

# Statistical Machine Learning

## Reading Assignment 1 Report

Alireza Sadeghi - Mohsen Shojaei

April 2, 2016

### 1 Introduction

As we've seen earlier, formal definition of probability is a triple  $P : (\Omega, \mathcal{F}, \mathbb{P})$ , in which  $\Omega$  denotes the sample space,  $\mathcal{F}$  denotes a  $\sigma$ -field over  $\Omega$  consisting of the desired events and finally,  $\mathbb{P}$  denotes a measure over the  $\sigma$ -field  $\mathcal{F}$ . A random variable  $X$  is also defined as follows:

$$X : \Omega \rightarrow \mathbb{R}$$

To be more implicit, a random variable  $X$  is defined as a measurable function over the desired  $\sigma$ -field, mapping events to real numbers, such that the following holds:

$$\forall x : \{\omega | X(\omega) \leq x\} \in \mathcal{F}$$

### 2 Random Process

A random process can be implicitly defined as follows:

$$\Pi : \Omega \rightarrow S^\infty$$

$S$  is the state space, which in most cases, considered to be the same as  $\mathbb{R}^n$ , and  $S^\infty$  denotes the collection of all countable subsets of  $S$ . In most real world cases, we are not interested in all subsets of  $S$ , instead we're interested in a smaller collections which we call *Test Sets*. Test sets usually is considered to be all open intervals in  $\mathbb{R}^n$ .

Since we are often interested in counting the occurrences of events in a specific interval, we move on to define a counting measure  $N$  over test sets:

$$N(A) = \#\{\Pi(\omega) \cap A\}$$

$N(A)$  counts the number of occurrences of the process points in a test set  $A$ . It's important to notice, that  $N(A)$  is a random variable, mapping outcomes in  $\Omega$  to non-negative integers. This drills down to  $N(A)$  being a measurable function defined as follows:

$$\forall n : \{\omega | N(A) = n\} \in \mathcal{F}$$

If for a random process  $\Pi$ , the resulted  $N(A)$  over test sets, follows a Poisson distribution and for every non-overlapping test sets  $A, B$  we have that random variables  $N(A)$  and  $N(B)$  are independent,  $\Pi$  is called a *Poisson process*.

It's important to notice that there are several other ways of defining a Poisson process, such as taking the distribution of inter arrival times to be of the Geometric distribution (Notice that in this case, we focus on inter event times instead of the points themselves).

### 3 Poisson Process

A Poisson process is a random process, parametrized by a deterministic measure on state space  $S$ . Notice that a random process can itself be defined as a random measure on  $S$ , this can be further explained using an example. Take for example the number of points in an area  $A$ . The number of points is a random measure, since there can be countably many points inside the the area, but the number itself follows a Poisson distribution whose parameter is the size of  $A$ , which is a deterministic measure (e.g. Area of  $A$ ).

Implicitly put, let  $\mu(A)$  be a measure on  $S$ , which we shall call the *mean measure*.  $\mu$  must be a non atomic measure in a sense that:

$$\forall x \in S, \mu(\{x\}) = 0$$

meaning, there can be no single point with a non zero mass.

The counting measure  $N(A)$  of a Poisson process, can be parametrized by a Poisson distribution with mean of  $\mu(A)$ . It's important to notice that the only piece of information, required to uniquely parametrize a Poisson process, is its mean measure  $\mu(A)$ .

To give a more thorough intuition on Poisson processes properties, it's good to notice that the parametrization above, is consistent with the initial definition in the previous section. Consider a sequence of disjoint test sets  $A_1, A_2, \dots$ , and let

$$A = \bigcup_i A_i \quad (1)$$

We plan to compute the distribution  $N(A)$  using two different approaches.

- Direct:

$$N(A) = \text{Poisson}(\mu(A)) \quad (2)$$

- Indirect: Since we have  $N(A) = \sum_i N(A_i)$  and according to the above parametrization  $N(A_i) = \text{Poisson}(\mu(A_i))$ , we can conclude that  $N(A) = \text{Poisson}(\sum_n \mu(A_n))$ . Considering  $\mu$  is a measure, we have that  $\mu(A) = \sum_n \mu(A_n)$  which is basically the same as 2.

### 3.1 Rate/Intensity

When  $S = \mathbb{R}^n$ , the mean measure is in most interesting cases given in terms of a rate or intensity. This is a positive measurable function  $\lambda$  with respect to d-dimensional measure:

$$\mu(A) = \int_A \lambda(x) dx \quad (3)$$

The term *rate* is used when we're dealing with 1D state space, which is usually the time axis.

In the special case when  $\lambda$  is a constant, so that

$$\mu(A) = \lambda|A|$$

the Poisson process is called *Uniform* or *Homogeneous*. In this case, stochastic properties of a test set  $A$ , are invariant to rotation and translation. The reverse argument also holds true, in which Poisson processes invariant to rotation and translation, are essentially *Homogeneous*.

### 3.2 Properties

In this section, we skim over three interesting properties of a Poisson process.

**Superposition:** Let  $\Pi_1, \Pi_2, \dots$  be a countable collection of independent Poisson processes on  $S$  and let  $\Pi_n$  have mean measure  $\mu_n$  for each  $n$ . then their superposition

$$\Pi = \bigcup_{n=1}^{\infty} \Pi_n$$

is a Poisson process with mean measure:

$$\mu = \sum_{n=1}^{\infty} \mu_n$$

**Restriction:** Let  $\Pi$  be a Poisson process with mean measure  $\mu$  on  $S$ , and let  $S_1$  be a measurable subset of  $S$ . Then the random countable

$$\Pi_1 = \Pi \cap S_1$$

can be regarded either as a Poisson process on  $S$  with mean measure

$$\mu_1(A) = \mu(A \cap S_1)$$

or as a Poisson process on  $S_1$  whose mean measure is the restriction of  $\mu$  to  $S_1$ .

**Mapping:** Let  $\Pi$  be a Poisson process on state space  $S$  whose measure  $\mu$  conforms to a specific condition ( $\sigma$ -finiteness), and let  $f : S \rightarrow T$  be a measurable function such that the induced measure is non atomic. Then  $f(\Pi)$  is a Poisson process on  $T$  having the induced measure  $\mu^*$  as its mean measure.

## 4 Marked Process

### 4.1 Coloring

Sometimes we want to model a situation in which the points can differ in some-way. For example, consider a case in which we want to model a router and we want to take into account, from which devices the packets are being received. Assuming there's only three different types of devices connecting to our router, points on the time axis can be of those three types.

We might be interested in analyzing this event types individually, or analyze them all together (keeping track of arrival times and types of the sender devices). If we define a multinomial distribution over different types, Coloring theorem provides us with a framework to analyze the incoming events.

**Coloring Theorem** Let  $\Pi$  be a Poisson process on  $S$  with mean measure  $\mu$ . Let the points of  $\Pi$  be colored randomly with  $k$  colors, the probability that

a point receives the  $i^{th}$  color being  $p_i$  and the colors of different points being independent (of one another and of the positions of the points). Let  $\Pi_i$  be the set of points with it  $i^{th}$  color. Then the  $\Pi_i$  are independent Poisson processes with means measures

$$\mu_i = p_i \mu$$

Note that this is consistent with the *Superposition* Theorem, since  $\Pi$  is the superposition of the independent Poisson processes  $\Pi_i$ .

This theorem seems to be too restrictive, in such a manner that it assumes lots of independencies and expects parameters to be fixed and non changeable. We might want our event to be dependent on its location or time. For example, take the case of modeling the traffic in a highway. To overcome these shortcomings, we define the notion of *Marked Poisson Process*.

## 4.2 Marked Poisson Process

We want to associate a new random variable  $M(x)$  with each point  $x$  in the  $S$ , describing the additional information we're aiming to model. The distribution of  $M(x)$  may depend on  $x$  but not on the other points of  $\Pi$  and  $M(x)$  for different  $x$ s are independent.

The pair  $(X, M(x))$  can then be regarded as a random point  $X^*$  in the product space  $S * M$ . In this space, a new random process  $\Pi^*$  can be defined as follows:

$$\Pi^* = \{(X, M(x)) | X \in \Pi\}$$

Under general conditions,  $\Pi^*$  turns out to be a poisson process.

**Marking Theorem** The random subset  $\Pi^*$  is a Poisson process on  $S * M$  with mean measure  $\mu^*$  given by

$$\mu^*(C) = \iint_{(x,m) \in C} \mu(dx) p(x, dm)$$

This model can be specifically beneficial for use cases in which we want to measure the intensity of an event. Measuring the intensity, alongside the frequency of earthquakes, thunderbolts is of the likes of this model's usecases. Another field of usecases is in trading and asset return estimation.