TD 2 - Temporal point processes

Beyond Poisson processes - Correction

1 Basic exercises

Exercise 1. Let $X_1, \ldots, X_k$ be independent random variables with respective hazard rate functions q_i, $i = 1, \ldots, k$. Let $Z = \min_{i=1,\ldots,k} X_i$.

Let $t > 0$ and compute, by independence between the X's,

$$\overline{F}_Z(t) := \mathbb{P}(Z > t) = \prod_{i=1}^{k} \mathbb{P}(X_i > t) = \prod_{i=1}^{k} \overline{F}_{X_i}(t).$$

Hence, $\ln \overline{F}_Z(t) = \sum_{i=1}^{k} \ln \overline{F}_{X_i}(t)$ and differentiating this equality gives the fact that the hazard rate function of Z is $q(t) = \sum_{i=1}^{k} q_i(t)$.

Exercise 2 (Life time k-sample)**.** Let $(\xi_1, \ldots, \xi_k)$ be i.i.d. positive random variables with common hazard rate function q and density function f. Let $N = \{\xi_1, \ldots, \xi_k\}$ denote the point process made of those k random variables. And let $(T_1, \ldots, T_k)$ denote the order statistics made from $(\xi_1, \ldots, \xi_k)$.

Here are below two ways to solve this exercise. The first one uses the densities of $(T_1, \ldots, T_n)$ for all $n \leq k$. The second one focuses on the computation of the conditional survival functions

$$\overline{F}(s; t_1, \ldots, t_n) = \mathbb{P}(S_{n+1} > s | T_1 = t_1, \ldots, T_n = t_n).$$

Possibility 1. Let $\sigma \in \mathfrak{S}_k$ denote the (random) permutation of $\{1, \ldots, k\}$ such that $\mathbf{T}_k = (T_1, \ldots, T_k) = (\xi_{\sigma(1)}, \ldots, \xi_{\sigma(k)})$ is such that $T_1 < \cdots < T_k$. The following statement is well-known and its proof can be found easily on the internet:

- the distribution of σ is the uniform distribution on $\mathfrak{S}_k$,
- the density of $(T_1, \ldots, T_k)$ is

$$f_{\mathbf{T}_k}(t_1, \ldots, t_k) = k! \prod_{i=1}^{k} f(t_i) \mathbf{1}_{t_1 < \cdots < t_k}. \tag{1}$$

Let us then prove by induction from $n = k$ to $n = 1$ that the density of the vector $\mathbf{T}_n = (T_1, \ldots, T_n)$ is

$$f_{\mathbf{T}_n}(t_1, \ldots, t_n) = \frac{k!}{(k-n)!} \prod_{i=1}^{n} f(t_i) \overline{F}(t_n)^{k-n} \mathbf{1}_{t_1 < \cdots < t_n}, \tag{2}$$

where $\overline{F}$ is the survival function of the ξ's. The initial case corresponds exactly to Equation (1). Then, assume that (2) is true for n such that $2 \leq n \leq k$ and prove that it is still true for $n - 1$. By definition of the marginal densities, we have

$$f_{\mathbf{T}_{n-1}}(t_1, \ldots, t_{n-1}) = \int_0^\infty f_{\mathbf{T}_n}(t_1, \ldots, t_n) dt_n = \frac{k!}{(k-n)!} \prod_{i=1}^{n-1} f(t_i) \int_{t_{n-1}}^\infty f(t_n) \overline{F}(t_n)^{k-n} dt_n.$$

Yet, one recognizes that the integrand above is the derivative of $\frac{1}{k-n+1}\overline{F}(t)^{k-n+1}$. Hence, Equation (2) holds true for $n-1$.

Then, it is clear that, for all $n=0,\dots,k-1$, the conditional density of T_{n+1} given $\mathbf{T}_n$ is

$$f_{T_{n+1}|\mathbf{T}_n}(t;t_1,\dots,t_n) = \frac{f_{\mathbf{T}_{n+1}}(t_1,\dots,t_n,t)}{f_{\mathbf{T}_n}(t_1,\dots,t_n)} = (k-n)\frac{f(t)\overline{F}(t)^{k-n-1}}{\overline{F}(t_n)^{k-n}},$$

and its associated survival function is

$$\overline{F}(t;t_1,\dots,t_n) = \frac{k-n}{\overline{F}(t_n)^{k-n}}\int_t^\infty f(s)\overline{F}(s)^{k-n-1}ds = (k-n)\left(\frac{\overline{F}(t)}{\overline{F}(t_n)}\right)^{k-n}.$$

Finally, the derivative of $-\ln\overline{F}(\cdot;t_1,\dots,t_n)$ gives the associated hazard rate and so

$$\lambda(t;t_1,\dots,t_n) = (k-n)\frac{d}{dt}(-\ln\overline{F})(t) = (k-n)q(t).$$

Finally, this gives the desired intensity $\lambda_t = \lambda(t;N\cap[0,t)) = (k-N_{t-})q(t)$.

Of course, it implicitly assumes the convention that a null hazard rate function corresponds to a random variable equal to $+\infty$ almost surely. This is the case for T_{k+1} in this exercise, so that $\lambda_t = 0$ as soon as $N_{t-} = k$.

Possibility 2. To compute the intensity of N, it suffices to compute $Q(s;T_1,\dots,T_n)$ (namely the hazard rate function of S_{n+1} given $T_1,\dots,T_n$) for $n=0,\dots,k$.

Step $n=0$. It is the (unconditional) hazard rate function of T_1, that is the minimum of $\xi_1,\dots,\xi_k$. By exercice 1, it is $Q(s;\emptyset) = \sum_{i=1}^k q(s) = kq(s) = (k-n)q(s)$.

Step $n=1$. Let $t_1>0$ and $s>0$. We first compute the conditional survival function

$$\overline{F}(s;t_1) = \mathbb{P}(S_2 > s|T_1 = t_1).$$

Then, we will take the derivative of its log to recover the hazard rate function.

Since T_1 is one of the ξ's, we have by the law of total probability

$$\overline{F}(s;t_1) = \sum_{i=1}^k \mathbb{P}(S_2 > s, T_1 = \xi_i|T_1 = t_1).$$

Then, by property of the conditional probabilities, we have

$$\mathbb{P}(S_2 > s, T_1 = \xi_i|T_1 = t_1) = \mathbb{P}(S_2 > s|T_1 = t_1, T_1 = \xi_i)\,\mathbb{P}(T_1 = \xi_i|T_1 = t_1)$$

The result above is classic when all the conditioning is of non zero probability, but it also holds true when the conditioning is of zero probability.

The conditioning event $\{T_1 = t_1, T_1 = \xi_i\}$ is equal to $\{\xi_i = t_1, \hat{\xi}_i > t_1\}$, where

$$\hat{\xi}_i = \min_{\substack{j=1,\dots,k\\ j\neq i}} \xi_j,$$

and $S_2 = \hat{\xi}_i - \xi_i$ on that event. Hence,

$$\mathbb{P}(S_2 > s|T_1 = t_1, T_1 = \xi_i) = \mathbb{P}\left(S_2 > s|\xi_i = t_1, \hat{\xi}_i > t_1\right) = \mathbb{P}\left(\hat{\xi}_i > s + t_1|\xi_i = t_1, \hat{\xi}_i > t_1\right).$$

By independence of the ξ's, we have independence between $\hat{\xi}_i$ and ξ_i so that

$$\mathbb{P}\left(S_2 > s | T_1 = t_1, T_1 = \xi_i\right) = \mathbb{P}\left(\hat{\xi}_i > s + t_1 | \hat{\xi}_i > t_1\right) = \frac{\overline{F}_\xi(s+t_1)^{k-1}}{\overline{F}_\xi(t_1)^{k-1}},$$

where $\overline{F}_\xi$ is the common survival function of the ξ's.

Coming back to $\overline{F}(s;t_1)$, we have

$$\overline{F}(s;t_1) = \sum_{i=1}^{k} \frac{\overline{F}_\xi(s+t_1)^{k-1}}{\overline{F}_\xi(t_1)^{k-1}} \mathbb{P}\left(T_1 = \xi_i | T_1 = t_1\right) = \frac{\overline{F}_\xi(s+t_1)^{k-1}}{\overline{F}_\xi(t_1)^{k-1}}.$$

Hence, $\ln \overline{F}(s;t_1) = (k-1)\left(\ln \overline{F}_\xi(s+t_1) - \ln \overline{F}_\xi(t_1)\right)$. By differentiating with respect to s, we get $Q(s;t_1) = (k-1)q(s+t_1)$ which is what we want (remind that t_1 is the last observation here so that we perform the change of variable $t = s + t_1$ and the number of observed events is $N_{t-} = 1$).

Step $n \geq 2$. In summary, the idea of the previous step is to restrict to the case where we "fix" the index $i \in \{1,\dots,k\}$ of the minimum of the ξ's. Then, we can use the mutual independence of the ξ's. One can adapt this argument to $n \geq 2$ by "fixing" the indices of the n lowest ξ's. It suffices then to compute this kind of conditional probabilities:

$$\mathbb{P}\left(S_{n+1} > s | T_1 = \xi_1, \dots, T_n = \xi_n \text{ and } T_1 = t_1, \dots, T_n = t_n\right)$$

The conditioning event $\{T_1 = \xi_1, \dots, T_n = \xi_n \text{ and } T_1 = t_1, \dots, T_n = t_n\}$ is equal to $\{\xi_1 = t_1, \dots, \xi_n = t_n \text{ and } \hat{\xi} > t_n\}$, where

$$\hat{\xi} = \min_{j=n+1,\dots,k} \xi_j,$$

and $S_{n+1} = \hat{\xi} - \xi_n$ on that event. Hence,

$$\mathbb{P}\left(S_{n+1} > s | T_1 = \xi_1, \dots, T_n = \xi_n \text{ and } T_1 = t_1, \dots, T_n = t_n\right) = \mathbb{P}\left(\hat{\xi} > s + t_n | \xi_1 = t_1, \dots, \xi_n = t_n \text{ and } \hat{\xi} > t_n\right).$$

By independence of the ξ's, we have independence between $\hat{\xi}$ and $\xi_1, \dots, \xi_n$ so that

$$\mathbb{P}\left(S_{n+1} > s | T_1 = \xi_1, \dots, T_n = \xi_n \text{ and } T_1 = t_1, \dots, T_n = t_n\right) = \frac{\overline{F}_\xi(s+t_n)^{k-n}}{\overline{F}_\xi(t_n)^{k-n}},$$

where $\overline{F}_\xi$ is the common survival function of the ξ's.

Coming back to $\overline{F}(s;t_1,\dots,t_n)$, we have

$$\overline{F}(s;t_1,\dots,t_n) = \frac{\overline{F}_\xi(s+t_n)^{k-n}}{\overline{F}_\xi(t_n)^{k-n}}.$$

Hence, $\ln \overline{F}(s;t_1,\dots,t_n) = (k-n)\left(\ln \overline{F}_\xi(s+t_n) - \ln \overline{F}_\xi(t_n)\right)$. By differentiating with respect to s, we get $Q(s;t_1,\dots,t_n) = (k-n)q(s+t_n)$ which is what we want (remind that t_n is the last observation here so that we perform the change of variable $t = s + t_n$ and the number of observed events is $N_{t-} = n$).

Exercise 3 (Poisson contamination)**.** Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ and $(N^i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. Poisson processes with intensity $h(t)$. Let $N_t = \sum_{i=0}^{\infty} N^i_{t-i}$.

1. The translated processes $\hat{N}^i$ defined by $\hat{N}^i_t = N^i_{t-i}$, or equivalently by $\hat{N}^i = \{T+i, T \in N^i\}$, are independent Poisson processes on $\mathbb{R}_+$ (the independence property of Poisson is clearly preserved by translation) with intensities $\lambda^i(t) = h(t-i)$ with the convention that $h(s) = 0$ for all $s < 0$ (the intensity is translated in the same way and completed by 0). Hence, the superposition Theorem for Poisson processes (stated for two processes but clearly valid for more than two processes by induction) yields that $N_t = \sum_{i=0}^{\infty} \hat{N}^i_t$ defines a Poisson process with intensity $\lambda(t) = \sum_{i=0}^{\infty} \lambda^i(t) = \sum_{i=0}^{\infty} h(t-i)$.

2. Let $\tilde{N} = \{\tilde{T}_i, i \in \mathbb{N}\}$ be a point process independent of $(N^i)_{i\in\mathbb{N}}$. Define N by $N_t = \sum_{i=0}^{\infty} N^i_{t-\tilde{T}_i}$.

   By independence assumption, if we work conditionally on $\tilde{N}$, the processes $(N^i)_{i\in\mathbb{N}}$ are still independent Poisson processes with intensity $h(t)$. The same arguments as in the previous question yield that $\hat{N}^i = \{T + \tilde{T}_i, T \in N^i\}$ are independent Poisson processes on $\mathbb{R}_+$ with intensities $\lambda^i(t) = h(t-\tilde{T}_i)$. Hence, given $\tilde{N}$, $N_t = \sum_{i=0}^{\infty} \hat{N}^i_t$ defines a Poisson process with intensity $\lambda(t) = \sum_{i=0}^{\infty} \lambda^i(t) = \sum_{i=0}^{\infty} h(t-\tilde{T}_i)$. In particular, N is a doubly stochastic Poisson process.

Exercise 4 (Thinning simulation)**.** See the Julia notebook.

Exercise 5 (Change-time simulation)**.** See the Julia notebook.

2 Intermediate exercises

Exercise 6 (Generalization of exercises 2 and 3)**.** *Most of the arguments are similar to those of Exercise 2*

Let N^1 and N^2 be two independent point processes with intensities $\lambda^1_t = \lambda^1(t; N^1 \cap [0,t))$ and $\lambda^2_t = \lambda^2(t; N^2 \cap [0,t))$. Denote $N = N^1 \cup N^2$ the superposition of the two processes. Assume that there exists a measurable function λ such that $\lambda^1_t + \lambda^2_t = \lambda(t; N \cap [0,t))$. Then, let us prove that $\lambda_t = \lambda(t; N \cap [0,t))$ is the intensity of N.

Let us denote $N^1 = \{T^1_1 < T^1_2 < \dots\}$, $N^2 = \{T^2_1 < T^2_2 < \dots\}$ and $N = \{T_1 < T_2 < \dots\}$. Let us prove that for all $n = 0, \dots, \infty$, $\lambda(\cdot; t_1, \dots, t_n)$ is the hazard rate function of T_{n+1} given $T_1 = t_1, \dots, T_n = t_n$. Moreover, let us denote $\overline{F}, \overline{F}^1$ and $\overline{F}^2$ the generalized survival functions associated with the generalized hazard rate functions λ, λ^1 and λ^2 respectively.

Step $n = 0$**.** Since T^1_1 and T^2_1 are independent, the (unconditional) hazard rate function of $T_1 = \min(T^1_1, T^1_2)$ is $\lambda^1(t, \emptyset) + \lambda^2(t, \emptyset)$ (see Exercise 1). This quantity is equal to $\lambda(t; \emptyset)$ by assumption.

Step $n \geq 1$**.** Let $t_1 < \dots < t_n$ and $t > t_n$. Let us compute the conditional survival function

$$\overline{F}(t; t_1, \dots, t_n) = \mathbb{P}(T_{n+1} > t | T_1 = t_1, \dots, T_n = t_n).$$

Next we specify for each time $T_1, \dots, T_n$ if it belongs to N^1 or N^2. More precisely, by the law of total probability, we have

$$\overline{F}(t; t_1, \dots, t_n) = \sum_{\varepsilon \in \{1,2\}^n} \mathbb{P}(T_{n+1} > t, T_1 \in N^{\varepsilon_1}, \dots, T_n \in N^{\varepsilon_n} | T_1 = t_1, \dots, T_n = t_n).$$

Now, fix $\varepsilon \in \{1,2\}^n$ and denote for $k = 1,2$, $n_k = \sum_{i=1}^n \mathbf{1}_{\varepsilon_i = k}$ the number of events that belong to N^k and $t^k_1 < \dots < t^k_{n_k}$ such that $\{t^k_1, \dots, t^k_{n_k}\} = \{t_i, i \in \{1, \dots, n\}$ s.t. $\varepsilon_i = k\}$, namely the event times that belong to N^k. Let us consider the following conditional survival function

$$A_\varepsilon(t) = \mathbb{P}(T_{n+1} > t | T_1 \in N^{\varepsilon_1}, \dots, T_n \in N^{\varepsilon_n} \text{ and } T_1 = t_1, \dots, T_n = t_n).$$

The conditioning event $\{T_1 \in N^{\varepsilon_1}, \dots, T_n \in N^{\varepsilon_n} \text{ and } T_1 = t_1, \dots, T_n = t_n\}$ is equal to

$$B_\varepsilon = \left\{T_1^1 = t_1^1, \dots, T_{n_1}^1 = t_{n_1}^1 \text{ and } T_1^2 = t_1^2, \dots, T_{n_2}^2 = t_{n_2}^2\right\},$$

and, on this event, $T_{n+1} = \min\left(T_{n_1+1}^1, T_{n_2+1}^2\right)$, so that

$$A_\varepsilon(t) = \mathbb{P}\left(T_{n_1+1}^1 > t, T_{n_2+1}^2 > t | B_\varepsilon\right).$$

By independence between N^1 and N^2, we can factorize $A_\varepsilon(t) = A_\varepsilon^1(t) \times A_\varepsilon^2(t)$, where

$$A_\varepsilon^k(t) = \mathbb{P}\left(T_{n_k+1}^k > t | T_1^k = t_1^k, \dots, T_{n_k}^k = t_{n_k}^k\right) = \overline{F}^k(t; t_1^k, \dots, t_{n_k}^k).$$

This survival function corresponds to the hazard rate $\lambda^k(t; t_1^k, \dots, t_{n_k}^k)$. Hence, by differentiating $-\ln A_\varepsilon$ we get that the hazard rate associated with the survival function A_ε is

$$\lambda^1(t; t_1^1, \dots, t_{n_1}^1) + \lambda^2(t; t_1^2, \dots, t_{n_2}^2).$$

By assumption, this quantity is equal to $\lambda(t; t_1, \dots, t_n)$ whatever ε is. In particular, the survival function $A = A_\varepsilon$ does not depend on ε so that the survival function we seek is

$$\overline{F}(t; t_1, \dots, t_n) = \sum_{\varepsilon \in \{1,2\}^n} A(t) \mathbb{P}(T_1 \in N^{\varepsilon_1}, \dots, T_n \in N^{\varepsilon_n} | T_1 = t_1, \dots, T_n = t_n) = A(t),$$

and its hazard rate function is $\lambda(\cdot; t_1, \dots, t_n)$.

All in all, it means that N admits $\lambda_t = \lambda(t; N \cap [0,t))$ as an intensity.

Exercise 7 (Thinning coupling)**.** See the homework

Exercise 8 (Change-time coupling)**.** Let $\lambda^1, \lambda^2 : \mathbb{R}_+ \to \mathbb{R}_+$ be two measurable functions. Let Π be a unit rate Poisson process on $\mathbb{R}_+$. By the change-time representation, we know that, for $k = 1, 2$, the point process N^k defined by, for all $t \geq 0$,

$$N_t^k = \Pi_{\Lambda^k(t)}, \quad \text{where } \Lambda^k(t) = \int_0^t \lambda^k(s) ds,$$

is a Poisson process with intensity λ^k.

In particular, we have $N_t^k = \int_0^\infty g_t^k(s) \Pi(ds)$ with the measurable test function $g_t^k(s) = \mathbf{1}_{s \leq \Lambda^k(t)}$, and, since the function $g_t^1 - g_t^2$ does not change its sign as s vary, we have

$$\mathbb{E}\left[|N_t^1 - N_t^2|\right] = \mathbb{E}\left[\int_0^\infty |g_t^1(s) - g_t^2(s)| \Pi(ds)\right] = \mathbb{E}\left[\int_0^\infty |g_t^1(s) - g_t^2(s)| ds\right] = \int_0^\infty |g_t^1(s) - g_t^2(s)| ds,$$

by Campbell Theorem for the unit rate Poisson process Π (remark that there is nothing random in the test functions g). Using again the fact that $g_t^1 - g_t^2$ does not change its sign, we get

$$\mathbb{E}\left[|N_t^1 - N_t^2|\right] = \left|\int_0^\infty g_t^1(s) - g_t^2(s) ds\right| = \left|\int_0^t \lambda^1(s) - \lambda^2(s) ds\right|. \tag{3}$$

Moreover, assume without loss of generality that $\Lambda^1(t) \leq \Lambda^2(t)$. Then,

$$\mathbb{P}\left(N_t^1 \neq N_t^2\right) = \mathbb{P}\left(\Pi_{\Lambda^2(t)} - \Pi_{\Lambda^1(t)} \neq 0\right) = \mathbb{P}\left(\Pi(\,[\Lambda^1(t), \Lambda^2(t))\,) \neq 0\right).$$

Yet, $\Pi(\,[\Lambda^1(t), \Lambda^2(t))\,)$ is a Poisson random variable with parameter $\Pi(\Lambda^2(t) - \Lambda^1(t))$ so that

$$\mathbb{P}\left(N_t^1 \neq N_t^2\right) = 1 - \exp\left(-\left|\int_0^t \lambda^1(s) - \lambda^2(s) ds\right|\right). \tag{4}$$

1. Assume that λ^2 is such that $|\lambda_t^1 - \lambda_t^2| \leq \varepsilon$ for all $t \geq 0$ and some $\varepsilon > 0$. Applying the last formula, we get

$$\mathbb{P}\left(N_t^1 \neq N_t^2\right) \leq 1 - \exp\left(-\int_0^t \left|\lambda^1(s) - \lambda^2(s)\right| ds\right) \leq 1 - e^{-t\varepsilon} \xrightarrow[\varepsilon \to 0]{} 0.$$

We have here a convergence result for the one time marginals N_t^1 and N_t^2. This is way weaker than the convergence result in total variation obtained in Exercise 7.

2. Let $\lambda^1(t) = \sum_{k=0}^{+\infty} \mathbf{1}_{[2k,2k+1)}(t)$ and $\lambda^2(t) = \sum_{k=0}^{+\infty} \mathbf{1}_{[2k+1,2k+2)}(t)$. The associated cumulative intensities are

$$\Lambda^1(t) = \int_0^t \lambda^1(s)ds = \begin{cases} k + (t - 2k), & t \in [2k, 2k+1), \\ k+1, & t \in [2k+1, 2k+2). \end{cases}$$

and

$$\Lambda^2(t) = \int_0^t \lambda^2(s)ds = \begin{cases} k, & t \in [2k, 2k+1), \\ k + (t - 2k - 1), & t \in [2k+1, 2k+2). \end{cases}$$

It is clear that $0 \leq \int_0^t \lambda^1(s) - \lambda^2(s)ds \leq 1$ for all t and so

$$\mathbb{E}\left[|N_t^1 - N_t^2|\right] \leq 1 \quad \text{and} \quad \mathbb{P}\left(N_t^1 \neq N_t^2\right) \leq 1 - e^{-1} \approx 0{,}63,$$

follow from Equations (3) and (4).

The generalized inverse functions of Λ^1 and Λ^2 are

$$(\Lambda^1)^{-1}(y) = y + \sum_{k=0}^{\infty} k\mathbf{1}_{(k,k+1]}(y) = y + \lfloor y \rfloor \text{ and } (\Lambda^2)^{-1}(y) = y + \lfloor y \rfloor + 1.$$

Let $(T_n)_{n\geq 1}$ denote the points of the Poisson process Π. By definition, we then have $N^1 = \{(\Lambda^1)^{-1}(T_n), n \in \mathbb{N}^*\}$ and $N^2 = \{(\Lambda^2)^{-1}(T_n), n \in \mathbb{N}^*\}$. The image of $(\Lambda^1)^{-1}$ is clearly the support of λ^1, that is $\cup_{k=0}^{\infty}[2k, 2k+1)$. And the image of $(\Lambda^2)^{-1}$ is clearly the support of λ^2, that is $\cup_{k=0}^{\infty}[2k+1, 2k+2)$. In particular, these are disjoints and so N^1 and N^2 are disjoints.

Moreover, since $(\Lambda^2)^{-1}(y) = 1 + (\Lambda^1)^{-1}(y)$ for all y, it is then obvious that $N^2 = \{T+1, T \in N^1\}$.

3 Advanced exercises

Exercise 9 (Thinning and renewal)**.** Let $q^1, q^2 : \mathbb{R}_+ \to \mathbb{R}_+$ be two left-continuous functions.

For $k = 1, 2$, let us define N^k by, for all $t \geq 0$,

$$N_t^k = \int_0^t \int \mathbf{1}_{[0, q^k(A_s^k)]}(z) \Pi(ds, dz),$$

where A^k is the age process associated with N^k. One could find measurable functions f^1 and f^2 such that for all $t \geq 0$, $A_t^k = f^k(t; \Pi \cap ([0,t) \times \mathbb{R}_+))$.

Now, let us fix $t \geq 0$ and denote $p(t) = \mathbb{P}\left(\left(N^1 \Delta N^2\right) \cap [0,t] \neq \emptyset\right)$. The test function (of s and z)

$$\mathbf{1}_{[0,t]}(s) \left| \mathbf{1}_{[0,q^1(A_s^1)]}(z) - \mathbf{1}_{[0,q^2(A_s^2)]}(z) \right| \mathbf{1}_{\forall r \leq s, A_r^1 = A_r^2}$$

can be expressed as a measurable function $g(s,z;\Pi\cap([0,s)\times\mathbb{R}_+))$. Hence, we can use it in Campbell's formula, that is $E_1=E_2$ where

$$E_1=\mathbb{E}\left[\int\int \mathbf{1}_{[0,t]}(s)\left|\mathbf{1}_{[0,q^1(A^1_s)]}(z)-\mathbf{1}_{[0,q^2(A^2_s)]}(z)\right|\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}\Pi(ds,dz)\right]=p(t),$$

and

$$E_2=\mathbb{E}\left[\int\int \mathbf{1}_{[0,t]}(s)\left|\mathbf{1}_{[0,q^1(A^1_s)]}(z)-\mathbf{1}_{[0,q^2(A^2_s)]}(z)\right|\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}dsdz\right].$$

Remark that $\int\left|\mathbf{1}_{[0,q^1(A^1_s)]}(z)-\mathbf{1}_{[0,q^2(A^2_s)]}(z)\right|dz=|q^1(A^1_s)-q^2(A^2_s)|$. The rest of the proof is divided in two steps: 1) proof that the function p is right differentiable, 2) upper-bound of $p(t)$ using some kind of Grönwall lemma argument.

Right differentiability. Let $h\ge 0$. By Campbell and the remark below the definition of E_2, we have

$$p(t+h)-p(t)=\mathbb{E}\left[\int_t^{t+h}|q^1(A^1_s)-q^2(A^2_s)|\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}ds\right]$$

As h goes to 0, the left-continuity of q^1 and q^2 and the fact that the probability to get a point in $[t,t+h]$ goes to 0, we can approximate the integrand above by its value for $s=t$ so that

$$p(t+h)-p(t)\underset{h\to 0^+}{\sim} h\mathbb{E}\left[|q^1(A^1_t)-q^2(A^2_t)|\mathbf{1}_{\forall r\le t, A^1_r=A^2_r}\right],$$

which gives the fact that p is right differentiable. Let us denote p' its right derivative.

Conclusion. Let us upper-bound $p(t)$ using Campbell formula and the assumption on the hazard rates q^1 and q^2. We have

$$\begin{aligned} p(t) &= \mathbb{E}\left[\int_0^t|q^1(A^1_s)-q^2(A^2_s)|\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}ds\right]\\ &\le \varepsilon\mathbb{E}\left[\int_0^t\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}ds\right]\\ &\le \varepsilon\int_0^t\mathbb{E}\left[\mathbf{1}_{\forall r\le s, A^1_r=A^2_r}\right]ds. \end{aligned}$$

Yet, the event $\{\forall r\le s, A^1_r=A^2_r\}$ is exactly the event $\{(N^1\Delta N^2)\cap[0,s)=\emptyset\}$. Hence, the inequality above writes as $p(t)\le\varepsilon\int_0^t 1-p(s)ds$. In particular, it implies that $p'(t)\le\varepsilon(1-p(t))$.

What follows is inspired from the proof of Grönwall lemma (in its differentiable form). Let us define $v(t)=e^{\varepsilon t}(1-p(t))$. The function v is clearly right differentiable and its right derivative is $v'(t)=e^{\varepsilon t}(\varepsilon(1-p(t))-p'(t))$. Hence, the inequality above implies that $v'(t)\ge 0$. Like for the usual derivative, the fact that $v'(t)\ge 0$ implies that v is non decreasing.

Since $v(0)=1$, we get that $v(t)\ge 1$, which exactly means that $p(t)\le 1-e^{-\varepsilon t}$.

Exercise 10 (Lebesgue-Stieltjes integral)**.** No correction given yet.

Exercise 11 (Optimal stopping time)**.** No correction given yet.