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 \(\mu = \int_t^{t+h} \lambda(u) \, du \) \(\forall t, h \geq 0\)
  
  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 \(\mu \triangleq \int_t^{t+h} \lambda(u) \mathcal{H}_u \, du \) \(\forall t, h \geq 0\)
  \[
  \lambda(t \mid \mathcal{H}_t) \triangleq \lim_{\Delta t \to 0} \frac{P(\Delta N(t, t+\Delta t] > 0 \mid \mathcal{H}_t)}{\Delta t}
  \]

where: \(\Delta N(t, t+h] \triangleq \text{increment in the interval } (t, t+h]\)

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

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

**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:

\[
\lambda(t) \triangleq \exp \left( \alpha - \frac{1}{2} (x(t) - \mu_x)^T Q^{-1} (x(t) - \mu_x) \right)
\]

\[
Q = \begin{pmatrix}
\sigma_{x1}^2 & 0 \\
0 & \sigma_{x2}^2 \\
\end{pmatrix}
\]

\(x(t) \triangleq \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}\) - position in a 2D space

\(\mu_x = \begin{bmatrix} \mu_{x1} \\ \mu_{x2} \end{bmatrix}\) and \(\alpha\) - parameters

In this case, \(X_i\) are i.i.d. \(\forall i\) because the next spike only depends on the position \(x\)
and the previous spike, i.e.:

\[ P(X_i \leq t, X_{i+1} \leq s) = P\left(\Delta N(\bar{s}_{i-1}, \bar{s}_{i-1}+t] = 1, \Delta N(\bar{s}_i, \bar{s}_{i+1}] = 1\right) = P(\Delta N(\bar{s}_{i-1}, \bar{s}_{i-1}+t]) = 1) P(\Delta N(\bar{s}_i, \bar{s}_{i+1}] = 1) \Rightarrow X_i, X_{i+1} \text{ are independent while}
\]

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

Also, we have:

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

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 \( X_i, Y_i \). In fact:

\[
X_i \sim f_X(x) = \sqrt{\frac{\alpha}{2\pi x^3}} e^{-\frac{(x-\mu)^2}{2\mu x^2}} \quad \text{Inverse Gaussian (} \alpha \text{ and } \mu \text{ are parameters to be estimated)}
\]

\[
P(\Delta N(t, t+h] = 0) = P\left(\frac{X_i > t+h}{X_i > t} = \frac{P(X_i > t+h)}{P(X_i > t)} = \frac{1 - F_X(t+h)}{1 - F_X(t)} \right)
\]

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

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

\[
= \log \left( 1 - F_X(t) \right) - \log \left( 1 - F_X(t+h) \right)
\]
\[ \lambda(t) = \frac{d}{dt} \left( -\log \left( 1 - F_X(t) \right) \right) = \frac{f_X(t)}{1 - F_X(t)} \quad \text{- Hazard Function} \]

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:

\[ \lambda(t) = \frac{f_X(t - s_{in}(t))}{1 - F_X(t - s_{in}(t))} \quad \text{where } s_{in}(t) \text{ is the last event occurring before } t \]

\[ \Rightarrow \lambda(t) = \lambda(t | \mathcal{H}_t) \quad \text{where, in this case, the history is limited to the last event before } t \]

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

\[ \lambda(t | \mathcal{H}_t) = g_0(t) \cdot \prod_{i=1}^{k} g_i(t - s_{in}(t)) \]

\[ \uparrow \text{history-independent} \]

\( (i \text{t could be constant}) \)

where:

\[ s_{1\#}(t) \triangleq \text{last event occurring before time } t \]

\[ s_{2\#}(t) \triangleq \text{last event occurring before time } s_{1\#}(t) \]

\[ s_{i\#}(t) \triangleq \text{last event occurring before time } s_{i-1\#}(t) \]
Note that, with this definition, \( \lambda(t \mid X(t)) \) is obtained by combining arbitrary functions \( g_i(t) \) \( i = 0, 1, 2, \ldots, k \) \( \Rightarrow \) Provided that these functions are positive and smooth enough, we can write:

\[
\log \lambda(t \mid X(t)) = \log g_0(t) + \sum_{i=1}^{k} \log g_i(t - s_i(t)) \Rightarrow \text{The log } \lambda \text{ function can be obtained via spline fitting as suggested for non-parametric regression.}
\]

\Rightarrow \text{This approach is useful when } \lambda(t \mid X(t)) \text{ is a smooth but non-monotonic function of time. In the case of the interspike intervals seen above, for instance, we had:}

\[
\begin{align*}
g_0(t) & = 1 \\
g_1(t - s_{11}(t)) & = \frac{f_X(t - s_{11}(t))}{1 - F_X(t - s_{11}(t))}
\end{align*}
\]

\( \Rightarrow \text{The combination of known log } \lambda \text{ functions can reduce the number of parameters to be estimated on the data.} \)

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:

\[
\begin{array}{ccc}
\text{NP-PP} & \rightarrow & \text{Sequence of Poisson RVs} \\
\downarrow & \text{asymptotic convergence} & \circ \text{ via ML method} \\
\text{Estimation on Data} & \text{via ML method} \\
\end{array}
\]

The approximation converges to the actual NP-PP for the sampling step \( \Delta t \rightarrow 0 \)
(⋆) Standard NP-PP problem: 
\[ \Delta N(0_t) \sim \mathcal{P}(\lambda_t) \quad \forall t \leq T \]
\[ \lambda_t = \int_0^t \lambda(u| H_t, \theta) \, du \]

\[ \log \lambda(t|H_t, \theta) = g(t, H_t, \theta) \]

where \( \theta \) is a vector of parameters to be estimated, and \( g(\cdot) \) is a known class of functions to be used.

Note: In the example of the place cells, \( g = g(t, \theta) \) nonlinear and history-independent, with \( \theta = [\alpha, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2]^T \).

In the example of the interspike intervals, \( g = \log \frac{f_X(t - s_{\text{int}}(t))}{1 - F_X(t - s_{\text{int}}(t))} \) and \( \theta = [\alpha, \mu_X]^T \), parameters of the IG RV.

In the example of the LMs, \( g = \log g_0(t) + \sum_{i=1}^{k_0} \log g_i(t - s_i(t)) \) and \( \theta \) is the vector of the parameters in \( g_0, g_1, \ldots, g_k \).

(⋆⋆) Approximated NP-PP problem: 
\[ Y_i \sim \mathcal{P}\left( \lambda(t_i|H_{t_i}, \theta) \Delta t \right) \]
\[ \log \lambda(t|H_t, \theta) = g(t, H_t, \theta) \]

\[ t_0 = 0 \quad t_1 \quad t_2 \ldots \quad t_m = T \]

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

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

The correspondence between problem (⋆) and problem (⋆⋆) extends to the
Joint probability function over the entire interval \((0, T]\):

\[
(*) \quad f(s_1, s_2, s_3, \ldots, s_n, n) = \prod_{i=1}^{n} \lambda(s_i | H_{\bar{s_i}}) \exp \left( \int_{0}^{T} \lambda(u | H_u) \, du \right) = \\
= \exp \left( \sum_{i=1}^{n} \log \lambda(s_i | H_{\bar{s}_i}) - \int_{0}^{T} \lambda(u | H_u) \, du \right)
\]

\[
(**) \quad f(y_1, y_2, y_3, \ldots, y_m) = \prod_{i=1}^{m} \left( \frac{\lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t}{1 - \lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t} \right)^{y_i} \left(1 - \lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t \right)^{1-y_i}
= \exp \left( \sum_{i=1}^{m} y_i \log \left( \frac{\lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t}{1 - \lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t} \right) + (1-y_i) \log \left(1 + \frac{\log(1+t) \Delta t}{\log(1+t) \Delta t} \right) \right)
\]

For small values \(\Delta t\) we approximate

\[
\exp \left( \sum_{i=1}^{m} y_i \log \left( \frac{\lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t}{1 - \lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t} \right) - (1-y_i) \lambda(\hat{t}_i | H_{\hat{t}_i}) \Delta t \right) \to \quad \text{as } \Delta t \to 0
\]

where we showed that \(f(y_1, y_2, y_3, \ldots, y_m) \sim \frac{1}{\Delta t^n} \to f(s_1, s_2, \ldots, s_n, n) \)

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

1) We consider the log-likelihood function:

\[
l(\theta) = \log f(y_1, y_2, \ldots, y_m) = \\
= \sum_{i=1}^{m} \left\{ y_i \log \left( \frac{\lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t}{1 - \lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t} \right) - (1-y_i) \lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t \right\}
\]

when the number \(n\) of events is

\[
n < m \quad \sum_{i=1}^{m} \left\{ y_i \log \left( \frac{\lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t}{1 - \lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t} \right) - \lambda(\hat{t}_i | H_{\hat{t}_i}, \theta) \Delta t \right\}
\]
2) We assume that the link function is:

\[
\log \left( \lambda(\hat{e}_i | \mathcal{H}_{e_i}, \theta) \Delta t \right) = \alpha + \sum_{j=1}^{i-1} \beta_j y_{i-j} = \left[ \begin{array}{c}
\alpha \\
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_{i-1}
\end{array} \right] \begin{array}{c}
y_{i-1} \\
y_i \\
y_{i-2} \\
\vdots \\
y_1
\end{array} = \gamma
\]

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

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

Similarly, because \( Y_i \neq \Delta N_{(t_i-1, t_i]} \), 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):

\[
\log \left( \lambda(\hat{e}_i | \mathcal{H}_{e_i}, \theta) \Delta t \right) = \alpha + \sum_{j=1}^{k} \beta_j \Delta N_{(t_i-j \Delta t, t_i-j \Delta t-1]} 
\]

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

\( k \) = number of consecutive intervals, each including \( \Delta \) bins

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

\[
\log \left( \lambda(\hat{e}_i | \mathcal{H}_{e_i}, \theta) \Delta t \right) = \hat{X}_i \theta \quad \Rightarrow \text{GLM for the Poisson RVs}
\]

3) Finally, we solve the regression problem:
\( ? \Theta : Y_i \sim p(\hat{\eta}_i) \) for \( i = 1, 2, \ldots, m \)

\[ \hat{\eta}_i = \lambda (\hat{t}_i | H_{t_i}) \Delta t \]

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

\[ [Y_1, Y_2, \ldots, Y_m]^T = [Y_1, Y_2, \ldots, Y_m]^T \]

\[ [\log \hat{\eta}_1, \log \hat{\eta}_2, \ldots, \log \hat{\eta}_m]^T = \left[ \begin{array}{c} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \\ \hat{X}_m \end{array} \right] \Theta \]

Note that the paradigm (1-2) can be extended to a more general case where \( \log \hat{\eta}_i \) depends on other explanatory random variables, e.g.:

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

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

\[ \lambda (t | H_t) \quad \lambda (t | X_t) \]

\( H_t \) is the history of the NP-PP up to time \( t \)

\( X_t \) is \( (H_t, H_{u_1}, H_{u_2}, \ldots, \) etc. \)

\( U_1, U_2, \ldots \) are called "covariates". \( U_2, U_2, \) etc. up to time \( t \), etc.
For instance, the case of the place cells can be formulated as a regression problem:

$$
\lambda(t) = \exp \left( \alpha - \frac{1}{2} \left( x(t) - \mu_x \right)^T \Theta^{-1} \left( x(t) - \mu_x \right) \right)
$$

$$\uparrow$$

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

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

$$
\log \lambda(t|x_e) = \left[ \begin{array}{c} 1 \\ \frac{1}{2} x_1^2(t) \\ \frac{1}{2} x_2^2(t) \end{array} \right] \left[ \begin{array}{c} \alpha_1 \\ \sigma_1^{-2} \\ \alpha_2 \\ \sigma_2^{-2} \end{array} \right] \Theta = \text{The form is as in (***)}
$$

In case $\mu_{x_1} \neq 0$ and/or $\mu_{x_2} \neq 0$, instead, the link function is not log-linear anymore. We can use a gradient-based maximization procedure to determine the MLE, i.e., we obtain the estimation of the parameter vector $\Theta$ iteratively, via the formula:

$$
\hat{\Theta}_k = \hat{\Theta}_{k-1} + \delta \nabla L(\hat{\Theta}_{k-1})
$$

Estimation at the k-th iteration

This approach can be used also when the parameters $\beta_i$'s and $\gamma_i$'s in (***), are replaced by basis functions:

$$
\log \hat{\eta}_i = \alpha(t_i) + \sum_{j=1}^{k} \begin{bmatrix} f_j(t_i) & f_j(t_{i-1}) & \ldots & f_j(t_{i-q}) \end{bmatrix} \begin{bmatrix} y_i \\ y_{i-1} \\ \vdots \\ y_{i-q} \end{bmatrix}
$$

$$
+ \sum_{r=1}^{h} \begin{bmatrix} \omega_r(t_i) & \omega_r(t_{i-1}) & \ldots & \omega_r(t_{i-p}) \end{bmatrix} \begin{bmatrix} u_i \\ u_{i-1} \\ \vdots \\ u_{i-p} \end{bmatrix}
$$
where the lengths $q$ and $p$ depend on the memory of the basis functions $f_j(t)$ and $\omega_r(t)$, respectively. We are replacing linear, moving-average filters in (77) with more generic filters $f_j(t)$ and $\omega_r(t)$. 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 $\omega_r(t)$ are:

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

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

$$f_j(t) = \frac{1}{1 + a(t - \Phi_j)} \quad \text{Parameters $a$ and $c$ are usually part of } \Theta, \text{ while } \Phi_j \text{ are fixed}$$

Eventually, the combination of filters in (77) can be replaced by a weighted combination. This was the case in the example with INIs.

* Goodness-of-fit and Residual Analysis

Let us assume that a solution to the problem (77) is determined and we want to assess the goodness of the fit on the data. One approach involves drawing the P-P plot:

$$X \sim \hat{F}_i(x) \quad \text{-theoretical CDF}$$

Random Variable

$$X_i \sim \hat{F}_i(x) \quad \forall i \quad \text{-empirical CDF}$$
The P-P plot reports the discrepancies between the empirical and theoretical CDF

\[ \lim_{n \to \infty} \hat{F}_n(x) = \hat{F}_x(x) \quad \forall x \]

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

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

**Time-Rescaling: \( S_k \rightarrow k - NP-PF \)**  
**Theorem:** \( f_{X_i}(x|S_{i-1}) > 0 \) and continuous on \((s_{i-1}, T] \quad \forall i \geq 1 \)

\[ Z_j = \int_0^{S_j} \lambda(u|H_u) \, du \quad \text{and} \quad Z_j = \int_{S_{j-1}}^{S_j} \lambda(u|H_u) \, du \quad j = 2, 3, \ldots, n \]

It results that \( Z_j \sim \text{Exp}(1) \quad j = 1, 2, 3, \ldots, n \) and i.i.d.

Because of this theorem, the value \( u_j = 1 - \exp\left(-\int_{S_{j-1}}^{S_j} \lambda(u|H_u) \, du\right) \) is the CDF of \( \text{Exp}(1) \) evaluated in \( S_j \). Because \( Z_j \) are i.i.d., the values \( u_j \), \( j = 1, 2, \ldots, n \), should be distributed homogeneously in \([0, 1]\), i.e., \( U_j \sim \text{Uniform}(0, 1) \quad j = 1, 2, \ldots, n \)

\( \Rightarrow \) The P-P plot is the plot of \( u_1, u_2, \ldots, u_n \) vs. the CDF of a uniform distribution

A measure of the closeness between the empirical cdf \( \hat{F}_n(x) \) and the theoretical cdf \( \hat{F}_x(x) \) is the Kolmogorov-Smirnov (KS) statistic:
\[ ks = \sup_x \left| \hat{F}_n(x) - \tilde{F}_n(x) \right| \]

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

**KS-Test:** The hypothesis \( \hat{F}_n(x) = \tilde{F}_n(x) \), where \( \hat{F}_n(x) \) is the empirical distribution function of \( n \) independent identically distributed observations, 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 \( \tilde{z}_j \) 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 \( u_j \), \( j = 1, 2, \ldots, n \)

\[ u_j = 1/\sqrt{n} \]

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

\[ \Phi(x) - CDF \text{ of a Gaussian } \Rightarrow \text{By invoking the theorem "From } U \text{ to } Y \text{" shown in lecture 4, we have that the RVs} \]

\[ W_j = \Phi^{-1}(u_j) \sim N(0, 1) \quad \forall j \]
Hence, denoted with \( w_j = \Phi_i^{-1}(u_j) \) the realization of \( W_j \), we have:

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

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

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

Hence, a good approximation of the independence can be considered when \( \text{ACF}(\tau) \) 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 \( \gamma_i \) \( i = 1, 2, \ldots, m \) \( \Rightarrow \) a way to test this is to check if residuals and covariates are independent \( \Rightarrow \) how do we define residuals?

Because \( \Delta N(\ell_i, b + h) \sim \mathcal{P}(\mu) \) (definition of NP-PP) and \( \mathbb{E}(\Delta N(\ell_i, b + h)) = \mu \), we can define the residual:

\[
r(i, h) = \sum_{j=i}^{i+h} y_j - \sum_{j=i}^{i+h} \lambda(\ell_j | \gamma_{\ell_j}) \Delta t
\]

It represents \( \Delta N(\ell_i, b + h \Delta t) \) \( \Rightarrow \)

\[
\mu = \int_{\ell_i}^{\ell_i + h \Delta t} \lambda(u | \gamma_{\ell}) du
\]

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: \( r(1, h), r(h+1, h), r(2h+1, h), \ldots \).
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:

- Truccolo et al. (2010), Nat. Neurosci., vol. 13, pp. 105-11