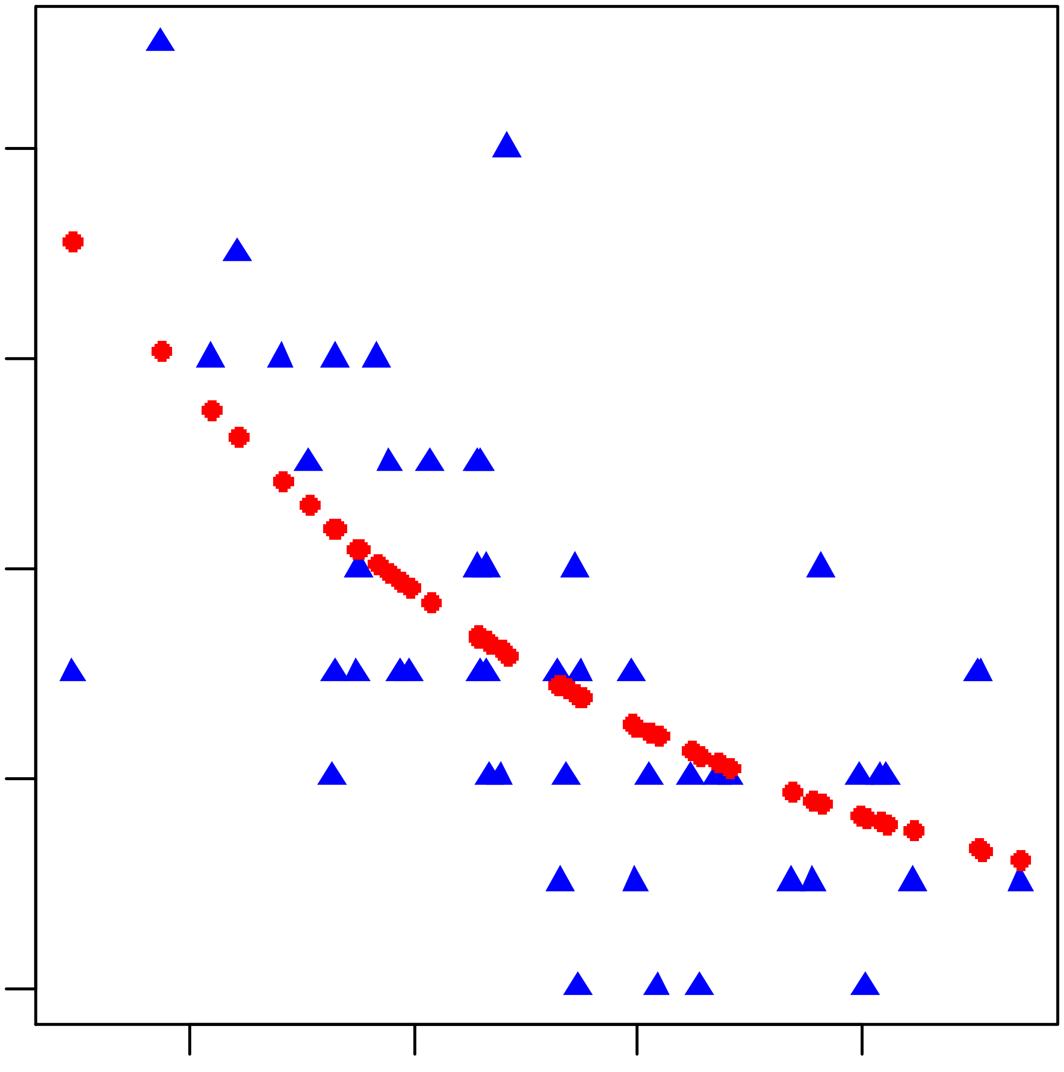
一般化線形モデル（GLM）

- 回帰や分散分析の一般化
- 正規分布以外の分布（指数分布族）を扱える
- データの特徴にあわせた柔軟な分析ができる（パラメータに関して非線形なモデリングも含む）
- ロジスティック回帰
- ポアソン回帰

例（ポアソン回帰）

```
a <- 1
b <- -0.3
x <- rnorm(50)
y <- rpois(50,exp(a+b*x))

res <- glm(y~x,family="poisson")
```

例（ポアソン回帰）



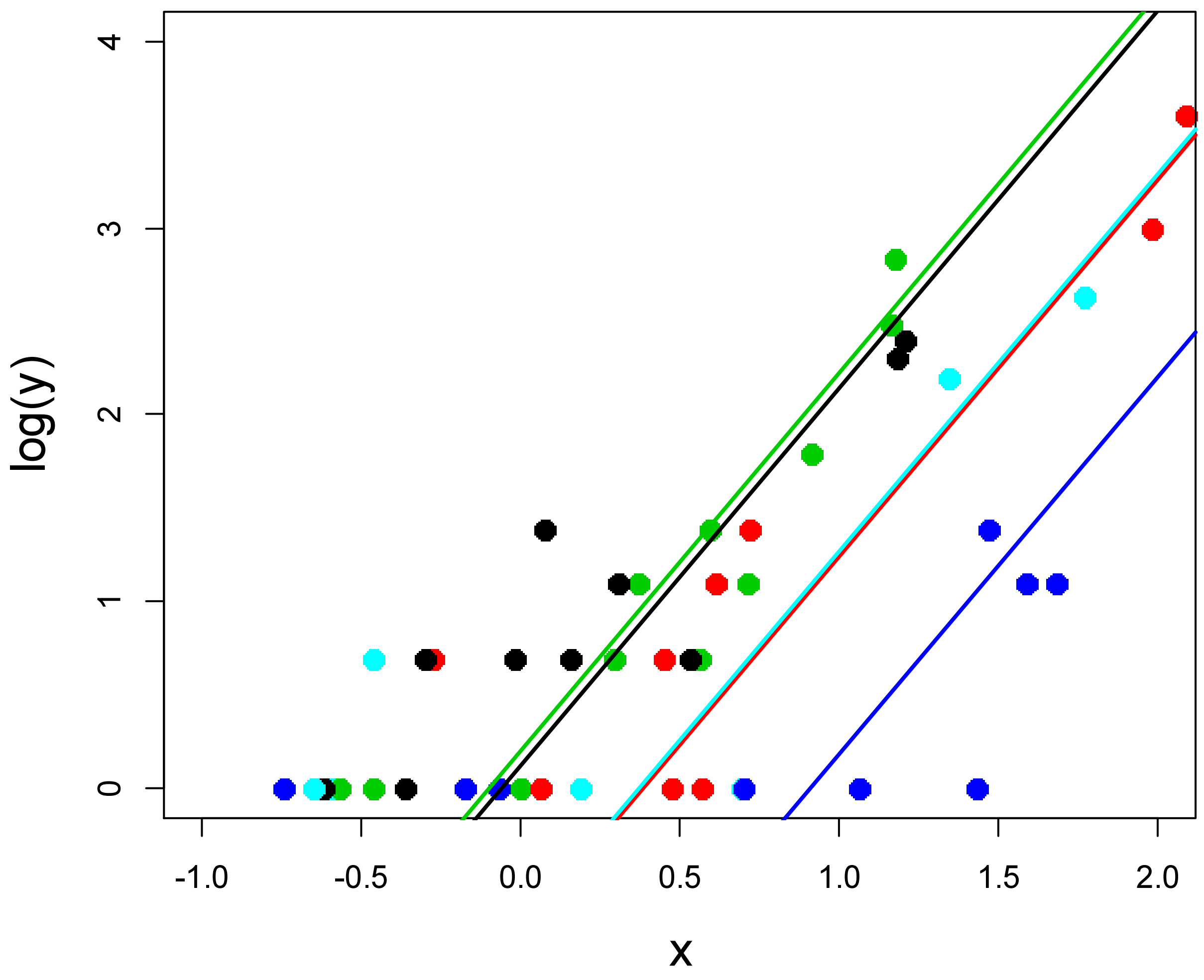
一般化線形混合モデル（GLMM）

- GLMにランダム効果を取り込んだ拡張版

```
library(lme4)
site <- rnorm(5)
z <- sample(5,100,replace=TRUE)
x <- rnorm(100)
a <- 0.5; b <- -2
y <- rpois(100,exp(-(a+site[z]+b*x)))
res <- glmer(y~x+(1|z),family="poisson")
```

ランダム効果

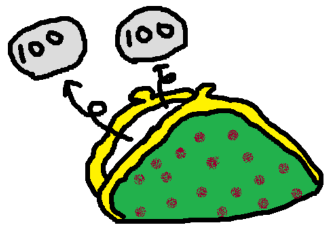
例（ポアソン混合モデル）



魚の年齢

- いつ、どれだけとればよいかを決めるには年齢の情報が不可欠

  ・・・年齢を間違うと資源評価・管理の失敗につながる (Campana 2001)
- 失敗例　　Beamish and McFarlane (1983)

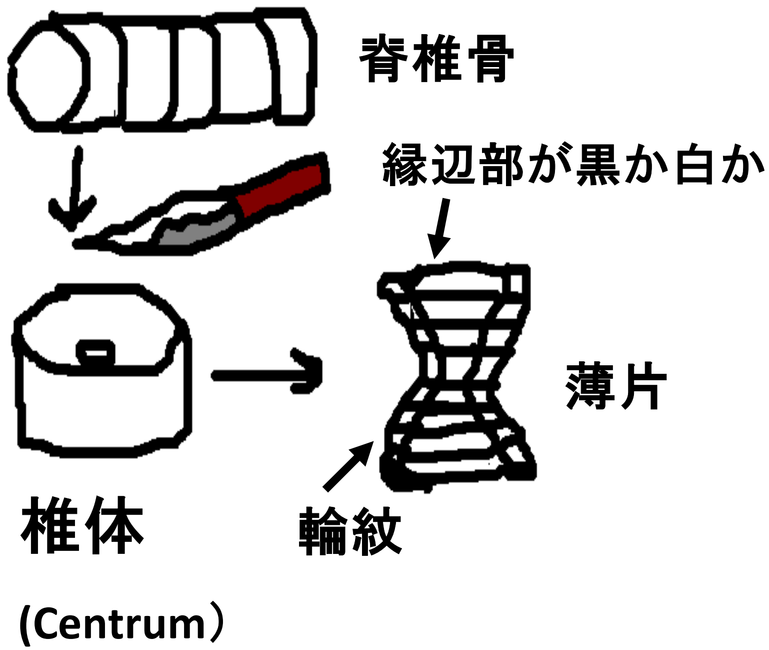


太平洋のパーチ（メヌケの一種）漁業では，年齢の過小推定により増加率を過大推定．結果，適正な漁獲率を誤ったために資源管理に失敗し数百万ドルの損失をこうむった

サメの輪紋周期性検証法
(centrum edge analysis)

- サメの場合は、脊椎骨に輪紋が形成される.
- 脊椎骨上の輪紋を観察し、縁辺部を黒(凹)か白(凸)かに分類.→月ごとにまとめる.

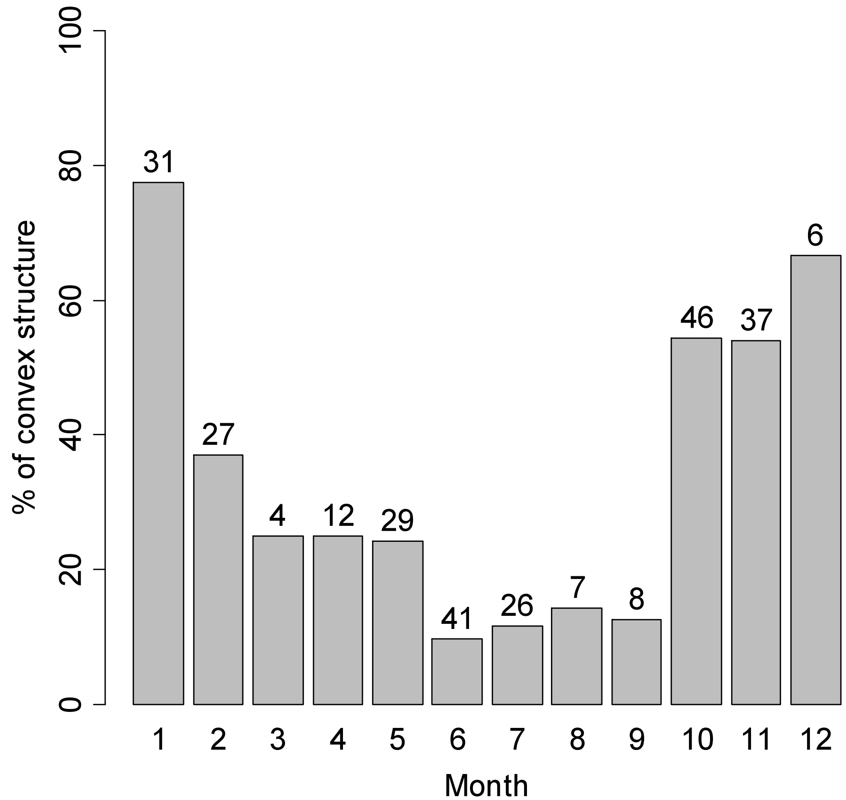
例：1月　黒 20　白5



CEAデータ

北太平洋アオザメの
輪紋検証データ

見た目では年に1本の
ように見える（右のヒストグラムのピークが1個なら年1輪紋，ピークが2個なら年2輪紋，など）

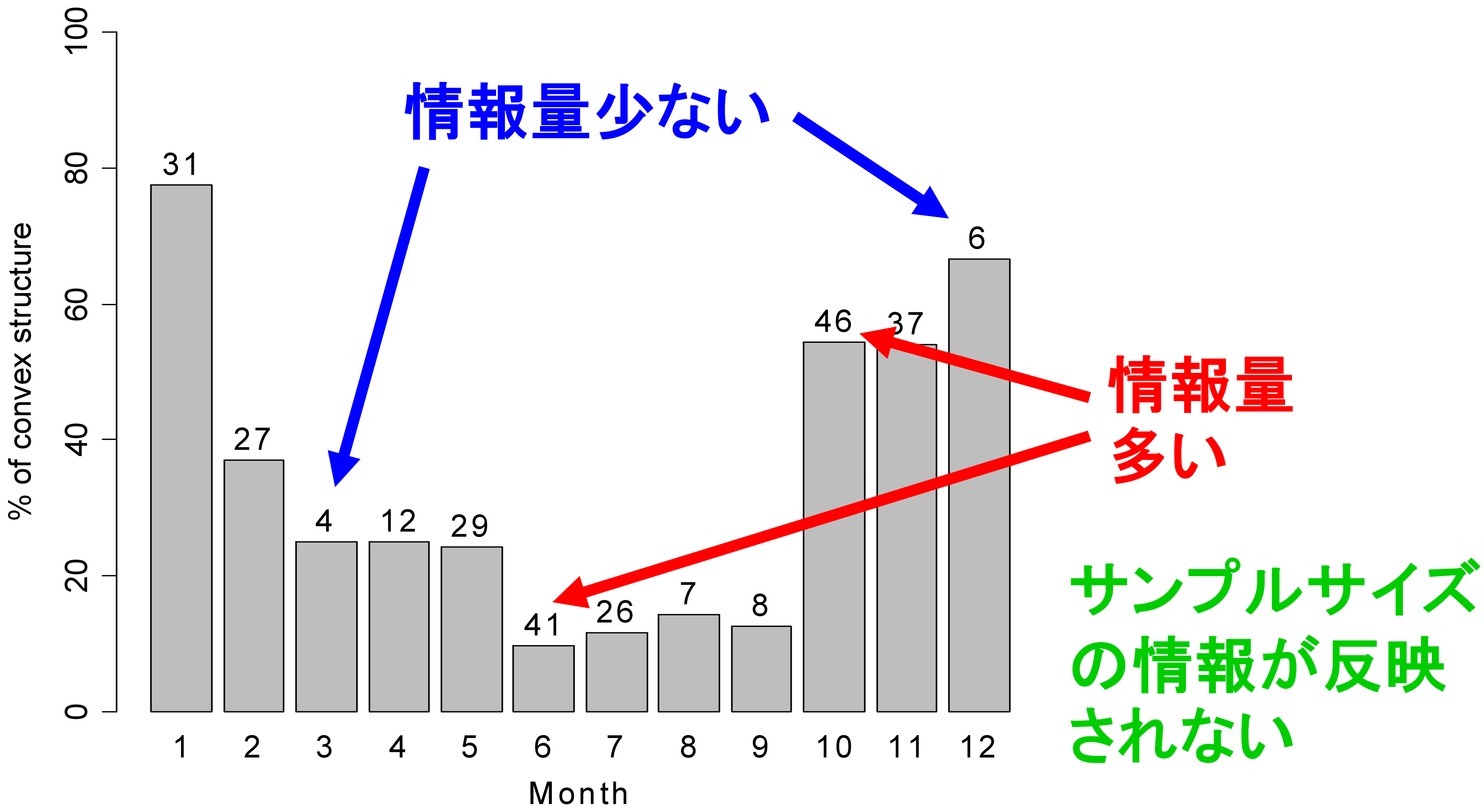


問題点

- 従来は目で見て，主観的に年1本と判断していた.
- 客観的に周期性を判定する手法がない.
- サメ類では年2本と考えられている種がある
  北大西洋のアオザメ Pratt and Casey 1983
  ウバザメ

年齢査定の誤り → 資源管理の失敗 → 生態的･経済的損失大

目で見て周期性を判定すると



新しい統計モデル

- 二項応答変数モデル+循環統計分布
- 輪紋周期性を仮定した上で，データにモデルをフィットして，最尤法によりパラメータを推定
- どの周期性が尤もらしいかは，赤池情報量規準（AIC）を用いて判定する．AICが小さいモデルは最も良いモデルである．

新しいモデル

$\mathbf{x} = (x_1, \ldots, x_{12})$　　各月のconvexの個数

$\mathbf{n} = (n_1, \ldots, n_{12})$　　各月のサンプルサイズ

各月の累積日数をラジアンに変換したもの

データが得られる確率は,

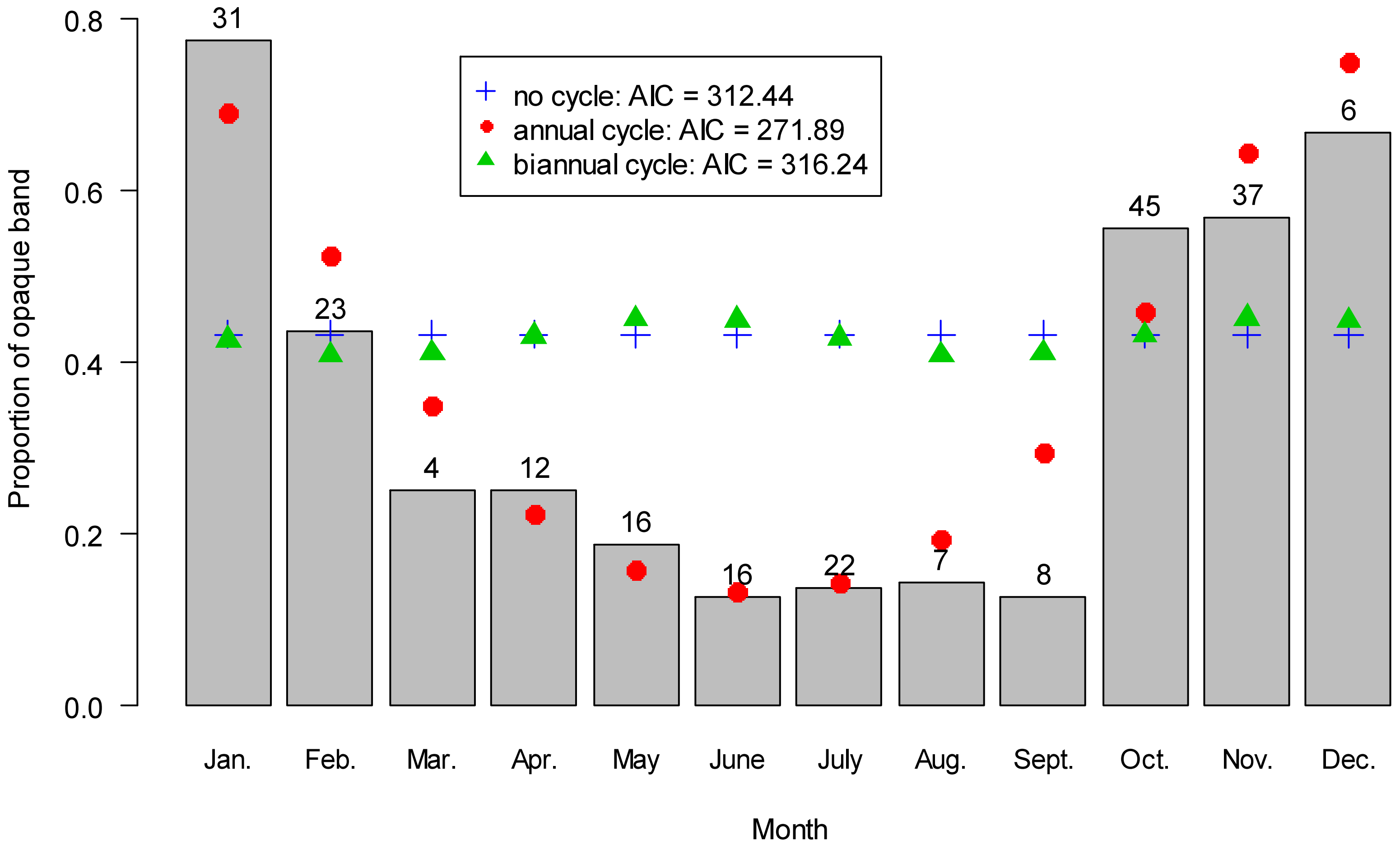
$$P(\mathbf{x}, \mathbf{n} \mid \theta) = \prod_{i=1}^{12} \{\gamma \int_{y_i}^{y_{i+1}} f(u \mid \alpha, \mathbf{m}, \mathbf{k}) du\}^{x_i} \{1 - \gamma \int_{y_i}^{y_{i+1}} f(u \mid \alpha, \mathbf{m}, \mathbf{k}) du\}^{n_i - x_i}$$

パラメータ　$\theta = (\gamma, \alpha, \mathbf{m}, \mathbf{k})$

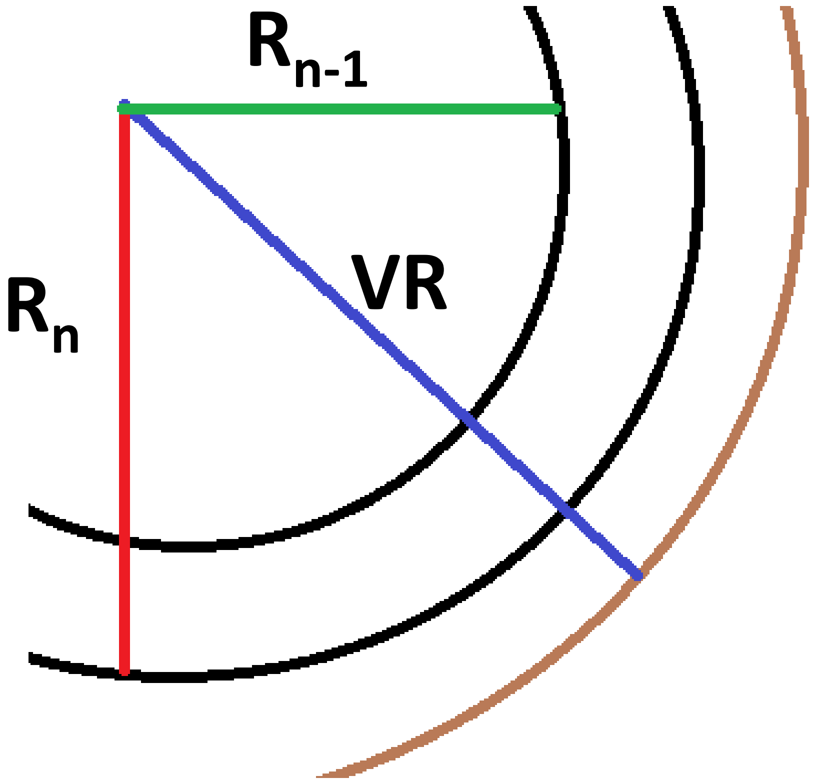
ここで, 各月のconvex structure（凸）の出現比率は,

von Mises分布の混合分布

北太平洋アオザメに対する適用結果



Marginal Increment Analysis (MIA)

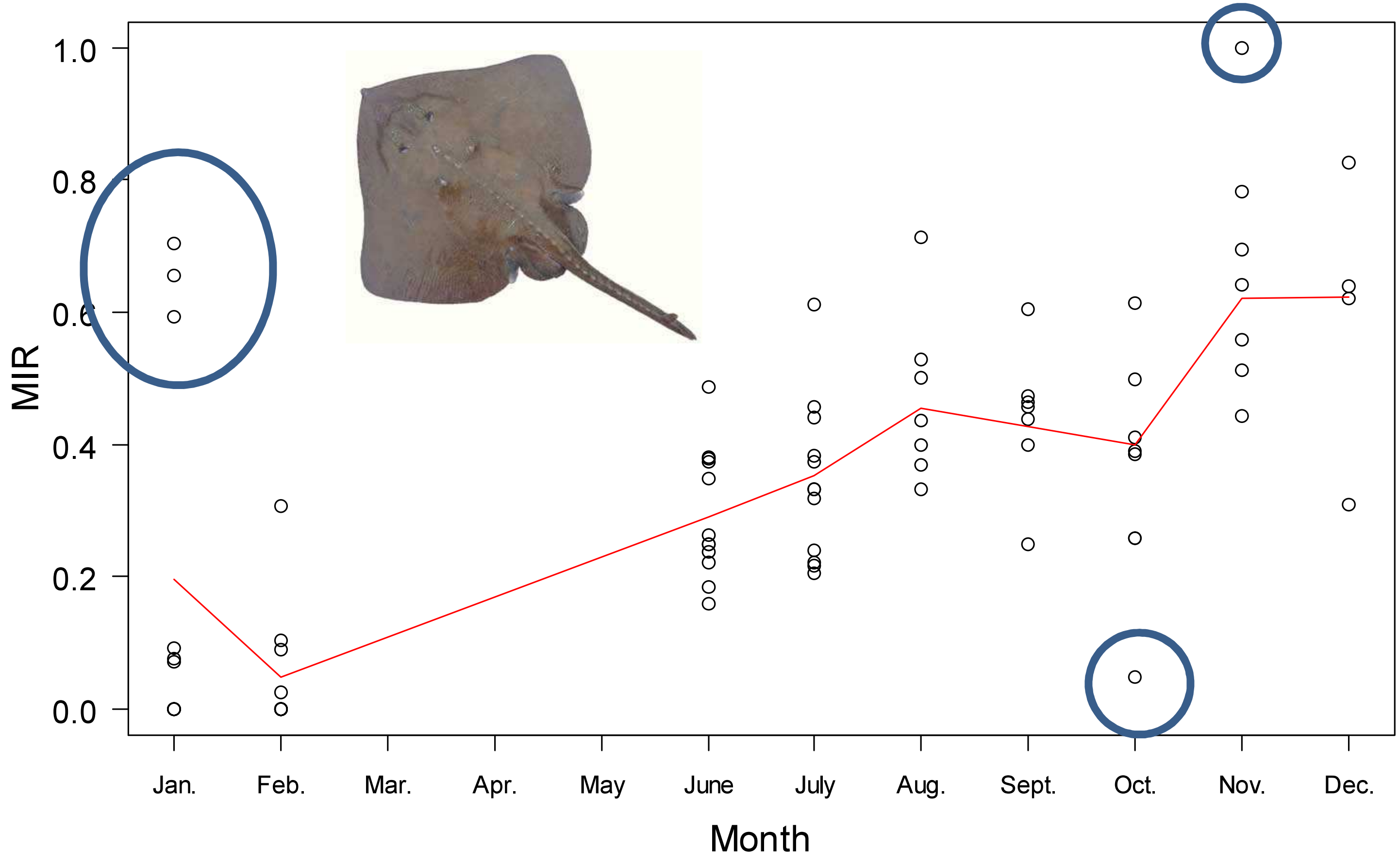


$$MIR = (VR - R_n)/(R_n - R_{n-1})$$

月に対してMIRをプロットして，もし周期的なピークが発見されれば，年齢形質は1年のサイクルで形成されると考える

MIAデータ

Alaska skate data

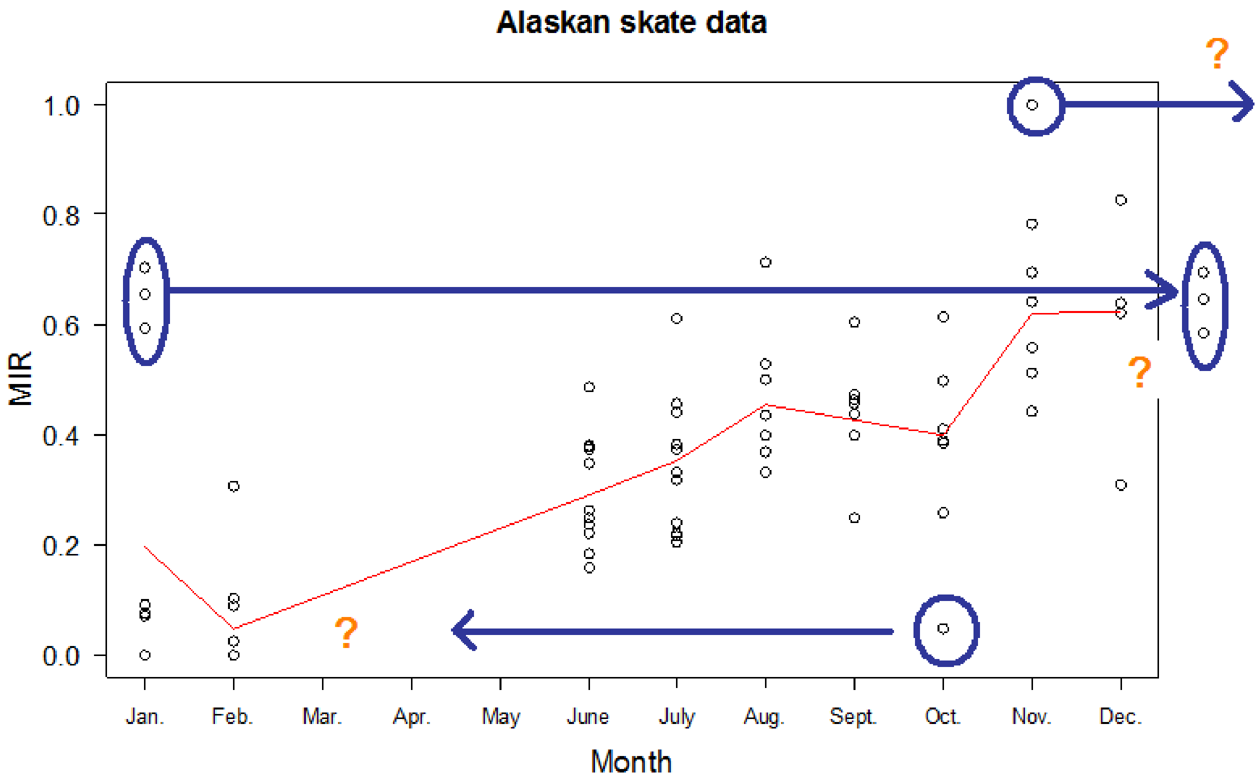


Cf. Matta and Gunderson (2006) EBF 80: 309-323

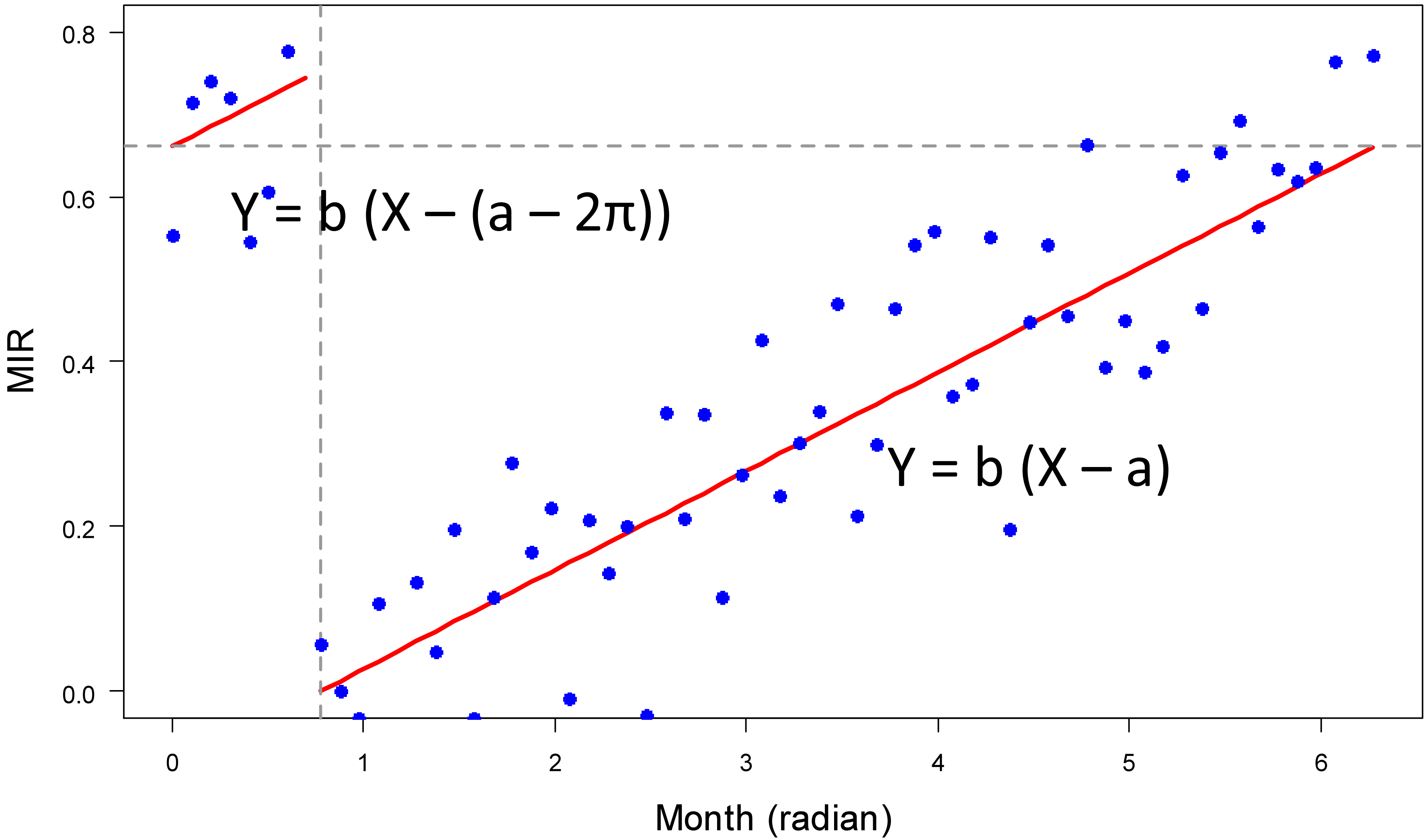
問題

月データは周期的.

伝統的な方法(ANOVA, K-W test)は適切であろうか？



ぐるっと回る線形回帰



周期1モデル

基本モデル $Y = b(X-a)+\varepsilon,\ \varepsilon \sim N(0, \sigma^2)$

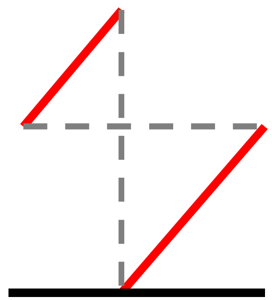
$a = a_0 + \theta$ where $0 < a_0 < 2\pi,\ \theta \sim TWC_{[-\pi/2,\ \pi/2]}(0, \rho)$, $(-\pi/2 < \theta < \pi/2)$

[*, *]の範囲のTruncated Wrapped Cauchy分布のランダム効果

この範囲の外側では輪紋形成は起こらないとする
（-0.5π ~ 0.5π = 約6ヶ月）

if $(a < 0)\ X = X - 2\pi$,
if $(a >= 2\pi)\ X = X + 2\pi$,

if $(X < a)\ \mu = b(X-(a-2\pi))$,
if $(X >= a\ \&\ X < a+2\pi)\ \mu = b(X-a)$,
if $(X >= a+2\pi)\ \mu = b(X-(a+2\pi))$.



尤度関数 $\int N(Y|\mu, \sigma^2)\ TWC(\theta|0, \rho)\ d\theta$

周期2モデル

基本モデル $Y = b(X-a) + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$

$a = a_0 + \theta$ where $0 < a_0 < \pi,\ \theta \sim TWC_{[-\pi/4,\ \pi/4]}(0, \rho)$ (**$-\pi/4 < \theta < \pi/4$**)

if $(a < 0)\ X = X - \pi$,

if $(a >= \pi)\ X = X + \pi$,

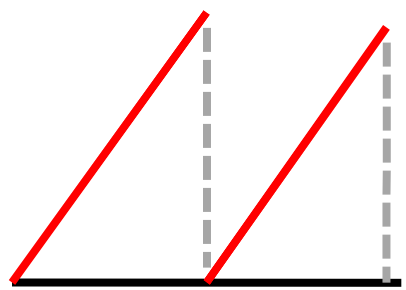
if $(X < a)\ \mu = b(X-(a-\pi))$,

if $(X >= a\ \&\ X < a+\pi)\ \mu = b(X-a)$,

if $(X >= a+\pi)\ \mu = b(X-(a+\pi))$.

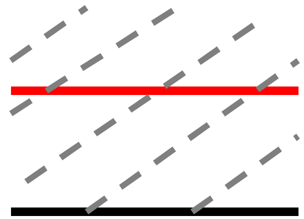
尤度関数 $\int N(Y|\mu, \sigma^2)\ TWC(\theta|0, \rho)\ d\theta$

同じパターンが半年ごとに繰り返される



周期0モデル

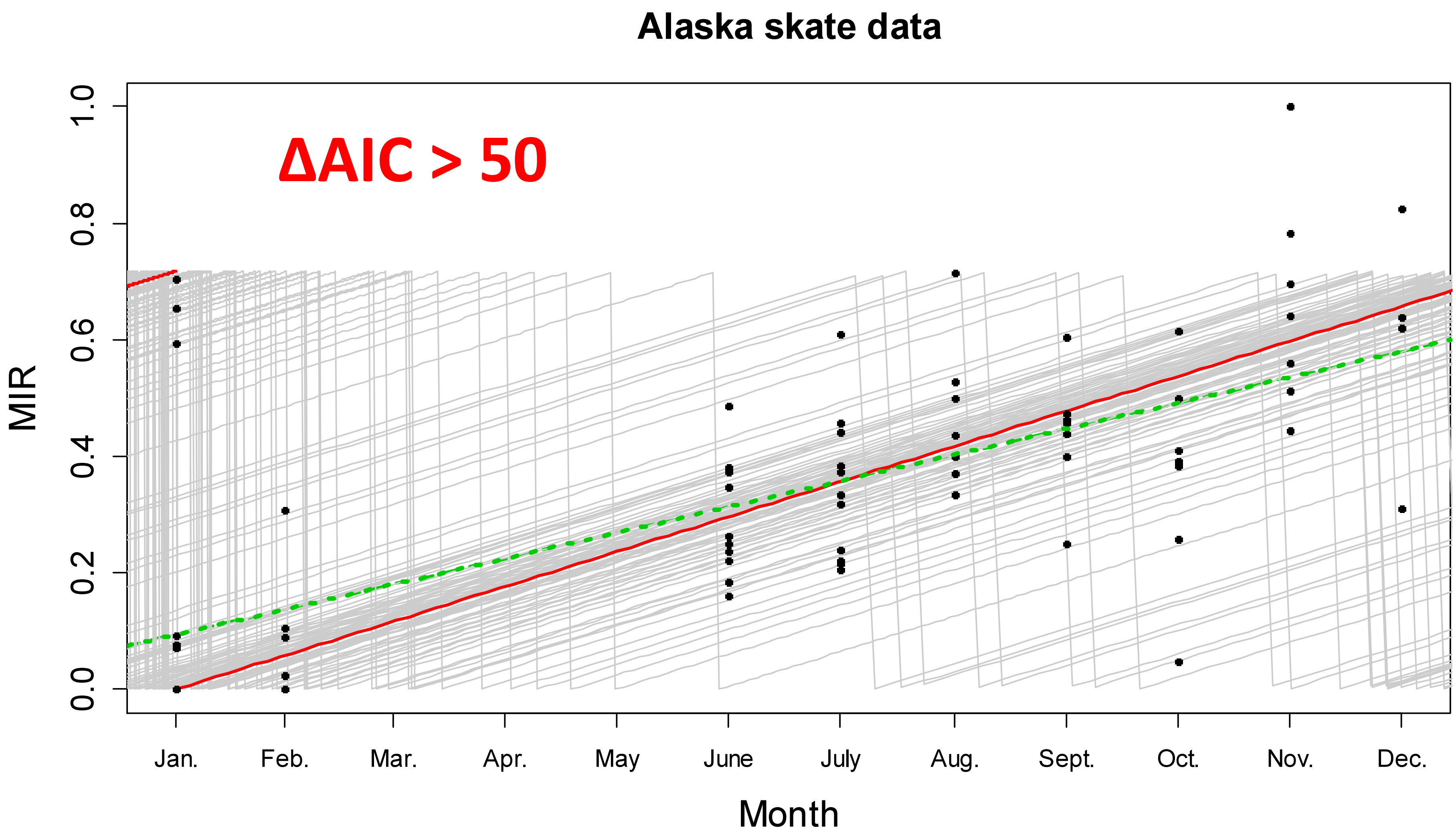
- ランダム効果部分を一様分布 U[-π, π]で置き換えたものが周期0に対応する（面全体を直線で塗りつぶすようなもの）. あとは周期1モデルと同じ.



基本モデル $Y = b(X-a)+\varepsilon,\ \varepsilon \sim N(0, \sigma^2)$

$a = a_0 + \theta$ where $0 < a_0 < 2\pi,\ \theta \sim U[-\pi, \pi],\ (\boldsymbol{-\pi < \theta < \pi})$

実データ分析結果



Alaska skate data
ΔAIC > 50
MIR
1.0
0.8
0.6
0.4
0.2
0.0
Jan.
Feb.
Mar.
Apr.
May
June
July
Aug.
Sept.
Oct.
Nov.
Dec.
Month

まとめ

- 周期データはいろいろな形で出てくるが，GLMと循環統計分布を組み合わせることにより，客観的な分析が可能となる
- より正確なパラメータ推定，より効率の良い調査デザイン設計などに役立つであろう
- 将来課題：多項分布への拡張，個体群動態モデルとの統合

文献

- Okamura, H. & Semba, Y. (2009) A novel statistical method for validating the periodicity of vertebral growth band formation in elasmobranch fishes. Can. J. Fish. Aquat. Sci. 66:771-780.
- Okamura, H., Punt, A.E., Semba, Y. & Ichinokawa, M. (2013) Marginal increment analysis: a new statistical approach of testing for temporal periodicity in fish age verification. J. Fish. Biol. 82: 1239-1249.