Rare Events: The Discrete Poisson & Continuous Exponential Random Variables / Distributions

Some examples: the occurrence of an earthquake, people entering a given establishment (bank, post office, gas station, etc.), the start of a war. What we require, in terms of assumptions, are 3: ① for small time intervals of time h, the probability that exactly 1 event occurs is $\lambda h + o(h)$ ($o(h)$ means $\lim_{h \to 0} \frac{o(h)}{h} = 0$, e.g. $o(h) = h^2$) for some constant $\lambda > 0$.

② The prob. of 2 or more events occuring in h is $o(h)$.

③ Whatever occurs in one time interval has no probability effect on what will occur in other, nonoverlapping intervals. Formally, for any $n, j_1, \ldots, j_n \in \mathbb{N}$ and any set of n nonoverlapping time intervals $h_1, \ldots, h_n > 0$, $h_i \cap h_j = \emptyset$ for $i \neq j$, let

$$\boxed{E_i \stackrel{\text{def}}{=} \text{event that exactly } j_i \text{ of the occurrences take place in } h_i}$$

Then we require that $E_1, \ldots, E_n$ are independent, i.e. $P(E_i \cap E_j) = P(E_i)P(E_j)$.

Let us show that under these assumptions, the Poisson (discrete) & exponential (continuous) distributions apply.

I. The Poisson approach: (in Section 4.7)

Let $N(t) = \#$ of occurrences of the rare event in the interval $[0, t]$.

To obtain $P(N(t) = k)$, divide the interval into n nonoverlapping time intervals of length t/n:

$h_1 = [0, t/n)$ $\quad [t/n, 2t/n) = h_2 \ldots$

so that

$$P(N(t) = k) = P(\overbrace{\{k \text{ subintervals contain exactly 1 event, other } n-k \text{ contain } 0\}}^{E_1} \sqcup \underbrace{\{N(t) = k \text{ \& at least one subinterval contains 2 or more}\}}_{E_2})$$

$$= P(E_1) + P(E_2)$$

Observe that if

$$\boxed{F = \text{at least one interval contains 2 or more occurrences}}$$

then

$$E_2 \subseteq F$$

and

$$P(E_2) \le P(F)$$

Boole's inequality:
$P(\bigcup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} P(A_i)$
(prob. 2.14, p. 55)

$$= P\Big(\bigcup_{i=1}^{n} \overbrace{\{\text{interval } h_i \text{ contains 2+ occurrences}\}}^{\text{call this } F_i}\Big)$$

$$\le \sum_{i=1}^{n} P(F_i)$$

assump. 2

$$= \sum_{i=1}^{n} o(t/n)$$

$$= n \cdot o(t/n)$$

$$= t \cdot \underbrace{\left(\frac{o(t/n)}{t/n}\right)}_{\lim_{n\to\infty} = 0 \text{ by def. of } o(h)} \longrightarrow 0$$

Hence, $\boxed{\lim_{n\to\infty} P(E_2) = 0}$ i.e. E_2 is $o(h)$

Next, we observe that

$$P(\{0 \text{ occurrences in } h\text{-length interval}\}) = 1 - \underset{E_1}{\underset{\uparrow}{\lambda h}} - \underset{\substack{E_2 \,\&\, E_1\text{'s tail} \\ (\text{assump. } 1)}}{\underset{\uparrow}{o(h)}}$$

$$\Longrightarrow \boxed{P(E_1) = \underbrace{\binom{n}{k}\left(\frac{\lambda t}{n} + o\left(\frac{t}{n}\right)\right)^k}_{\substack{k \text{ occurrences in} \\ k \text{ subintervals}}} \overbrace{\left(1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right)\right)^{n-k}}^{n-k \text{ of these}}}$$

Finally, $\boxed{\lim_{n \to \infty} P(E_1) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}}$ exactly

as in the binomial case (see the Poisson notes), and therefore

$$\boxed{\lim_{n \to \infty} P(N(t) = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}}$$

or, for large n, $\boxed{P(N(t) = k) \approx e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}}$

ex. Earthquakes occur in western part of US in line with our 3 assumptions, & with $\lambda = 2$ and t = one week. Find

(a) $P(N(2) \geq 3)$

(b) F(time fr. now until next earthquake)

A: (a)
$$P(N(2) \geq 3) = 1 - P(N(2) < 3)$$
$$= 1 - P(N(2) = 0) - P(N(2) = 1) - P(N(2) = 2)$$

$t\lambda$

$$= 1 - e^{-2\cdot 2} - e^{-2\cdot 2}\cdot\frac{2\cdot 2}{1!} - e^{-2\cdot 2}\cdot\frac{(2\cdot 2)^2}{2!}$$

$$= 1 - e^{-4}(1 + 4 + 8)$$
$$= 1 - 13e^{-4}$$
$$\approx 0.7619$$

(b) Let X = time (in weeks) till next earthquake, then

$$P(X > t) = P(N(t) = 0)$$
$$= e^{-\lambda t}$$

$$\Longrightarrow \boxed{\begin{aligned} F(t) &= P(X \le t) \\ &= 1 - P(X > t) \\ &= 1 - e^{-\lambda t} \\ &= 1 - e^{-2t} \end{aligned}}$$

II. The exponential approach: (Sec. 5.5)

The exponential random variables / distributions are characterized by their probability density functions,

$$\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}}$$ (* for $\lambda > 0$)

That is,

$$\boxed{P(X \in B) = \int_B \lambda e^{-\lambda x}\, dx}$$

The basic properties of such distributions are:

(1) $F(a) = P(X \leq a) = 1 - e^{-\lambda a}$ $(a \geq 0)$

(2) $E[X] = \dfrac{1}{\lambda}$

(3) $Var(X) = \dfrac{1}{\lambda^2}$ (see activity 3, or example 5a, p. 198)

One of the principal applications of the exponential random variables is to model the distribution of the amount of time until some specific occurrence takes place (e.g. earthquake, new war breaking out, etc.)

ex. Time you have to wait (e.g. on hold) behind someone in front of you. If $\lambda = 1/10$,

(a) $\underbrace{P(X > 10)}_{\text{prob. you way more than 10 min}} = 1 - P(X \le 10)$

$= 1 - F(10)$

$= 1 - (1 - e^{-10/10})$

$= \boxed{e^{-1} \approx 0.368}$

(b) $\underbrace{P(10 < X < 20)}_{\text{prob. you have to wait bet. 10 \& 20 min}} = P(X \le 20) - P(X \le 10)$

$= F(20) - F(10)$

$= (1 - e^{-20/10}) - (1 - e^{-10/10})$

$= \boxed{e^{-1} - e^{-2} \approx 0.233}$