Statstical models, selection effects and sampling bias

Peter McCullagh

Department of Statistics
University of Chicago

Gainesville, January 2010

Outline

Conventional regression model

Fixed index set $\mathcal{U}$ (always infinite): $u_1, u_2, \ldots$ subjects, plots...
Covariate $x(u_1), x(u_2), \ldots$ (non-random, vector-valued)
Response $Y(u_1), Y(u_2), \ldots$ (random, real-valued)

Regression model:

Sample = finite ordered subset $u_1, \ldots, u_n$ (distinct in $\mathcal{U}$)
For each sample with configuration $\mathbf{x} = (x(u_1), \ldots, x(u_n))$
Response distribution $p_{\mathbf{x}}(\mathbf{y})$ on $\mathcal{R}^n$ depends on $\mathbf{x}$

Kolmogorov consistency condition on distributions $p_{\mathbf{x}}(\cdot)$
$\mathbf{x}$ is not a random variable
$p_{\mathbf{x}}(\cdot)$ is not the conditional distribution given $\mathbf{x}$

Gaussian regression model

Covariate $x(u) = (x_1(u), x_2(u))$ partitioned into two components

$x_1(u) = (\text{variety}(u), \text{treatment}(u), \ldots)$ affecting the mean
$x_2(u) = (\text{block}(u), \text{coordinates}(u))$ affecting covariances

Example: response distribution for a *fixed* sample of n plots

$$p_{\mathbf{x}}(\mathbf{y} \in A; \theta) = N_n(X_1\beta,\ \sigma_0^2 I_n + \sigma_1^2 K[\mathbf{x}_2])(A)$$

$A \subset \mathcal{R}^n$, $K[\mathbf{x}] = \{K(x_i, x_j)\}$
block-factor models: $K(i,j) = 1$ if $\text{block}(i) = \text{block}(j)$
spatial/temporal models: $K(i,j) = \exp(-|x_2(i) - x_2(j)|/\tau)$
Generalized random field: $K(i,j) = -\,\text{ave}_{x \in i,\, x' \in j} \log|x - x'|$

Equivalent to $Y(u) = \mu(u) + \epsilon(u) + \eta(u)$ with $\epsilon \perp \eta$,...

Binary regression model (GLMM)

Units: $u_1, u_2, \ldots$ subjects, patients, plots (labelled)
Covariate $x(u_1), x(u_2), \ldots$ (non-random, $\mathcal{X}$-valued)
Latent process η on $\mathcal{X}$ (Gaussian, for example)
Responses $Y(u_1), \ldots$ conditionally independent given η

$$\operatorname{logit} \operatorname{pr}(Y(u) = 1 \mid \eta) = \alpha + \beta x(u) + \eta(x(u))$$

Joint distribution for sample having configuration **x**

$$p_{\mathbf{x}}(\mathbf{y}) = E_\eta \prod_{i=1}^{n} \frac{e^{(\alpha+\beta x_i+\eta(x_i))y_i}}{1 + e^{\alpha+\beta x_i+\eta(x_i)}}$$

parameters α, β, K, $K(x, x') = \operatorname{cov}(\eta(x), \eta(x'))$.

Binary regression model: computation

GLMM computational problem:

$$p_{\mathbf{x}}(\mathbf{y}) = \int_{\mathcal{R}^n} \prod_{i=1}^{n} \frac{e^{(\alpha+\beta x_i+\eta(x_i))y_i}}{1+e^{\alpha+\beta x_i+\eta(x_i)}}\,\phi(\eta; K)\,d\eta$$

Options:

Taylor approx: Laird and Ware; Schall; Breslow and Clayton, McC and Nelder, Drum and McC,...
Laplace approximation: Wolfinger 1993; Shun and McC 1994
Numerical approximation: Egret
E.M. algorithm: McCulloch 1994 for probit models
Monte Carlo: Z&L,...

But, . . ., wait a minute...

$p_{\mathbf{x}}(\mathbf{y})$ is the distribution for each *fixed* sample of n units.

But ... the sample might not be predetermined

- volunteer samples in clinical trials;
- pre-screening of patients to increase compliance;
- behavioural studies in ecology;
- marketing studies, with purchase events as units;
- public policy: crime type with crime events as units

Ergo, $\mathbf{x}$ is also random, so we need a joint distribution.

Q1: For a bivariate process, what does $p_{\mathbf{x}}(\mathbf{y})$ represent?
Q2: Is it necessarily the case that $p_{\mathbf{x}}(\mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})$?
... even if sample size is random?

Problems in the application of conventional models

Clinical trials / market research / traffic studies / crime...

(i) Operational interpretation of a sample as a fixed subset
or as a random subset independent of the process

(ii) Sample units generated by a random process
sequential recruitment, purchase events, traffic studies...

(iii) Population also generated by a random process in time
animal populations, purchase events, crime events,...

(iv) Samples: random, sequential, quota,...

(v) Conditional distribution given observed random $\mathbf{x}$
versus stratum distribution for fixed $\mathbf{x}$

Illustration: Kentucky traffic accidents

Units: traffic accidents in Kentucky
  Response: seat belt used? (Y or N)
  Explanatory: $x(s)$ road class at site $s \in$ Kentucky

$$\text{logit}\,\text{pr}(Y(s) = 1 \mid \eta, \text{event at } s) = \eta(s) + \beta x(s)$$
$$\text{cov}(\eta(s), \eta(s')) = K(s, s')$$

$$\begin{aligned}\text{pr}(Y(s) = 1 \mid \text{event at } s) &= E_\eta\left(\frac{e^{\eta(s)+\beta x(s)}}{1 + e^{\eta(s)+\beta x(s)}}\right) \\ &\simeq \frac{e^{\beta^* x(s)}}{1 + e^{\alpha(s)+\beta^* x(s)}}\end{aligned}$$

$|\beta^*| \le \beta$ (attenuation)

Kentucky traffic accidents: Poisson version

Log intensity of accidents w/o restraint $\eta(s,0) + \beta_0 x(s)$
Log intensity of accidents with restraint $\eta(s,1) + \beta_1 x(s)$

$$\begin{aligned}
\operatorname{logit} \operatorname{pr}(Y(s) = 1 \mid \text{event at } s, \eta) &= \eta(s,1) - \eta(s,0) + (\beta_1 - \beta_0) x(s) \\
&= \eta(s) + \beta x(s) \\
\operatorname{pr}(Y(s) = 1 \mid \text{event at } s) &= E_\eta \left(\frac{e^{\eta(s)+\beta x(s)}}{1 + e^{\eta(s)+\beta x(s)}} \right) \\
&\simeq \frac{e^{\alpha(s)+\beta^* x(s)}}{1 + e^{\alpha(s)+\beta^* x(s)}}
\end{aligned}$$

Same as logistic model with additive random effect!

Kentucky accidents: alternative Poisson version

Intensity of accidents given η:
- w/o restraint: $e^{\eta(s,0)} \exp(\beta_0 x(s))$
- with restraint $e^{\eta(s,1)} \exp(\beta_1 x(s))$

$$\mathrm{pr}(\text{accident w/o restraint in } ds \,|\, \eta) = e^{\eta(s,0)} e^{\beta_0 x(s)}\, ds$$

$$\mathrm{pr}(\text{accident w/o restraint in } ds) = E(e^{\eta(s,0)}) e^{\beta_0 x(s)}\, ds = m(s,0) e^{\beta_0 x(s)}$$

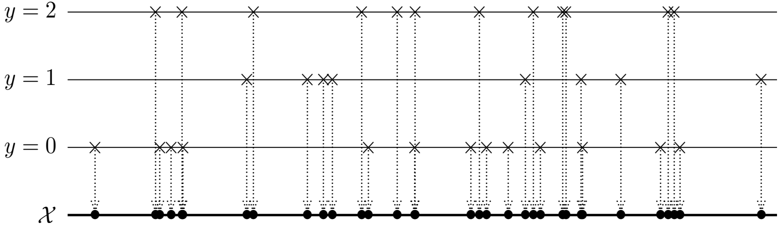
$$\mathrm{pr}(\text{accident with restraint in } ds) = E(e^{\eta(s,1)}) e^{\beta_1 x(s)}\, ds$$

$$\mathrm{pr}(Y = 1 | \text{accident in } ds) = \frac{m(s,1) e^{\beta_1 x(s)}}{m(s,0) e^{\beta_0 x(s)} + m(s,1) e^{\beta_1 x(s)}}$$

$$\log \mathrm{odds}(Y = 1 \mid \ldots) = \log(m(s,1)/m(s,0)) + (\beta_1 - \beta_0) x(s)$$

No approximation, no attenuation!

Point process model for auto-generated units



A point process on $\mathcal{C} \times \mathcal{X}$ for $\mathcal{C} = \{0, 1, 2\}$, and the superposition process on $\mathcal{X}$.

Intensity $\lambda_r(x)$ for class r: $r = 0, 1, 2$.

x-values auto-generated by the superposition process with intensity $\lambda_{\cdot}(x)$.

To each auto-generated unit there corresponds an x-value and a y-value.

Binary point process model

Intensity process $\lambda_0(x)$ for class 0, $\lambda_1(x)$ for class 1
Log ratio: $\eta(x) = \log \lambda_1(x) - \log \lambda_0(x)$
Events form a PP with intensity λ on $\{0, 1\} \times \mathcal{X}$.
Conventional GLMM calculation (Bayesian and frequentist):

$$\mathrm{pr}(Y = 1 \mid x, \lambda) = \frac{\lambda_1(x)}{\lambda_.(x)} = \frac{e^{\eta(x)}}{1 + e^{\eta(x)}}$$

$$\mathrm{pr}(Y = 1 \mid x) = E\left(\frac{\lambda_1(x)}{\lambda_.(x)}\right) = E\left(\frac{e^{\eta(x)}}{1 + e^{\eta(x)}}\right)$$

GLMM calculation is correct in a sense, but irrelevant. . .
. . . there might not be an event at x!

Correct calculation for auto-generated units

$$\begin{aligned}
\text{pr}(\text{event of type } r \text{ in } dx \mid \lambda) &= \lambda_r(x)\, dx + o(dx) \\
\text{pr}(\text{event of type } r \text{ in } dx) &= E(\lambda_r(x))\, dx + o(dx) \\
\text{pr}(\text{event in SPP in } dx \mid \lambda) &= \lambda_.(x)\, dx + o(dx) \\
\text{pr}(\text{event in SPP in } dx) &= E(\lambda_.(x))\, dx + o(dx)
\end{aligned}$$

$$\begin{aligned}
\text{pr}(Y(x) = r \mid \text{SPP event at } x) &= \frac{E\lambda_r(x)}{E\lambda_.(x)} \\
&= E\left(\frac{\lambda_r(x)}{\lambda_.(x)} \;\Big|\; x \in \mathrm{SPP}\right) = \frac{E\lambda_r(x)}{E\lambda_.(x)} \neq E\left(\frac{\lambda_r(x)}{\lambda_.(x)}\right)
\end{aligned}$$

Sampling bias:

Distn for fixed x versus distn for autogenerated x.

Two ways of thinking

First way: waiting for Godot!

Fix $x \in \mathcal{X}$ and wait for an event to occur in $(x, x + dx)$

$\mathrm{pr}(Y = 1 \mid \lambda, x) = \frac{\lambda_1(x)}{\lambda_\bullet(x)}$

$\mathrm{pr}(Y = 1; x) = E\left(\frac{\lambda_1(x)}{\lambda_\bullet(x)}\right) = E(Y_i \mid i\colon X_i = x)$

Conventional, mathematically correct, but seldom relevant

Second way: come what may!

SPP event occurs at x, a random point in $\mathcal{X}$

joint density at (y, x) proportional to $E(\lambda_y(x)) = m_y(x)$

x has marginal density proportional to $E(\lambda_.(x)) = m_.(x)$

$$\mathrm{pr}(Y = 1 \mid x \in \mathrm{SPP}) = \frac{E\lambda_1(x)}{E\lambda_.(x)} \neq E\left(\frac{\lambda_1(x)}{\lambda_.(x)}\right) = E(Y_i \mid i\colon X_i = x)$$

Gaussian spatial models

Standard Gaussian model: $\mathcal{U} = \mathcal{R}^2$ (population of units)

Sample: finite ordered subset $\mathbf{x} = (x_1, \ldots, x_n)$ (sites)

Response process $Y(x) \in \mathcal{R}$ observed at $x \in \mathbf{x}$

Joint distribution for fixed $\mathbf{x}$

$$p_{\mathbf{x}}(A) = N(\mathbf{1}\mu,\ \sigma_0^2 I_n + \sigma_1^2 K[\mathbf{x}])(A)$$

$(K[\mathbf{x}])_{ij} = K(x_i, x_j)$, for example $\exp(-\|x_i - x_j\|/\tau)$

Used for:

parameter estimation via likelihood

prediction of the value at unobserved sites given $Y[\mathbf{x}]$

$p_{\mathbf{x},x'}(Y(x') = y' \mid Y[\mathbf{x}]), \quad E(Y(x') \mid Y[\mathbf{x}])$ (Kriging)

Conditional distribution is Gaussian

Preferential sampling of spatial processes

Following Diggle, Menezes and Su (RSS discussion paper)
Environmental monitoring:
sites selected where pollution levels are thought to be high
Drilling:
the most promising sites are selected first

How does preferential sampling affect
(i) likelihood and parameter estimation?
(ii) predictions and conditional distributions?

Condition for benign sampling: $\mathbf{x}^{(r+1)} \perp\!\!\!\perp Y \mid Y[\mathbf{x}^{(r)}]$

Preferential sampling (contd.)

Point process model for preferential sampling:

$$S \sim GP(0, K) \quad \text{on } \mathcal{X} = \mathcal{R}^2$$

$$\lambda(x, y) = e^{\alpha+\beta S(x)} \tau^{-1} \phi\left(\frac{y - \mu(x) - S(x))}{\tau}\right) \quad \text{at } (x, y)$$

$$\lambda_{.}(x) = e^{\alpha+\beta S(x)} \quad \text{marginal intensity at } x \in \mathcal{X}$$

(preferential sampling if $\beta \neq 0$)

Point process density on $A \subset \mathcal{R}^2$

$$p_A(\mathbf{x}, \mathbf{y}) = E_\lambda\left(e^{-\Lambda_{.}(A)} \prod_{\mathbf{x}} \lambda(x_i, y_i)\right)$$

Spatial preferential sampling (contd.)

Diggle's Gaussian model sampled preferentially

$S \sim GP(0, K)$ (ground truth)
$Y(x) = \mu(x) + S(x) + \epsilon(x)$ (observed at certain sites in A)

sites $\mathbf{x} \subset \mathcal{R}^2$ generated with intensity $\exp(\alpha + \beta S(x))$

Implications:

$E(Y(x)) = \mu(x)$ for each fixed x
$E(Y(x) \mid x \in \mathbf{x}) = \mu(x) + \beta K(x, x)$
$E(Y(x) \mid x, x' \in \mathbf{x}) = \mu(x) + \beta K(x, x) + \beta K(x, x')$
$\mathrm{cov}(Y(x), Y(x') \mid x, x' \in \mathbf{x}) = K(x, x')$

Logistic illustration of sampling bias

$$\begin{aligned}
\eta_0(x) &\sim GP(0, K), & \lambda_0(x) &= \exp(\eta_0(x)) \\
\eta_1(x) &\sim GP(\alpha+\beta x, K), & \lambda_1(x) &= \exp(\eta_1(x)) \\
\eta(x) = \eta_1(x) - \eta_0(x) &\sim GP(\alpha+\beta x, 2K), & K(x,x) &= \sigma^2
\end{aligned}$$

One-dimensional sampling distributions:

$$\rho(x) = p_x(Y=1) = E\left(\frac{e^{\eta(x)}}{1+e^{\eta(x)}}\right) \quad \text{(fixed } x\text{)}$$

$$\operatorname{logit}(\rho(x)) \simeq \alpha^* + \beta^* x \qquad (|\beta^*| < |\beta|)$$

$$\pi(x) = \mathrm{pr}(Y=1 \mid x \in \mathrm{SPP}) = \frac{E\lambda_1(x)}{E\lambda_.(x)} = \frac{e^{\alpha+\beta x+\sigma^2/2}}{e^{\sigma^2/2} + e^{\alpha+\beta x+\sigma^2/2}}$$

$$\operatorname{logit}\mathrm{pr}(Y=1 \mid x \in \mathrm{SPP}) = \alpha + \beta x$$

No approximation; no attenuation

Conditional joint distributions of $Y[\mathbf{x}]$ given $\mathbf{x}$

Quota sampling with fixed $\mathbf{x}$

$$p_{\mathbf{x}}(\mathbf{y}) = E\left(\prod \frac{\lambda_{y_i}(x_i)}{\lambda_{\bullet}(x_i)}\right) = E\left(\prod \frac{e^{y_i\eta(x_i)}}{1 + e^{\eta(x_i)}}\right)$$

coincides with standard GLMM model

Sequential sampling fixed time: $\#\mathbf{x}$ random

$$p(\mathbf{y} \mid \mathbf{x}) = \frac{E \prod \lambda_{y_i}(x_i)\, e^{-\int \lambda_{\bullet}(x)\nu(dx)}}{E \prod \lambda_{\bullet}(x_i)\, e^{-\int \lambda_{\bullet}(x)\nu(dx)}}$$

Alternative formulation: auto-selection

To each $u \in \mathcal{U}$ there corresponds a random intensity $\lambda^{(u)}_{y,t}$

$$\begin{array}{cc} & \begin{array}{cc} t=0 & t=1 \end{array} \\ \begin{array}{c} y=0 \\ y=1 \end{array} & \begin{pmatrix} \lambda^{(u)}_{00} & \lambda^{(u)}_{01} \\ \lambda^{(u)}_{10} & \lambda^{(u)}_{11} \end{pmatrix} \end{array}$$

$$\begin{aligned} \mathrm{pr}(u \in \text{Sample} \mid \lambda) &= \lambda^{(u)}_{..} \qquad \text{(random and small)} \\ \mathrm{pr}(u \in S \,\&\, t_u = t \mid \lambda) &= \lambda^{(u)}_{.t} \\ \mathrm{pr}(Y_u = 1 \mid u \in S \,\&\, t_u = t, \lambda) &= \lambda^{(u)}_{1t} / \lambda^{(u)}_{.t} \end{aligned}$$

Randomization implies $\lambda^{(u)}_{.0} = \lambda^{(u)}_{.1}$

Auto-selection and volunteer samples (contd)

$$\mathrm{pr}(u \in \mathrm{Sample} \mid \lambda) = \lambda^{(u)}_{\bullet\bullet} \qquad \text{(volunteer intensity)}$$

$$\mathrm{pr}(u \in S \,\&\, t_u = t \mid \lambda) = \lambda^{(u)}_{\bullet t}$$

$$\mathrm{pr}(Y_u = 1 \mid u \in S \,\&\, t_u = t, \lambda) = \frac{\lambda^{(u)}_{1t}}{\lambda^{(u)}_{\bullet t}}$$

$$\mathrm{pr}(Y_u = 1 \mid u \in S \,\&\, t_u = t) = \frac{E(\lambda^{(u)}_{1t})}{E(\lambda^{(u)}_{\bullet t})} \qquad (PP)$$

$$\mathrm{pr}(Y_u = 1 \mid t_u = t) = E\left(\frac{\lambda^{(u)}_{1t}}{\lambda^{(u)}_{\bullet t}}\right) \qquad \text{(GLMM for fixed } u\text{)}$$

Defining the treatment effect $\tau(x)$

Classical definition involves two fixed units $u \neq u'$ such that

$$x(u) = x(u') = x \quad \text{and} \quad t(u) = 0, t(u') = 1$$

Definitions 1 & 1':

$$\frac{\text{odds}(Y(u') = 1)}{\text{odds}(Y(u) = 1)} = e^{\tau(x)} = \frac{\text{odds}(Y(u) = 1 \mid t(u) = 1)}{\text{odds}(Y(u) = 1 \mid t(u) = 0)}$$

Exchangeability: $\longrightarrow$ Ratio same for all pairs u, u'

Classical definition 2: (also for fixed units as above)

$$\frac{\text{pr}(Y(u') = 1,\ Y(u) = 0)}{\text{pr}(Y(u') = 0,\ Y(u) = 1)} = e^{\tau'(x)}$$

$Y(u) \perp\!\!\!\perp Y(u')$ implies $\tau(x) = \tau'(x)$

But τ' may depend on the relationship between u, u'

Defining the treatment effect (contd)

Given $u, u' \in S_x$ such that $x(u) = x(u') = x$
PP definition 1:

$$\frac{\mathrm{odds}(Y(u') = 1 \mid u' \in S_x,\ t(u') = 1)}{\mathrm{odds}(Y(u) = 1 \mid u \in S_x,\ t(u) = 0)} = e^{\tau(x)}$$

PP definition 2: (explicitly involving pairs)

$$\frac{\mathrm{pr}(Y(u') = 1, Y(u) = 0 \mid u, u' \in S_x,\ t(u) = 0,\ t(u') = 1)}{\mathrm{pr}(Y(u') = 0, Y(u) = 1 \mid u, u' \in S_x,\ t(u) = 0,\ t(u') = 1)} = e^{\tau'(x)}$$

If $N_{rs} = \#\{u \in S_x : Y(u) = r, t(u) = s\}$
$N_{00}N_{11} - e^{\tau'(x)}N_{01}N_{10}$ has exactly zero expectation

Notation: meaning of $E(Y_i \mid X_i = x)$

Exchangeable sequence $(Y_1, X_1), (Y_2, X_2), \ldots$ with binary Y *implies* conditionally iid given λ

Stratum x: $\mathcal{U}_x = \{i : X_i = x\}$ an infinite random subsequence
Stratum average: $\mathrm{ave}\{Y_i : i \in \mathcal{U}_x\} = \lambda_1(x)/\lambda_{.}(x)$
Stratum mean = expected value of stratum average:

$$\rho(x) = E\left(\frac{\lambda_1(x)}{\lambda_{.}(x)}\right) \simeq \frac{e^{\alpha^* + \beta^* x}}{1 + e^{\alpha^* + \beta^* x}}$$

is declared target in much biostatistical work (PA)

Correct calculation for a random stratum in SPP:

$$\pi(x) = E\left(\frac{\lambda_1(x)}{\lambda_{.}(x)} \;\middle|\; x \in \mathrm{SPP}\right) = \frac{E(\lambda_1(x))}{E(\lambda_{.}(x))} = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

Conditional mean $\pi(x) = E(Y_i \mid X_i = x)$ versus stratum mean

Consequences of ambiguous notation

Sample $(Y_1, X_1), (Y_2, X_2), \ldots$ observed sequentially

$$\pi(x) = E\left(\frac{\lambda_1(x)}{\lambda_.(x)} \;\middle|\; x \in \mathrm{SPP}\right) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} = E(Y_i \,|\, X_i = x)$$

$$\rho(x) = E\left(\frac{\lambda_1(x)}{\lambda_.(x)}\right) \simeq \frac{e^{\alpha^*+\beta^* x}}{1 + e^{\alpha^*+\beta^* x}} = E(Y_i \,|\, i\colon X_i = x)$$

($\rho(x)$ computed by logistic-normal integral)

Conventional PA estimating function $Y_i - \rho(x_i)$ is such that

$$E(Y_1 - \rho(x) \,|\, X_1 = x) = \pi(x) - \rho(x) \neq 0$$

and same for $Y_2, \ldots$.

Estimating functions done correctly

Mean intensity for class r: $m_r(x) = E(\lambda_r(x))$
$\pi_r(x) = m_r(x)/m_{\cdot}(x); \quad \rho_r(x) = E(\lambda_r(x)/\lambda_{\cdot}(x))$

$E(Y_i) = \rho(x)$ for each i in $\mathcal{U}_x = \{u : X_u = x\}$

For autogenerated x, $E(Y|x \in \mathrm{SPP}) = \pi(x) \neq \rho(x)$

$$T(\mathbf{x}, \mathbf{y}) = \sum_{x \in \mathrm{SPP}} h(x)(Y(x) - \pi(x))$$

has zero mean for auto-generated configurations **x**.
Note: $E(T \mid \mathbf{x}) \neq 0$; average is also over configurations

Explanation of unbiasedness

$\mathbf{z} = \{(x_1, y_1), \ldots\}$ configuration generated in $(0, t)$.
$\mathbf{x} = \{x_1, \ldots,\}$ marginal configuration (SPP)

$\mathbf{z}$ is a random measure with mean $t\, m_r(x)\, \nu(dx)$ at (r, x)
$\mathbf{x}$ is marginal random measure with mean $t\, m_.(x)\, \nu(dx)$ at x
$\pi_r(x) = m_r(x)/m_.(x)$

Hence $E(\mathbf{z}(r, dx)) = \pi_r(x) E(\mathbf{x}(dx))$ for all (r, x) implies

$$\begin{aligned} T(\mathbf{x}, \mathbf{y}) &= \sum_{x \in \mathrm{SPP}} h(x)(Y(x) - \pi(x)) \\ &= \int_{\mathcal{X}} h(x) \Big(\mathbf{z}(r, dx) - \pi_r(x) \mathbf{x}(dx) \Big) \end{aligned}$$

has zero expectation. But $E(T \mid \mathbf{x}) \neq 0$.

variance calculation: binary case

$(\mathbf{y}, \mathbf{x})$ generated by point process;

$$T(\mathbf{x}, \mathbf{y}) = \sum_{x \in \mathrm{SPP}} h(x)(Y(x) - \pi(x))$$

$E(T(\mathbf{x}, \mathbf{y})) = 0; \qquad E(T \mid \mathbf{x}) \neq 0$

$$\begin{aligned} \mathrm{var}(T) = & \int_{\mathcal{X}} h^2(x)\pi(x)(1 - \pi(x))\, m_{.}(x)\, dx \\ & + \int_{\mathcal{X}^2} h(x)h(x')\, V(x, x')\, m_{..}(x, x')\, dx\, dx' \\ & + \int_{\mathcal{X}^2} h(x)h(x')\Delta^2(x, x') m_{..}(x, x')\, dx\, dx' \end{aligned}$$

V: spatial or within-cluster correlation;
Δ: interference

What is interference?

Physical/biological interference:
distribution of $Y(u)$ depends on $x(u')$

Sampling interference for autogenerated units

$$m_r(x) = E(\lambda_r(x)); \quad m_{rs}(x, x') = E(\lambda_r(x)\lambda_s(x'))$$

Univariate distributions:

$$\pi_r(x) = m_r(x)/m_.(x) = \mathrm{pr}(Y(x) = r \mid x \in \mathrm{SPP})$$

Bivariate distributions: $\pi_{rs}(x, x') = m_{rs}(x, x')/m_{..}(x, x')$

$$\pi_{rs}(x, x') = \mathrm{pr}(Y(x) = r, \; Y(x') = s \mid x, x' \in \mathrm{SPP})$$

Hence $\pi_{r.}(x, x') = \mathrm{pr}(Y(x) = r \mid x, x' \in \mathrm{SPP})$

$$\Delta_r(x, x') = \pi_{r.}(x, x') - \pi_r(x)$$

No second-order sampling interference if $\Delta_r(x, x') = 0$

Inference: Conventional Gaussian model

$$\text{Model} \quad p_{\mathbf{x}}(A) = N_n(X\beta,\ \Sigma_{\mathbf{x}} = \sigma_0^2 I_n + K[\mathbf{x}])(A)$$
$$\text{Rationalization} \quad Y(i) = x_i'\beta + \epsilon_i + \eta(x(i))$$

Stratum average: $\bar{Y}(\mathcal{U}_x) = \text{ave}\{Y_i \mid i\colon x_i = x\} = x'\beta + \eta(x)$

Conditional distribution of Y_u for $u \in \mathcal{U}_x$ given observation y, X

$$Y_u \mid \text{data} \sim N(x'\beta + k'\Sigma_{\mathbf{x}}^{-1}(y - X\beta),\ \Sigma_{uu} - k'\Sigma_{\mathbf{x}}^{-1}k)$$
$$\bar{Y}(\mathcal{U}_x) \mid \text{data} \sim N(x'\beta + k'\Sigma_{\mathbf{x}}^{-1}(y - X\beta),\ K(x,x) - k'\Sigma_{\mathbf{x}}^{-1}k)$$

$k_i = K(x, x_i)$, (such as $e^{-|x-x_i|}$ or $|x - x_i|^3$)

Stratum average is a random variable, not a parameter
Estimate is a distribution (not a function of sufficient statistic)
Likewise for GLMMs

Inference and predicton for the PP model

For a sequential sample
Observation $(\mathbf{x}, \mathbf{y}) \equiv (\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(k-1)})$
Product density $m_r(\mathbf{x}^{(r)}) = E(\prod_{x \in \mathbf{x}^{(r)}} \lambda_r(x))$ for class r
Conditional distribution as a random labelled partition of $\mathbf{x}$:

$$p(\mathbf{y} \,|\, \mathbf{x}) \propto m_0(\mathbf{x}^{(0)}) \cdots m_{k-1}(\mathbf{x}^{(k-1)})$$

For a subsequent autogenerated event

$$p(Y(x') = r \,|\, \text{data}, x' \in \text{SPP}) \propto m_r(\mathbf{x}^{(r)}, x')/m_r(\mathbf{x}^{(r)})$$

Use likelihood or estimating function to estimate parameters
Use conditional distribution for inference/prediction

Brief summary of conclusions

(i) Reasonable case for fixed-population model in certain areas
laboratory work; field trials; veterinary trials;...

(ii) Good case for autogenerated units in other areas
clinical trials; marketing; crime; animal behaviour

(iii) The choice matters in random-effects models

(iv) $\pi(x) = E(Y_i \mid X_i = x)$ versus $\rho(x) = E(Y_i \mid i\colon X_i = x)$
attenuation or non-attenuation

(v) What is modelled and estimated by PA?
claims to estimate $\rho(x)$ but actually estimates $\pi(x)$

(vi) What does the GLMM likelihood estimate?
Difficult to say; probably neither