NagoyaStat#6

データ解析のための統計モデリング入門 11章

空間構造のある階層ベイズモデル

2017.5.24 Fri. Yahoo名古屋

@nonsabotage

はじめに

久保先生からのメッセージ

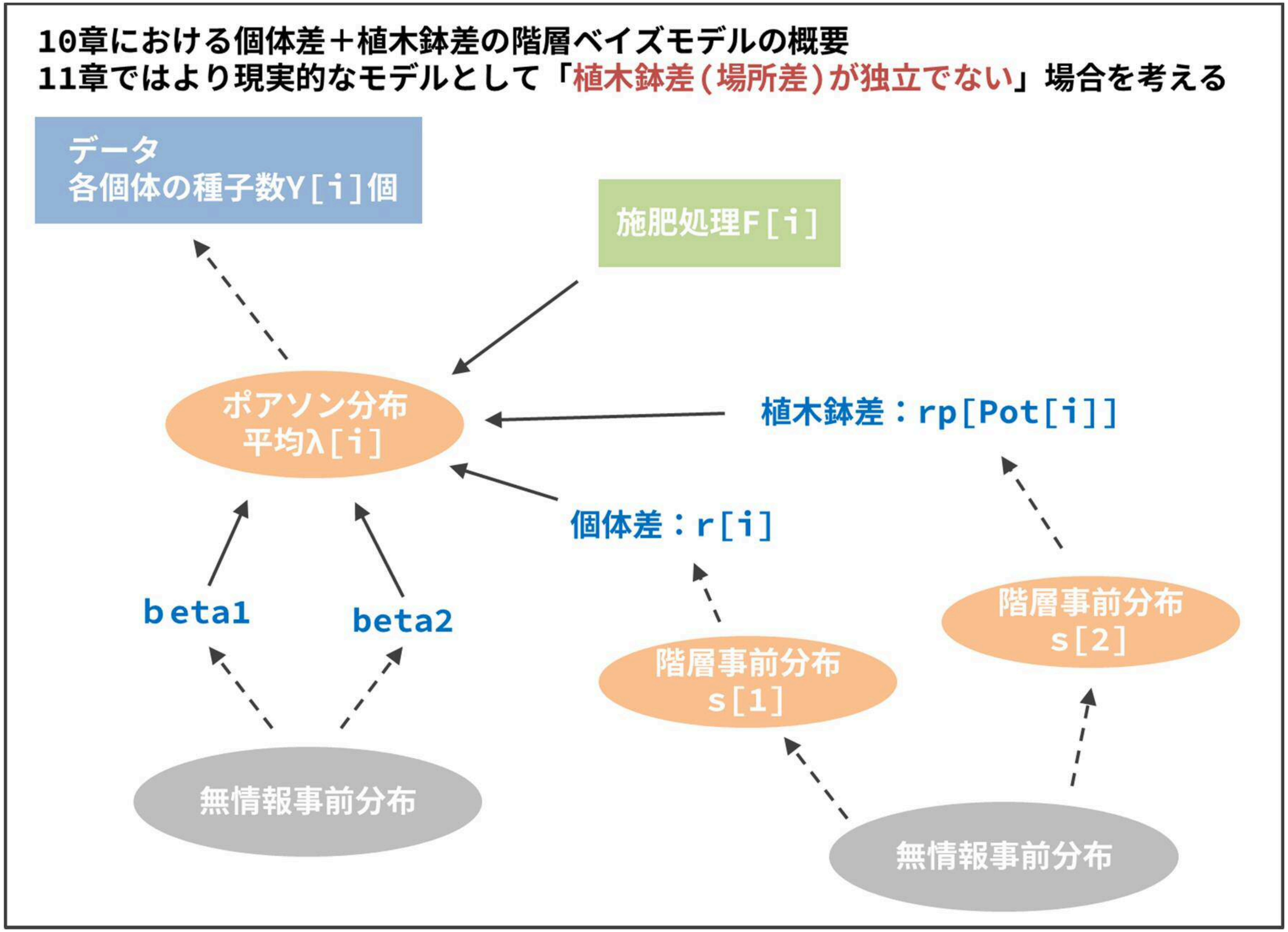
全体ではばらついているけれど、「近所」では似ているというのが空間相関のある場所差です。**事前分布を工夫して**、このような場所差をいれた階層ベイズモデルをつくってみましょう

11章の扉ページより（p.241）

11章への導入

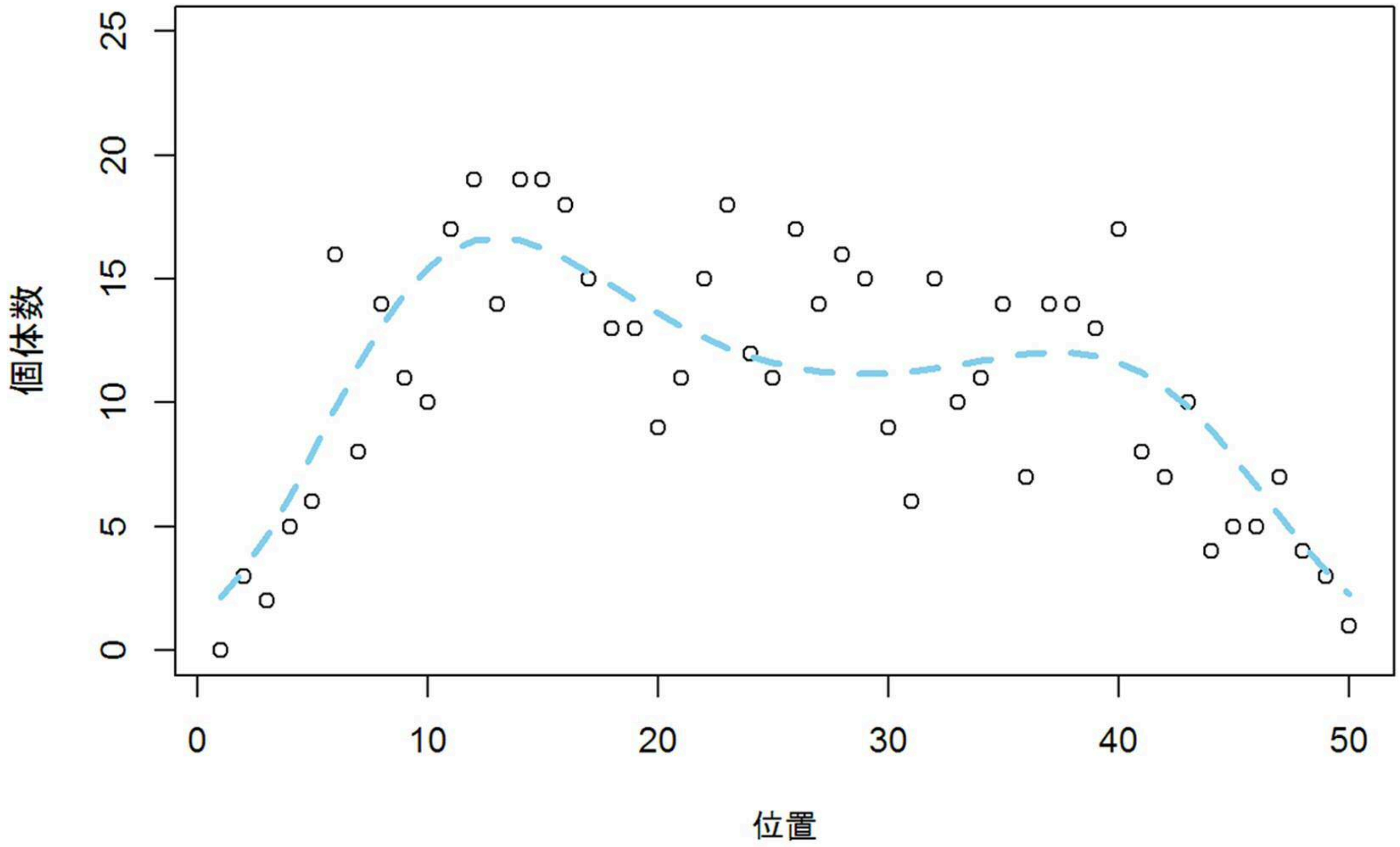
- 10章では場所差が独立に決まるというモデルを考えた
- 実際には場所差が独立に決まるデータ構造とは限らない
- 空間相関を考慮したモデルを考えるには？

10章で考えたモデル


10章における個体差＋植木鉢差の階層ベイズモデルの概要
11章ではより現実的なモデルとして「植木鉢差(場所差)が独立でない」場合を考える
データ
各個体の種子数Y[i]個
施肥処理F[i]
ポアソン分布
平均λ[i]
植木鉢差：rp[Pot[i]]
個体差：r[i]
beta1
beta2
階層事前分布
s[1]
階層事前分布
s[2]
無情報事前分布
無情報事前分布

11章で考えるデータ

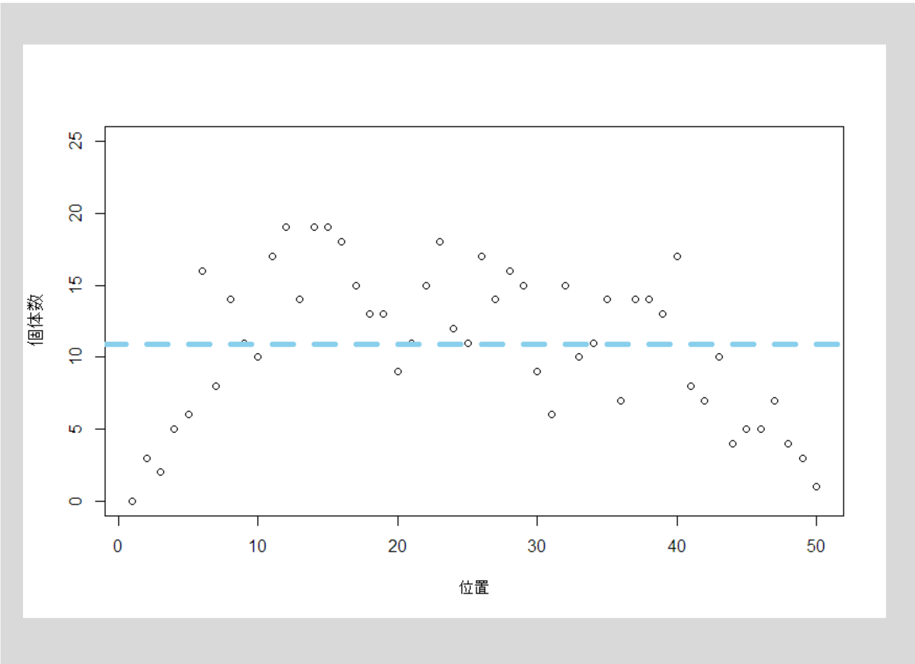
11章で考える一次元空間上で観測した個体数のカウントデータ（仮想）
局所密度（破線）が場所ごとで変化、それも滑らかにしている
⇒ 場所差を考慮する必要がある
⇒ 場所差が独立と仮定できない



定数モデルだと

- 場所差で変化しない
- 今回のデータには**過分散**

$$\log \lambda = \beta$$

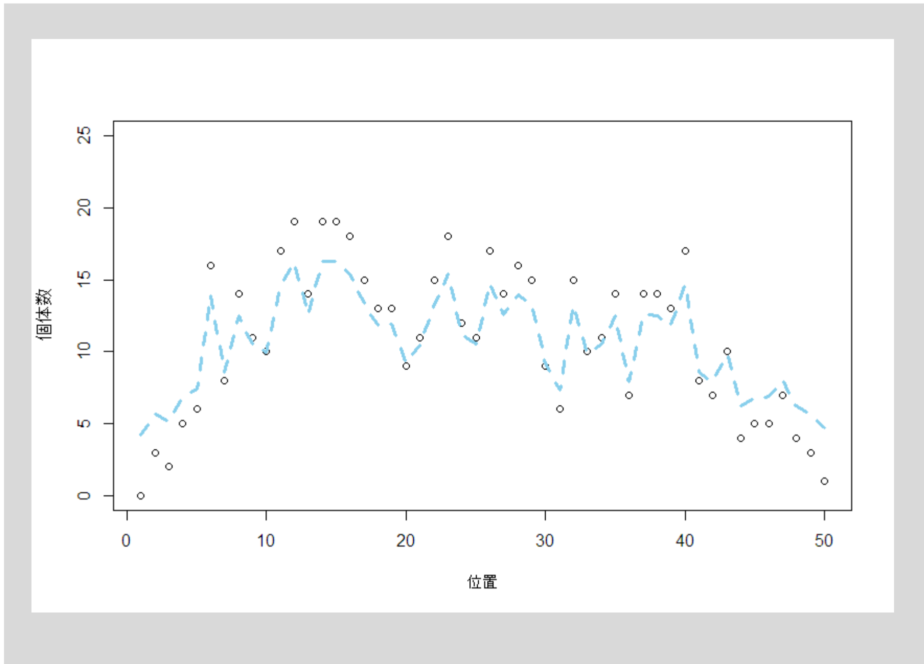


独立な場所差のモデルだと

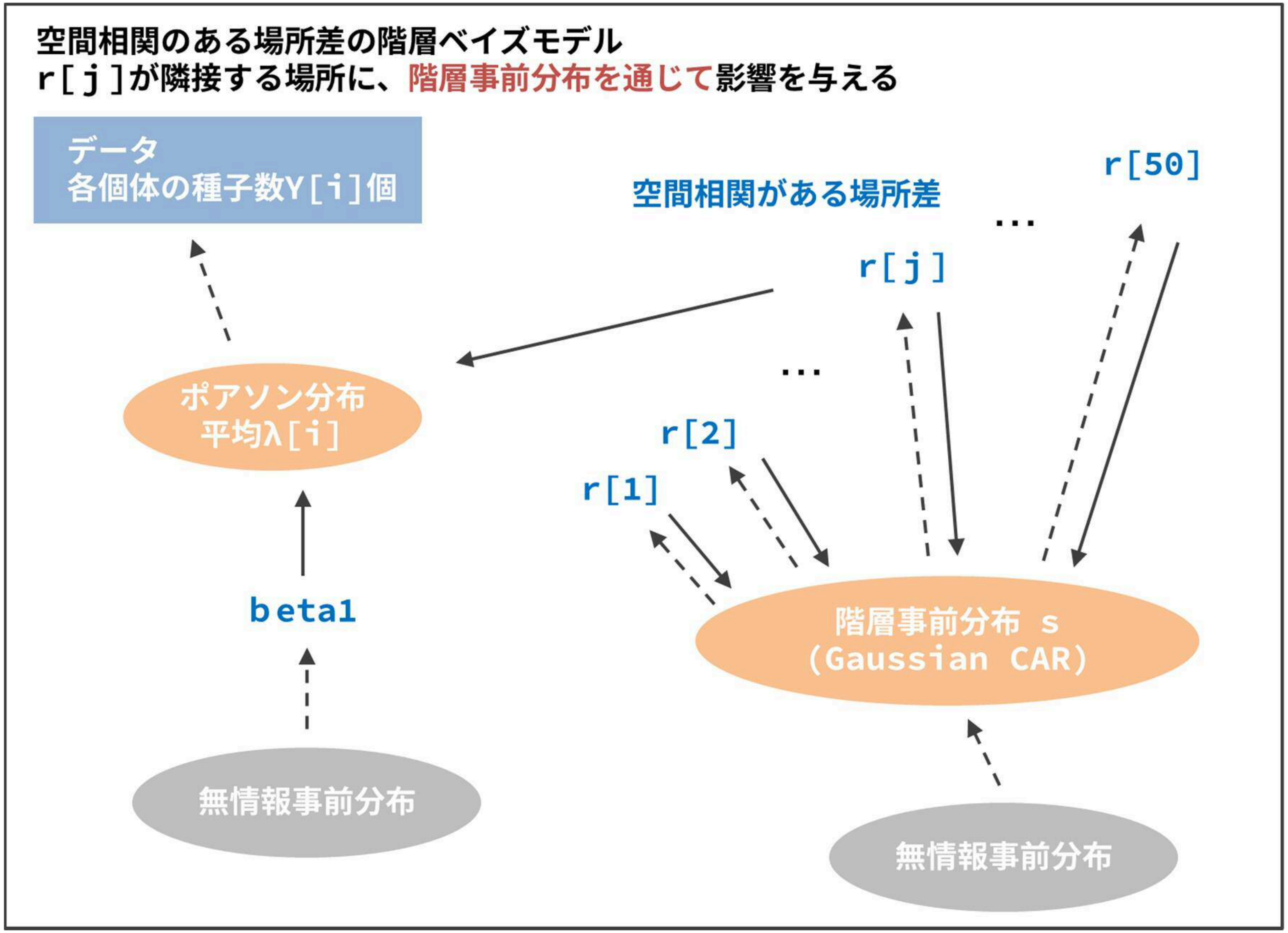
- 滑らかに変化しない
- 場所差へ**独立な事前分布**を設定

$$\log \lambda_j = \beta + r_j$$

$$p(r_j|s) = \frac{1}{\sqrt{2\pi s^2}} \exp\left\{\frac{1}{2}(r_j^2/s^2)\right\}$$



11章で考えるモデル



空間相関のある場所差の階層ベイズモデル
r[j]が隣接する場所に、階層事前分布を通じて影響を与える
データ
各個体の種子数Y[i]個
空間相関がある場所差
r[50]
…
r[j]
…
r[2]
r[1]
ポアソン分布
平均λ[i]
beta1
階層事前分布 s
(Gaussian CAR)
無情報事前分布
無情報事前分布

空間構造がある階層事前分布

空間構造は事前分布で表現

- みどり本では次の３つを仮定してCARモデルによる事前分布を設計
  - 場所差は「近傍」の場所差にしか影響しない
  - 近傍から受ける影響は等しい
  - 近傍数は有限

1階差分の場合

$$p(\{r_j\}|s) \propto \exp\left\{\frac{1}{2s^2}\sum_{j\sim\prime j}(r_j - r_{\prime j})^2\right\}$$

2階差分の場合

$$p(r_j|\mu_j, s) = \sqrt{\frac{n_j}{2\pi s^2}}\exp\left\{-\frac{n_j}{2s^2}(r_j - \mu_j)^2\right\}$$

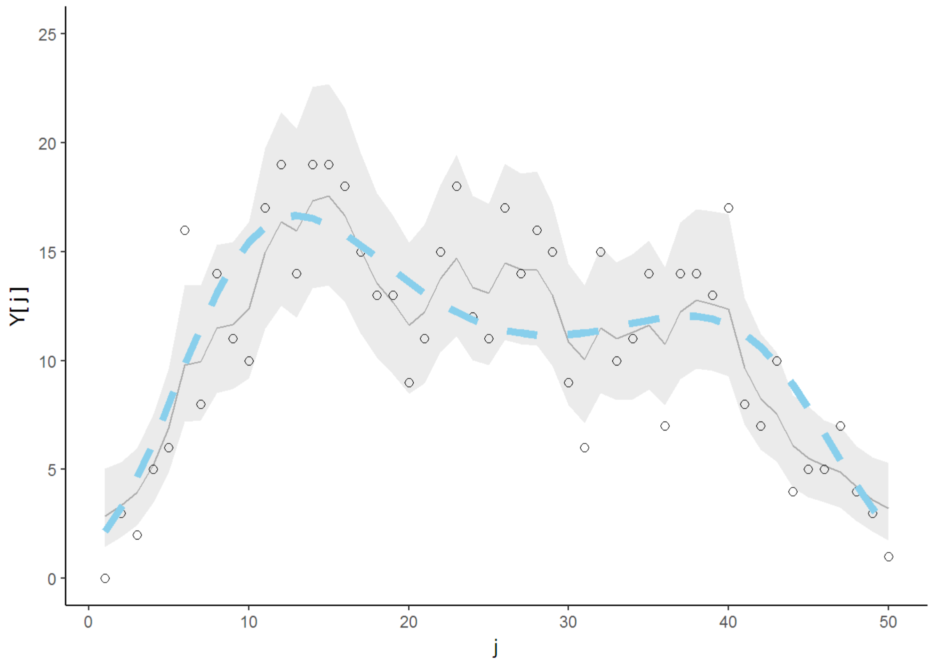
$$\mu_j = \frac{r_{j-1} - r_{j+1}}{2}$$

※　1階差分の式でｊに着目した
フルコンディションで2階差分の式が導出

1階差分の場合

- 2階差分に比べて**MCMCが高速**
- なるだけ定数っぽく

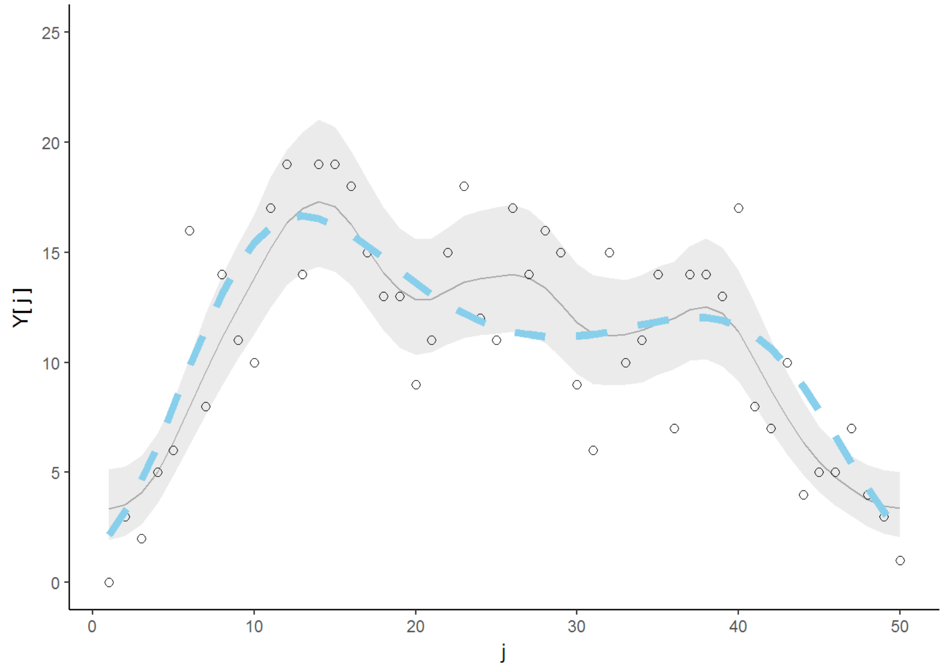
$$r_j - r_{j-1} \sim \mathcal{N}(0, s)$$



2階差分の場合

- 1階差分に比べて**滑らか**
- なるだけ1次関数っぽく

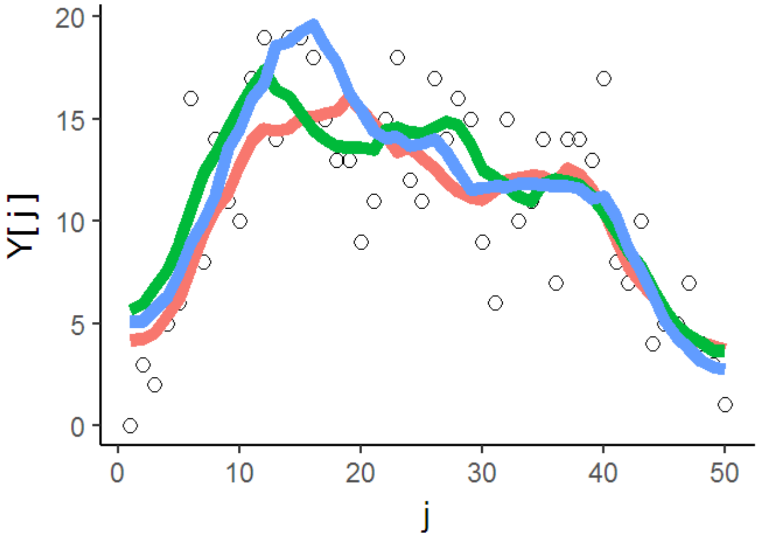
$$(r_{j+1} - r_j) - (r_j - r_{j-1}) \sim \mathcal{N}(0, s)$$



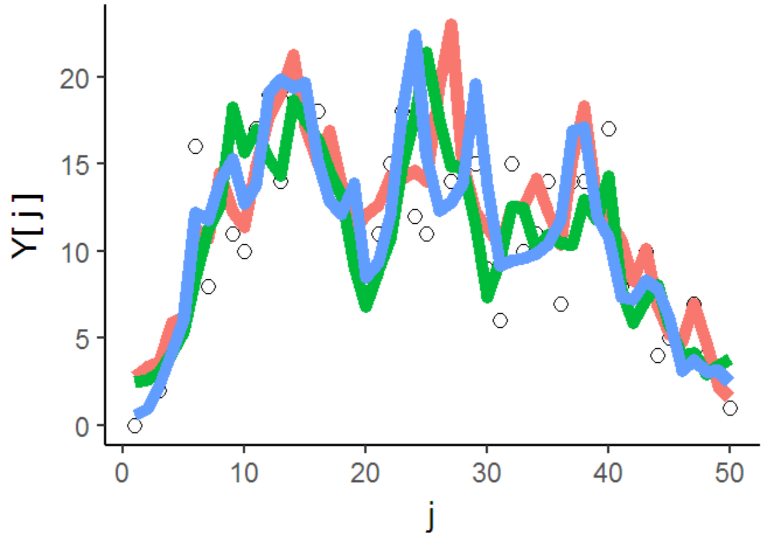
分散パラメータ s が確率場に与える影響

MCMCサンプルで確認

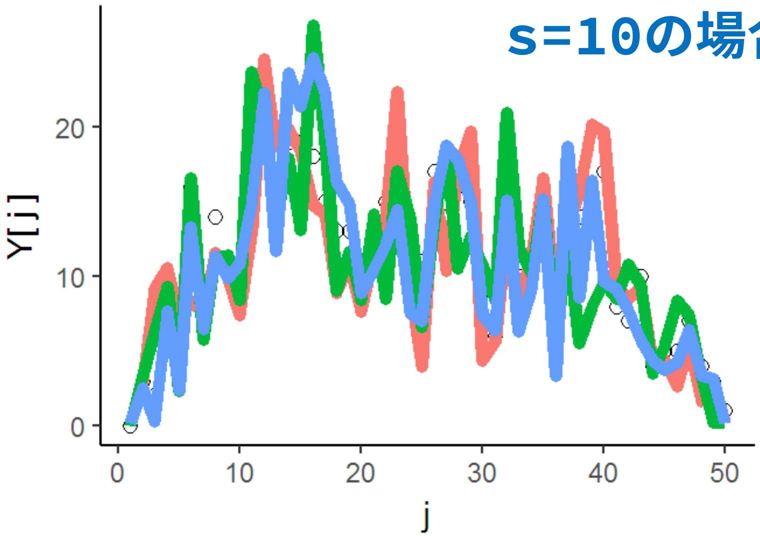
s=0.03の場合



s＝0.3の場合



s=10の場合



分散パラメータが大きくなると？

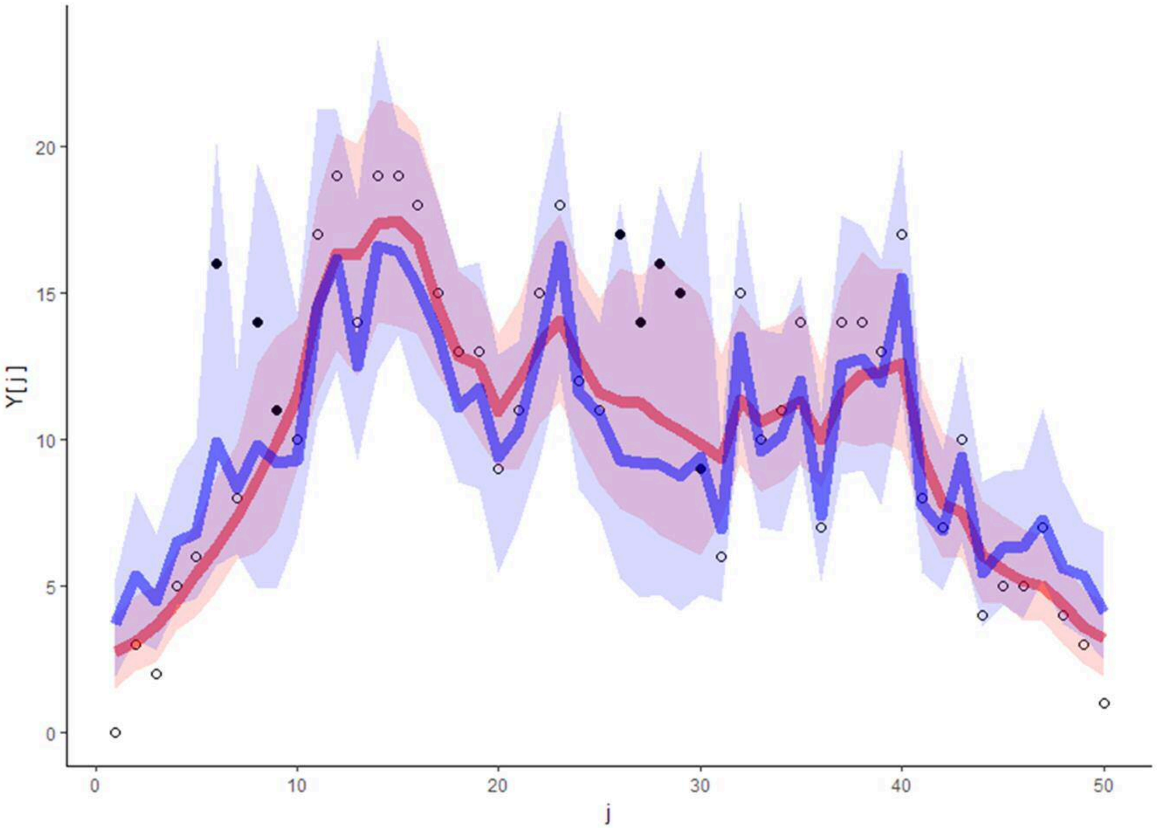
- 確率場のランダム性が大きくなる
- 「近隣が似ている」が表現できない

分散パラメータの調整

- 滑らかさを調整している

欠測があるデータには空間相関が有効

空間相関を考慮した事前分布を使用した場合（赤）と、独立の場合（青）の比較
考慮した場合の方がなめらかで、**欠測がある場所でも事後分布が広くなりにくい**
※1　図では事後分布の80%区間と中央値を表示
※2　黒丸が欠測させた場所の観測値



参考文献

久保緑本11章の
マルコフ場モデル（空間構造のあるベイズモデル）

- http://statmodeling.hatenablog.com/entry/markov-field-model

MCMCサンプルを{dplyr}で操る

- http://statmodeling.hatenablog.com/entry/using-mcmc-samples-with-dplyr

#みどりぼん　最終回
「データ解析のための統計モデリング入門」読書会
あの日作ったモデルの名前を僕達はまだ知らない

- https://www.youtube.com/watch?v=M7rN6paP99g