1 Stochastic processes

Definition 1. A stochastic process $X = \{X(t) : t \in T\}$ is a collection of random variables defined over a common probability space $(\Omega, \mathcal{F})$. The random variables $\{X(t) : t \in T\}$ are indexed by a set $T$ that we call the index set.

$t \in T$ is often interpreted as time and $X(t)$ as the state of the process at time $t$. If the index set $T$ is countable we say that $X$ is a discrete-time stochastic process, and if $T$ is a continuum (e.g., $\mathbb{R}$ or an interval $[a, b]$), we say that $X$ is a continuous-time stochastic process.

Example 1. If $\{X_n\}_{n \in \mathbb{N}}$ is a sequence of random variables, then $X = \{X_n : n \in \mathbb{N}\}$ is a discrete-time stochastic process.

Example 2. Let $\{T_i\}_{0 \leq i < n}$ be a collection of $n$ random variables taking values in $\mathbb{R}$. For each $t \in \mathbb{R}$ define $X(t) = |\{0 \leq i < n : T_i \leq t\}| = \text{number of random variables } T_i \text{ having a value less than } t$.

$X = \{X(t) : t \in \mathbb{R}\}$ is a continuous-time stochastic process indexed by $\mathbb{R}$. One can show that for every $t \leq t'$, we have $X(t) \leq X(t')$ with probability 1.

A stochastic process $X = \{X(t) : t \in T\}$ induces a random function $X : T \to \mathbb{R}$ from the index set $T$ to $\mathbb{R}$. In other words, the stochastic process $X$ can be thought of as a random entity taking values in the space of functions $T \to \mathbb{R}$.

Sometimes we would like to say something about the “global behavior” of this random function. For instance, in Example 2 it is obvious that all the possible realizations of $\{X(t) : t \in \mathbb{R}\}$ satisfy $\lim_{t \to -\infty} X(t) = 0$, $\lim_{t \to \infty} X(t) = n$ and $X(t)$ is piecewise constant. Therefore, we should be able to say that $\mathbb{P}\left( \big\{ \lim_{t \to -\infty} X(t) = 0 \big\} \right) = 1$ and $\mathbb{P}(\{X(t) \text{ is piecewise constant}\}) = 1$.

However, subsets of $\Omega$ like $\{\omega \in \Omega : \lim_{t \to -\infty} X(t)(\omega) = 0\}$, $\{\omega \in \Omega : X(t)(\omega) \text{ is continuous in } t\}$ and $\{\omega \in \Omega : X(t)(\omega) \text{ is piecewise constant in } t\}$ might not belong to $\mathcal{F}$ in general because they involve uncountably many indices $t \in T$.

Therefore, before trying to compute the probability that $X$ has a certain “global property” like the ones mentioned above, we have to make sure that this property is measurable. Otherwise, the probability of such an “event” will not be well defined.

Note that in this lecture we will only consider very simple stochastic processes, so we will not worry a lot about these formal difficulties.

2 The Poisson process

The Poisson process is one of the widely used models in science and engineering. But before defining the Poisson process, let us recall the memoryless property of the exponential distribution.
Definition 2. A positive random variable $X$ is said to be memoryless if it satisfies:

- $\mathbb{P}(\{X > 0\}) > 0$.
- For every $s, t \geq 0$ we have
  \[
  \mathbb{P}(\{X \geq s + t\}) = \mathbb{P}(\{X \geq s\}) \mathbb{P}(\{X \geq t\}).
  \]

In other words, if we condition on $\{X \geq s\}$, then the distribution of $X - s$ has exactly the same distribution as the original unconditioned distribution of $X$. So if $X$ represents the time of an event, then no matter how long we waited for the event to happen, the remaining time has always the same distribution.

Note that we added the condition $\mathbb{P}(\{X > 0\}) > 0$ to avoid the trivial case where $X = 0$ with probability 1.

Proposition 1. A positive random variable $X$ is memoryless if and only if it follows an exponential distribution $\mathcal{E}(\lambda)$ for some $\lambda > 0$.

Proof. Let $X \sim \mathcal{E}(\lambda)$. We have $\mathbb{P}(\{X > 0\}) = 1 > 0$. Moreover, for every $s, t > 0$ we have:

\[
\mathbb{P}(\{X \geq s + t\}|\{X \geq s\}) = \frac{\mathbb{P}(\{X \geq s + t\})}{\mathbb{P}(\{X \geq s\})} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X \geq t).
\]

Conversely, suppose that $X$ is a positive memoryless random variable. Let us first show that $\mathbb{P}(\{X \geq 1\}) > 0$.

Since $\mathbb{P}(\{X \leq 0\}) < 1$ and since the cdf is continuous on the right, there exists $\alpha > 0$ such that $\mathbb{P}(\{X \leq \alpha\}) < 1$. Let $k > 0$ be such that $\frac{1}{k} < \alpha$. We have:

\[
\mathbb{P}\left(\left\{ X \geq \frac{1}{k} \right\}\right) \geq \mathbb{P}(\{X > \alpha\}) > 0.
\]

Now by the memoryless property we have $\mathbb{P}(\{X \geq 1\}) = \mathbb{P}(\{X \geq \frac{1}{k}\})^k > 0$.

Let $\lambda = -\ln \mathbb{P}(\{X \geq 1\})$ so that $\mathbb{P}(\{X \geq 1\}) = e^{-\lambda}$. We will first show that $\mathbb{P}(\{X \geq t\}) = e^{-\lambda t}$ for every $t \in \mathbb{Q}^+$.

Let $x = \frac{n}{m} \in \mathbb{Q}^+$. We have:

- $\mathbb{P}(\{X \geq n\}) = \mathbb{P}(\{X \geq x\})^m$.
- $\mathbb{P}(\{X \geq n\}) = \mathbb{P}(\{X \geq 1\})^n = e^{-\lambda}$.

Therefore, $\mathbb{P}(\{X \geq x\}) = e^{-\frac{n}{m}\lambda} = e^{-\lambda x}$.

Now let $t \in \mathbb{R}^+$. We have:

\[
\mathbb{P}(\{X \geq t\}) \leq \inf_{x < t} \mathbb{P}(\{X \geq x\}) = \inf_{x < t} e^{-\lambda x} = e^{-\lambda t},
\]

and

\[
\mathbb{P}(\{X \geq t\}) \geq \sup_{x > t} \mathbb{P}(\{X \geq x\}) = \sup_{x > t} e^{-\lambda x} = e^{-\lambda t}.
\]

Therefore, $\mathbb{P}(\{X \geq t\}) = e^{-\lambda t}$ for every $t \in \mathbb{R}^+$ and so $X \sim \mathcal{E}(\lambda)$. \hfill \square
Definition 3. Let \( \{N(t) : t \geq 0\} \) be a stochastic process indexed by \( \mathbb{R}^+ \). \( \{N(t) : t \geq 0\} \) is said to be a counting process if it satisfies the following properties:

- \( N(t) \) takes values in \( \mathbb{N} \).
- If \( s < t \), then \( N(s) \leq N(t) \).
- Every realization of \( N(t) \) is continuous on the right.

Remark 1. One can think of a counting process as representing “events” that happen at random times: \( N(t) \) is the number of “events” that has happened until time \( t \), or more precisely, in the interval \([0, t]\). \( N(t + s) - N(t) \) is the number of “events” that happened between time \( t \) and time \( t + s \), or more precisely, in the interval \((t, t + s)\).

Definition 4. We say that a counting process \( \{N(t) : t \geq 0\} \) has independent increments if for every \( 0 < t_1 < \ldots < t_k, \ N(t_1), \ N(t_2) - N(t_1), \ldots, \ N(t_k) - N(t_{k-1}) \) are independent (i.e., the number of events occurring in disjoint intervals are independent).

We say that a counting process \( \{N(t) : t \geq 0\} \) has stationary increments if \( N(t_2 + s) - N(t_1 + s) \) has the same distribution as \( N(t_2) - N(t_1) \) for every \( 0 \leq t_1 \leq t_2 \) and every \( s \geq 0 \), (i.e., the distribution of the number of events occurring in any interval of time depends only on the length of the interval).

Definition 5. Let \( \{N(t) : t \geq 0\} \) be a counting process. For every \( n \in \mathbb{N} \), the waiting time until the \( n \)th event is defined as:
\[
S_n = \inf\{t \geq 0 : N(t) > n\}.
\]
For every \( n \in \mathbb{N}^* \), the \( n \)th interarrival time is defined as:
\[
X_n = S_n - S_{n-1}.
\]
We have the following:

- \( S_0 = 0 \).
- \( S_n = \sum_{i=1}^{n} X_i \) for \( n > 0 \).
- \( N(t) = \max\{n : S_n \leq t\} \). This equation shows that we can define a counting process using its interarrival times.

Remark 2. For every \( n \in \mathbb{N} \) and every \( t \geq 0 \), we have
\[
S_n \leq t \iff N(t) \geq n.
\]
Therefore, for every \( n_1, \ldots, n_k \in \mathbb{N} \) and every \( t_1, \ldots, t_k \geq 0 \), we have:
\[
\{S_{n_1} \leq t_1, \ldots, S_{n_k} \leq t_k\} = \{N(t_1) \geq n_1, \ldots, N(t_k) \geq n_k\}.
\]
Hence,
\[
\mathbb{P}(\{S_{n_1} \leq t_1, \ldots, S_{n_k} \leq t_k\}) = \mathbb{P}(\{N(t_1) \geq n_1, \ldots, N(t_k) \geq n_k\}).
\]
This means that the distribution of \( \{N(t) : t \geq 0\} \) can be determined by the distribution of \( \{S_n : n \in \mathbb{N}\} \) and vice versa.

Moreover, since \( S_n = \sum_{i=1}^{n} X_i \) and \( X_n = S_n - S_{n-1} \), the distribution of \( \{N(t) : t \geq 0\} \) can be determined by the distribution of \( \{X_n : n \in \mathbb{N}^*\} \) and vice versa.
Now we are in position to define the Poisson process:

**Definition 6.** A Poisson process of rate \( \lambda > 0 \) is a counting process \( \{ N(t) : t \geq 0 \} \) with the following properties:

- \( N(0) = 0 \).
- The process has independent increments.
- For every \( s, t \geq 0 \), the distribution of \( N(s + t) - N(s) \) is \( \mathcal{P}(\lambda t) \). I.e.,
  \[
  \mathbb{P}(\{ N(s + t) - N(s) = n \}) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.
  \]

Note that the third condition implies that the process has stationary increments.

A Poisson process is memoryless in the sense that what happens after any instant \( t \) is completely independent of what happened before it. So we should expect the distribution of interarrival times to be exponential since it is the only memoryless distribution.

**Theorem 1.** A counting process is a Poisson process of rate \( \lambda > 0 \) if and only if the sequence of interarrival times is i.i.d. \( \mathcal{E}(\lambda) \).

In the light of Remark 2, we only need to prove one direction. In the sequel, we will prove the sufficiency of the condition because it is easier and simpler. If you are curious about how one can directly prove the necessity part, please refer to the Appendix.

We need two lemmas:

**Lemma 1.** Let \( \{ N(t) : t \geq 0 \} \) be a counting process such that the sequence of interarrival times is i.i.d. \( \mathcal{E}(\lambda) \). Then \( N(t) \sim \mathcal{P}(\lambda t) \).

**Proof.** For \( n = 0 \) we have:

\[
\mathbb{P}(\{ N(t) = 0 \}) = \mathbb{P}(\{ X_1 > t \}) = e^{-\lambda t}.
\]

Now let \( n > 0 \) and suppose that:

\[
\mathbb{P}(\{ N(t) = n - 1 \}) = \mathbb{P}(\{ S_{n-1} \leq t, S_n > t \}) = e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.
\]

Define \( S'_{n-1} = \sum_{i=2}^{n} X_i \) and \( S'_n = \sum_{i=2}^{n+1} X_i \). We have \( S_n = X_1 + S'_{n-1} \) and \( S_{n+1} = X_1 + S'_n \).

Moreover, since \( (S_{n-1}, S_n) \) and \( (S'_{n-1}, S'_n) \) have the same distribution, it follows that

\[
\mathbb{P}(\{ S'_{n-1} \leq t, S'_n > t \}) = e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.
\]

Therefore,

\[
\mathbb{P}(\{ N(t) = n \}) = \mathbb{P}(\{ S_n \leq t, S_{n+1} > t \})
= \int_0^t \mathbb{P}(\{ S_n \leq t, S_{n+1} > t \} \mid \{ X_1 = x_1 \}) \lambda e^{-\lambda x_1} dx_1
= \int_0^t \mathbb{P}(\{ S'_{n-1} \leq t - x_1, S'_n > t - x_1 \} \mid \{ X_1 = x_1 \}) \lambda e^{-\lambda x_1} dx_1
= \int_0^t e^{-\lambda(t-x_1)} \frac{\lambda^{n-1}(t-x_1)^{n-1}}{(n-1)!} \lambda e^{-\lambda x_1} dx_1 = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.
\]
It remains to show that \( \{N(t) : t \geq 0\} \) has independent and stationary increments.

**Lemma 2.** Let \( N(t) \) be a counting process such that the sequence of interarrival times is i.i.d. \( \mathcal{E}(\lambda) \). For every \( s, t > 0 \), the distribution of \( N(t+s) - N(s) \) is \( \mathcal{P}(\lambda t) \). Moreover, \( \{N(t+s) - N(s) : t > 0\} \) is independent of \( \{N(s') : s' \leq s\} \)

**Proof.** Fix \( s > 0 \). Note that the process \( N(t) = N(t+s) - N(s) \) is a counting process.

For every \( k \geq 0 \) define \( W_k = \inf\{t > 0 : N(s+t) \geq N(s)+k\} = \inf\{t > 0 : N'(t) \geq k\} \). This is the amount of time we wait after \( s \) until the next \( k \)th event. Now for every \( k > 0 \) let \( Z_k = W_k - W_{k-1} \) be the \( k \)th interarrival time of \( \{N(t) : t \geq 0\} \).

Let \( n > 0, 0 < s_1 < \ldots < s_n < s \) and \( t_1, \ldots, t_k > 0 \). We have:

\[
\mathbb{P}(\{Z_1 > t_1, \ldots, Z_k > t_k\}|\{N(s) = n, S_1 = s_1, \ldots, S_n = s_n\})
= \mathbb{P}(\{X_{n+1} > t_1 + s - s_n, X_{n+2} > t_2, \ldots, X_{n+k} > t_k\}|\{X_{n+1} > s - s_n, S_n = s_n, \ldots, S_1 = s_1\})
\overset{(a)}{=} \mathbb{P}(\{X_{n+1} > t + s - s_n, X_{n+2} > t_2, \ldots, X_{n+k} > t_k\}|\{X_{n+1} > s - s_n\})
\overset{(b)}{=} \mathbb{P}(\{X_{n+1} > t_1 + s - s_n\}|\{X_{n+1} > s - s_n\}) \mathbb{P}(\{X_{n+2} > t_2, \ldots, X_{n+k} > t_k\})
\overset{(c)}{=} \prod_{i=1}^{k} \mathbb{P}(\{X_{n+i} > t_i\}) = \prod_{i=1}^{k} e^{-\lambda t_i},
\]

where (a) holds because \( \{X_{n+i}, 1 \leq i \leq k\} \) is independent of \( \{S_i, 1 \leq i \leq n\} \). (b) is true because \( X_{n+1} \) is independent of \( \{X_{n+i}, 2 \leq i \leq k\} \). (c) follows from the memoryless property of the exponential distribution.

We conclude that \( \{Z_k\}_{k \geq 1} \) is i.i.d. \( \mathcal{E}(\lambda) \) and independent of \( \{N(s') : s' \leq s\} \).

Now since the sequence of interarrival times of \( N'(t) = N(t+s) - N(s) \) is i.i.d. \( \mathcal{E}(\lambda) \), it follows from Lemma 1 that \( N(t+s) - N(s) \sim \mathcal{P}(\lambda t) \). Moreover, since \( \{Z_k\}_{k \geq 1} \) is independent of \( \{N(s') : s' \leq s\} \), we deduce that \( \{N(t+s) - N(s) : t \geq 0\} \) is independent of \( \{N(s') : s' \leq s\} \).

The sufficient condition part of Theorem 1 is then a direct corollary of Lemma 2.

### 3 Renewal theory

Renewal processes generalize Poisson processes: They are counting processes with independent and identically distributed interarrival times but the distribution is now arbitrary.

**Definition 7.** Let \( \{X_n : n \in \mathbb{N}^*\} \) be a sequence of i.i.d positive random variables such that \( \mathbb{P}(\{X_1 = 0\}) < 1 \). Let \( S_0 = 0 \) and \( S_n = \sum_{i=1}^{n} X_i \) for \( n > 0 \). Let \( F \) be the cdf of \( X_1 \), and for every \( n \in \mathbb{N}^* \) let \( F_n \) be the cdf of \( S_n \).

The renewal process with interarrival times \( \{X_n : n \in \mathbb{N}^*\} \) is the counting process \( \{N(t) : t \geq 0\} \) defined as

\[
N(t) = \max\{n : S_n \leq t\}.
\]
The condition \( P(\{X_1 = 0\}) < 1 \) was added to avoid the trivial case where all events happen at \( t = 0 \). One question that might be asked is whether it is possible to have an infinite number of events that occur in a finite interval of time. The condition \( P(\{X_1 = 0\}) < 1 \) can be used to show that almost surely, this does not happen:

**Proposition 2.** With probability 1, the number of events occurring inside any finite interval of time is finite.

**Proof.** For every \( n \geq 1 \), let \( X'_n = \min\{X_n, 1\} \) and \( S'_n = \sum_{i=1}^{n} X'_i \).

Since \( X'_1 \geq 0 \) and \( P(\{X'_1 = 0\}) = P(\{X_1 = 0\}) < 1 \), we have \( \mu' = \mathbb{E}(X'_1) > 0 \). Fix \( t > 0 \) and let \( n > \left\lceil \frac{2t}{\mu'} \right\rceil \). We have:

\[
P(\{N(t) = \infty\}) \leq P(\{N(t) > n\}) = P(\{S_n < t\}) \leq P(\{S'_n < t\}) = P\left(\left\{ \frac{S'_n}{n} < \frac{t}{n} \right\}\right)
\leq P\left(\left\{ \frac{S'_n}{n} < \frac{\mu'}{2} \right\}\right) = P\left(\left\{ \frac{\mu'}{n} - \frac{S'_n}{n} > \frac{\mu'}{2} \right\}\right) \xrightarrow{S.L.L.N.} 0.
\]

\( \square \)

**Proposition 3.** The distribution of \( N(t) \) can be obtained as follows:

\[
P(\{N(t) = n\}) = F_n(t) - F_{n+1}(t).
\]

**Proof.**

\[
P(\{N(t) = n\}) = P(\{N(t) \geq n\}) - P(\{N(t) \geq n + 1\}) = P(\{S_n \leq t\}) - P(\{S_{n+1} \leq t\}) = F_n(t) - F_{n+1}(t).
\]

\( \square \)

**Proposition 4.** All the moments of \( N(t) \) are finite for every \( t \geq 0 \).

The first moment \( m(t) = \mathbb{E}(N(t)) \) of \( N(t) \) — which is the expected number of events occurring before time \( t \) — is particularly important and it is called the renewal function of the process. We have:

\[
m(t) = \sum_{n=1}^{\infty} F_n(t).
\]

**Proof.** Since \( F_1(0) = P(\{X_1 = 0\}) < 1 \) and since the \( F_1 \) is continuous on the right, there exists \( \alpha > 0 \) such that \( F_1(\alpha) < 1 \). Therefore, \( P(\{X_1 \geq \alpha\}) \geq P(\{X_1 > \alpha\}) = 1 - F_1(\alpha) > 0 \).

For every \( n \geq 1 \) define:

\[
\tilde{X}_n = \begin{cases} 
0 & \text{if } X_n < \alpha \\
\alpha, & \text{if } X_n \geq \alpha,
\end{cases}
\]

and let \( \tilde{S}_n = \sum_{i=1}^{n} \tilde{X}_n \). Define \( \tilde{N}(t) = \max\{n : \tilde{S}_n \leq t\} \). Now since \( \tilde{S}_n \leq S_n \) for all \( n \), we have \( \tilde{N}(t) \geq N(t) \) for every \( t \). Therefore, for every \( k > 0 \) we have \( \mathbb{E}(N^k(t)) \leq \mathbb{E}(\tilde{N}^k(t)) \).

For every \( k > 0 \), let \( A_k = \tilde{N}(k\alpha) - \tilde{N}((k-1)\alpha) \). Observe that \( \{A_k\}_{k \geq 0} \) are i.i.d. geometric random variables with mean \( \mathbb{E}(A_1) = \frac{1}{P(\{X_1 \leq \alpha\})} \). Therefore, \( \tilde{N}(k\alpha) = \sum_{i=1}^{k} A_i \) is the sum of \( k \) i.i.d. geometric random
variables with finite mean. This shows that all the moments of $\bar{N}(k\alpha)$ are finite. We conclude that for every $t > 0$ and every $r > 0$ we have:

$$\mathbb{E}(N^r(t)) \leq \mathbb{E}(\bar{N}^r(t)) \leq \mathbb{E}(\bar{N}^r(t)) < \infty.$$ 

Let us now compute the first moment of $N(t)$. We have $N(t) = \sum_{n=1}^{\infty} 1_{\{S_n \leq t\}}$. Therefore,

$$m(t) = \mathbb{E}(N(t)) = \mathbb{E}\left(\sum_{n=1}^{\infty} 1_{\{S_n \leq t\}}\right) = \sum_{n=1}^{\infty} \mathbb{P}(\{S_n \leq t\}) = \sum_{n=1}^{\infty} F_n(t),$$

where (a) is valid because $1_{\{S_n \leq t\}}$ is positive.

**Lemma 3.** $\mathbb{P}\left(\left\{ \lim_{t \to \infty} N(t) = \infty \right\}\right) = 1.$

**Proof.** For every $n > 0$, we have:

$$\mathbb{P}\left(\left\{ \lim_{t \to \infty} N(t) \geq n \right\}\right) = \mathbb{P}(\{\exists t > 0, N(t) \geq n\}) = \mathbb{P}(\{\exists k \in \mathbb{N}, N(k) \geq n\}) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} \{N(k) \geq n\}\right)$$

$$= \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} \{S_n \leq k\}\right) = \lim_{k \to \infty} \mathbb{P}(\{S_n \leq k\}) = \lim_{k \to \infty} F_n(k) = 1.$$ 

Since this is true for every $n > 0$, we conclude that

$$\mathbb{P}\left(\left\{ \lim_{t \to \infty} N(t) = \infty \right\}\right) = \mathbb{P}\left(\bigcap_{n>0} \left\{ \lim_{t \to \infty} N(t) \geq n \right\}\right) = 1.$$ 

**Proposition 5.** We have

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu} \text{ almost surely.}$$

**Proof.** Notice that $S_{N(t)} \leq t < S_{N(t)+1}$. Therefore,

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)}.$$ 

From Lemma 3 we have $\lim_{t \to \infty} N(t) = \infty$ with probability 1. On the other hand, the strong law of large numbers implies that both $\frac{S_n}{n}$ and $\frac{S_{n+1}}{n}$ converge to $\mu$ with probability 1. We conclude that $\frac{t}{N(t)}$ converges almost surely to $\mu$. Hence,

$$\mathbb{P}\left(\left\{ \lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu} \right\}\right) = 1.$$ 


Proposition 5 shows that \( \frac{N(t)}{t} \) converges almost surely to \( \frac{1}{\mu} \). One can ask whether \( \frac{E(N(t))}{t} \) converges to \( \frac{1}{\mu} \) as well. The next theorem answers this question in the positive:

**Theorem 2. (The elementary renewal theorem)** \( \lim_{t \to \infty} \frac{m(t)}{t} = \frac{1}{\mu} \).

In order to prove Theorem 2, we need *Wald’s equation*:

**Theorem 3. (Wald’s equation)** Let \( \{X_n\}_{n \geq 1} \) be i.i.d. random variables with \( E(|X_1|) < \infty \). Let \( T \) be a stopping time with respect to \( \{X_n\}_{n \geq 1} \) (i.e., \( \{T = n\} \) is \( \sigma\{X_1, \ldots, X_n\}\)-measurable). If \( E(T) < \infty \) then

\[
E(S_T) = E \left( \sum_{n=1}^T X_n \right) = E(T)E(X_1).
\]

**Proof.** We have:

\[
S_T = \sum_{n=1}^T X_n = \sum_{n=1}^\infty X_n \cdot 1_{\{T \geq n\}}.
\]

On the other hand, for every \( n > 0 \) we have:

\[
\{T \geq n\} = \{T < n\}^c = \bigcap_{1 \leq i < n} \{T = i\}^c \in \sigma(\{X_1, \ldots, X_{n-1}\}).
\]

Now since \( X_n \) is independent of \( \{X_1, \ldots, X_{n-1}\} \), we conclude that \( X_n \) is independent of \( 1_{\{T \geq n\}} \).

Therefore,

\[
E(|S_T|) \leq \sum_{n=1}^\infty E(|X_n| \cdot 1_{\{T \geq n\}}) \overset{(a)}{=} \sum_{n=1}^\infty E(|X_n|)E(1_{\{T \geq n\}}) = \sum_{n=1}^\infty E(|X_1|)P(\{T \geq n\}) = E(|X_1|)E(T) < \infty,
\]

where (a) follows from the independence of \( X_n \) and \( 1_{\{T \geq n\}} \). We conclude that \( S_T \) has a well defined expectation which can now be computed as follows:

\[
E(S_T) = E \left( \sum_{n=1}^\infty X_n \cdot 1_{\{T \geq n\}} \right) \overset{(a)}{=} \sum_{n=1}^\infty E(X_n \cdot 1_{\{T \geq n\}}) \overset{(b)}{=} E(X_1)E(T),
\]

where (a) is justified by Lebesgue dominated convergence and (b) follows by the same reasoning as before. \( \Box \)

**Corollary 1.** Let \( \{X_n\}_{n \geq 1} \) be the interarrival times of a renewal process \( \{N(t) : t \geq 0\} \). If \( \mu = E(X_1) < \infty \) then

\[
E(S_{N(t)+1}) = \mu \cdot (m(t) + 1).
\]
Proof. Let \( T = N(t) + 1 \). For every \( n > 0 \), we have:

\[
\{ T = n \} = \{ N(t) + 1 = n \} = \{ N(t) = n - 1 \} = \{ S_{n-1} \leq t, \ S_n > t \} \in \sigma(\{X_1, \ldots, X_n\}).
\]

Therefore, \( T \) is a stopping time. From Wald’s equation we get:

\[
\mathbb{E}(S_{N(t)+1}) = \mathbb{E}(X_1)\mathbb{E}(N(t) + 1) = \mu(m(t) + 1).
\]

Now we are ready to prove the elementary renewal theorem:

Proof of Theorem 2. Suppose that \( \mu < \infty \). By taking expectations and applying Corollary 1 to \( S_{N(t)+1} > t \) we get \((m(t) + 1)\mu > t\) which implies that

\[
\liminf_{t \to \infty} \frac{m(t)}{t} \geq \frac{1}{\mu}.
\]

Now fix \( M > 0 \) and let

\[
\tilde{X}_n = \min\{X_n, M\}.
\]

Let \( \tilde{S}_n = \sum_{i=1}^{n} \tilde{X}_i, \ \tilde{N}(t) = \max\{n : \tilde{S}_n \leq t\}, \ \mu_M = \mathbb{E}(\tilde{X}_n) \) and \( \tilde{m}(t) = \mathbb{E}(\tilde{N}(t)) \).

Notice that \( \tilde{S}_{N(t)+1} \leq t + X_{N(t)+1} \leq t + M \) and so by Corollary 1 we have \( \mu_M(\tilde{m}(t) + 1) \leq t + M \).

Therefore,

\[
\limsup_{t \to \infty} \frac{\tilde{m}(t)}{t} \leq \frac{1}{\mu_M}.
\]

On the other hand, since \( \tilde{X}_n \leq X_n \) for every \( n > 0 \), we have \( \tilde{S}_n \leq S_n \) for every \( n > 0 \) which implies that \( \tilde{N}(t) \geq N(t) \) and so \( \tilde{m}(t) \geq m(t) \). Hence,

\[
\limsup_{t \to \infty} \frac{m(t)}{t} \leq \limsup_{t \to \infty} \frac{\tilde{m}(t)}{t} \leq \frac{1}{\mu_M}.
\]

But this is true for every \( M > 0 \). Therefore,

\[
\limsup_{t \to \infty} \frac{m(t)}{t} \leq \lim_{M \to \infty} \frac{1}{\mu_M} = \lim_{M \to \infty} \frac{1}{\mathbb{E}(X_1) \cdot 1_{\{X_1 \leq M\}}} = \frac{1}{\mathbb{E}(X_1)} = \frac{1}{\mu}.
\]

We conclude that

\[
\lim_{t \to \infty} \frac{m(t)}{t} = \frac{1}{\mu}.
\]

Now suppose that \( \mu = \infty \). By using the truncation (1) and exactly the same reasoning as above we get:

\[
\limsup_{t \to \infty} \frac{m(t)}{t} \leq \lim_{M \to \infty} \frac{1}{\mathbb{E}(X_1 \cdot 1_{\{X_1 \leq M\}})}.
\]

But \( \lim_{M \to \infty} \mathbb{E}(X_1 \cdot 1_{\{X_1 \leq M\}}) = \mathbb{E}(X_1) = \mu = \infty \), so \( \lim_{t \to \infty} \frac{m(t)}{t} = 0 \). 

\[9\]
Theorem 4. Let $\sigma^2$ be the variance of an interarrival time. If $\mu$ and $\sigma^2$ are finite then
\[
\lim_{t \to \infty} P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right) = \frac{1}{2\pi} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} \, dx.
\]

Proof. Let $r_t = t/\mu + y \sigma \sqrt{t/\mu^3}$ and $n_t = \lfloor r_t \rfloor$. Clearly $\lim_{t \to \infty} n_t = \infty$. We have:
\[
P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right) = P\{N(t) < r_t\}.
\]
Now since $n_t \leq r_t < n_t + 1$, we have:
\[
P\{N(t) < n_t\} \leq P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right) \leq P\{N(t) < n_t + 1\}.
\]
Hence,
\[
P\{S_{nt} > t\} \leq P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right) \leq P\{S_{nt+1} > t\}.
\]
and so
\[
P\left(\left\{ \frac{S_{nt} - n_t \mu}{\sigma \sqrt{n_t}} > \frac{t - n_t \mu}{\sigma \sqrt{n_t}} \right\} \right) \leq P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right)
\]
\[
\leq P\left(\left\{ \frac{S_{nt+1} - (n_t + 1) \mu}{\sigma \sqrt{n_t + 1}} > \frac{t - (n_t + 1) \mu}{\sigma \sqrt{n_t + 1}} \right\} \right).
\]
By the central limit theorem $\frac{S_n - n\mu}{\sigma \sqrt{n}}$ converges in distribution to $N(0, 1)$. Moreover, one can show that
\[
\lim_{t \to \infty} \frac{t - n_t \mu}{\sigma \sqrt{n_t}} = \lim_{t \to \infty} \frac{t - (n_t + 1) \mu}{\sigma \sqrt{n_t + 1}} = -y.
\]
Therefore,
\[
\lim_{t \to \infty} P\left(\left\{ \frac{N(t) - t/\mu}{\sigma \sqrt{t/\mu^3}} < y \right\} \right) = P\{N(0, 1) > -y\} = \frac{1}{2\pi} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} \, dx = \frac{1}{2\pi} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} \, dx.
\]
A  Direct proof of the necessary condition part of Theorem 1

We need here a few lemmas:

**Lemma 4.** Let $N(t)$ be a Poisson process of rate $\lambda > 0$. For every $n > 1$, there exists a mapping $g_n : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that:

- $\mathbb{P}(\{N(a) = n - 2, N(b) = n - 1\}|\{N(a) < n - 1, N(b) \geq n - 1\}) \geq g_n(a, b - a)$.
- For every $a > 0$, $\lim_{\epsilon \to 0} g_{a,n}(\epsilon) = 1$.
- If $0 < a' \leq a \leq a''$, then $g_n(a, \epsilon) \geq \min\{g_n(a', \epsilon), g_n(a'', \epsilon)\}$.

**Proof.** Let $0 < a < b$ and $\epsilon = b - a$. We have:

\[
\mathbb{P}(\{N(a) = n - 2, N(a + \epsilon) = n - 1\}) = \mathbb{P}(\{N(a) = n - 2, N(a + \epsilon) - N(a) = 1\})
= e^{-\lambda a} \frac{(\lambda a)^{n-2}}{(n-2)!} \cdot e^{-\lambda \epsilon} \cdot \epsilon = f_n(a)e^{-\lambda \epsilon} \epsilon,
\]

where, $f_n(a) = e^{-\lambda a} \frac{(\lambda a)^{n-2}}{(n-2)!}$. Note that the function $f_n$ is increasing on $[0, \frac{n-2}{\lambda}]$ and then decreasing on $[\frac{n-2}{\lambda}, \infty)$. This implies that for every $0 < a' \leq a \leq a''$, we have $f_n(a) \geq \min\{f_n(a'), f_n(a'')\}$.

On the other hand, we have

\[
\mathbb{P}(\{N(a) < n - 1, N(a + \epsilon) \geq n - 1\}) \\
\leq \mathbb{P}(\{N(a) = n - 2, N(a + \epsilon) = n - 1\}) + \mathbb{P}(\{N(a + \epsilon) - N(a) \geq 2\}) \\
= f_n(a)e^{-\lambda \epsilon} + (1 - e^{-\lambda \epsilon}(1 + \epsilon)).
\]

Therefore,

\[
\mathbb{P}(\{N(a) = n - 2, N(b) = n - 1\}|\{N(a) < n - 1, N(b) \geq n - 1\}) \\
= \frac{\mathbb{P}(\{N(a) = n - 2, N(a + \epsilon) = n - 1\})}{\mathbb{P}(\{N(a) < n - 1, N(a + \epsilon) \geq n - 1\})} \\
\geq \frac{f_n(a)e^{-\lambda \epsilon}}{f_n(a)e^{-\lambda \epsilon} + (1 - e^{-\lambda \epsilon}(1 + \epsilon))} \\
= \frac{1}{1 + \frac{1}{f_n(a)} \frac{e^{-\lambda \epsilon} - (1 + \epsilon)}{\epsilon}} = g_n(a, \epsilon).
\]

It is easy now to check that $g_n$ satisfies the conditions of the lemma. \hfill \square

Now we are in position to prove the necessary condition part of Theorem 1.

**Proof.** Let $N(t)$ be a Poisson process of rate $\lambda > 0$. Fix $n \geq 1$, and recall that $X_n = S_n - S_{n-1}$ is the $n^\text{th}$ interarrival time. For $n = 1$ we have $\mathbb{P}(\{X_1 > t\}) = \mathbb{P}(\{N(t) = 0\}) = e^{-\lambda t}$ and so $X_1 \sim \mathcal{E}(\lambda)$.\hfill 11
Now suppose \( n > 1 \). Let \( t > 0 \), \( 0 < a < b \) and suppose that \( b - a < t \). We have:

\[
\mathbb{P}(\{X_n > t\} \mid \{a < S_{n-1} \leq b\}) \leq \mathbb{P}(\{S_{n-1} + X_n > t + a\} \mid \{a < S_{n-1} \leq b\})
\]
\[
= \mathbb{P}(\{S_n > t + a\} \mid \{a < S_{n-1} \leq b\})
\]
\[
= \mathbb{P}(\{N(t + a) < n\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\leq \mathbb{P}(\{N(t + a) - N(b) < 1\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\leq \mathbb{P}(\{N(t + a) - N(b) = 0\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
= e^{-\lambda(t-a)}.
\]

On the other hand,

\[
\mathbb{P}(\{X_n > t\} \mid \{a < S_{n-1} \leq b\}) \geq \mathbb{P}(\{S_{n-1} + X_n > t + b\} \mid \{a < S_{n-1} \leq b\})
\]
\[
= \mathbb{P}(\{S_n > t + b\} \mid \{a < S_{n-1} \leq b\})
\]
\[
= \mathbb{P}(\{N(t + b) < n\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\geq \mathbb{P}(\{N(t + b) < n, N(b) - N(a) = 1\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
= \mathbb{P}(\{N(t + b) < n\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\times \mathbb{P}(\{N(b) - N(a) = 1\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
= \mathbb{P}(\{N(t + b) < n\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\times \mathbb{P}(\{N(b) = n - 1\} \mid \{N(a) < n - 1, N(b) \geq n - 1\})
\]
\[
\geq e^{-\lambda g_n(a,b-a)},
\]

where \((a)\) follows from Lemma 4. We conclude that for every \( t > 0 \), \( 0 < a < b \) satisfying \( b - a < t \) we have:

\[
e^{-\lambda t} g_n(a,b-a) \leq \mathbb{P}(\{X_n > t\} \mid \{a < S_{n-1} \leq b\}) \leq e^{-\lambda t} e^{\lambda(b-a)}.
\]

(2)

Now fix \( 0 < a < b \). Let \( t > 0 \) and \( k > \frac{b-a}{\min\{a,t\}} \). For every \( 0 \leq i \leq k \) define \( a_i = a + i \frac{b-a}{k} \) (so \( a = a_0 < a_1 < \ldots < a_k = b \)). Note that for every \( 0 \leq i < k \) we have \( a_{i+1} - a_i = \frac{b-a}{k} < t \). Therefore, (2) implies that:

\[
e^{-\lambda t} g_n \left(a_i, \frac{b-a}{k}\right) \leq \mathbb{P}(\{X_n > t\} \mid \{a_i < S_{n-1} \leq a_{i+1}\}) \leq e^{-\lambda t} e^{\frac{b-a}{k}}.
\]

Now since \( a \leq a_i \leq b \), we have \( g_n(a_i, \frac{b-a}{k}) \geq h(\frac{b-a}{k}) \), where \( h(\epsilon) = \min\{g_n(a,\epsilon),g_n(b,\epsilon)\} \) for \( \epsilon > 0 \). We conclude that for every \( 0 \leq i < n \) we have:

\[
e^{-\lambda t} h \left( \frac{b-a}{k}\right) \leq \mathbb{P}(\{X_n > t\} \mid \{a_i < S_{n-1} \leq a_{i+1}\}) \leq e^{-\lambda t} e^{\frac{b-a}{k}}.
\]

12
Now since
\[ P(\{X_n > t\}|\{a < S_{n-1} \leq b\}) = \frac{\sum_{i=1}^k P(\{X_n > t\}|\{a_{i-1} < S_{n-1} \leq a_i\})P(\{a_{i-1} < S_{n-1} \leq a_i\})}{\sum_{i=1}^k P(\{a_{i-1} < S_{n-1} \leq a_i\})}, \]

It follows that
\[ e^{-\lambda t} h \left( \frac{b-a}{k} \right) \leq P(\{X_n > t\}|\{a < S_{n-1} \leq b\}) \leq e^{-\lambda t} e^{\lambda \frac{b-a}{k}}. \]

But \( k \) can be arbitrarily large and \( \lim_{\epsilon \to 0} h(\epsilon) = 1 \). We conclude that for every \( t > 0 \) and every \( 0 < a < b \) we have:
\[ P(\{X_n > t\}|\{a < S_{n-1} \leq b\}) = e^{-\lambda t}. \]

This shows that \( X_n \sim \mathcal{E}(\lambda) \) and that \( X_n \) is independent of \( S_{n-1} \).

One can show using the same technique that for every \( 2(n-1) \) positive numbers \( 0 < a_1 < b_1 < a_2 < b_2 < \ldots < a_{n-1} < b_{n-1} \) and every \( t > 0 \) we have
\[ P(\{X_n > t\}|\{a_i < S_i \leq b_i, 1 \leq i < n\}) = e^{-\lambda t}. \]

This shows that \( X_n \) is independent of \( \{S_i\}_{1 \leq i < n} \), which implies that \( X_n \) is independent of \( \{X_i\}_{1 \leq i < n} \).

We conclude that the sequence \( \{X_n\}_{n \geq 1} \) is i.i.d. \( \mathcal{E}(\lambda) \). \( \Box \)