Densities for Poisson processes

(The distribution of) A point process \( X \) has density \( f \) (is absolutely continuous) wrt. (the distribution of) a point process \( Y \) if

\[
P(X \in F) = \mathbb{E}1[Y \in F]f(Y)
\]

Suppose \( \rho_1(\cdot) \) and \( \rho_2(\cdot) \) are intensity functions so that \( \mu_1(S) \) and \( \mu_2(S) \) are finite and that \( \rho_1(u) > 0 \Rightarrow \rho_2(u) > 0 \). Then Poisson\( (S, \rho_1) \) has density

\[
f(x) = \exp(\mu_2(S) - \mu_1(S)) \prod_{u \in x} \frac{\rho_1(u)}{\rho_2(u)}
\]

wrt. Poisson\( (S, \rho_2) \)

In particular, for bounded \( S \), Poisson\( (S, \rho) \) has density

\[
\exp(|S| - \mu(S)) \prod_{u \in x} \rho(u)
\]

wrt. standard (unit-rate) Poisson process Poisson\( (S, 1) \).

The log likelihood function and derivatives I

\( x \) observation of \( X \) Poisson\( (W, \lambda(\cdot; \beta)) \).

Density wrt. unit rate Poisson process:

\[
f(x; \beta) = \exp(|W| - \int_W \lambda(u; \beta)du) \prod_{u \in x} \lambda(u; \beta) = \exp(|W| - \int_W \lambda(u; \beta)du) \exp(t \beta^T)
\]

where \( t = \sum_{u \in x} z(u) \) (exponential family density)

log likelihood function:

\[
l(\beta) = \sum_{u \in x} z(u)\beta^T - \int_W \lambda(u; \beta)du
\]

score function:

\[
u(\beta) = \sum_{u \in x} z(u) - \int_W z(u)\lambda(u; \beta)du
\]

Note: by Slivnyak-Mecke, score function has expectation zero.
The log likelihood function and derivatives II

Observed information (=Fisher information) \((p \times p\) matrix):
\[
j(\beta) = \int_W z(u)^T z(u) \lambda(u; \beta) du
\]

NB: by (extended) Slivnyak-Mecke
\[
\text{Var} u(\beta) = \text{Var} \sum_{u \in X} z(u) = j(\beta)
\]

Suppose \(j(\beta)\) positive definite (typically the case).
Newton-Raphson for finding root of likelihood equation:
\[
\beta^{m+1} = \beta^m + u(\beta^m)j(\beta^m)^{-1}
\]
(see Sections 8.2, 8.3 and 8.5)

Quadrature schemes

Simple quadrature scheme: Riemann approximation \((w(u)\) area of cell corresponding to \(u \in Q\)).

\text{spatstat}: Q = D \cup X. Two types of weights:
1. (grid) \(w(u) = \frac{|C_u|}{\sum_{u \in D} |C_u|}\) where \(W = \cup_{u \in D} C_u\) disjoint subdivision with \(u \in C_u\).
2. (Dirichlet) \(w(u)\) area of Dirichlet cell for \(u\) in Dirichlet tessellation generated by \(Q\) (page 264 in M&W).

Numerical approximation of the likelihood function

Problem: often impossible to evaluate integral \(\int_W z(u) \lambda(u; \beta) du\) in score function (e.g. if covariates only observed at finite set of locations \(Q \subset W\)).
Numerical quadrature (R package \text{spatstat})
\[
u(\beta) \approx \sum_{u \in X} z(u) - \sum_{u \in Q} z(u) \lambda(u; \beta) w(u)
\]
\((w(u)\) quadrature weight for quadrature point \(u \in Q\), page 174 in M&W).

Relation to generalized linear models and iterative weighted least squares

Let \(Z\) be the matrix with rows \(z(u), u \in X \cup D\), and \(V\) the diagonal matrix with diagonal entries \(w(u) \lambda(u; \beta)\). Then we may rewrite the \text{spatstat} approximation of score function as
\[
Z^T (1[u \in x] - w(u) \lambda(u; \beta))_{u \in X \cup D} = Z^T V y
\]
with corresponding information matrix
\[
Z^T V Z
\]
where \(y = V^{-1} (1[u \in x] - w(u) \lambda(u; \beta))_{u \in X \cup D}\).
Approximate score function formally equivalent to score function for Poisson regression (generalized linear model) and Newton-Raphson steps:
\[
Z^T V Z (\beta^{m+1} - \beta^m) = Z^T V y
\]
equivalent to iterative weighted least squares.
The conditions for consistency and asymptotic normality are that

1. \( c_n a_n^{-2} \to 0. \)
2. \( u_n(\theta) = 0 \) has almost surely a unique solution \( \hat{\theta}_n \) for each \( n \).
3. \( j_n/a_n^2 \to F \) in probability under \( P \) for a positive definite matrix \( F \).
4. For all \( c > 0 \), \( \sup_{\|\theta - \theta^*\| \leq c} \| j_n(\theta) - j_n \| / a_n^2 = \gamma_n c \to 0 \) in probability under \( P \).
5. The normalized score function \( u_n/c_n \) is asymptotically zero-mean normal with covariance matrix \( \Sigma \).

The use of different normalizing sequences \( a_n \) and \( c_n \) is not standard in the literature on asymptotics for estimating functions. Often \( a_n = c_n = \sqrt{n} \).

The second condition is assumed for ease of exposition and can be relaxed in view of the third condition.

### Distribution of parameter estimates: a general result

Consider a parametrized family of probability measures \( P_\theta, \theta \in \mathbb{R}^p \), and a sequence of estimating functions \( u_n : \mathbb{R}^p \to \mathbb{R}^p \), \( n \geq 1 \). The distribution of \( \{u_n(\theta)\}_{n \geq 1} \) is governed by \( P = P_{\theta^*} \) where \( \theta^* \) denotes the ‘true’ parameter value. Similarly, \( u_n = u_n(\theta^*) \) and \( j_n = j_n(\theta^*) \) where \( j_n(\theta) \) is the derivative of \( -u_n(\theta) \). For a matrix \( A = [a_{ij}] \), \( \|A\| = \max_{ij} |a_{ij}| \).

The following theorem ensures \( a_n^2/c_n \) consistency of \( \hat{\theta}_n \) and asymptotic normality where \( a_n \) and \( c_n \) are normalizing sequences so that \( c_n a_n^{-2} \to 0 \).

**Theorem** Under conditions to be stated on the next slide, for each \( \epsilon > 0 \), there exists a \( c > 0 \) such that

\[
P(\|\hat{\theta}_n - \theta^*\|^2 / c_n < c) > 1 - \epsilon
\]

whenever \( n \) is sufficiently large. Moreover, for matrices \( F \) and \( \Sigma \),

\[
(\hat{\theta}_n - \theta^*) a_n^2 / c_n \to N(0, F^{-1}\Sigma F^{-1})
\]

### Heuristic considerations

**Consistency:** follows from convergence of \( u_n \) (asymptotically unbiased).

**Asymptotic normality:** Taylor expansion \( (a_n = c_n = \sqrt{n}) \):

\[
u_n(\theta^*)/\sqrt{n} \approx u_n(\hat{\theta}_n)\sqrt{n} + \sqrt{n}(\hat{\theta}_n - \theta^*) [j_n(\lambda_n) / n]
\]

where \( \lambda_n \) between \( \theta^* \) and \( \hat{\theta}_n \). Hence,

\[
\sqrt{n}(\hat{\theta}_n - \theta^*) [j_n(\lambda_n) / n] = \sqrt{n}(\hat{\theta}_n - \theta^*) F + \sqrt{n}(\hat{\theta}_n - \theta^*) [j_n(\lambda_n) / n - F] \to N(0, \Sigma)
\]

Using \( \sqrt{n} \) consistency

\[
\sqrt{n}(\hat{\theta}_n - \theta^*) F \approx N(0, \Sigma)
\]

(we return to detailed proof in case \( p = 1 \) next time)

(alternative (Azzalini, \( p = 1 \)): second order Taylor and boundedness of second derivative of \( u_n \))
What is $n$ for a spatial Poisson process

- increasing intensity: $\lambda_n(\cdot; \theta) = n \exp(z(u)\beta^T)$
- increasing domain: sequence of increasing observation windows $W_n$ of increasing area $n$.

Asymptotic normality straightforward for increasing intensity. Also for increasing domain provided

$$\frac{1}{n} j_n(\theta) = \frac{1}{n} \int_{W_n} z(u)^T z(u) \exp(z(u)\beta^T) du$$

converges (Lindeberg/Lyapounov conditions).

Asymptotic normality of MLE

$$\sqrt{n} (\hat{\beta}_n - \beta^*) \rightarrow N(0, I)$$

where (increasing intensity)

$$I = \int_W z(u)^T z(u) \exp(z(u)(\beta^*)^T) du$$

or (increasing domain)

$$I = \lim_{n \to \infty} \frac{1}{n} \int_{W_n} z(u)^T z(u) \exp(z(u)(\beta^*)^T) du$$

In practice, $I$ is consistently estimated by observed information

$$j_n(\hat{\beta}_n)/n = \int_W z(u)^T z(u) \exp(z(u)\hat{\beta}_n^T) du$$

Correcting for incomplete covariate data

Consider $n$ random uniform independent uniform dummy points on $W$.

Monte Carlo approximation of integral in score function:

$$\frac{|W|}{\sqrt{n}} \left[ \sum_{u \in D} z(u) \exp(z(u)\beta^T) - \int_W z(u)^T z(u) \exp(z(u)\beta^T) du \right] \rightarrow N(0, G_g)$$

where $g(u) = z(u) \exp(z(u)\beta^T)$ and

$$G_r = \int_W f(u)^T f(u) du - \frac{1}{|W|} \int_W f(u)^T du \int_W f(u) du.$$

Monte Carlo approximation of score function:

$$u_n^{mc}(\beta) = \sum_{u \in X} z(u) - \frac{1}{n} \sum_{u \in D} z(u) \lambda(u; \beta) = u_n(\beta) + \left[ \int_W z(u) \lambda(u; \beta) du - \frac{1}{n} \sum_{u \in D} z(u) \lambda(u; \beta) \right]$$

Hence,

$$u_n^{mc}(\beta^*)/\sqrt{n} \rightarrow N(0, I + G_g)$$

and

$$\sqrt{n}(\hat{\beta}_n^{mc} - \beta^*)I \rightarrow N(0, I + G_g)$$

Note $\hat{\beta}_n^{mc}$ approximate MLE obtained from estimating function $u_n^{mc}(\beta)$.
Note: today considered following general topics:

- maximum likelihood inference
- generalized linear models
- iterative weighted least squares
- estimating functions
- asymptotic inference