MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Lecture 20

THE BASICS OF STOCHASTIC PROCESSES

Contents

We now turn to the study of some simple classes of stochastic processes. Examples and a more leisurely discussion of this material can be found in the corresponding chapter of [BT].

A discrete-time stochastic is a sequence of random variables $\{X_n\}$ defined on a common probability space $(\quad , \mathcal{F}, \mathbb{P})$. In more detail, a stochastic process is a function X of two variables n and ω. For every n, the function $\omega \mapsto X_n(\omega)$ is a random variable (a measurable function). An alternative perspective is provided by fixing some $\omega \in \quad$ and viewing $X_n(\omega)$ as a function of n (a "time function," or "sample path," or "trajectory").

A continuous-time stochastic process is defined similarly, as a collection of random variables $\{X_t\}$ defined on a common probability space $(\quad , \mathcal{F}, \mathbb{P})$, where t varies over non-negative real values $\mathbb{R}_+$.

1 SPACES OF TRAJECTORIES: $\mathbb{R}^\infty$ and $\mathbb{R}^{[0,\infty)}$

1.1 σ-algebras on spaces of trajectories

Recall that earlier we defined the Borel σ-algebra $\mathcal{B}^n$ on $\mathbb{R}^n$ as the smallest σ algebra containing all measurable rectangles, i.e. events of the form

$$B_1 \times \cdots \times B_n = \{\mathbf{x} \in \mathbb{R}^n : x_j \in B_j \quad \forall j \in [n]\}$$

where B_j are (1-dimensional) Borel subsets of $\mathbb{R}$. A generalization is the following:

Definition 1. *Let T be an arbitrary set of indices. The product space $\mathbb{R}^T$ is defined as*

$$\mathbb{R}^T \triangleq \prod_{t \in T} \mathbb{R} = \{(x_t, t \in T)\} .$$

A subset $\mathcal{J}_S(B)$ of $\mathbb{R}^T$ is called a cylinder with base B on time indices $S = \{s_1, \dots, s_n\}$ if

$$\mathcal{J}_S(B) = \{(x_t) : (x_{s_1}, \dots, x_{s_n}) \in B\} , \qquad B \subset \mathbb{R}^n , \tag{1}$$

with $B \in \mathcal{B}^n$. The product σ-algebra $\mathcal{B}^T$ is the smallest σ-algebra containing all cylinders:

$$\mathcal{B}^T = \sigma\{\mathcal{J}_S(B) : \forall S\text{-finite and } B \in \mathcal{B}^S\} .$$

For the special case $T = \{1, 2, \dots, \}$ the notation $\mathbb{R}^\infty$ and $\mathcal{B}^\infty$ will be used.

The following are measurable subsets of $\mathbb{R}^\infty$:

$$E_0 = \{x \in \mathbb{R}^\infty : x_n\text{–converges}\}$$

The following are measurable subsets of $\mathbb{R}^{[0,\infty)}$:

$$E_1 = \{x \in \mathbb{R}^{[0,\infty)} : x_t = 0 \quad \forall t \in \mathbb{Q}\} \tag{2}$$

$$E_2 = \{x \in \mathbb{R}^{[0,\infty)} : \sup_{t \in \mathbb{Q}} x_t > 0\} \tag{3}$$

The following are not measurable subsets of $\mathbb{R}^{[0,\infty)}$:

$$E_1' = \{x \in \mathbb{R}^{[0,\infty)} : x_t = 0 \quad \forall t\} \tag{4}$$

$$E_2' = \{x \in \mathbb{R}^{[0,\infty)} : \sup_{t} x_t > 0\} \tag{5}$$

$$E_3 = \{x \in \mathbb{R}^{[0,\infty)} : x_t\text{–continuous}\} \tag{6}$$

Non-measurability of E_1' and E_2' will follow from the next result. We mention that since $E_1 \cap E_3 = E_1' \cap E_3$, then by considering a trace of $\mathcal{B}^{[0,+\infty)}$ on E_3 sets E_1' and E_2' can be made measurable. This is a typical approach taken in the theory of continuous stochastic processes.

Proposition 1. *The following provides information about $\mathcal{B}^T$:*

(i) For every measurable set $E \in \mathcal{B}^T$ there exists a countable set of time indices $S = \{s_1, \dots\}$ and a subset $B \in \mathcal{B}^\infty$ such that

$$E = \{(x_t) : (x_{s_1}, \dots, x_{s_n}, \dots) \in B\} \tag{7}$$

(ii) *Every measurable set* $E \in \mathcal{B}^T$ *can be approximated within arbitrary* ϵ *by a cylinder:*

$$\mathbb{P}[E \triangle \mathcal{J}_S(B)] \le \epsilon\,,$$

where $\mathbb{P}$ *is any probability measure on* $(R^T, \mathcal{B}^T)$.

(iii) *If* $\{X_t, t \in T\}$ *is a collection of random variables on* $(\quad, \mathcal{F})$*, then the map*

$$X : \quad \to \mathbb{R}^T\,, \tag{8}$$

$$\omega \mapsto (X_t(\omega), t \in T) \tag{9}$$

is measurable with respect to $\mathcal{B}^T$.

Proof: For (i) simply notice that collection of sets of the form (7) contains all cylinders and closed under countable unions/intersections. To see this simply notice that one can without loss of generality assume that every set in, for example, union $F = \bigcup E_n$ correspond to the same set of indices in (7) (otherwise extend the index sets S first).

(ii) follows from the next exercise and the fact that $\{\mathcal{J}_S(B), B \in \mathcal{B}^S\}$ (under fixed finite S) form a σ-algebra. For (iii) note that it is sufficient to check that $X^{-1}(\mathcal{J}_S(B)) \in \mathcal{F}$ (since cylinders generate $\mathcal{B}^T$). The latter follows at once from the definition of a cylinder (1) and the fact that

$$\{(X_{s_1}, \dots, X_{s_n}) \in B\}$$

are clearly in $\mathcal{F}$. □

Exercise 1. Let $\mathcal{F}_\alpha$, $\alpha \in S$ be a collection of σ-algebras and let $\mathcal{F} = \bigvee_{\in S} \mathcal{F}_\alpha$ be the smallest σ-algebra containing all of them. Call set B finitary if $B \in \bigvee_{\in S_1} \mathcal{F}_\alpha$, where S_1 is a finite subset of S. Prove that every $E \in \mathcal{F}$ is finitary approximable, i.e. that for every $\epsilon > 0$ there exists a finitary B such that

$$\mathbb{P}[E \triangle B] \le \epsilon\,.$$

(Hint: Let $\mathcal{L} = \{E : E\text{–finitary approximable}\}$ and show that $\mathcal{L}$ contains the algebra of finitary sets and closed under monotone limits.)

With these preparations we are ready to give a definition of stochastic process:

Definition 2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. A stochastic process with time set* T *is a measurable map* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^T, \mathcal{B}^T)$*. The pushforward* $\mathbb{P}_X \triangleq \mathbb{P} \circ X^{-1}$ *is called the law of* X.

1.2 Probability measures on spaces of trajectories

According to Proposition 1 we may define probability measures on $\mathbb{R}^T$ by simply computing an induced measure along a map (9). An alternative way to define probabilities on $\mathbb{R}^T$ is via the following construction.

Theorem 1 (Kolmogorov)**.** *Suppose that for any finite $S \subset T$ we have a probability measure $\mathbb{P}_S$ on $\mathbb{R}^S$ and that these measures are consistent. Namely, if $S' \subset S$ then*

$$\mathbb{P}_{S'}[B] = \mathbb{P}_S[B \times \mathbb{R}^{S \setminus S'}] .$$

Then there exists a unique probability measure $\mathbb{P}$ on $\mathbb{R}^T$ such that

$$\mathbb{P}[\mathcal{J}_S(B)] = \mathbb{P}_S[B]$$

for every cylinder $\mathcal{J}_S(B)$.

Proof (optional): As a simple exercise, reader is encouraged to show that it suffices to consider the case of countable T (cf. Proposition 1.(i)). We thus focus on constructing a measure on $\mathbb{R}^\infty$. Let $\mathcal{A} = \bigcup_{n\geq 1} \mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-algebra of all cylinders with time indices $\{1, \ldots, n\}$. Clearly $\mathcal{A}$ is an algebra. Define a set-function on $\mathcal{A}$ via:

$$\forall E = \{(x_1, \ldots, x_n) \in B\} : \quad \mathbb{P}[E] \triangleq \mathbb{P}_{\{1,\ldots,n\}}[B] .$$

Consistency conditions guarantee that this assignment is well-defined and results in a finitely additive set-function. We need to verify countable additivity. Let

$$E_n \searrow \emptyset \tag{10}$$

By repeating the sets as needed, we may assume $E_n \in \mathcal{F}_n$. If we can show that

$$\mathbb{P}[E_n] \searrow 0 \tag{11}$$

then Caratheodory's extension theorem guarantees that $\mathbb{P}$ extends uniquely to $\sigma(\mathcal{A}) = \mathcal{B}^\infty$.

We will use the following facts about $\mathbb{R}^n$:

1. Every finite measure μ on $(\mathbb{R}^n, \mathcal{B}^n)$ is *inner regular*, namely for every $E \in \mathcal{B}^n$
$$\mu[E] = \sup_{K \subset E} \mu[K] , \tag{12}$$
supremum over all compact subsets of E.

2. Every decreasing sequence of non-empty compact sets has non-empty intersection:
$$K_n \neq \emptyset, K_n \searrow K \quad \Rightarrow \quad K \neq \emptyset \tag{13}$$

3. If $f : \mathbb{R}^n \to \mathbb{R}^k$ is continuous, then $f(K)$ is compact for every compact K.

Then according to (12) for every E_n and every $\epsilon > 0$ there exists a compact subset $K'_n \subset \mathbb{R}^n$ such that such that

$$\mathbb{P}[E_n \setminus \mathcal{J}_{1,\dots,n}(K'_n)] \le \epsilon 2^{-n} \,.$$

Then, define by induction

$$K_n = K'_n \cap (K_{n-1} \times \mathbb{R}) \,.$$

(Note that $K_{n-1} \subset \mathbb{R}^{n-1}$ and the set $K_{n-1} \times \mathbb{R}$ is simply an extension of K_{n-1} into $\mathbb{R}^n$ by allowing arbitrary last coordinates.) Since $E_n \subset E_{n-1}$ we have

$$\mathbb{P}[E_n \setminus \mathcal{J}_{1,\dots,n}(K_n)] \le \epsilon 2^{-n} + \mathbb{P}[E_{n-1} \setminus \mathcal{J}_{1,\dots,n-1}(K_{n-1})] \,.$$

Thus, continuing by induction we have shown that

$$\mathbb{P}[E_n \setminus \mathcal{J}_{1,\dots,n}(K_n)] \le \epsilon(2^{-1} + \cdots 2^{-n}) < \epsilon \tag{14}$$

We will show next that $K_n = \emptyset$ for all n large enough. Since by construction

$$E_n \supset \mathcal{J}_{1,\dots,n}(K_n) \tag{15}$$

we then have from (14) and $K_n = \emptyset$ that

$$\limsup_{n\to\infty} \mathbb{P}[E_n] < \epsilon \,.$$

By taking ϵ to 0 we have shown (11) and the Theorem.

It thus remains to show that $K_n = \emptyset$ for all large enough n. Suppose otherwise, then by construction we have

$$K_n \subset K_{n-1} \times \mathbb{R} \subset K_{n-2} \times \mathbb{R}^2 \subset \cdots \subset K_1 \times \mathbb{R}^{n-1} \,.$$

Thus by projecting each K_n onto first coordinate we get a decreasing sequence of non-empty compacts, which by (13) has non-empty intersection. Then we can pick a point $x_1 \in \mathbb{R}$ such that

$$x_1 \in \mathrm{Proj}_{n\to 1}(K_n) \quad \forall n \,.$$

Repeating the same argument but projecting onto first two coordinates, we can now pick $x_2 \in \mathbb{R}$ such that

$$(x_1, x_2) \in \mathrm{Proj}_{n\to 2}(K_n) \quad \forall n\,.$$

By continuing in this fashion we will have constructed the sequence

$$(x_1, x_2, \ldots) \in \mathcal{J}_{1,\ldots,n}(K_n) \quad \forall n\,.$$

By (15) then we have

$$(x_1, x_2, \ldots) \in \bigcap_{n\geq 1} E_n$$

which contradicts (10). Thus, one of K_n must be empty. $\square$

1.3 Tail σ-algebra and Kolmogorov's 0/1 law

Definition 3. *Consider $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ and let $\mathcal{F}_n^\infty$ be a sub-σ-algebra generated by all cylinders $\mathcal{J}_{s_1,\ldots,s_k}(B)$ with $s_j \geq n$. Then the σ-algebra*

$$\mathcal{T} \triangleq \bigcap_{n>0} \mathcal{F}_n^\infty$$

is called a tail σ-algebra on $\mathbb{R}^\infty$. If $X : \Omega \to \mathbb{R}^\infty$ is a stochastic process, then σ-algebra $X^{-1}\mathcal{T}$ is called a tail σ-algebra of X.

Examples of tail events:

$$E_1 = \{\text{sequence } X_n \text{ converges}\} \tag{16}$$

$$E_2 = \{\text{series } \sum X_n \text{ converges}\} \tag{17}$$

$$E_3 = \{\limsup_{n\to\infty} X_n > 0\}\,, \tag{18}$$

An example of the event which is not a tail event:

$$E_4 = \{\limsup_{n\to\infty} \sum_{k=1}^{n} X_k > 0\}$$

Theorem 2 (Kolmogorov's 0/1 law)**.** *If $X_j, j = 1, \ldots$ are independent then any event in the tail σ-algebra of X has probability* 0 *or* 1*.*

Proof: Let $\mathbb{P}_X$ be the law of X (so that $\mathbb{P}_X$ is a measure on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$). Take $E \in \mathcal{T}$, then $E \in \mathcal{F}_n^\infty$ for every n. Thus under $\mathbb{P}_X$ event E is independent of every cylinder:

$$\mathbb{P}_X[E \cap \mathcal{J}_{s_1,\ldots,s_k}(B)] = \mathbb{P}_X[E]\mathbb{P}_X[\mathcal{J}_{s_1,\ldots,s_k}(B)] \tag{19}$$

On the other hand, by Proposition 1 every element of $\mathcal{B}^\infty$ can be arbitrarily well approximated with cylinders. Taking a sequence of such approximations converging to E in (19) we derive that E must be independent of itself:

$$\mathbb{P}_X[E \cap E] = \mathbb{P}_X[E]\mathbb{P}_X[E]\,,$$

implying $\mathbb{P}_X[E] = 0$ or 1. □

2 THE BERNOULLI PROCESS

In the Bernoulli process, the random variables X_n are i.i.d. Bernoulli, with common parameter $p \in (0,1)$. The natural sample space in this case is $\quad = \{0,1\}^\infty$.

Let $S_n = X_1 + \cdots + X_n$ (the number of "successes" or "arrivals" in n steps). The random variable S_n is binomial, with parameters n and p, so that

$$p_{S_n}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1 \ldots, n,$$

$$\mathbb{E}[S_n] = np, \qquad \mathrm{var}(S_n) = np(1-p).$$

Let T_1 be the time of the first success. Formally, $T_1 = \min\{n \mid X_n = 1\}$. We already know that T_1 is geometric:

$$p_{T_1}(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots; \qquad \mathbb{E}[T_1] = \frac{1}{p}.$$

2.1 Stationarity and memorylessness

The Bernoulli process has a very special structure. The discussion below is meant to capture some of its special properties in an abstract manner.

Consider a Bernoulli process $\{X_n\}$. Fix a particular positive integer m, and let $Y_n = X_{m+n}$. Then, $\{Y_n\}$ is the process seen by an observer who starts watching the process $\{X_n\}$ at time $m+1$, as opposed to time 1. Clearly, the process $\{Y_n\}$ also involves a sequence of i.i.d. Bernoulli trials, with the same parameter p. Hence, it is also a Bernoulli process, and has the same distribution as the process $\{X_n\}$. More precisely, for every k, the distribution of $(Y_1, \ldots, Y_k)$

is the same as the distribution of $(X_1,\dots,X_k)$. This property is called **stationarity** property.

In fact a stronger property holds. Namely, even if we are given the values of $X_1,\dots,X_m$, the distribution of the process $\{Y_n\}$ does not change. Formally, for any measurable set $A \subset$, we have

$$\begin{aligned}\mathbb{P}((X_{n+1},X_{n+2},\dots)\in A \mid X_1,\dots,X_n) &= \mathbb{P}((X_{n+1},X_{n+2},\dots)\in A)\\ &= \mathbb{P}((X_1,X_2\dots,\dots)\in A).\end{aligned}$$

We refer to the first equality as a **memorylessness** property. (The second inequality above is just a restatement of the stationarity property.)

2.2 Stopping times

We just discussed a situation where we start "watching" the process at some time $m+1$, where m is an integer constant. We next consider the case where we start watching the process at some random time $N+1$. So, let N be a nonnegative integer random variable. Is the process $\{Y_n\}$ defined by $Y_n = X_{N+n}$ a Bernoulli process with the same parameter? In general, this is not the case. For example, if $N = \min\{n \mid X_{n+1}=1\}$, then $\mathbb{P}(Y_1=1)=\mathbb{P}(X_{N+1}=1)=1\neq p$. This inequality is due to the fact that we chose the special time N by "looking into the future" of the process; that was determined by the future value X_{n+1}.

This motivates us to consider random variables N that are determined causally, by looking only into the past and present of the process. Formally, a nonnegative random variable N is called a **stopping time** if, for every n, the occurrence or not of the event $\{N=n\}$ is completely determined by the values of $X_1,\dots,X_n$. Even more formally, for every n, there exists a function h_n such that

$$I_{\{N=n\}} = h_n(X_1,\dots,X_n).$$

We are now a position to state a stronger version of the memorylessness property. If N is a stopping time, then for all n, we have

$$\begin{aligned}\mathbb{P}((X_{N+1},X_{N+2},\dots)\in A \mid N=n, X_1,\dots,X_n) &= \mathbb{P}((X_{n+1},X_{n+2},\dots)\in A)\\ &= \mathbb{P}((X_1,X_2\dots,\dots)\in A).\end{aligned}$$

In words, the process seen if we start watching right after a stopping time is also Bernoulli with the same parameter p.

2.3 Arrival and interarrival times

For $k \geq 1$, let Y_k be the k**th arrival time**. Formally, $Y_k = \min\{n \mid S_n = k\}$. For convenience, we define $Y_0 = 0$. The k**th interarrival time** is defined as $T_k = Y_k - Y_{k-1}$.

We already mentioned that T_1 is geometric. Note that T_1 is a stopping time, so the process $(X_{T_1+1}, X_{T_1+2}, \ldots)$ is also a Bernoulli process. Note that the second interarrival time T_2, in the original process is the first arrival time in this new process. This shows that T_2 is also geometric. Furthermore, the new process is independent from $(X_1, \ldots, X_{T_1})$. Thus, T_2 (a function of the new process) is independent from $(X_1, \ldots, X_{T_1})$. In particular, T_2 is independent from T_1.

By repeating the above argument, we see that the interarrival times T_k are i.i.d. geometric. As a consequence, Y_k is the sum of k i.i.d. geometric random variables, and its PMF can be found by repeated convolution. In fact, a simpler derivation is possible. We have

$$\begin{aligned}
\mathbb{P}(Y_k = t) &= \mathbb{P}(S_{t-1} = k-1 \text{ and } X_t = 1) = \mathbb{P}(S_{t-1} = k-1) \cdot \mathbb{P}(X_t = 1) \\
&= \binom{t-1}{k-1} p^{k-1}(1-p)^{t-k} \cdot p = \binom{t-1}{k-1} p^k (1-p)^{t-k}.
\end{aligned}$$

The PMF of Y_k is called a **Pascal** PMF.

2.4 Merging and splitting of Bernoulli processes

Suppose that $\{X_n\}$ and $\{Y_n\}$ are independent Bernoulli processes with parameters p and q, respectively. Consider a "merged" process $\{Z_n\}$ which records an arrival at time n if and only if one or both of the original processes record an arrival. Formally,

$$Z_n = \max\{X_n, Y_n\}.$$

The random variables Z_n are i.i.d. Bernoulli, with parameter

$$\mathbb{P}(Z_n = 1) = 1 - \mathbb{P}(X_n = 0,\ Y_n = 0) = 1 - (1-p)(1-q) = p + q - pq.$$

In particular, $\{Z_n\}$ is itself a Bernoulli process.

"Splitting" is in some sense the reverse process. If there is an arrival at time n (i.e., $X_n = 1$), we flip an independent coin, with parameter q, and record an arrival of "type I" or "type II", depending on the coin's outcome. Let $\{X_n\}$ and $\{Y_n\}$ be the processes of arrivals of the two different types. Formally, let $\{U_n\}$ be a Bernoulli process with parameter q, independent from the original process $\{Z_n\}$. We then let

$$X_n = Z_n \cdot U_n, \qquad Y_n = Z_n \cdot (1 - U_n).$$

Note that the random variables X_n are i.i.d. Bernoulli, with parameter pq, so that $\{X_n\}$ is a Bernoulli process with parameter pq. Similarly, $\{Y_n\}$ is a Bernoulli process with parameter $p(1-q)$. Note however that the two processes are dependent. In particular, $\mathbb{P}(X_n = 1 \mid Y_n = 1) = 0 \neq pq = \mathbb{P}(X_n = 1)$.

3 THE POISSON PROCESS

The Poisson process is best understood intuitively as a continuous-time analog of the Bernoulli process. The process starts at time zero, and involves a sequence of arrivals, at random times. It is described in terms of a collection of random variables $N(t)$, for $t \geq 0$, all defined on the same probability space, where $N(0) = 0$ and $N(t)$, $t > 0$, represents the number of arrivals during the interval $(0, t]$.

If we fix a particular outcome (sample path) ω, we obtain a time function whose value at time t is the realized value of $N(t)$. This time function has discontinuities (unit jumps) whenever an arrival occurs. Furthermore, this time function is right-continuous: formally, $\lim_{\tau \downarrow t} N(\tau) = N(t)$; intuitively, the value of $N(t)$ incorporates the jump due to an arrival (if any) at time t.

We introduce some notation, analogous to the one used for the Bernoulli process:

$$Y_0 = 0, \quad Y_k = \min\{t \mid N(t) = k\}, \qquad T_k = Y_k - Y_{k-1}.$$

We also let

$$P(k; t) = \mathbb{P}(N(t) = k).$$

The Poisson process, with parameter $\lambda > 0$, is defined implicitly by the following properties:

(a) The numbers of arrivals in disjoint intervals are independent. Formally, if $0 < t_1 < t_2 < \cdots < t_k$, then the random variables $N(t_1)$, $N(t_2) - N(t_1), \ldots, N(t_k) - N(t_{k-1})$ are independent. This is an analog of the independence of trials in the Bernoulli process.

(b) The distribution of the number of arrivals during an interval is determined by λ and the length of the interval. Formally, if $t_1 < t_2$, then

$$\mathbb{P}(N(t_2) - N(t_1) = k) = \mathbb{P}(N(t_2 - t_1) = k) = P(k; t_2 - t_1).$$

(c) There exist functions o_1, o_1, o_3 such that

$$\lim_{\delta \downarrow 0} \frac{o_k(\delta)}{\delta} = 0, \;\; k = 1, 2, 3,$$

and

$$\begin{aligned}
P(0;\delta) &= 1-\lambda\delta+o_1(\delta)\\
P(1;\delta) &= \lambda\delta+o_2(\delta),\\
\sum_{k=2}^{\infty} P(k;\delta) &= o_3(\delta),
\end{aligned}$$

for all $\delta > 0$.

The o_k functions are meant to capture second and higher order terms in a Taylor series approximation.

3.1 The distribution of $N(t)$

Let us fix the parameter λ of the process, as well as some time $t > 0$. We wish to derive a closed form expression for $P(k;t)$. We do this by dividing the time interval $(0,t]$ into small intervals, using the assumption that the probability of two or more arrivals in a small interval is negligible, and then approximate the process by a Bernoulli process.

Having fixed $t > 0$, let us choose a large integer n, and let $\delta = t/n$. We partition the interval $[0,t]$ into n "slots" of length δ. The probability of at least one arrival during a particular slot is

$$p = 1 - P(0;\delta) = \lambda\delta + o(\delta) = \frac{\lambda t}{n} + o(1/n),$$

for some function o that satisfies $o(\delta)/\delta \to 0$.

We fix k and define the following events:

A: exactly k arrivals occur in $(0,t]$;

B: exactly k slots have one or more arrivals;

C: at least one of the slots has two or more arrivals.

The events A and B coincide unless event C occurs. We have

$$B \subset A \cup C, \qquad A \subset B \cup C,$$

and, therefore,

$$\mathbb{P}(B) - \mathbb{P}(C) \le \mathbb{P}(A) \le \mathbb{P}(B) + \mathbb{P}(C).$$

Note that

$$\mathbb{P}(C) \le n \cdot o_3(\delta) = (t/\delta) \cdot o_3(\delta),$$

which converges to zero, as $n \to \infty$ or, equivalently, $\delta \to 0$. Thus, $\mathbb{P}(A)$, which is the same as $P(k;t)$ is equal to the limit of $\mathbb{P}(B)$, as we let $n \to \infty$.

The number of slots that record an arrival is binomial, with parameters n and $p = \lambda t/n + o(1/n)$. Thus, using the binomial probabilities,

$$\mathbb{P}(B) = \binom{n}{k}\Big(\frac{\lambda t}{n} + o(1/n)\Big)^k \Big(1 - \frac{\lambda t}{n} + o(1/n)\Big)^{n-k}.$$

When we let $n \to \infty$, essentially the same calculation as the one carried out in Lecture 6 shows that the right-hand side converges to the Poisson PMF, and

$$P(k;t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This establishes that $N(t)$ is a Poisson random variable with parameter λt, and $\mathbb{E}[N(t)] = \mathrm{var}(N(t)) = \lambda t$.

3.2 The distribution of T_k

In full analogy with the Bernoulli process, we will now argue that the interarrival times T_k are i.i.d. exponential random variables.

3.2.1 First argument

We have

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = P(0;t) = e^{-\lambda t}.$$

We recognize this as an exponential CDF. Thus,

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \qquad t > 0.$$

Let us now find the joint PDF of the first two interarrival times. We give a heuristic argument, in which we ignore the probability of two or more arrivals during a small interval and any $o(\delta)$ terms. Let $t_1 > 0$, $t_2 > 0$, and let δ be a small positive number, with $\delta < t_2$. We have

$$\begin{aligned}
\mathbb{P}(t_1 \le T_1 \le t_1 + \delta, \quad & t_2 \le T_2 \le t_2 + \delta) \\
&\approx P(0;t_1) \cdot P(1;\delta) \cdot P(0; t_2 - t_1 - \delta) \cdot P(1;\delta) \\
&= e^{-\lambda t_1} \lambda \delta e^{-\lambda (t_2 - \delta)} \lambda \delta.
\end{aligned}$$

We divide both sides by δ^2, and take the limit as $\delta \downarrow 0$, to obtain

$$f_{T_1,T_2}(t_1,t_2) = \lambda e^{-\lambda t_1} \lambda e^{-\lambda t_2}. \qquad t_1, t_2 > 0.$$

This shows that T_2 is independent of T_1, and has the same exponential distribution. This argument is easily generalized to argue that the random variables T_k are i.i.d. exponential, with common parameter λ.

3.2.2 Second argument

We will first find the joint PDF of Y_1 and Y_2. Suppose for simplicity that $\lambda = 1$. let us fix some s and t that satisfy $0 < s \le t$. We have

$$\begin{aligned}
\mathbb{P}(Y_1 \le s,\ Y_2 \le t) &= \mathbb{P}\big(N(s) \ge 1,\ N(t) \ge 2\big) \\
&= \mathbb{P}(N(s) = 1)\mathbb{P}(N(t) - N(s) \ge 1) + \mathbb{P}(N(s) \ge 2) \\
&= se^{-s}(1 - e^{-(t-s)}) + (1 - e^{-s} - se^{-s}) \\
&= -se^{-t} + 1 - e^{-s}.
\end{aligned}$$

Differentiating, we obtain

$$f_{Y_1,Y_2}(s,t) = \frac{\partial^2}{\partial t \partial s}\mathbb{P}(Y_1 \le s,\ Y_2 \le t) = e^{-t}, \qquad 0 \le s \le t.$$

We point out an interesting consequence: conditioned on $Y_2 = t$, Y_1 is uniform on $(0, t)$; that is given the time of the second arrival, all possible times of the first arrival are "equally likely."

We now use the linear relations

$$T_1 = Y_1, \qquad T_2 = Y_2 - Y_1.$$

The determinant of the matrix involved in this linear transformation is equal to 1. Thus, the Jacobian formula yields

$$f_{T_1,T_2}(t_1, t_2) = f_{Y_1,Y_2}(t_1, t_1 + t_2) = e^{-t_1}e^{-t_2},$$

confirming our earlier independence conclusion. Once more this approach can be generalized to deal with ore than two interarrival times, although the calculations become more complicated

3.2.3 Alternative definition of the Poisson process

The characterization of the interarrival times leads to an alternative, but equivalent, way of describing the Poisson process. Start with a sequence of independent exponential random variables $T_1, T_2, \ldots$, with common parameter λ, and record an arrival at times T_1, $T_1 + T_2$, $T_1 + T_2 + T_3$, etc. It can be verified that starting with this new definition, we can derive the properties postulated in our original definition. Furthermore, this new definition, being constructive, establishes that a process with the claimed properties does indeed exist.

3.3 The distribution of Y_k

Since Y_k is the sum of k i.i.d. exponential random variables, its PDF can be found by repeating convolution.

A second, somewhat heuristic, derivation proceeds as follows. If we ignore the possibility of two arrivals during a small interval, We have

$$\mathbb{P}(y \le Y_k \le y+\delta) = P(k-1; y)P(1;\delta) = \frac{\lambda^{k-1}}{(k-1)!} y^{k-1} e^{-\lambda y} \lambda\delta.$$

We divide by δ, and take the limit as $\delta \downarrow 0$, to obtain

$$f_{Y_k}(y) = \frac{\lambda^{k-1}}{(k-1)!} y^{k-1} e^{-\lambda y} \lambda, \qquad y > 0.$$

This is called a **Gamma** or **Erlang** distribution, with k degrees of freedom.

For an alternative derivation that does not rely on approximation arguments, note that for a given $y \ge 0$, the event $\{Y_k \le y\}$ is the same as the event

$$\big\{\text{number of arrivals in the interval } [0, y] \text{ is at least } k\big\}.$$

Thus, the CDF of Y_k is given by

$$F_{Y_k}(y) = \mathbb{P}\big(Y_k \le y\big) = \sum_{n=k}^{\infty} P(n, y) = 1 - \sum_{n=0}^{k-1} P(n, y) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda y)^n e^{-\lambda y}}{n!}.$$

The PDF of Y_k can be obtained by differentiating the above expression, and moving the differentiation inside the summation (this can be justified). After some straightforward calculation we obtain the Erlang PDF formula

$$f_{Y_k}(y) = \frac{d}{dy} F_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}.$$