LECTURE 6

Two representations of Non-Poisson Point Processes (NP-PP) have been considered:

- Renewal Processes: a) $\Delta N_{(t,t+h]} \sim \mathcal{P}(\mu)$ with $\mu \triangleq \int_t^{t+h} \lambda(u)\,du \quad \forall t, h \geq 0$

  b) X_i i.i.d. $\forall i$ and

  $P(X_i > t+h \mid X_i > t) \neq P(X_i > s+h \mid X_i > s) \quad \forall h>0,\ t \neq s$

- History-dependent Point Processes: $\Delta N_{(t,t+h]} \sim \mathcal{P}(\mu)$ with $\mu \triangleq \int_t^{t+h} \lambda(u|\mathcal{H}_u)\,du \quad \forall t, h \geq 0$

$$\lambda(t|\mathcal{H}_t) \triangleq \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t]} > 0 \mid \mathcal{H}_t)}{\Delta t}$$ Conditional Intensity Function (CIF)

where: $\Delta N_{(t,t+h]} \triangleq$ increment in the interval $(t, t+h]$

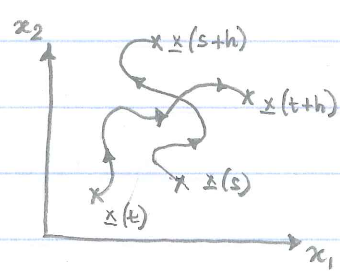
$X_i \triangleq S_i - S_{i-1}$ inter-event interval between arrival times S_{i-1} and S_i

$\mathcal{H}_t \triangleq (s_1, s_2, \ldots, s_n, n)$ realization of the vector of RVs: $[S_1, S_2, \ldots, S_n, N(t)]^T$

Examples of Renewal Processes:

1) Hippocampal Place Cell $\Rightarrow$ It is an example of how a renewal process can be obtained from the definition of $\lambda(t)$. In fact:

$$\lambda(t) \triangleq \exp\left(\alpha - \frac{1}{2}\left(\underline{x}(t) - \underline{\mu}_x\right)^T Q^{-1} \left(\underline{x}(t) - \underline{\mu}_x\right)\right)$$



$Q \triangleq \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ $\quad \underline{x}(t) \triangleq \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$ - position in a 2D space

$\underline{\mu}_x = \begin{bmatrix} \mu_{x1} \\ \mu_{x2} \end{bmatrix}$ and α - parameters

In this case, X_i are i.i.d. $\forall i$ because the next spike only depends on the position $\underline{x}$

2

and the previous spike, i.e.:

$$\forall i,\ P(X_i \le t, X_{i+1} \le s) = P\left(\Delta N_{(s_{i-1}, s_{i-1}+t]}=1,\ \Delta N_{(s_i, s_i+s]}=1\right) =$$

$$= P\left(\Delta N_{(s_{i-1}, s_{i-1}+t]}=1\right) P\left(\Delta N_{(s_i, s_i+s]}=1\right) \Rightarrow X_i, X_{i+1} \text{ are independent while}$$

the identical distribution is a consequence of the definition of $\lambda(t)$

Also, we have:

$$P(X_i > t+h \mid X_i > t) = P\left(\Delta N_{(t,t+h]}=0\right) = \exp\left(-\int_t^{t+h} \lambda(u)\,du\right)$$

$$P(X_i > s+h \mid X_i > s) = P\left(\Delta N_{(s,s+h]}=0\right) = \exp\left(-\int_s^{s+h} \lambda(u)\,du\right)$$

These two are different, provided that the trajectories are different (see figure)

2) Inter-Spike Intervals $\Rightarrow$ It is an example of how a renewal process can be obtained from the definition of $X_i\ \forall i$. In fact:

$$X_i \sim f_X(x) \triangleq \sqrt{\frac{\alpha}{2\pi x^3}}\, e^{-\frac{\alpha(x-\mu_X)^2}{2\mu_X^2 x}}$$

- Inverse Gaussian (α and μ_X are parameters to be estimated)

$$\Downarrow$$

$$P\left(\Delta N_{(t,t+h]}=0\right) \underset{\substack{\uparrow \\ \text{there must} \\ \text{be one "}i\text{"}}}{=} P(X_i > t+h \mid X_i > t) = \frac{P(X_i > t+h)}{P(X_i > t)} = \frac{1 - F_X(t+h)}{1 - F_X(t)}$$

with $F_X(x)$ - cdf of the RV X_i. Because of the definition of the process:

$$P\left(\Delta N_{(t,t+h]}=0\right) = \exp\left(-\int_t^{t+h} \lambda(u)\,du\right) \Rightarrow \int_t^{t+h} \lambda(u)\,du = -\log\frac{1-F_X(t+h)}{1-F_X(t)}$$

$$= \log(1 - F_X(t)) - \log(1 - F_X(t+h))$$

$$\Rightarrow \lambda(t) = \frac{d}{dt}\left(-\log\left(1-F_x(t)\right)\right) = \frac{f_x(t)}{1-F_x(t)} \quad \text{- Hazard Function}$$

Note, however, that the definition of $\lambda(t)$ as hazard function holds as long as the interval $(t, t+h]$ does not include any arrival time $\Rightarrow$ $\lambda(t)$ must be "resetted" after each arrival time

Examples of History-dependent Point Processes:

1) Renewal Process $\Rightarrow$ It is an example of how a Renewal Process can actually be formulated in terms of CIF. In fact, the "resetting" of $\lambda(t)$ can be encompassed in this formulation:

$$\lambda(t) = \frac{f_x(t - s_{1*}(t))}{1 - F_x(t - s_{1*}(t))}$$ where $s_{1*}(t)$ is the last event occurring before t

$\Rightarrow \lambda(t) = \lambda(t \mid \mathcal{H}_t)$ where, in this case, the history is limited to the last event before t

2) Inhomogeneous Markov Intervals (IMIs) $\Rightarrow$ It is a generalization of the example above in case a longer history is considered:

$$\lambda(t \mid \mathcal{H}_t) = \underset{\substack{\uparrow \\ \text{history-independent} \\ \text{(it could be constant)}}}{g_0(t)} \cdot \prod_{i=1}^{k} g_i\left(t - s_{i*}(t)\right)$$

where: $s_{1*}(t) \triangleq$ last event occurring before time t

$s_{2*}(t) \triangleq$ last event occurring before time $s_{1*}(t)$

$s_{i*}(t) \triangleq$ last event occurring before time $s_{i-1*}(t)$

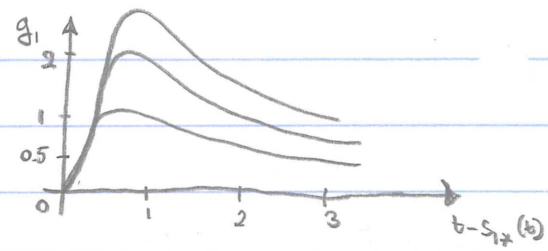
Note that, with this definition, $\lambda(t|\mathcal{H}_t)$ is obtained by combining arbitrary functions $g_i(\cdot)$ $i=0,1,2,\dots,k$ $\Rightarrow$ Provided that these functions are positives and smooth enough, we can write:

$$\log \lambda(t|\mathcal{H}_t) = \log g_0(t) + \sum_{i=1}^{k} \log g_i(t - s_{i*}(t)) \Rightarrow \text{The } \log \lambda \text{ function}$$

can be obtained via spline fitting as suggested for non-parametric regression $\Rightarrow$ This approach is useful when $\lambda(t|\mathcal{H}_t)$ is a smooth but non-monotonic function of time. In the case of the interspike intervals seen above, for instance, we had:

$$g_0(t) = 1$$

$$g_1(t - s_{1*}(t)) = \frac{f_X(t - s_{1*}(t))}{1 - F_X(t - s_{1*}(t))}$$



$\Rightarrow$ The combination of known log-functions can reduce the number of parameters to be estimated on the data. □

From a practical standpoint, we are interested in NP-PPs because we can determine a discrete-time, quantized approximation of these processes, which can be fitted on data by using the ML method:

NP-PP $\longrightarrow$ Sequence of Poisson RVs

$\longleftarrow$ asymptotic convergence

Estimation on Data via ML method

The approximation converges to the actual NP-PP for the sampling step $\Delta t \to 0$

(*) Standard NP-PP problem: $\Delta N_{(0,t]} \sim \mathcal{P}(\mu_t) \quad \forall t \leq T$

$$\mu_t = \int_0^t \lambda(u|\mathcal{H}_u, \vartheta)\, du$$

$$\log \lambda(t|\mathcal{H}_t, \vartheta) = g(t, \mathcal{H}_t, \vartheta)$$

where ϑ is a vector of parameters to be estimated and $g(\cdot)$ is a know class of functions to be used.

Note: In the example of the place cells, $g = g(t, \vartheta)$ nonlinear and history-indep., with $\vartheta = [\alpha\ \mu_{x_1}\ \mu_{x_2}\ \sigma_1^2\ \sigma_2^2]^T$.

In the example of the interspike intervals, $g = \log \dfrac{f_x(t - s_{1*}(t))}{1 - F_x(t - s_{1*}(t))}$ and $\vartheta = [\alpha\ \mu_x]^T$, parameters of the IG RV.

In the example of the IMIs, $g = \log g_0(t) + \sum_{i=1}^{k} \log g_i(t - s_{i*}(t))$ and ϑ is the vector of the parameters in $g_0, g_1, \ldots, g_k$.

(**) Approximated NP-PP problem: $Y_i \sim \mathcal{P}(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i}, \vartheta)\Delta t)$

$$\log \lambda(t|\mathcal{H}_t, \vartheta) = g(t, \mathcal{H}_t, \vartheta)$$

$\Downarrow$

This is a sequence of RVs whose distribution functions belong to the very same class

$\Downarrow$

We can formulate the problem as the standard regression problem we saw before

$t_i - t_{i-1} = \Delta t$ - const $\quad i = 1, 2, 3, \ldots, m$

$\hat{t}_i \triangleq$ mid-point in $]t_{i-1}, t_i[\quad i = 1, 2, \ldots, m$

$Y_i \triangleq \Delta N_{(t_{i-1}, t_i]}$

The correspondence between problem (*) and problem (**) extends to the

⑥

joint probability function over the entire interval $(0,T]$

Dependency on parameter vector ϑ is dropped for simplicity here

$$(*)\quad f(s_1, s_2, s_3, \dots, s_n, n) = \prod_{i=1}^{n} \lambda(s_i|\mathcal{H}_{s_i}) \exp\left(-\int_0^T \lambda(u|\mathcal{H}_u)\,du\right) =$$

$$= \exp\left(\sum_{i=1}^{n} \log \lambda(s_i|\mathcal{H}_{s_i}) - \int_0^T \lambda(u|\mathcal{H}_u)\,du\right)$$

$$(**)\quad f(y_1, y_2, y_3, \dots, y_m) = \prod_{i=1}^{m} \left(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right)^{y_i} \left(1-\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right)^{1-y_i}$$

$$= \exp\left(\sum_{i=1}^{m} y_i \log\left(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right) + (1-y_i)\log\left(1 + - \lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right)\right)$$

$\log(1+t) \cong t$ for small values t

$$\cong \exp\left(\sum_{i=1}^{m} y_i \log\left(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right) - (1-y_i)\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i})\Delta t\right)$$

where we showed that $f(y_1, y_2, y_3, \dots, y_m)\dfrac{1}{\Delta t^n} \xrightarrow{\Delta t \to 0} f(s_1, s_2, \dots, s_n, n)$ $\square$

* GLM-based NP-PPs and solution via ML-method

1) We consider the log-likelihood function:

$$\ell(\vartheta) = \log f(y_1, y_2, \dots, y_m) =$$

$$= \sum_{i=1}^{m} \left\{ y_i \log\left(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i}, \vartheta)\Delta t\right) - (1-y_i)\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i}, \vartheta)\Delta t \right\}$$

when the number n of events is $n \ll m$

$$\cong \sum_{i=1}^{m} \left\{ y_i \log\left(\lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i}, \vartheta)\Delta t\right) - \lambda(\hat{t}_i|\mathcal{H}_{\hat{t}_i}, \vartheta)\Delta t \right\}$$

2) We assume that the link function is:

$$\log\left(\lambda(\hat{t}_i \mid \mathcal{H}_{\hat{t}_i}, \theta)\Delta t\right) = \alpha + \sum_{j=1}^{i-1} \beta_j y_{i-j} = \underbrace{\begin{bmatrix} 1 & Y_{i-1} & Y_{i-2} & \dots & Y_1 \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{i-1} \end{bmatrix}}_{\theta} = Y\theta$$

Alternatively, one can assume that only a finite window of history affects the current value of the CIF:

$$\log\left(\lambda(\hat{t}_i \mid \mathcal{H}_{\hat{t}_i}, \theta)\Delta t\right) = \alpha + \sum_{j=1}^{k} \beta_j y_{i-j} \quad \text{with } k < i \text{ and } k\text{-fixed } \forall i$$

Similarly, because $Y_i \triangleq \Delta N_{(t_{i-1}, t_i]}$, one can replace Y_i's with increments over longer intervals (i.e., one assumes that the integral of the past history affects the current value of the link function):

$$\log\left(\lambda(\hat{t}_i \mid \mathcal{H}_{\hat{t}_i}, \theta)\Delta t\right) = \alpha + \sum_{j=1}^{k} \beta_j \Delta N_{(t_i - jq\Delta t,\, t_i - (j-1)q\Delta t - 1]}$$

with $q \triangleq$ size of the interval in number of bins (e.g., $q = 10$)

$k \triangleq$ number of consecutive intervals, each including q bins

Hence, denoted with $\hat{X}_i$ the vector of history values, i.e., $\hat{X}_i = [1\ Y_{i-1}\ Y_{i-2} \dots Y_1]$ in the first case, $\hat{X}_i = [1\ Y_{i-1}\ Y_{i-2} \dots Y_{i-k}]$ in the second case, and finally $\hat{X}_i = [1\ \Delta N_{(t_i - q\Delta t, t_i - 1]}\ \Delta N_{(t_i - 2q\Delta t, t_i - q\Delta t - 1]} \dots]$, we have:

$$\log\left(\lambda(\hat{t}_i \mid \mathcal{H}_{\hat{t}_i}, \theta)\Delta t\right) = \hat{X}_i\theta \Rightarrow \text{GLM for the Poisson RVs}$$

3) Finally, we solve the regression problem:

$$\left.\begin{array}{l} ?\hat{\theta}: \quad Y_i \sim \mathcal{P}(\hat{\eta}_i) \\ \log \hat{\eta}_i = \hat{X}_i \vartheta \\ i = 1, 2, \dots, m \\ \hat{\eta}_i \triangleq \lambda(\hat{t}_i | \mathcal{H}_{t_i}) \Delta t \end{array}\right\} \Rightarrow$$

MLE

The solution can be numerically determined by searching for $\hat{\theta}$ that maximizes $\ell(\vartheta)$ constrained to the conditions:

$$\underbrace{[Y_1 \; Y_2 \; \dots \; Y_m]^T}_{Y} = [y_1 \; y_2 \; \dots \; y_m]^T$$

$$[\log \hat{\eta}_1 \; \log \hat{\eta}_2 \dots \log \hat{\eta}_m]^T = \underbrace{\begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_m \end{bmatrix}}_{X} \vartheta$$

Note that the paradigm (1-3) can be extended to a more general case where $\log \hat{\eta}_i$ depends on other explanatory random variables, e.g.:

$$\log \hat{\eta}_i = \alpha + \sum_{j=1}^{k} \beta_j y_{i-j} + \sum_{r=1}^{h} \gamma_r u_{i-r} \qquad (***)$$

where u_i is the realization of the RV U_i in the i-th bin $(i=1,2,3,\dots,m)$, with U_i describing an exogenous input (e.g., position in space, activity of another system, stimulus, etc.) $\Rightarrow$ We can generalize the CIF definition as reported here:

$\lambda(t | \mathcal{H}_t)$ $\rightsquigarrow$ $\lambda(t | X_t)$

$\mathcal{H}_t \triangleq$ history of the NP-PP up to time t

$X_t \triangleq (\mathcal{H}_t, \mathcal{H}_{u_1 t}, \mathcal{H}_{u_2 t}, \dots)$

$\mathcal{H}_{u_1 t} \triangleq$ history of the exogenous process U_1 up to time t

$U_1, U_2, \dots$ are called "covariates"

$\mathcal{H}_{u_2 t} \triangleq$ history of the exogenous process U_2 up to time t, etc.

For instance, the case of the place cells can be formulated as a regression problem:

$$\lambda(t) = \exp\left(\alpha - \frac{1}{2}\left(\underline{x}(t) - \underline{\mu}_x\right)^T Q^{-1}\left(\underline{x}(t) - \underline{\mu}_x\right)\right)$$

$$\Updownarrow$$

$$\log \lambda(t) = \alpha - \frac{1}{2}\sigma_1^{-2}\left(x_1(t) - \mu_{x_1}\right)^2 - \frac{1}{2}\sigma_2^{-2}\left(x_2(t) - \mu_{x_2}\right)^2$$

Assuming $\mu_{x_1} = \mu_{x_2} = 0$, the function can be formulated as:

$$\log \lambda(t | x_t) = \overbrace{\left[\, 1 \quad \frac{1}{2}x_1^2(t) \quad \frac{1}{2}x_2^2(t) \,\right]}^{\hat{X}} \overbrace{\begin{bmatrix} \alpha \\ \sigma_1^{-2} \\ \sigma_2^{-2} \end{bmatrix}}^{\vartheta}$$: The form is as in (***)

In case $\mu_{x_1} \neq 0$ and/or $\mu_{x_2} \neq 0$, instead, the link function is not log-linear anymore

$\Rightarrow$ We can use a gradient-based maximization procedure to determine the MLE, i.e., we obtain the estimation of the parameter vector ϑ iteratively, via the formula:

$$\text{Estimation at the k-th iteration} \rightarrow \hat{\vartheta}_k = \hat{\vartheta}_{k-1} + \underbrace{\delta}_{\text{increment}} \underbrace{\nabla \ell(\hat{\vartheta}_{k-1})}_{\text{gradient of } \ell(\vartheta)}$$

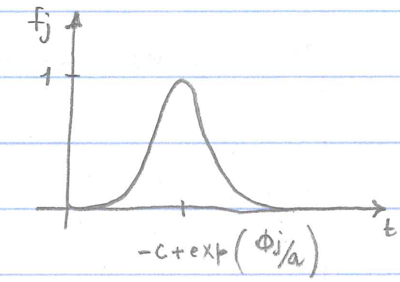
This approach can be used also when the parameters β_j's and γ_r's in (***) are replaced by basis functions:

$$\log \hat{\eta}_i = \alpha(t_i) + \sum_{j=1}^{k} \left[f_j(t_i) \; f_j(t_{i-1}) \ldots f_j(t_{i-q}) \right] \begin{bmatrix} y_i \\ y_{i-1} \\ \vdots \\ y_{i-q} \end{bmatrix} + \sum_{r=1}^{h} \left[w_r(t_i) \; w_r(t_{i-1}) \ldots w_r(t_{i-p}) \right] \begin{bmatrix} u_i \\ u_{i-1} \\ \vdots \\ u_{i-p} \end{bmatrix} \qquad (\Delta\nabla)$$

where the lengths q and p depend on the memory of the basis functions $f_j(t)$ and $w_r(t)$, respectively $\Rightarrow$ We are replacing linear, moving-average filters in (***) with more generic filters $f_j(t)$ and $w_r(t)$ $\Rightarrow$ This solution may help when $\log \lambda(t)$ depends on a very long window of past history and we want to limit the number of parameters to be estimated.

Typical choices for $f_j(t)$ and $w_r(t)$ are:

$$f_j(t) = \frac{1}{2} + \frac{1}{2}\cos\left(a \cdot \log(t+c) - \phi_j\right)$$



$$f_j(t) = (t-\phi_j)^2 \log(t-\phi_j)$$

$$f_j(t) = \frac{1}{1+a(t-\phi_j)}$$

Parameters a and c are usually part of θ, while ϕ_j are fixed

Eventually, the combination of filters in ($\Delta\nabla$) can be replaced by a weighted combination $\Rightarrow$ This was the case in the example with IMIs

* Goodness-of-fit and Residual Analysis

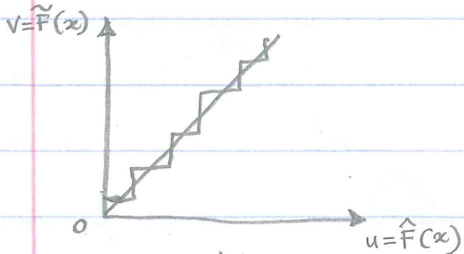
Let us assume that a solution to the problem (**) is determined and we want to assess the goodness of the fit on the data $\Rightarrow$ One approach involves drawing the P-P plot:

$\boxed{X} \longrightarrow x_1, x_2, x_3 \ldots$

Random Variable

$X \sim \tilde{F}_X(x)$ - theoretical CDF

$X_i \sim \hat{F}_i(x)$ $\forall i$ - empirical CDF



P-P plot

The P-P plot reports the discrepancies between the empirical and theoretical CDF

$\Downarrow$

Provided that X_i are i.i.d. $\forall i$ and that the number n of samples is large, we have:

$$\lim_{n\to\infty} \hat{F}_n(x) = \tilde{F}_X(x) \quad \forall x$$

In our case, we can depict the P-P plot for the inter-event waiting times X_i (if i.i.d.) and have $\tilde{F}_X(x)$ - CDF determined by the point process. If X_i are NOT i.i.d., instead, one can use the following result:

Time-Rescaling Theorem

$\{S_k\}_k$ - NP-PP with CIF: $\lambda(t|\mathcal{H}_t)$ over $(0,T]$

$f_{X_i}(x|S_{i-1}) > 0$ and continuous on $(s_{i-1}, T]$ $\forall i \geq 1$

pdf of the inter-event intervals

$$z_1 \triangleq \int_0^{s_1} \lambda(u|\mathcal{H}_u)\,du \quad \text{and} \quad z_j \triangleq \int_{s_{j-1}}^{s_j} \lambda(u|\mathcal{H}_u)\,du \quad j = 2,3,\dots,n$$

It results that $z_j \sim \text{Exp}(1)$ $j = 1,2,3,\dots,n$ and i.i.d.

II

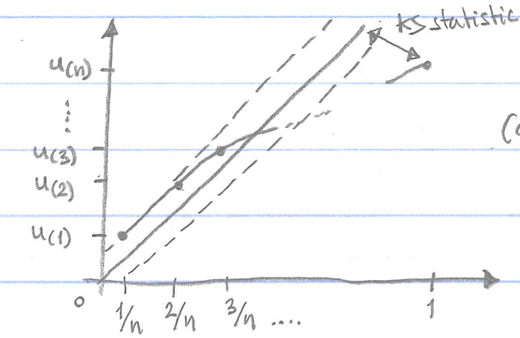
Because of this theorem, the value $u_j = 1 - \exp\left(-\int_{s_{j-1}}^{s_j} \lambda(u|\mathcal{H}_u)\,du\right)$ is the CDF of Exp(1) evaluated in z_j. Because z_j are i.i.d., the values u_j $j = 1,2,\dots,n$ should be distributed homogeneously in $[0,1]$, i.e., $U_j \sim \text{Uniform}(0,1)$ $j = 1,2,\dots,n$

$\Rightarrow$ The P-P plot is the plot of $u_1, u_2, \dots, u_n$ vs. the CDF of a uniform distribution

A measure of the closeness between the empirical cdf $\hat{F}_n(x)$ and the theoretical cdf $\tilde{F}_X(x)$ is the Kolmogorov-Smirnov (KS) statistic:

$$KS \triangleq \sup_{x} \left| \hat{F}_n(x) - \tilde{F}_X(x) \right|$$

Based on the value of this statistic, one can assess the goodness-of-fit of the point process:



P-P plot
(a.k.a. "KS plot")

KS-Test: The hypothesis $\hat{F}_X(x) = \tilde{F}_X(x)$, where $F_X(x) \triangleq \lim_{n\to\infty} \hat{F}_n(x)$, is rejected with 95% confidence if n is large and the KS statistic is

$$KS > 1.36/\sqrt{n}$$

Hence, a good fitting (i.e., p-value $p < 0.05$) is obtained when $KS < 1.36/\sqrt{n}$

Note: Passing the KS-Test only means that the rescaled times z_j have identical distribution but it does not say anything about independence $\Rightarrow$ A way to show independence is by calculation of the auto-correlation function (ACF) for the rescaled values u_j $j = 1, 2, \dots, n$

$$\Downarrow$$

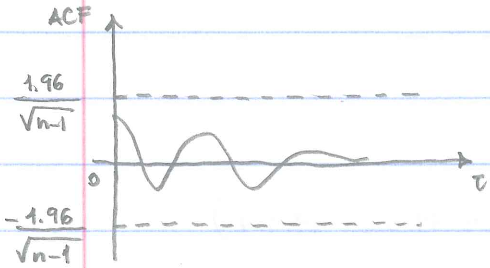
In order to create a test for independence, though, it may be useful to transform the RVs U_j $j = 1, 2, \dots, n$ into RVs for which confidence intervals on the ACF are known:

$\Phi(x)$ - CDF of a Gaussian distribution with $\mu = 0$ and $\sigma = 1$ $\Rightarrow$ By invoking the theorem "From U to Y" shown in lecture 1, we have that the RVs

$$W_j = \Phi^{-1}(U_j) \sim N(0,1) \quad \forall j$$

Hence, denoted with $w_j = \phi^{-1}(u_j)$ the realization of W_j, we have:

$$ACF(\tau) = \frac{1}{n-\tau}\sum_{i=1}^{n-\tau} w_i w_{i+\tau}$$



Correlation Test: The hypothesis "W_j are independent $\forall j$" is rejected with 95% confidence (i.e., p-value $p < 0.05$) if

$$\max_{\tau} |ACF(\tau)| > \frac{1.96}{\sqrt{n-1}}$$

Hence, a good approximation of the independence can be considered when $ACF(\tau)$ is within the confidence bounds $\pm 1.96/\sqrt{n-1}$ $\forall \tau$ □

Finally, note this: A solution to the problem (**) may not completely capture the relationship between covariates and observations Y_i $i = 1, 2, \dots, m$ $\Rightarrow$ A way to test this is to check if residuals and covariates are independent $\Rightarrow$ How do we define residuals?

Because $\Delta N_{(t, t+h]} \sim \mathcal{P}(\mu)$ (definition of NP_PP) and $E(\Delta N_{(t,t+h]}) = \mu$, we can define the residual:

$$r(i, h) \triangleq \underbrace{\sum_{j=i}^{i+h} y_j}_{\substack{\text{It represents} \\ \Delta N(t_i,\, t_i + h\Delta t]}} - \underbrace{\sum_{j=i}^{i+h} \lambda(\hat{t}_j \mid \mathcal{H}_{\hat{t}_j})\Delta t}_{\substack{\text{It approximates} \\ \mu = \int_{t_i}^{t_i + h\Delta t} \lambda(u \mid \mathcal{H}_u)\, du}}$$

Hence, we can consider a set of non-overlapping windows, each of size $h\Delta t$, and compute the residual in each one of them: $r(1, h)$, $r(h+1, h)$, $r(2h+1, h)$, etc.

Then, one can look at the cross-correlation between the sequence of residuals and the covariates of interest ▫

References:

Textbook: ch 19 (sections 19.3.4, 19.3.5, 19.3.6, 19.3.7)
ch 10 (section 10.3.7)

Truccolo et al. (2005), J. Neurophysiol., vol. 93, pp. 1074-89 ⇒ A copy is on HuskyCT

For examples of GLM-based and non GLM-based NP-PPs fitted on data, consider:
- Cajigas et al. (2012), J. Neurosci. Methods, vol. 211, pp. 245-64
- Truccolo et al. (2010), Nat. Neurosci., vol. 13, pp. 105-11
- Pillow et al. (2008), Nature, vol. 454, pp. 995-99
- Lepage & MacDonald (2015), J. Comput. Neurosci., vol. 38, pp. 499-519