Lecture 6: The Poisson Family of Distributions (Continued)

Lecturer: Prof. F. Jalayer, 19 February 2021
Notes prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. Gamma Probability Distributions

The Gamma distribution models the time until observing k events. It actually shows the distribution of the time needed to arrive to the kth success. Accordingly, let the statement A be defined as follows:

$A \equiv$ the time until observing k events is t_k

$\mathbf{I} \equiv$ the mean rate of occurrence of the event is λ

Considering the fact that the time to the kth event, denoted as t_k, can be written as the sum of the intermediate IAT's corresponding to the previous events:

$$t_k = IAT_1 + IAT_2 + \cdots + IAT_k = \sum_{i=1}^{k} IAT_i \tag{20}$$

Hence, the probability expression, $P\left(A|\mathbf{I}\right)$, can be defined based on the *i.i.d.* IAT_i's as follows:

$$P\left(A|\mathbf{I}\right) = p\left(t_k|\lambda\right) \equiv p\left(\sum_{i=1}^{k} IAT_i|\lambda\right) \tag{21}$$

Accordingly, one can simply calculate the expected value and variance of the Gamma distribution as follows based on our previous knowledge about the properties of the exponential distribution:

$$\begin{aligned} \mathbb{E}\left(t_k|\lambda\right) &\equiv \mathbb{E}\left(\sum_{i=1}^{k} IAT_i|\lambda\right) = \frac{k}{\lambda} \\ \mathbb{VAR}\left(t_k|\lambda\right) &\equiv \mathbb{VAR}\left(\sum_{i=1}^{k} IAT_i|\lambda\right) = \frac{k}{\lambda^2} \end{aligned} \tag{22}$$

Let us derive the PDF and CDF of the Gamma distribution for k=2. By applying the Total Probability Theorem and considering that $t_2=IAT_1+IAT_2$ and that IAT_1 and IAT_2 are independent:

$$\begin{aligned} P\left(t_2 < t \middle| \lambda\right) &= \int_0^t P\left(t_2 < t \middle| IAT_1 = x, \lambda\right) p\left(IAT_1 = x \middle| \lambda\right) \mathrm{d}x \\ &= \int_0^t P\left(t_2 - IAT_1 < t - x \middle| \lambda\right) f_{IAT}\left(x \middle| \lambda\right) \mathrm{d}x = \int_0^t \left[1 - e^{-\lambda(t-x)}\right]\left[\lambda e^{-\lambda x}\right] \mathrm{d}x \\ &= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} \end{aligned} \tag{23}$$

Thus,

$$p\left(t_2 \middle| \lambda\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - e^{-\lambda t} - \lambda t e^{-\lambda t}\right)\Big|_{t=t_2} = \lambda^2 t_2 e^{-\lambda t_2} \tag{24}$$

The CDF for the Gamma distribution can be expressed as the probability that the time to the kth event is less than t. In general, this is equivalent to calculating the probability of having at least k events in time t (using the Poisson distribution):

$$P\left(t_k < t \middle| \lambda\right) = P\left(r \geq k \middle| \lambda, t\right) = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!} \tag{25}$$

The PDF for Gamma Distribution can be calculated as the derivative of the CDF:

$$p(t_k | \lambda) = \frac{d}{dt} P\left(t_k < t \middle| \lambda\right) = \frac{d}{dt}\left[1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!}\right] = \frac{\lambda\left(\lambda t\right)^{k-1} e^{-\lambda t}}{\left(k-1\right)!} \tag{26}$$

2. Filtered Poisson

Consider that we have a Poisson process with the mean rate of occurrence of the events denoted as λ_0. As a result, the probability distribution of the number of events, *n*, which can take place in the time interval *T*, can be defined with the Poisson PMF given by:

$$P\left(n\left|\lambda_0 t\right.\right)=\frac{\left(\lambda_0 t\right)^n e^{-\lambda_0 t}}{n!},\ n\in\left[0,+\infty\right) \tag{27}$$

Now, we want to accept each event with a probability of $p_F = P\left(F\left|\text{event}\right.\right)$; i.e., probability of success of each event in satisfying the condition *F*. Therefore, we have actually applied a binary filter to the corresponding Poisson distribution. Hence, using the Total Probability Theorem:

$$\begin{aligned}
P\left(r\left|p_F,\lambda_0 t\right.\right) &= \sum_{n=r}^{+\infty} P\left(r\left|n,p_F,\lambda_0 t\right.\right)P\left(n\left|\lambda_0 t\right.\right) \\
&= \sum_{n=r}^{+\infty}\left[\binom{n}{r}p_F^r\left(1-p_F\right)^{n-r}\right]\left[\frac{\left(\lambda_0 t\right)^n e^{-\lambda_0 t}}{n!}\right] \\
&= \frac{p_F^r e^{-\lambda_0 t}\left(\lambda_0 t\right)^r}{r!}\sum_{n-r=0}^{+\infty}\frac{\left(\lambda_0 t\right)^{n-r}\left(1-p_F\right)^{n-r}}{\left(n-r\right)!} \\
&= \frac{p_F^r\left(\lambda_0 t\right)^r e^{-\lambda_0 t}}{r!}e^{\lambda_0 T\left(1-p_F\right)} = \frac{\left[\left(p_F\lambda_0\right)t\right]^r e^{-\left(p_F\lambda_0\right)t}}{r!}
\end{aligned} \tag{28}$$

According to this Equation, the new distribution is actually a Poisson distribution with the rate:

$$\lambda = p_F\lambda_0 \tag{29}$$

A practical application of filtered Poisson is in calculation of Probabilistic seismic hazard analysis (PSHA). Assuming a Poisson distribution for the events with magnitudes greater than a threshold M_{min}, $M \geq M_{\min}$, having a rate $\lambda\left(M > M_{\min}\right)$. In order to estimate the rate of events having ground motion intensity measure, *IM*, greater than a value *x*, $\lambda\left(IM > x\right)$, one can directly use a filtered Poisson as follows:

$$\lambda\left(IM > x\right) = \lambda\left(M > M_{\min}\right) \cdot P\left[IM > x \,\middle|\, M > M_{\min}\right] \tag{30}$$

3. Sum of independent Poisson random variables

It is noteworthy that the sum of independent Poisson random variables is Poisson. In order to prove this statement, let X and Y be two independent Poisson random variables with rate parameters λ_X and λ_Y, respectively. Then the PDF for the random variable $Z = X + Y$ is:

$$\begin{aligned} P\left(Z=n \middle| \lambda_X,\lambda_Y,t\right) &\overset{\text{using TPT}}{=} \sum_{k=0}^{n} P\left(Z=n \middle| X=k\right) P\left(X=k\right) \\ &= \sum_{k=0}^{n} P\left(Z-X=n-k \middle| X=k\right) P\left(X=k\right) \\ &= \sum_{k=0}^{n} \underbrace{P\left(Y=n-k \middle| X=k\right)}_{X \text{ and } Y \text{ are independent}} P\left(X=k\right) \\ &= \sum_{k=0}^{n} P\left(Y=n-k\right) P\left(X=k\right) = \sum_{k=0}^{n} \left(\frac{\left(\lambda_Y t\right)^{n-k} e^{-\lambda_Y t}}{\left(n-k\right)!} \cdot \frac{\left(\lambda_X t\right)^{k} e^{-\lambda_X t}}{k!} \right) \\ &= \frac{e^{-\lambda_X t} e^{-\lambda_Y t}}{n!} \sum_{k=0}^{n} \left(\frac{n!}{k!\left(n-k\right)!} \left(\lambda_X t\right)^{k} \left(\lambda_Y t\right)^{n-k} \right) = \frac{\left(\lambda_X+\lambda_Y\right)^{n} e^{-\left(\lambda_X+\lambda_Y\right)}}{n!} \end{aligned} \tag{31}$$

Hence, the result is a Poisson process with the rate equal to $\lambda_X + \lambda_Y$. As a practical example of this specific property, assume that a number of N_s seismic sources affect the site of interest. In order to estimate $\lambda\left(IM > x\right)$ for the desired site, one need to add the rate of $IM>x$ associated with all seismic sources as follows:

$$\lambda\left(IM > x\right) \triangleq \lambda_{IM} = \sum_{i=1}^{N_s} \lambda_i \left(M > M_{\min}\right) \cdot P_i \left[IM > x \middle| M > M_{\min}\right] \tag{32}$$

Now, suppose that we are interested in calculating the probability of exceeding IM in time interval t. This is equal to calculating the probability of having at least one event in the time interval t:

$$P\left(IM > x \middle| \lambda_{IM}, t\right) = 1 - e^{-\lambda_{IM} t} \tag{33}$$