Poisson Process and Poisson Distribution

1 Poisson Process

A Poisson process is the stochastic process in which events occur continuously and independently of one another. Examples that are well-modeled as Poisson processes include the radioactive decay of atoms, the number of mutations in a given sequence of DNA, telephone calls arriving at a switchboard, page view requests to a website, and rainfall.

The Poisson process is a collection $\{N(t) : t \geq 0\}$ of random variables, where $N(t)$ is the number of events that have occurred up to time t (starting from time 0). The number of events between time a and time b is given as $N(b) - N(a)$ and has a **Poisson distribution**.

2 Poisson distribution

The Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event.

Each realization of the process $\{N(t)\}$ is a non-negative integer-valued step function that is non-decreasing, but for intuitive purposes it is usually easier to think of it as a point pattern on $[0, \infty)$ (the points in time where the step function jumps, i.e. the points in time where an event occurs). At any point in time, the probability of one more new event is the mean of the Poisson distribution.

3 Binomial and Poisson distribution

In our case of the arrival of innovations—the events being counted are actually the outcomes of discrete trials, and would more precisely be modelled using the binomial distribution. However, the binomial distribution with parameters n and λ/n, i.e., the probability distribution of the number of successes in n trials, with probability λ/n of success on each trial, approaches the Poisson distribution with expected value λ as n approaches infinity. So we can approximate these random variables using the Poisson distribution rather than the more-cumbersome binomial distribution.

To show how the binomial distribution approaches the Poisson distribution in limit, first, recall from calculus

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda}$$

and the definition of the Binomial distribution

$$P\left(X=k\right)=\binom{n}{k}p^{k}(1-p)^{n-k}$$

If the binomial probability can be defined such that $p=\lambda/n$, we can evaluate the limit of P as n goes large:

$$\begin{aligned}
\lim_{n\to\infty}P\left(X=k\right) &= \lim_{n\to\infty}\binom{n}{k}p^{k}(1-p)^{n-k}\\
&= \lim_{n\to\infty}\left(\frac{n!}{k!(n-k)!}\right)\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k}\\
&= \lim_{n\to\infty}\left(\frac{n!}{n^{k}(n-k)!}\right)\frac{\lambda^{k}}{k!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}\\
&\approx \frac{\lambda^{k}e^{-\lambda}}{k!}
\end{aligned}$$

4 Exponentially-distributed inter-arrival times

To illustrate the exponentially-distributed inter-arrival times property, consider a Poisson process $N(t)$ with rate parameter λ, and let T_k be the time of the k^{th} arrival, for $k=1,2,3,\ \ldots$. Clearly the number of arrivals before some fixed time t is less than k if and only if the waiting time until the k^{th} arrival is more than t. In symbols, the event $[N(t)<k]$ occurs if and only if the event $[T_k>t]$ occurs. Consequently the probabilities of these events are the same:

$$P(T_k>t)=P\left(N(t)<k\right)$$

In particular, consider the waiting time until the first arrival. Clearly that time is more than t if and only if the number of arrivals before time t is

0. Combining this latter property with the above probability distribution for the number of homogeneous Poisson process events in a fixed interval gives

$$
\begin{aligned}
P(T_1 &> t) = P\left(N(t) = 0\right) \\
&= P\left(N(t) - N(0) = 0\right) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}
\end{aligned}
$$

The cumulative function of an exponential distribution is

$$
F(t;\lambda) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} .
$$

Consequently, the waiting time until the first arrival T_1 has an exponential distribution, and is thus memoryless. One can similarly show that the other interarrival times $T_k - T_{k-1}$ share the same distribution.

$$
\begin{aligned}
P(T_k - T_{k-1} &> t) = P\left[N(T_{k-1} + t) - N(T_{k-1}) = 0\right] \\
&= \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}
\end{aligned}
$$

Hence, they are independent, identically-distributed (i.i.d.) random variables, namely exponential distribution with mean $1/\lambda$. For example, if the average rate of arrivals is 5 per minute, then the average waiting time between arrivals is $1/5$ minute.