データ解析のための統計学入門
第11章
空間構造のある
階層ベイズモデル

大内　克之

空間構造

　これまでの章では、「空間構造」は考えなくて良いものとして統計モデルを作ってきた。

　しかし、現実には「データを取る場合の空間配置の影響」を無視できない場合がある。

個体数分布

　現実の調査データの空間構造では2次元や3次元となる場合が多いが、ここでは分かりやすい1次元の空間構造を扱う。

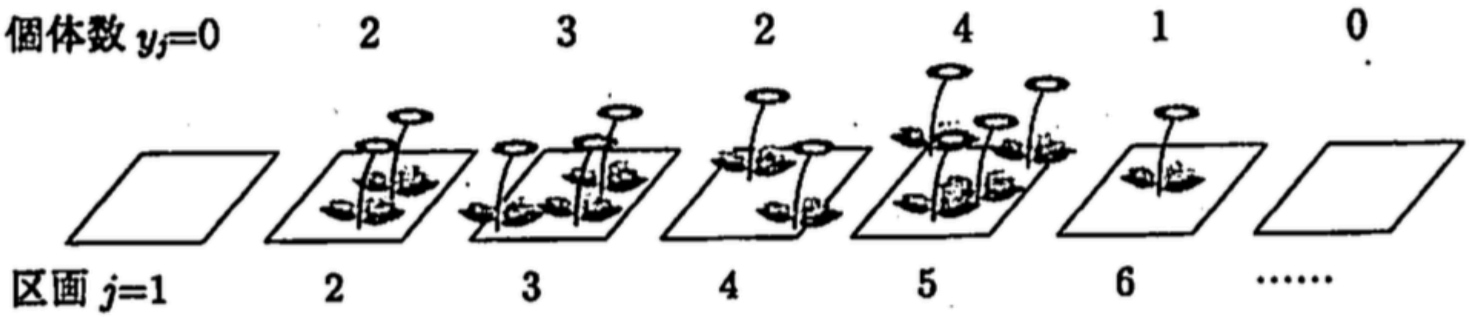
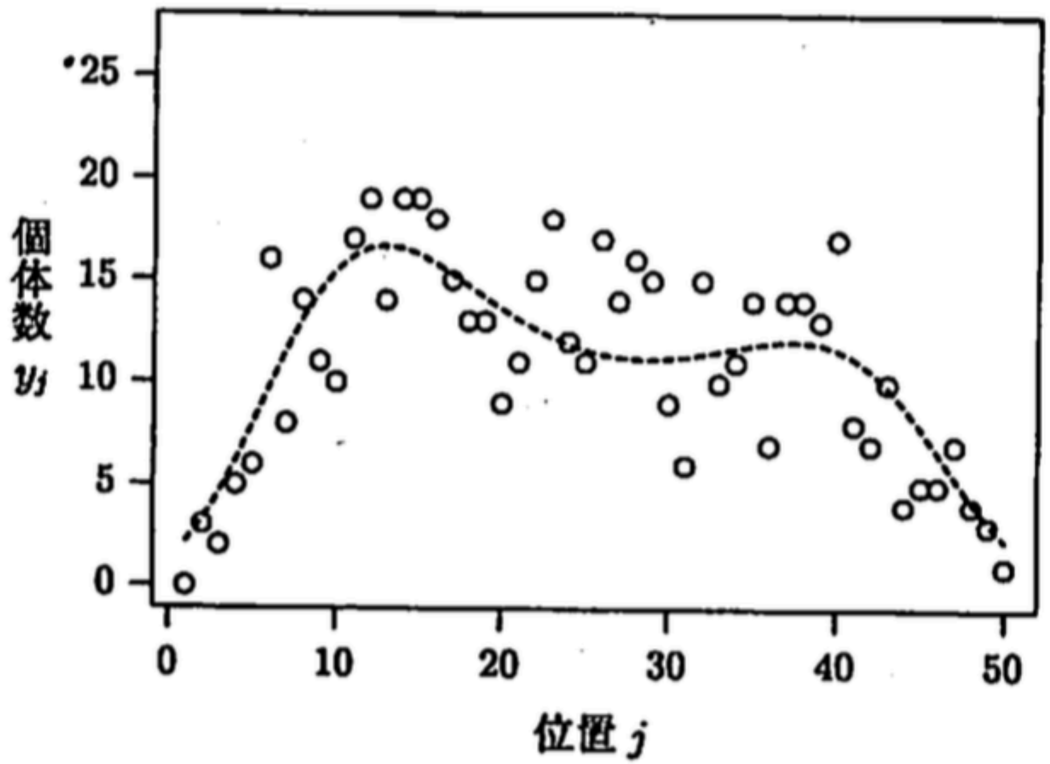


図 11.1　一次元の空間構造をもつ調査地の一例．ある調査ラインにそって 50 個の調査区画($j \in \{1, 2, \cdots, 50\}$)があり，それぞれにおいて架空植物の個体数 y_j がカウントされた．

個体数分布



空間構造の組み込み

例題がポアソン分布に従うとする。

$$p(y_j|\lambda) = \frac{\lambda_{yj} exp{-\lambda}}{y_j!}$$

標本平均は10.9なので、平均値λも分散も同じくらいになると期待できる。しかし、標本分散は27.4であり、過分散であることが分かる。

そこで、平均個体数λ_jを下のように表す。

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

空間構造の組み込み

場所差r_jが位置によって少しずつ変化する様子を表現するために、以下のように仮定する。

- 区画の場所差は「近傍」区画の場所差にしか影響されない
- 区画jの近傍の個数n_jは有限個であり、どの区画が近傍であるかはモデル設計者が指定する
- 近傍の直接の影響はどれも等しく$1/n_j$

空間構造の組み込み

近傍数を2(接している区画のみ)とする。

r_jの条件付き事前分布が、

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

であるとして、μ_jを近傍の平均とし、標準偏差を$s/\sqrt{nj}$であるとする。事前分布は

$$p(\{r_j\}|s) \propto exp\{\frac{1}{2s^2}\sum_{\{j,j'\}}(r_j - r_{j'})\}$$

となる。

データに当てはめる

これで、モデルの設計は終了した。

事後確率は、

$$p(\beta, s, \{r_j\}|Y) \propto p(\{r_j\}|s)p(s)p(\beta)\prod_j p(y_j|\lambda_j)$$

となる。

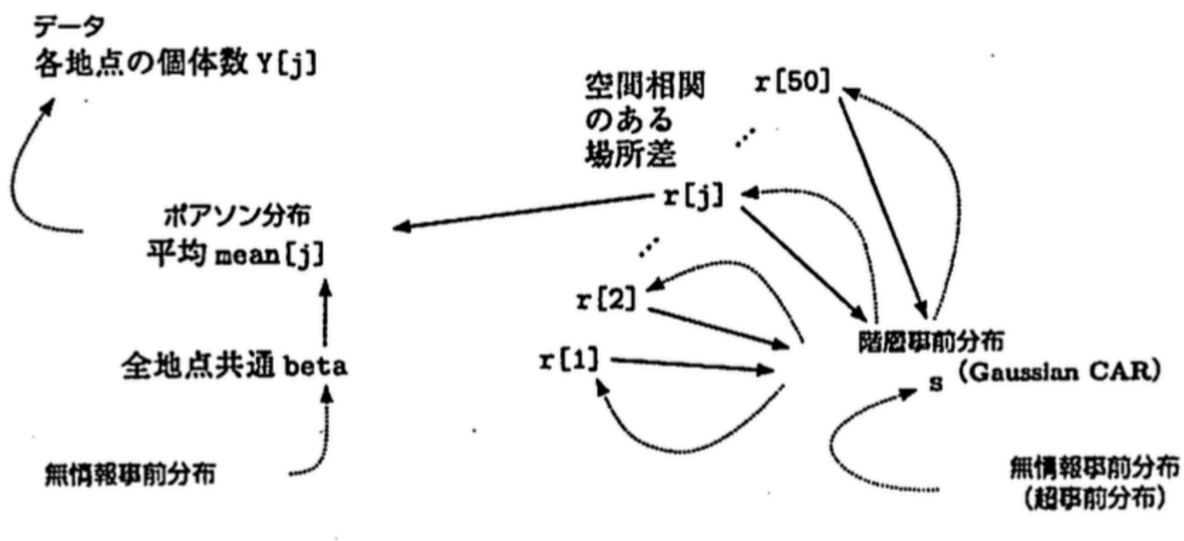
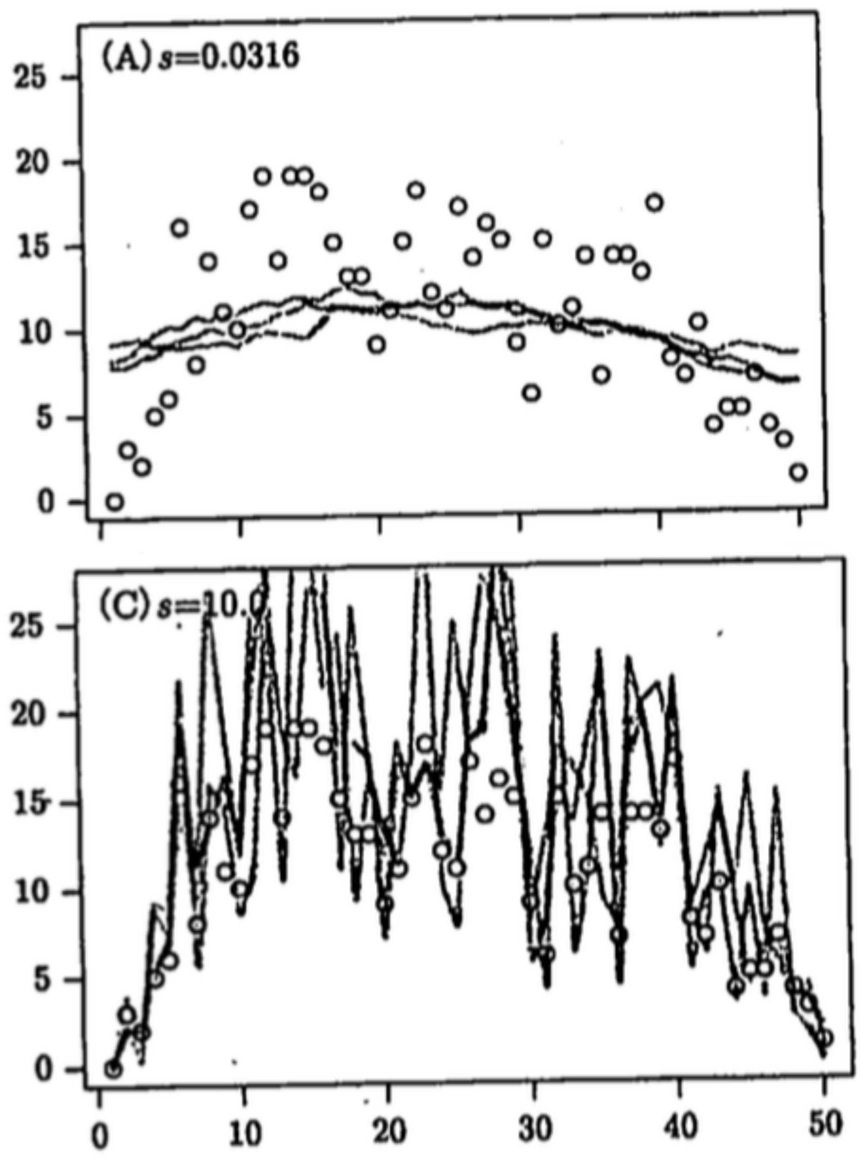
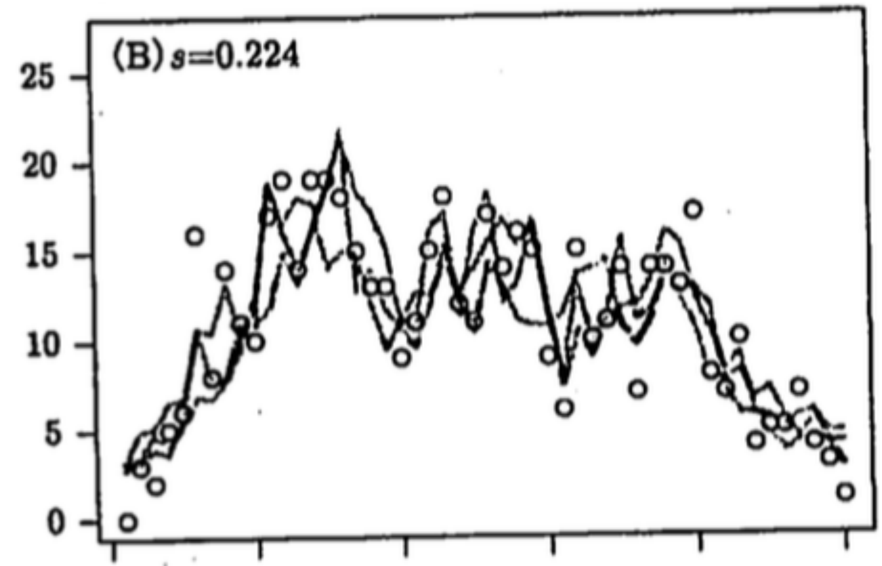


図 11.3　空間相関のある場所差の階層ベイズモデルの概要.

確率場





縦軸：個体数y_i
横軸：位置j_i

欠測のある観測データ

空間相関を組み込んだモデルは、「欠測データ」に強い。