On the classification problem for Poisson Point Processes.

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, UdelaR

Escuela CIMPA - La Habana
1 Binary Classification
   - On the classification problem

2 Classification of P.P.P.
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3 Simulations
1. **Binary Classification**
   - On the classification problem

2. **Classification of P.P.P.**
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3. **Simulations**
On the classification problem

**Goal:**

From a training sample \( \mathcal{D}_n = \{ (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \} \) i.i.d. of \( (X, Y) \in \mathcal{F} \times \{0, 1\} \) we want a predictor \( g : \mathcal{F} \to \{0, 1\} \), which minimize \( P(g(X) \neq Y) \).

Let us denote

1) \( \eta(x) = \mathbb{E}(Y|X = x) = P(Y = 1|X = x) \) the regression function.

2) \( g^*(x) = I_{\{\eta(x) > 1/2\}} \) the Bayes rule.

3) \( L^* = P(g^*(X) \neq Y) \) the optimal Bayes risk.

If \( \eta_n(x) : \mathcal{E} \to [0, 1] \) and \( g_n(x) = I_{\{\eta_n(x) > 1/2\}} \)

\[
0 \leq P(g_n(X) \neq Y) - L^* \leq \mathbb{E}|\eta(X) - \eta_n(X)|^2.
\]

The classical estimator of \( \eta(x) \) is \( \eta_n(X) = \sum_{i=1}^n W_{ni}(X)Y_i \) for some weights \( W_{ni}(X) = W_{ni}(X, X_1, \ldots, X_n) \).

**k-NN in \( \mathbb{R}^d \)**

For \( W_{ni}(X) = \frac{1}{k}I_{\{X_i \in k_n(X)\}} \), where \( X_i \in k_n(X) \) if \( X_i \) is one of the \( k \) nearest neighbours of \( X \), conditions 1 to 5 holds if \( n \to \infty \) and \( k_n/n \to 0 \).
Motivation
Motivation

![Image Description]

- sad
- happy
- aff
- mixed
- anger
- fear
- disgust
- surprise
Given,

- \((S, \rho)\) a separable bounded metric space.
- \(\nu\) a Borel measure.
- \(S^\infty = \{x \subset S : \#x < \infty\}\).
- \(\lambda : S \to \mathbb{R}^+\) integrable.
- \((\Omega, \mathcal{A}, \mathbb{P})\) a probability space.

\(X : \Omega \to S^\infty\) is a Poisson Point Process on \(S\) with intensity \(\lambda\), \(X \sim \text{Poisson}(S, \lambda)\), if

- \(N_A : \Omega \to \{0, 1, \ldots, \infty\}, N_A(\omega) = \#(X(\omega) \cap A)\) are random variables for all \(A \in \mathcal{B}(S)\).
- given \(n\) disjoint Borel subsets \(A_1, \ldots, A_n\) of \(S\), \(N_{A_1}, \ldots, N_{A_n}\) are independent.
- \(N_A\) has Poisson distribution with mean \(\mu(A)\), being

\[
\mu(A) = \int_A \lambda(x)d\nu(x).
\]
1. Binary Classification
   - On the classification problem

2. Classification of P.P.P.
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3. Simulations
Radon Nikodym for P.P.P.

**Theorem**

Let $X_1 \sim \text{Poisson}(S, \lambda_1)$ and $X_2 \sim \text{Poisson}(S, \lambda_2)$ being $(S, \rho)$ a non-empty, bounded metric space. such that $\mu_i(S) < \infty$, $i = 1, 2$. Suppose that $\lambda_1(\xi) > 0 \Rightarrow \lambda_2(\xi) > 0$. Then, $P_{X_1} \ll P_{X_2}$, with density

$$f(x) = \exp \left[ \mu_2(S) - \mu_1(S) \right] \prod_{\xi \in x} \frac{\lambda_1(\xi)}{\lambda_2(\xi)},$$

with $0/0 = 0$.

**Corollary**

If $X_2 \sim \text{Poisson}(S, 1)$, then, for all $X \sim \text{Poisson}(S, \lambda)$ we have $P_X \ll P_{X_2}$ and

$$f(x) = \exp \left[ \nu(S) - \mu(S) \right] \prod_{\xi \in x} \lambda(\xi),$$

being $\mu(S) = \int_S \lambda d\nu$. 
Bayes rule

\[
\mathbb{P}(Y = 1|X = x) = \frac{f_{X_1}(x)\mathbb{P}(Y = 1)}{f_{X_1}(x)\mathbb{P}(Y = 1) + f_{X_0}(x)\mathbb{P}(Y = 0)},
\]

then,

\[
\mathbb{P}(Y = 1|X = x) > \frac{1}{2} \iff \exp \left[ \mu_0(S) - \mu_1(S) \right] \prod_{\xi \in x} \frac{\lambda_1(\xi)}{\lambda_0(\xi)} > \frac{(1 - p)}{p},
\]

where \( p = \mathbb{P}(Y = 1) \).

For two homogeneous Poisson processes

\[
\exp \left( (\lambda_0 - \lambda_1)\nu(S) \right) \left( \frac{\lambda_1}{\lambda_0} \right)^n > \frac{1 - p}{p}, \tag{1}
\]

where \( n = \#x \).
Bayes rule: estimating the intensity

Given a realization \( \{\xi_1, \ldots, \xi_n\} \) of \( X \), the intensity \( \lambda(x) \) for \( x \in S \) can be estimated by

\[
\hat{\lambda}(x) = \frac{1}{K_\sigma(x)} \sum_{i=1}^{n} k \left( \frac{\rho(x, \xi_i)}{\sigma} \right),
\]

where \( K_\sigma(x) = \int_S k \left( \frac{\rho(x, \xi)}{\sigma} \right) d\nu(\xi), \)

with \( k : \mathbb{R}^d \to \mathbb{R} \) a symmetric kernel, \( \sigma > 0 \) a smoothing parameter.

If we have \( \{X_1, \ldots, X_m\} \), where \( X_j = \{\xi_1, \ldots, \xi_{n(j)}\} \) \( j = 1, \ldots, m \), let us define \( k_{\sigma_m}(\cdot) = k(\cdot/\sigma_m) \),

\[
\hat{\lambda}_m(x) = \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j^m(x), \quad \hat{\lambda}_j^m(x) = \frac{1}{K_{\sigma_m}(x)} \sum_{i=1}^{n(j)} k_{\sigma_m}(\rho(x, \xi_i)).
\]

**Theorem**

If \( \sigma_m \to 0 \), we have that

\[
\lim_{m \to \infty} \sup_{x \in S} |\hat{\lambda}_m(x) - \lambda(x)| = 0 \quad a.s.
\]
Binary Classification
- On the classification problem

Classification of P.P.P.
- Bayes rule
- k-NN
- Other distances for k-NN

Simulations
A suitable distance for k-NN

**Hausdorff distance.**

Given two non-empty compact sets $A, C \subset S$,

$$d_H(A, C) = \max \left\{ \sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A) \right\},$$

(2)

where $d(a, C) = \inf \{ \rho(a, c) : c \in C \}$.

**Remark:** $(S^\infty, d_H)$ is separable.
Binary Classification
- On the classification problem

Classification of P.P.P.
- Bayes rule
- k-NN
- Other distances for k-NN

Simulations
Theorem, Consistency of k-NN

Let us consider

- \((X, Y) \in S^\infty \times \{0, 1\}\).
- \(X_1 \overset{\sim}{=} X|Y = 1 \sim Poisson(\lambda_1), \quad X_0 \overset{\sim}{=} X|Y = 0 \sim Poisson(\lambda_0)\).
- \(\lambda_1\) and \(\lambda_0\) continuous functions of \(\rho\).
- \(\nu\) does not have atoms (i.e. \(\nu(\{z\}) = 0\) for all \(z \in S\)).

Then \(k - NN\) is consistent.
Simulations

The model

We generate $N = 200$ data ($N/2$ for training), half of them (the 0-class) with Poisson distribution on $[0, 1]^2$ and intensity

$$\lambda_0(x, y) = c_0 \exp(-d_0((x - 1/2)^2 + (y - 1/2)^2)),$$

where we fixed $c_0 = 500$ and $d_0 = 20$.

$$\lambda_1((x, y), c_1, d_1, a) = c_1 \exp(-d_1((x - 1/2 - a_1)^2 + (y - 1/2 - a_2)^2)).$$
Level sets of the estimation of the densities

Figure: In all cases $d_0 = 20$, $c_0 = 500$. **First row:** $x = 0.03$, $x = 0.04$, $x = 0.05$. $d_1 = 20$, $c_1 = 500$. **Second row:** $x = 0$, $d_1 = 20, 30, 40$, $c_1 = 600$. **Third row:** $x = 0$, $d_1 = 20, 30, 40$, $c_1 = 700$. 