LECTURE 5

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

Random Variable

$X \sim f_X(x)$

$X_i \sim f_{X_i}(x) \quad i = 1, 2, 3, \dots$

In the problems considered thus far we assume that samples are obtained from a RV (e.g., estimation problem) or a class of RVs (e.g., regression problem) AND that each sample is the outcome of one repetition of the given experiment (i.e., a trial) $\Rightarrow$ The notion of "time" is not involved

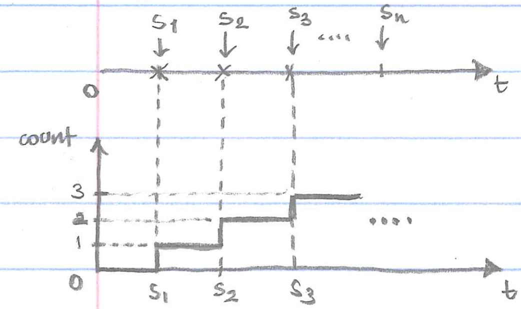
Ex.: $X \sim \mathcal{P}(\lambda) \Rightarrow$ The event $\{X = k\}$ means that the count of a certain event over an observation horizon (which is assigned by the experiment) is k in a certain repetition of the experiment

Let us now consider the case of a set $(s_1, s_2, s_3, \dots, s_n)$ of samples that are collected SEQUENTIALLY during the SAME experiment. In particular, let us consider the case where:

$s_i \triangleq$ arrival time of the i-th event (e.g., spike, heartbeat, lightbeam, etc.) $\Rightarrow$ It can be envisioned as a realization of the RV S_i, $i = 1, 2, 3, \dots, n$

$n \triangleq$ number of events occurring in the observed time interval $(0, t] \Rightarrow$ It can envisioned as a realization of the RV N_t

In this scenario, for any new time t, we will have a new RV N_t and, in general, we can expect that the knowledge of $N_{t_1}, N_{t_2}, N_{t_3}, \dots, N_{t_K}$ with $t_1 < t_2 < \dots < t_K$ can allow the prediction (up to some uncertainty) of $N_{t_{K+1}}$



In this example, if we pick $t_1 < s_1$, $s_1 < t_2 < s_2$, and $s_2 < t_3 < s_3$ we know that $N_{t_1} = 0$, $N_{t_2} = 1$, and $N_{t_3} = 2$

$\Downarrow$

Hence we know that $\forall\, t_k > t_3,\ N_{t_k} \geq 2$

The properties of the sequence $N_{t_1}, N_{t_2}, N_{t_3}, \ldots$ can be explained by noticing that:

$N_t : \xi \longrightarrow f(t, \xi) \Rightarrow$ N_t is called a "Stochastic Process"

outcome of an experiment (i.e., count of events $s_1, s_2, s_3, \ldots$)

function of time t

In particular: $f(t,\xi)$ defined for $t \in \mathcal{R} \Rightarrow$ N_t is a continuous-time process

$f(t,\xi)$ defined for $t \in \mathcal{Z} \Rightarrow$ N_t is a discrete-time process

$f(t,\xi)$ countable for any given $t \Rightarrow$ N_t is a discrete-state process

In our case, N_t is a continuous-time discrete-state process and is called "Counting process"

Note this: From the sequence of arrival times one can define:

$$\left.\begin{aligned} X_1 &\triangleq S_1 \\ X_2 &\triangleq S_2 - S_1 \\ &\vdots \\ X_i &\triangleq S_i - S_{i-1} \\ &\vdots \end{aligned}\right\}$$ Inter-event waiting times $\Rightarrow$ We can define: $S_k \triangleq \sum_{i=1}^{k} X_i$ - stochastic process

discrete time

continuous state

$\{S_k\}_k$ is called "point process"

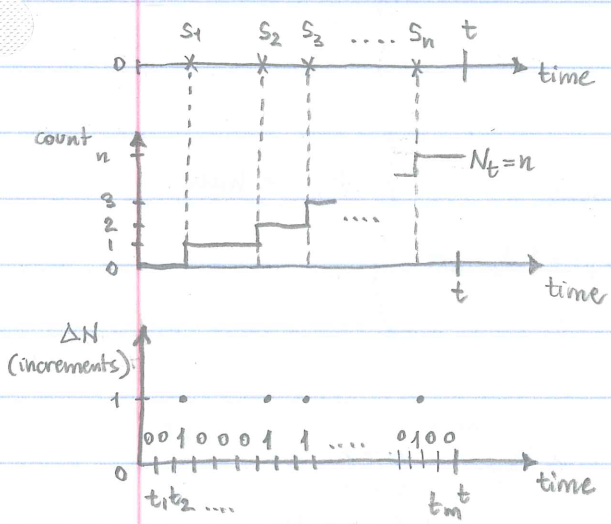
Note this: Processes $\{N_t\}_t$ and $\{S_k\}_k$ are equivalent representations of the same observation. In fact, we have that the same event can be equivalently stated in term of N_t or S_k, e.g.:

$\{N_t < j\} = \{S_j > t\}$ and

$\{t : N_t < j\} = \{t : S_j > t\}$

$\uparrow$ domain of N_t $\qquad$ $\uparrow$ co-domain of S_k

However, both $\{N_t\}$ and $\{S_k\}$ involves a continuous time horizon (either as domain or codomain) $\Rightarrow$ This is not practical for sake of model fitting and estimation $\Rightarrow$ A discrete-time formulation can be obtained as follows:



- We divide the time horizon $(0, t]$ into small time bins, all of equal size $\Delta t \triangleq t_j - t_{j-1} = \text{const.} \ \forall j$

- We choose Δt such that at most one arrival time S_k falls in a bin

- We define the RVs: $Y_i \triangleq \begin{cases} N_{t_i} - N_{t_{i-1}} & i > 1 \\ N_{t_i} & i = 1 \end{cases}$

Note that Y_i denotes the increment in count from time t_{i-1} to time t_i, i.e., $Y_i = \Delta N_{(t_{i-1}, t_i]}$, and – given the definition of Δt – it can be $Y_i = 0$ or $Y_i = 1$ $\Rightarrow$ $Y_i \sim \text{Bernoulli}(p_i)$ $\Rightarrow$ Because Y_i $i = 1, 2, 3, \dots$ are defined out of a stochastic process, p_i may change with i and be dependent on previous variables Y_{i-1}, Y_{i-2}, etc. $\Rightarrow$ $\{Y_k\}_k$ is a sequence of INHOMOGENEOUS RVs

In order to develop an appropriate model for p_i, let us consider cases of incremental complexity:

* Homogeneous Poisson Processes

$\{S_k\}_k$ - Homogeneous Poisson Process (HPP) $\overset{DEF}{\Longleftrightarrow}$

a) $\forall (t, t+T]$, $\Delta N_{(t,t+T]} \sim \mathcal{P}(\lambda T)$ with λ assigned and constant

b) $\forall t_1 < t_2 < t_3 < t_4$, we have: $\Delta N_{(t_1,t_2]}$, $\Delta N_{(t_3,t_4]}$ are independent RVs

Part a) of the definition can be interpreted by observing:

- Divide $(t, t+T]$ into $m > 0$ intervals, each one of size Δt
- Define $Y_i \sim$ Bernoulli (p_i) the increment in the interval of length Δt $(t_{i-1}, t_i]$, $i = 1, 2, \dots, m$
- Define: $p_i = \dfrac{\lambda T}{m} = \lambda \Delta t$

$\Rightarrow$ We have:

$$P\left(\Delta N_{(t,t+T]} = k\right) = \binom{m}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{m-k} \xrightarrow[\substack{\Delta t \to 0 \\ (\text{i.e., } m \to \infty)}]{} e^{-\lambda T} \frac{(\lambda T)^k}{k!}$$

This means that, with HPP, we assume that the unit increments are independent, with the same probability, and with this probability being independent of previous history.

Moreover, condition a) implies that the longer the window T, the larger the expected value of the increment and that the increment is independent of T $\Rightarrow$

$\forall t_1 \neq t_2$, $\Delta N_{(t_1, t_1+T)}$ and $\Delta N_{(t_2, t_2+T]}$ have the same distribution $\Rightarrow$ HPP is TIME-INVARIANT and increments $\Delta N_{(t, t+T]}$ are STATIONARY

Furthermore, condition b) implies that the increments are INDEPENDENT random variables

Finally, one can observe:

$$P(X_i > t) = P(\Delta N_{(S_{i-1}, S_{i-1}+t]} = 0) = e^{-\lambda t} \iff F_{X_i}(t) = 1 - e^{-\lambda t} \iff X_i \sim Exp(\lambda)$$

X_i: inter-event waiting time

S_{i-1}: arrival time of $(i-1)$-th event

This implies that HPPs are memoryless (*)

* Inhomogeneous Poisson Processes

$\{S_k\}$ - Inhomogeneous Poisson Process (IPP) $\overset{DEF}{\iff}$

a) $\forall (t, t+T]$, $\Delta N_{(t, t+T]} \sim \mathcal{P}(\mu)$ with: $\mu \triangleq \int_t^{t+T} \lambda(\tau) d\tau$ and $\lambda(t)$ - known function

b) $\forall t_1 < t_2 < t_3 < t_4$, we have: $\Delta N_{(t_1, t_2]}$, $\Delta N_{(t_3, t_4]}$ are independent RVs

Because of b), IPPs keep the increments independent and condition (*) still holds (i.e., IPPs are memoryless). However, since $\lambda(t)$ may be different over two distinct time intervals of equal size (e.g., $(t_1, t_1+T]$ and $(t_2, t_2+T]$ with $t_1 \neq t_2$) because of a), we have that the increments are NOT stationary any more

Note: for $\Delta t = dt$ (infinitesimal) we have:

HPP: $P(\text{event in } (t, t+dt]) = \lambda dt$ IPP: $P(\text{event in } (t, t+dt]) \cong \lambda(t) dt$

$\Rightarrow \lambda(t)$ is an intensity function that determines the probability of the Bernoulli process associated with the interval $(t, t+dt] \Rightarrow$ The probability is now time-varying

The introduction of a function $\lambda(t)$ that varies with t leads to numerous useful properties. In particular, note:

$S_i \triangleq$ arrival time of i-th event $\Rightarrow$ We can write for the conditional pdf of S_i:

$$\underbrace{P(S_i > t \mid S_{i-1} = s)}_{\text{with } t>s} = P\left(\Delta N_{(s,t]} = 0\right) = \exp\left(-\int_s^t \lambda(\tau)d\tau\right) \Rightarrow$$

$$F_{S_i}(t \mid S_{i-1} = s) = 1 - P(S_i > t \mid S_{i-1} = s) = 1 - \exp\left(-\int_s^t \lambda(\tau)d\tau\right) \Rightarrow$$

$$f_{S_i}(t \mid S_{i-1} = s) = \frac{dF_{S_i}}{dt} = \exp\left(-\int_s^t \lambda(\tau)d\tau\right)\lambda(t) \qquad (**)$$

From (**) we can derive the joint pdf of the entire sequence $S_1, S_2, S_3, \ldots, S_n$ over a fixed interval $(0,T]$:

$$f(s_1, s_2) = f_{S_2}(s_2 \mid S_1 = s_1) f_{S_1}(s_1)$$

$$f(s_1, s_2, s_3) = f_{S_3}(s_3 \mid S_1 = s_1, S_2 = s_2) f(s_1, s_2) = f_{S_3}(s_3 \mid S_1 = s_1, S_2 = s_2) f_{S_2}(s_2 \mid S_1 = s_1) f_{S_1}(s_1)$$

Because these processes are memoryless, we have:

$$f_{S_3}(s_3 \mid S_1 = s_1, S_2 = s_2) = f_{S_3}(s_3 \mid S_2 = s_2)$$

Therefore, we can write:

$$f(s_1, s_2) = \lambda(s_2) \exp\left(-\int_{s_1}^{s_2} \lambda(\tau) d\tau\right) \cdot \lambda(s_1) \exp\left(-\int_{0}^{s_1} \lambda(\tau) d\tau\right)$$

$$f(s_1, s_2, s_3) = \lambda(s_3) \exp\left(-\int_{s_2}^{s_3} \lambda(\tau) d\tau\right) \cdot \lambda(s_2) \exp\left(-\int_{s_1}^{s_2} \lambda(\tau) d\tau\right) \cdot \lambda(s_1) \exp\left(-\int_{0}^{s_1} \lambda(\tau) d\tau\right)$$

$$= \prod_{i=1}^{3} \lambda(s_i) \exp\left(-\int_{0}^{s_3} \lambda(\tau) d\tau\right)$$

Analogously, we can write for any $n > 0$ and time T:

$$f(s_1, s_2, \dots, s_n, T) = \prod_{i=1}^{n} \lambda(s_i) \exp\left(-\int_{0}^{T} \lambda(\tau) d\tau\right) \qquad (***)$$

where in $(***)$ we take into account that $T > s_n$ and we use $P\left(\Delta N_{(s_n, T]} = 0\right) =$
$= \exp\left(-\int_{s_n}^{T} \lambda(\tau) d\tau\right)$

Formula $(***)$ can also be interpreted as the asymptotic probability of a binary time series. In fact, we can proceed as for HPPs and divide the interval $(0, T]$ into $m > 0$ intervals, each one of size Δt, and define the increment in the interval of length Δt $(t_{i-1}, t_i]$ as $Y_i \sim \text{Bernoulli}(p_i)$ with $p_i \triangleq \lambda(\hat{t}_i)\Delta t$ $\hat{t}_i \triangleq \frac{t_{i-1} + t_i}{2}$;

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m) = \underbrace{\prod_{i=1}^{m} p_i^{y_i} (1 - p_i)^{1 - y_i}}_{\text{Each factor here returns } p_i \text{ if } y_i = 1 \text{ or } (1-p_i) \text{ if } y_i = 0} = \prod_{j \in S} p_j \cdot \prod_{k \in \bar{S}} (1 - p_k)$$

where $S \triangleq$ set of intervals for which $Y_j = 1$

$\bar{S} \triangleq$ complement of S to the m intervals

Provided that n events occur in the interval $(0,T]$, we have:

$$\frac{1}{\Delta t^n}\prod_{j\in S} p_j = \frac{1}{\Delta t^n}\prod_{j\in S}\lambda(\hat{t}_j)\Delta t \xrightarrow[\Delta t\to 0]{} \prod_{i=1}^{n}\lambda(s_i)$$

$$\prod_{k\in\bar{S}}\left(1-\lambda(\hat{t}_k)\Delta t\right) = \exp\left(\sum_{k\in\bar{S}}\log\left(1-\lambda(\hat{t}_k)\Delta t\right)\right) \underset{\substack{\uparrow\\ \text{for small}\\ \text{values } \Delta t}}{\cong} \exp\left(\sum_{k\in\bar{S}} -\lambda(\hat{t}_k)\Delta t\right) \xrightarrow[\Delta t\to 0]{} e^{-\int_0^T \lambda(\tau)d\tau}$$

Because of the Taylor series expansion:

$$\log(1+t) = t + O(t)$$

Therefore we can interprete: $f(s_1, s_2, \dots, s_n, T) = \lim_{\Delta t\to 0}\frac{1}{\Delta t^n}\prod_{j=1}^{m}\left(\lambda(\hat{t}_j)\Delta t\right)^{y_j}\left(1-\lambda(\hat{t}_j)\Delta t\right)^{1-y_j}$

$\Rightarrow$ We can treat an IPP approximately as a series of Bernoulli trials

* Non-Poisson Point Processes

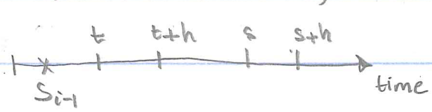
$$\text{HPP}\begin{Bmatrix}\text{constant probability}\\ \text{memoryless}\end{Bmatrix} \xrightarrow{\text{stationariety}} \text{IPP}\begin{Bmatrix}\text{memoryless}\end{Bmatrix} \xrightarrow{\text{independency}} \text{NP-PP}\begin{Bmatrix}\text{time-varying probability}\\ \text{history-dependent}\end{Bmatrix}$$

One way to proceed toward NP-PP is via inter-event waiting times. Note that, with HPPs, we showed that: $X_i \sim Exp(\lambda)\ \forall i \Rightarrow \forall t, s, h > 0$ with $t \neq s$ we have:

$$\left.\begin{aligned} P(X_i > t+h \mid X_i > t) &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h}\\ P(X_i > s+h \mid X_i > s) &= \frac{e^{-\lambda(s+h)}}{e^{-\lambda s}} = e^{-\lambda h}\end{aligned}\right\} \Rightarrow P(X_i > t+h \mid X_i > t) = P(X_i > s+h \mid X_i > s)$$

i.e., X_i is memoryless

We can define a NP-PP by imposing that: $P(X_i > t+h \mid X_i > t) \neq P(X_i > s+h \mid X_i > s)$ $\forall t \neq s \Rightarrow$ In this case, we have:



$$P\left(\Delta N_{(t,t+h]} = 0\right) = P\left(X_i > t+h-s_{i-1} \mid X_i > t-s_{i-1}\right)$$

$$P(\Delta N_{(s,s+h]}=0) = P(X_i > s+h-s_{i-1} \mid X_i > s-s_{i-1})$$

$$P(\Delta N_{(t,t+h]}=0 \cap \Delta N_{(s,s+h]}=0) = P(X_i > s+h-s_{i-1} \mid X_i > s-s_{i-1})$$

$\Rightarrow P(AB) \neq P(A)P(B)$ where: $A \triangleq \{\Delta N_{(s,s+h]}=0\}$, $B \triangleq \{\Delta N_{(t,t+h]}=0\}$ $\Rightarrow$ The independency does NOT hold any more

However, variables $X_1, X_2, \ldots, X_i, \ldots$ are still independent.

DEFINITION: $\{S_k\}_k$ - Renewal Process $\overset{DEF}{\Leftrightarrow}$ $\{X_i\}_i$ are i.i.d.

Note that HPPs have $X_i \sim Exp(\lambda)$ $\forall i \Rightarrow$ HPP is a particular Renewal Process
The example given above, instead, is a NP-PP (not a HPP) but it is a Renewal Process nonetheless

Ex.: $X_i \sim$ Inv. Gauss. $\Rightarrow$ $\{S_k\}_k$ - is a Renewal Process and NP-PP
$X_i \sim$ Gamma $\Rightarrow$ $\{S_k\}_k$ - is a Renewal Process and NP-PP

Interestingly, the CV can be used as a measure of regularity of a Renewal Process:

HPP $\Rightarrow X_i \sim Exp(\lambda)$ $\forall i \Rightarrow CV = 1$
NP-PP $\Rightarrow X_i$ are i.i.d. but non-Exp $\forall i \Rightarrow CV \neq 1$
$\Rightarrow$ Depending on how close to 1 the CV is, one can conclude how similar to a HPP the process is

Moreover: $CV < 1 \Rightarrow$ The process is more regular than a HPP (e.g., oscillatory neuron)
$CV > 1 \Rightarrow$ The process is less regular than a HPP (e.g., bursting neuron)

Interestingly, the results shown for the probability of a binary event in HPPs extend to Renewal Processes:

Renewal Theorem

$$\left.\begin{array}{l} \{S_k\}_k \text{ - Renewal Process} \\ X_i \sim f_{X_i}(x) \quad \forall i \text{ where:} \\ \underset{X_i}{E}(X_i) = \mu \text{ - const and} \\ f_{X_i}(x) \text{ - continuous function} \end{array}\right\} \Rightarrow \lim_{t\to\infty} E\left(\Delta N_{(t,t+h]}\right) = \frac{h}{\mu} \quad \forall h \geq 0$$

Based on this theorem we can conclude:

$$\left.\begin{array}{l} h = dt \text{ (infinitesimal)} \\ \lambda \triangleq 1/\mu \text{ ; } t\text{ - large} \end{array}\right\} P\left(\text{event in } (t,t+dt]\right) = E\left(\Delta N_{(t,t+dt]}=1\right) \cong \lambda dt$$

as we had before for HPPs and IPPs $\Rightarrow$ We can treat a NP-PP as a HPP provided that we wait long enough $\Rightarrow$ This has particular relevance when multiple renewal processes are combined □

Another (more general) way to proceed toward NP-PP is via conditional intensity functions. Let us consider again part a) of the definition of IPP:

$$\forall (t,t+T], \ \Delta N_{(t,t+T]} \sim \mathcal{P}(\mu) \quad \text{with } \mu = \int_t^{t+T} \lambda(\tau)d\tau$$

$\Downarrow$ We derived:

$T = dt$ (infinitesimal), then: $P\left(\text{event in } (t,t+dt]\right) \cong \lambda(t)\, dt$

Let us now assume that $\lambda(t)$ is a function of both time and previous arrival times; i.e., let us consider:

$$\lambda\left(t \mid S_1 = s_1, S_2 = s_2, \dots, S_n = s_n, N(t) = n\right)$$ - Conditional Intensity Function (CIF)

A practical way to define the CIF stems from the analogy with IPPs:

$$\text{IPP}: \lambda(t) = \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1)}{\Delta t} \Rightarrow \text{NP-PP}: \lambda(t|\mathcal{H}_t) \triangleq \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1 | \mathcal{H}_t)}{\Delta t}$$

$$\mathcal{H}_t \triangleq (s_1, s_2, \ldots, s_n, n) \qquad (\Delta)$$

Note that definition (Δ) requires that no more than one event can occur in $(t, t+\Delta t]$ for Δt small enough $\Rightarrow$ Mathematically, this means that $\{S_k\}_k$ is ORDERLY.

Note this:

$$\overbrace{P(X_i \in (t, t+\Delta t) | X_i > t, S_1 = s_1, S_2 = s_2, \ldots, S_{i-1} = s_i, N(t) = i-1)}^{(\alpha)} = P(\Delta N_{(t,t+\Delta t)} > 0 | \mathcal{H}_t)$$

$$\Rightarrow \lim_{\Delta t \to 0} \frac{(\alpha)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t)} > 0 | \mathcal{H}_t)}{\Delta t} \underset{\substack{\uparrow \\ \text{Because of} \\ \text{orderly condition}}}{=} \lim_{\Delta t \to 0} \frac{P(\Delta N_{(t,t+\Delta t)} = 1 | \mathcal{H}_t)}{\Delta t} = \lambda(t|\mathcal{H}_t)$$

Also note: $P(X_i \in (t, t+\Delta t) | X_i > t, \mathcal{H}_t) = \dfrac{F_{X_i}(t+\Delta t|\mathcal{H}_t) - F_{X_i}(t|\mathcal{H}_t)}{1 - F_{X_i}(t|\mathcal{H}_t)}$

with $F_{X_i}(\cdot|\mathcal{H}_t)$ - conditional CDF of the inter-event waiting time RV $\Rightarrow$ We can write:

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta t \to 0} \frac{P(X_i \in (t, t+\Delta t) | X_i > t, \mathcal{H}_t)}{\Delta t} = \frac{f_{X_i}(t|\mathcal{H}_t)}{1 - F_{X_i}(t|\mathcal{H}_t)}$$

with: $f_{X_i}(\cdot|\mathcal{H}_t)$ - conditional PDF of X_i and, by analogy with the IPP case, we have:

$$f_{X_i}(t|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)\left(1 - F_{X_i}(t|\mathcal{H}_t)\right) = \lambda(t|\mathcal{H}_t) \exp\left(-\int_{s_{i-1}}^{t} \lambda(u|\mathcal{H}_u)\, du\right)$$

Also, by following the same steps taken in (***) for the IPP case, we can write:

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

joint probability function

Note that, because of the use of conditional probabilities, if we consider the unit increment RV: $Y_i \triangleq \Delta N_{(t_{i-1}, t_i]}$ with $\Delta t \triangleq t_i - t_{i-1}$ $\forall i$, we can still write:

$Y_i \sim \text{Bernoulli}(p_i)$ with $p_i \triangleq \int_{t_{i-1}}^{t_i} \lambda(u|\mathcal{H}_u)\,du$ - but, because of $\mathcal{H}_t$ (history), in general we have: $P(Y_i = 1 | Y_{i-1} = 1) \neq p_i \Rightarrow$ The binary events are no more independent

Moreover, note this:

$\mathcal{H}_t = (s_1, s_2, \dots, s_n, n)$ is itself a realization of the finite process $\{S_k\}_{k=1}^{n}$ and the RV $N(t) \Rightarrow \lambda(t|\mathcal{H}_t)$ will assume different values at time t, depending on the value $\mathcal{H}_t \Rightarrow$ We can define a MARGINAL intensity function:

$$\hat{\lambda}(t) = \underset{\mathcal{H}_t}{E}\left(\lambda(t|\mathcal{H}_t)\right)$$

$$\Downarrow$$

We can use: $\lambda(t|\mathcal{H}_t) \rightarrow$ for intra-trial instantaneous event probability

$\hat{\lambda}(t) \rightarrow$ for across-trial averaged time-varying event probability ∎

* ML Estimation of IPPs and NP-PPs

The pivotal result for IPPs is that:

$$\underbrace{P(Y_1 = y_1, Y_2 = y_2, \dots, Y_m = y_m)}_{\text{discretization of the process}} \xrightarrow[\Delta t \to 0]{} \underbrace{f(s_1, s_2, \dots, s_n, T)}_{\text{joint-probability}}$$

Therefore, provided that Δt is small, we can approximate:

$$\ell(\text{PP}) \cong -n \log \Delta t + \sum_{j=1}^{m} y_j \log\left(\lambda(\hat{t}_j)\Delta t\right) + \sum_{j=1}^{m} (1-y_j) \log\left(1-\lambda(\hat{t}_j)\Delta t\right)$$

log-likelihood of the process on $(0,T]$

and we can fit the sequence of binary values $(y_1, y_2, \dots, y_m)$ by introducing a class of functions for λ and using ML estimation. In particular, we can observe:

$$Y_i \sim \text{Bernoulli}\left(\lambda(\hat{t}_i)\Delta t\right)$$
$$\lambda(t) = g(t|\theta)$$

$\Rightarrow$ Given the class of functions $g(\cdot|\theta)$, we can estimate the parameters θ by using the ML method

Also, it may be possible: $\lambda(t) = g(x(t)|\theta)$ with x being the realization of the explanatory RV X (e.g., $\eta(\lambda(t)) = X(t)\,\theta$)

Ex.: For hippocampal place cells, we can write:

$$\lambda(t) = \exp\left\{\alpha - \frac{1}{2}\begin{bmatrix} x(t)-\mu_x & y(t)-\mu_y \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x(t)-\mu_x \\ y(t)-\mu_y \end{bmatrix}\right\}$$

where: $x(t)$ is the realization at time t of R.V. X (position along the x-axis)

$y(t)$ is the realization at time t of R.V. Y (position along the y-axis)

$$\Sigma \triangleq \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$$ - covariance matrix

$$\theta \triangleq \left[\alpha\ \mu_x\ \mu_y\ \sigma_x^2\ \sigma_y^2\ \sigma_{xy}\right]^T$$

$$g(t|\theta) = g\left(\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \Big| \theta\right)$$

□

In a similar way, we can show for a NP-PP that:

$$\frac{1}{\Delta t^n} \prod_{j=1}^{m} \left(\lambda(\hat{t}_j | \mathcal{H}_{\hat{t}_j}) \Delta t \right)^{y_j} \left(1 - \lambda(\hat{t}_j | \mathcal{H}_{\hat{t}_j}) \Delta t \right)^{1-y_j} \xrightarrow[\Delta t \to 0]{} f(s_1, s_2, \dots, s_n, T)$$

Therefore, we can approximate the log-likelihood function ℓ(NP-PP) as done for the IPP case, with $\lambda(t|\mathcal{H}_t)$ replacing $\lambda(t)$, and use ML estimation to fit the model:

$$Y_i \sim \text{Bernoulli}\left(\lambda(\hat{t}_i | \mathcal{H}_{\hat{t}_i}) \Delta t \right)$$

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

The important thing here is that λ must depend on $\mathcal{H}_t$. One way to satisfy this is to notice: $Y_i = \Delta N_{(t_{i-1}, t_i]}$ $\Rightarrow$ One can consider, for instance:

$$\log\left(\lambda(t|\mathcal{H}_t)\right) = \alpha + \sum_k \beta_k \Delta N_{(t-(k+1)\Delta t,\, t-k\Delta t]}$$

$$\Updownarrow$$

$\lambda(t|\mathcal{H}_t)$ is a function of the observations obtained up to time t (t excluded)

□

References:

Textbook: ch. 19

ch. 14 (Example 14.5)