Definition of Poisson process

A point process X on $\mathbb{R}^2$ is a **Poisson process** with **intensity function** $\lambda : \mathbb{R}^2 \rightarrow [0, \infty)$ if

- $N_X(A)$ is Poisson distributed with mean
$$\Lambda(A) = \int_A \lambda(x) dx$$
for every bounded set $A \subset \mathbb{R}^2$;

- for any k disjoint bounded sets $A_1, \ldots, A_k$, $k \in \mathbb{N}$, the random variables $N_X(A_1), \ldots, N_X(A_k)$ are independent.

Poisson process – conditionally independent points

Let X be a Poisson process on $\mathbb{R}^2$ with intensity function λ and $A \subset \mathbb{R}^2$ bounded with $\Lambda(A) > 0$.

Claim: Conditionally on the event $\{N_X(A) = n\}$, X restricted to A consists of n **independent** points scattered according to the pdf

$$\frac{\lambda(x)}{\int_A \lambda(a) da}, \quad x \in A.$$

Proof

Clearly, conditionally on $N_X(A) = n$, $p_n = 1$ and $p_m = 0$ for $m \neq n$.

Fix $x_1, \ldots, x_n \in A$ and set $B_i = B(x_i, \epsilon)$ for ϵ small enough to make the B_i disjoint. Then

$$\mathbb{P}(N_X(B_1) = 1; \cdots; N_X(B_n) = 1 | N_X(A) = n)$$

$$= \frac{\mathbb{P}(N_X(B_1) = 1; \cdots; N_X(B_n) = 1; N_X(A \setminus \cup B_i) = 0)}{\mathbb{P}(N_X(A) = n)}.$$

Using the two defining properties, we get

$$\frac{\Lambda(B_1)e^{-\Lambda(B_1)} \cdots \Lambda(B_n)e^{-\Lambda(B_n)}e^{-\Lambda(A \setminus \cup B_i)}}{\Lambda(A)^n e^{-\Lambda(A)}/n!} = \frac{n!}{\Lambda(A)^n} \prod_{i=1}^{n} \Lambda(B_i).$$

Upon dividing by $n!$, the number of permutations of n points, and letting ϵ go to zero, the probability that the first random point falls in dx_1, the second one in dx_2, the last one in dx_n, is

$$j_n(x_1, \ldots, x_n | N_X(A) = n) = \frac{1}{\Lambda(A)^n} \Lambda(B_1) \cdots \Lambda(B_n) = \frac{1}{\Lambda(A)^n} \lambda(x_1)dx_1 \cdots \lambda(x_n)dx_n$$

in accordance with the claim.

The Poisson process – likelihood

Let $\mathbf{x}$ be a realisation of a Poisson process with intensity function λ observed in window $W \subset \mathbb{R}^2$. Suppose that $\lambda(x) = \lambda(x;\theta)$ depends on an unknown parameter θ.

Since the likelihood is

$$e^{-\Lambda(W)} \frac{\Lambda(W)^{n(\mathbf{x})}}{n(\mathbf{x})!} n(\mathbf{x})! \prod_{x \in \mathbf{x}} \frac{\lambda(x)}{\Lambda(W)},$$

inference may be based on the **log likelihood**

$$L(\theta) = L(\mathbf{x};\theta) = \sum_{x \in \mathbf{x}} \log \lambda(x;\theta) - \int_W \lambda(w;\theta) dw.$$

Inference for the homogeneous Poisson process

Let $\mathbf{x}$ be a realisation of a homogeneous Poisson process in W.

Then

$$\lambda(x;\theta) = \theta, \qquad \theta > 0,$$

and the log likelihood reads

$$L(\theta) = n(\mathbf{x}) \log \theta - \theta |W|.$$

Maximum likelihood estimator of intensity

The score function is

$$S(\theta) = \frac{\partial}{\partial\theta} L(\theta) = \frac{n(\mathbf{x})}{\theta} - |W|.$$

Solving $S(\theta) = 0$ yields the **maximum likelihood estimate**

$$\widehat{\theta} = \frac{n(\mathbf{x})}{|W|}.$$

The sign of the second derivative confirms that $\widehat{\theta}$ is the unique maximiser of $L(\theta)$.

Maximum likelihood estimator – moments

Let X be a homogeneous Poisson process on W with intensity $\theta > 0$.

Then the estimator $\widehat{\theta}(X) = n_X(W)/|W|$ is **unbiased** and

$$\text{Var}_\theta(\widehat{\theta}(X)) = \frac{\theta}{|W|}.$$

To see the latter, recall that

$$\begin{aligned}\text{Var}_\theta(N_X(W)) &= \mu^{(2)}(W \times W) - (\mu^{(1)}(W))^2 \\ &= \theta^2|W|^2 + \theta|W| - (\theta|W|)^2 = \theta|W|.\end{aligned}$$

Efficiency

An estimator $\widehat{\theta}(X)$ of θ is **minimum variance unbiased** (aka **efficient**) if

- $\mathbb{E}_\theta(\widehat{\theta}(X)) = \theta$ for all θ in the parameter space Θ;
- $\text{Var}_\theta(\widehat{\theta}(X)) \leq \text{Var}_\theta T(X)$ for all $\theta \in \Theta$ and all unbiased estimators T of θ.

It can be shown that (under some technical conditions) any unbiased estimator $\widehat{\theta}(X)$ that attains the **Cramér–Rao** lower bound, i.e.

$$\text{Var}_\theta(\widehat{\theta}(X)) = I(\theta)^{-1},$$

is efficient.

Maximum likelihood estimator – efficiency

Let X be a homogeneous Poisson process on W with intensity $\theta > 0$.

Then the **Fisher information**

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}L(X;\theta)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}L(X;\theta)\right]$$

reads

$$\mathbb{E}_\theta\left[\left(\frac{N_X(W)}{\theta} - |W|\right)^2\right] = \frac{1}{\theta^2}\mathrm{Var}_\theta(N_X(W)) = \frac{|W|}{\theta}$$

so $\widehat{\theta}(X) = n_X(W)/|W|$ is **efficient**.

Approximate confidence intervals

By the **normal approximation** to a Poisson,

$$\mathbb{P}_\theta\left(\phi_{0.025} < \frac{N_X(W) - \theta|W|}{\sqrt{\theta|W|}} < \phi_{0.975}\right) \approx 0.95$$

where ϕ_α is the α-quantile of the standard normal distribution.

The denominator $\sqrt{\theta|W|}$ depends on θ. Using a **second approximation**

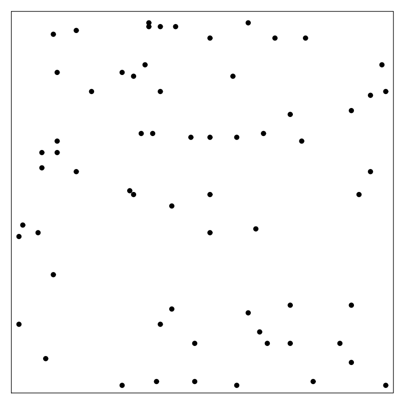
$$\sqrt{\theta|W|} \approx \sqrt{\widehat{\theta}|W|} = \sqrt{N_X(W)},$$

one obtains an approximate confidence interval

$$\left(\widehat{\theta} - \phi_{0.975}\sqrt{\frac{N_X(W)}{|W|^2}}, \quad \widehat{\theta} + \phi_{0.025}\sqrt{\frac{N_X(W)}{|W|^2}}\right).$$

Example: Numata pine data

`japanesepines` lists the locations of 65 Japanese black pine saplings in a 5.7×5.7 metre square in a natural forest near Choshi, Japan, rescaled to the unit square as collected by Numata (1961).



Example: Numata pine data

```
fitHom <- ppm(japanesepines ~ 1)
print(fitHom)

Stationary Poisson process
Intensity: 65
             Estimate      S.E.  CI95.lo  CI95.hi
log(lambda) 4.174387 0.1240347 3.931284 4.417491
```

Caution: The `ppm`-function calculates the maximum likelihood estimator of $\log\theta$. Since the transformed mle is the mle of the transformed parameter,

$$\widehat{\theta} = \exp(4.17438) = 65.$$

Delta method for standard errors

Idea: Use the Taylor expansion of $\log x$ around the mean

$$\log \widehat{\theta} \approx \log \theta + \frac{1}{\theta}(\widehat{\theta} - \theta)$$

to get an approximation of the variance of $\log \widehat{\theta}$,

$$\mathrm{Var}_\theta(\log \widehat{\theta}) \approx \frac{1}{\theta^2}\, \mathrm{Var}_\theta(\widehat{\theta}).$$

Plugging in $\mathrm{Var}_\theta(\widehat{\theta}) = \theta/|W|$, we find that the variance of $\log \widehat{\theta}$ is approximately $1/(\theta|W|)$.

Example: Numata pine data

For the pine data, the approximate standard error of $\log \widehat{\theta}$ is $\sqrt{1/65} = 0.1240347$.

The approximate confidence interval on the log scale reads

$$\left(\log \widehat{\theta} - \phi_{0.975}\sqrt{1/65}, \quad \log \widehat{\theta} + \phi_{0.025}\sqrt{1/65}\right)$$

$$= (3.931284, \quad 4.417491)\,.$$

Inference for baseline model

Let $\mathbf{x}$ be a realisation of a Poisson process with intensity function

$$\lambda(x;\theta) = \theta b(x)$$

in W, where $\theta > 0$ and $b : W \to [0, \infty)$ is a known function (e.g. population density) for which $0 < \int_W b(w)dw < \infty$.

Then

$$L(\theta) = \sum_{x \in \mathbf{x}} \log(\theta b(x)) - \theta \int_W b(w)dw.$$

Maximum likelihood estimator baseline model

The score function is

$$S(\theta) = \frac{n(\mathbf{x})}{\theta} - \int_W b(w)dw.$$

Solving $S(\theta) = 0$ yields the **maximum likelihood estimate**

$$\widehat{\theta} = \frac{n(\mathbf{x})}{\int_W b(w)dw}.$$

The sign of the second derivative confirms that $\widehat{\theta}$ is the unique maximiser of $L(\theta)$.

Maximum likelihood estimator – moments

Let X be a Poisson process on W with intensity function $\theta b(x)$. Write $B(W) = \int_W b(w)dw$.

Then the estimator $\widehat{\theta}(X) = n_X(W)/B(W)$ is **unbiased** and

$$\mathrm{Var}_\theta(\widehat{\theta}(X)) = \frac{\theta}{B(W)}.$$

To see the latter, recall that

$$\mathrm{Var}_\theta(N_X(W)) = \mu^{(2)}(W \times W) - (\mu^{(1)}(W))^2$$

$$= \theta^2 B(W)^2 + \theta B(W) - (\theta B(W))^2 = \theta B(W).$$

Maximum likelihood estimator – efficiency

Let X be a Poisson process on W with intensity function $\theta b(x)$.

Then the **Fisher information**

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}L(X;\theta)\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}L(X;\theta)\right]$$

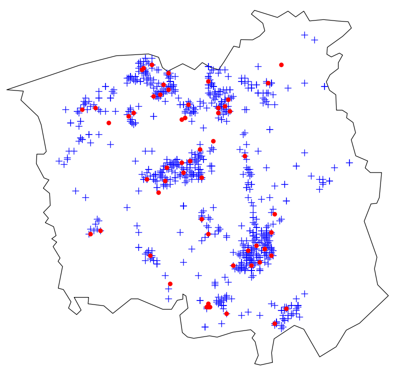
reads

$$\mathbb{E}_\theta\left[\left(\frac{N_X(W)}{\theta} - B(W)\right)^2\right] = \frac{1}{\theta^2}\mathrm{Var}_\theta(N_X(W)) = \frac{B(W)}{\theta}$$

so $\widehat{\theta}(X) = n_X(W)/B(W)$ is **efficient**.

Example: Chorley–Ribble data

`chorley` contains the locations of 58 cases of larynx (Dutch: strottenhoofd) and 978 of lung cancer in a Lancashire Health Authority (about 315 square km) recorded between 1974 and 1983 (Diggle, 1990).



Chorley–Ribble data: Baseline model

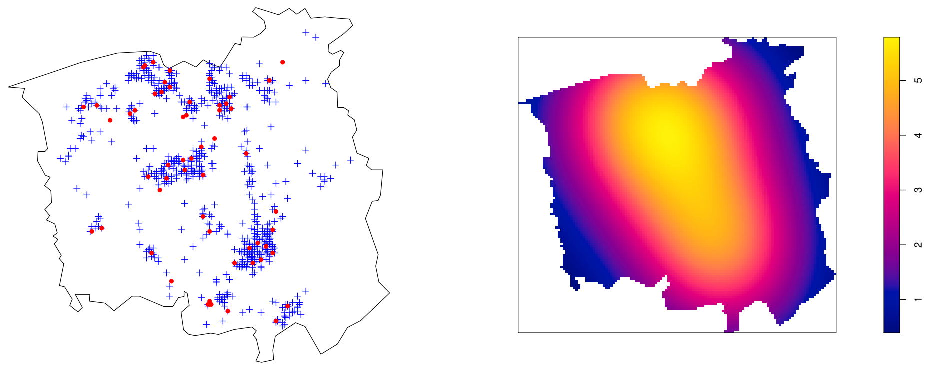
Assume that the rare larynx cancer cases follow a Poisson process with intensity function

$$\lambda(x; \theta) = \theta b(x),$$

where $b(x)$ is the density at x of the susceptible population.

Since b is not given, we may use the smoothed **intensity function** of the more common lung cancer cases as a **proxy**.

Chorley–Ribble data: Intensity function



```
lung <- split(chorley)$lung
larynx <- split(chorley)$larynx
blung <- density.ppp(lung, sigma=4.0, at="pixels",
    leaveoneout=FALSE, edge=TRUE, diggle=TRUE)
```

Example: Chorley–Ribble data

The bandwidth $h = 4.0$ was chosen by the Cronie–Van Lieshout method. A tighter fit around the lung cancer cases may be obtained by lowering the bandwidth ($h = 1.0$). Since local edge correction is mass preserving, $\widehat{\theta} = 58/978 = 0.06$, or $\log \widehat{\theta} = -2.83$.

```
ppm(larynx ~ offset(log(blung)))
```

gives

```
Fitted trend coefficient:  (Intercept) = -2.826217

              Estimate      S.E.   CI95.lo   CI95.hi
(Intercept) -2.826217 0.1313064 -3.083573 -2.568861
```

Log-linear models

The general log-linear model reads

$$\log \lambda(x;\theta) = \theta_0 + b(x) + \sum_{j=1}^{p} \theta_j C_j(x),$$

where θ_j are the model parameters, b is the known offset and C_j are known covariates.

Upon observation of $\mathbf{x} = \{x_1, \ldots, x_n\}$ in W, the log likelihood function $L(\theta)$ is

$$n\theta_0 + \sum_{j=1}^{p} \sum_{i=1}^{n} \theta_j C_j(x_i) - e^{\theta_0} \int_W \exp\left[b(w) + \sum_{j=1}^{p} \theta_j C_j(w) \right] dw.$$

Maximum likelihood estimator

The score equations are

$$\int_W C_j(w)\lambda(w;\theta)dw = \sum_{i=1}^{n} C_j(x_i)$$

for $j = 0, \ldots, p$ under the convention that $C_0 \equiv 1$.

The Hessian matrix $H(\theta)$ of second order partial derivatives has entries

$$H(\theta)_{ij} = \frac{\partial^2}{\partial\theta_i \partial\theta_j} L(\theta) = -\int_W C_i(w)C_j(w)\lambda(w;\theta)dw.$$

Note: $H(\theta)$ does not depend on $\mathbf{x}$.

Maximum likelihood estimator – remarks

- In general, the score equations cannot be solved explicitly.

- Any $\widehat{\theta}$ that solves the score equations and for which $H(\widehat{\theta})$ is negative definite is a local maximum of the log likelihood function $L(\theta)$.

- The integral in $L(\theta)$ must be approximated numerically.

Maximum likelihood estimator – asymptotics

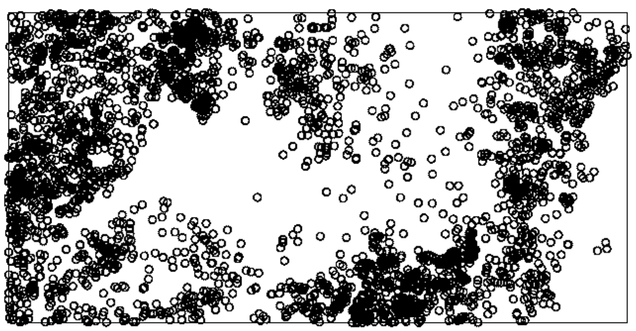
Little is known about the small sample properties of $\widehat{\theta}$.

Theorem (Kutoyants, 1998) Under suitable regularity conditions (p.49-51), when the window W grows to $\mathbb{R}^2$, $\widehat{\theta}$ is approximately multivariate normally distributed with mean θ and covariance matrix $I(\theta)^{-1}$, where the **Fisher information** is

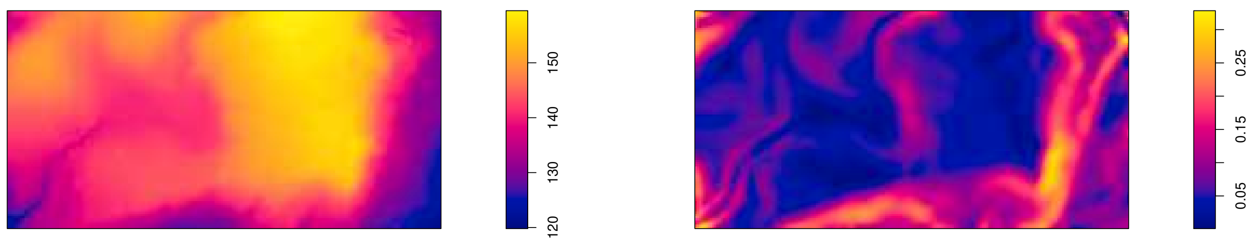
$$I(\theta) = -\mathbb{E}_\theta \left[\nabla^2 L(X;\theta) \right] = -H(\theta).$$

Example: Barro Colorado data

`bei` contains the locations of 3604 Beilschmiedia trees in a 1000×500 metre region in the tropical rainforest of Barro Colorado Island (Hubbell and Foster, 1983).



Barro Colorado: Covariates



Left: elevation; Right: gradient norm. Images from `bei.extra`.

Barro Colorado: Model with covariates

Write $E(x)$ for the elevation at x, $G(x)$ for the norm of the gradient. The model

$$\log \lambda(x) = \theta_0 + \theta_1 E(x) + \theta_2 G(x)$$

can be fitted by

```
fitbei <- ppm(bei ~ elev + grad, data=bei.extra)
```

which yields $\widehat{\theta}_0 = -8.6$, $\widehat{\theta}_1 = 0.02$ and $\widehat{\theta}_2 = 5.8$.

Interpretation: There tend to be more trees at higher elevation and steeper slopes.

Barro Colorado: Heterogeneous model

A model in which $\log \lambda(x, y)$ is a fourth order polynomial can be fitted by

```
fitbeiXY <- ppm(bei ~ polynom(x,y,4))
```

Note: For this model, the Fisher information matrix is singular.

Model validation by residuals

A residual analysis compares a kernel estimate of the intensity function to that of the fitted model ($\hat{\lambda}(\cdot) = \lambda_{\hat{\theta}}(\cdot)$):

$$s(x) = h^{-2} \sum_{y \in \mathbf{x}} \kappa \left(\frac{x-y}{h} \right) w_h(x,y)^{-1}$$

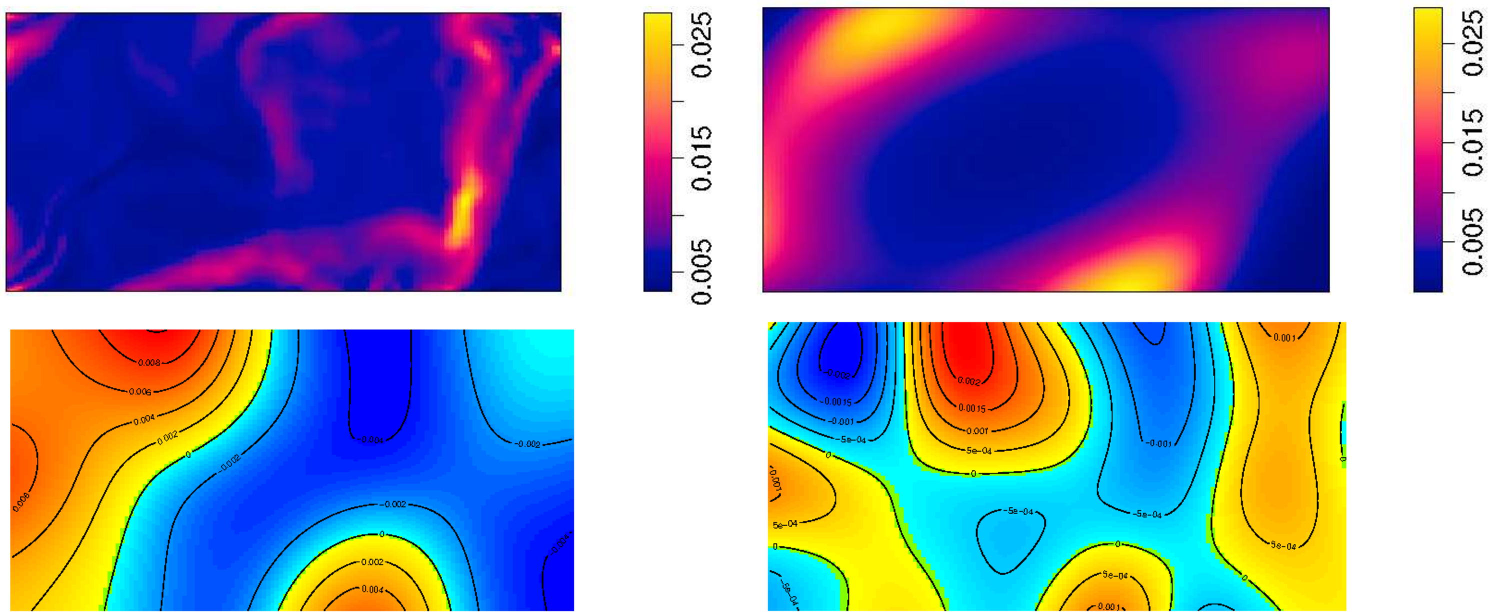
$$-h^{-2} \int_W \kappa \left(\frac{x-w}{h} \right) w_h(x,w)^{-1} \hat{\lambda}(w) dw,$$

where κ is a probability density function and w_h an edge correction factor.

In **spatstat**, use

```
plot(predict(fitbei))
diagnose.ppm(fitbei, which="smooth", sigma=100)
```

Results



Left: covariate model; Right: heterogeneous model.

Updating a model

Model building may be an iterative process. This is facilitated by the function `update.ppm`:

```
fitbeiCSR <- ppm(bei ~ 1, data=bei.extra)
fitbeiGrad <- update.ppm(fitbeiCSR, ~ . + grad)
fitbeiGradElev <- update.ppm(fitbeiGrad, ~ . + elev)
fitbeiAll <- update.ppm(fitbeiGradElev,
   ~ . + grad:elev)
```

The `dot` symbol means 'what was already there'. Instead of adding terms, one may delete terms too.

Factor covariates

Consider the model

$$\log \lambda(x) = \theta + \alpha_{A(x)},$$

where $A(x) \in \{1, 2\}$ is the level of the covariate at x.

The model is overparametrised, as there are three parameters θ, α_1 and α_2, but only two values ($\theta + \alpha_1$ and $\theta + \alpha_2$).

Note: Boolean values are treated as factors.

Example: Barro Colorado data

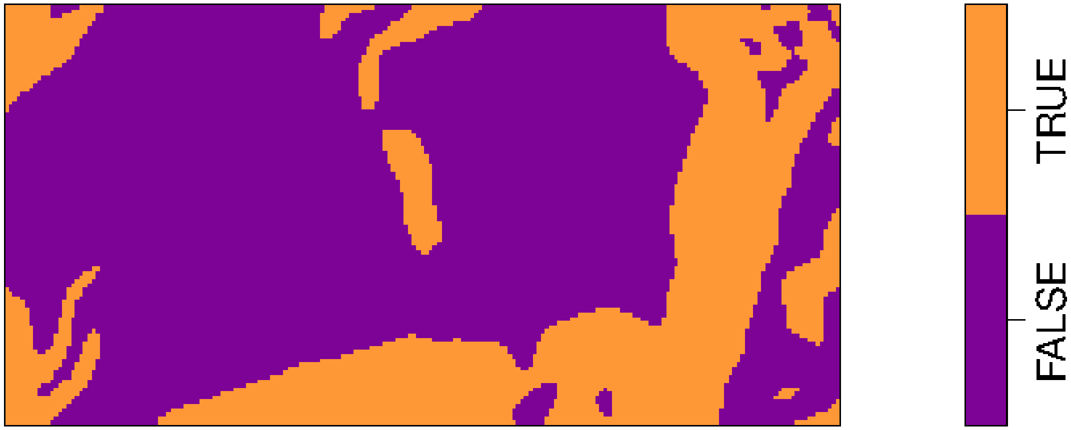
```
Slope <- (bei.extra$grad > 0.1)
ppm(bei ~ Slope)
```

yields

```
               Estimate       S.E.   CI95.lo    CI95.hi
(Intercept) -5.1338304 0.02194756 -5.1768468 -5.0908140
SlopeTRUE    0.5621068 0.03370676  0.4960428  0.6281709
```

so that

$$\widehat{\log \lambda(x)} = -5.13 + 0.56 \times 1\{G(x) > 0.1\}.$$


TRUE
FALSE

Two factor models

For the model

$$\log \lambda(x) = \theta + \alpha_{A(x)} + \beta_{B(x)} + \gamma_{A(x),B(x)},$$

$A(x) \in \{1, 2, \ldots, n_A\}$ and $B(x) \in \{1, 2, \ldots, n_B\}$, the default in **R** is to set

$$\alpha_1 = \beta_1 = 0$$

and

$$\gamma_{1,j} = 0, \quad j = 1, \ldots, n_B; \quad \gamma_{i,1} = 0, \quad i = 2, \ldots, n_A.$$

Likelihood ratio test

Return to the general log-linear model

$$\log \lambda(x;\theta) = \theta_0 + b(x) + \sum_{j=1}^{p} \theta_j C_j(x),$$

where θ_j are the model parameters, b is the known offset and C_j are known covariates.

Aim: Test whether the data depend significantly on the covariate function C_j.

Likelihood ratio test

To test

$$\mathcal{H}_0 : \theta_j = 0$$

against the alternative

$$\mathcal{H}_1 : \theta_j \neq 0,$$

use the test statistic

$$\Lambda(X) = \frac{\sup\{f(X;\theta) : \theta_j = 0\}}{f(X;\widehat{\theta})}$$

where $\widehat{\theta}$ is the **maximum likelihood estimator** and

$$f(X;\theta) = \exp\left[L(X;\theta)\right].$$

Composite likelihood ratio test

To test

$$\mathcal{H}_0 : \theta_1 = \cdots = \theta_p = 0,$$

use the test statistic

$$\Lambda(X) = \frac{\sup\{f(X;\theta) : \theta_1 = \cdots = \theta_p = 0\}}{f(X;\widehat{\theta})}.$$

Claim: $-2\log\Lambda(X)$ is approximately χ^2-distributed.

The number of degrees of freedom is equal to the difference in dimensionality between $\mathcal{H}_0$ and $\mathcal{H}_1$. Reject for large values.

Example: Barro Colorado data – component

$$\log \lambda(x;\theta) = \theta_0 + \theta_1 G(x)$$

```
> anova.ppm(fitbeiCSR, fitbeiGrad, test="LR")
Analysis of Deviance Table

Model 1: ~1   Poisson
Model 2: ~grad   Poisson
  Npar Df Deviance  Pr(>Chi)
1    1
2    2  1   383.11 < 2.2e-16 ***
```

rejects $\mathcal{H}_0 : \theta_1 = 0$ since $\mathbb{P}(\chi^2_1 > 383.11) \approx 0$.

Example: Barro Colorado data – composite

$$\log \lambda(x;\theta) = \theta_0 + \theta_1 G(x) + \theta_2 E(x)$$

```
> anova.ppm(fitbeiCSR, fitbeiGradElev, test="LR")
Analysis of Deviance Table

Model 1: ~1   Poisson
Model 2: ~grad + elev   Poisson
  Npar Df Deviance  Pr(>Chi)
1    1
2    3  2   472.81 < 2.2e-16 ***
```

rejects $\mathcal{H}_0 : \theta_1 = \theta_2 = 0$ since $\mathbb{P}(\chi^2_2 > 472.81) \approx 0$.

Model selection

Models that are not nested may be compared using the **Akaike Information Criterion** (AIC)

$$-2L(\widehat{\theta}) + 2\#(p)$$

which balances the fitted likelihood against the number of parameters in the model.

The model with the **lower** AIC is preferred.

Example: Barro Colorado data

```
> AIC(fitbei)
[1] 42295.11
> AIC(fitbeiXY)
[1] 40715.24
```

The heterogeneous model

$$\log\lambda(x,y) = \theta_0 + \theta_1 x + \theta_2 y + \theta_3 x^2 + \theta_4 xy + \theta_5 y^2 +$$
$$\theta_6 x^3 + \theta_7 x^2 y + \theta_8 xy^2 + \theta_9 y^3 + \theta_{10} x^4 + \theta_{11} x^3 y +$$
$$\theta_{12} x^2 y^2 + \theta_{13} xy^3 + \theta_{14} y^4$$

is preferred over the covariate model

$$\log\lambda(x,y) = \theta_0 + \theta_1 E(x,y) + \theta_2 G(x,y).$$

Numerical considerations

Baddeley and Turner (1998) proposed to approximate

$$\int_W \lambda(w;\theta)dw \approx \sum_{j=1}^{n} \lambda(u_j;\theta)\, w_j,$$

where u_j are **dummy points** in W and w_j are **quadrature weights**.

Then

$$L(\theta) \approx \sum_{x\in\mathbf{x}} \log \lambda(x;\theta) - \sum_{j=1}^{n} \lambda(u_j;\theta)\, w_j.$$

Generalised log-linear Poisson regression model

Add the **data points** $x \in \mathbf{x}$ to the set of dummies to form $\{u_j : j = 1, \ldots, n + n(\mathbf{x}) = m\}$. Then

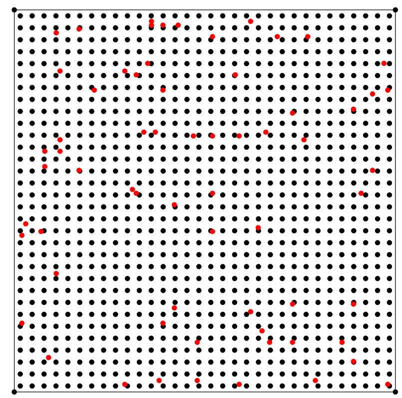
$$L(\theta) \approx \sum_{j=1}^{m} (y_j \log \lambda_j - \lambda_j)\, w_j,$$

where $\lambda_j = \lambda(u_j; \theta)$, $y_j = z_j / w_j$ and

$$z_j = \begin{cases} 1, & \text{if } u_j \in \mathbf{x} \text{ is a data point,} \\ 0, & \text{if } u_j \notin \mathbf{x} \text{ is a dummy point.} \end{cases}$$

Quadrature weights – counting weights

```
dj <- quadscheme(japanesepines, method="grid")
```



The weight $w_j = v/n_j$ where v is the volume of each cell and n_j is the number of points (dummy or data) in the same cell as u_j.

Quadrature weights – adaptive weights

Waagepetersen (2008) proposed to use the area of the Voronoi cells of **data and dummy** as weights:

$$\int_W \lambda(w;\theta)dw \approx \sum_{j=1}^{n} \frac{\lambda(u_j;\theta)}{\lambda(u_j;\theta)+n/|W|}$$

where the area of the Voronoi cell around u_j is approximated by $1/(\lambda(u_j;\theta)+n/|W|)$.

Implemented in `spatstat` by the argument `method="logi"` to the `ppm` function.

References

1. A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics* 42:283–322, 2000.

2. Y. A. Kutoyants. Statistical inference for spatial Poisson processes. Springer, 1998.

3. R. Waagepetersen. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika* 95:351–363.

Assessment

The `R`-package **spatstat** contains the dataset `chorley` and `chorley.extra`. The latter contains the location of an incinerator.

One is interested in the question whether the presence of the incinerator increases the risk of cancer of the larynx. Formulate a model to test this assumption and validate your results.