

STATISTICAL DISTRIBUTIONS FOR PROBABILISTIC SEISMIC HAZARD ANALYSIS

FLORIN KÖPE, CRISTIAN ONOFREI, PAUL OLTEANU

Abstract. The study of statistical distributions in the context of probabilistic seismic hazard analysis (PSHA) provides a convenient support for understanding and application of PSHA specific procedures. This approach also provides a useful insight on relevant aspects of probability theory with application to seismic hazard. The geometric, exponential, binomial and Poisson distributions along with their proofs are introduced within the seismic hazard framework. These distributions represent the main tools for handling probabilistic hazard data as they are essential for assessing the probability of exceedance of the hazard parameter within a certain exposure period of structural system.

Key words: seismic hazard, PSHA, binomial, geometric exponential and Poisson distributions, Poisson process, return period.

1. INTRODUCTION

The probabilistic seismic hazard analysis (PSHA) is a complex methodology built on the basis of several branches such as mathematics of probabilities, structural dynamics and seismology. The highly specialized publications available in this field generally discourage the structural engineer acquainted with physical intuitive relations of mechanics. This paper presents the relevant probability distributions involved in PSHA emphasizing the relationship between Poisson and exponential distributions which represents the central concept of modern hazard studies. Firstly, the relevant concepts of conditional probabilities and statistically independent events are briefly introduced using convenient representations and later on the geometric and binomial distributions are presented along with the relationships to their limiting cases *i.e.* exponential and Poisson distributions, respectively. This approach provides a suitable perspective over PSHA specific procedures and return period.

Antiseismic Structural Engineering, Romania; E-mails: kope@antiseismic.com, onofrei@antiseismic.com, paul@antiseismic.com

Rev. Roum. Sci. Techn. – Méc. Appl., Tome 57, N° 1, P. 85–99, Bucarest, 2012

2. PROBABILITIES

Consider a sequence of n repeated trials of a random experiment. The probability (or relative frequency) of occurrence of the event consisting of n_E successes denoted $P(E)$ is given by the ratio of occurrences n_E to the total number of trials n [1]:

$$P(E) = \lim_{n \rightarrow \infty} \frac{n_E}{n}. \quad (1)$$

The sample space S_S represents the set of all possible outcomes of a random experiment. The mathematics of probability rest on three basic assumptions or axioms [2]:

- (1) $P(E) \geq 0$,
- (2) $P(S_S) = 1$ or $0 \leq P(E) \leq 1$,
- (3) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ if $P(E_1 E_2) = \emptyset$.

The first two axioms are apparent by using equation (1). The third one is based on the definition of mutually exclusive events (the occurrence of any event precludes the success of the other), thus the probability of their joint occurrences $E_1 \cap E_2$ is void. The following notation was used for the probability of intersection of two events E_1 and E_2 : $P(E_1 \cap E_2) = P(E_1 E_2)$. Since an event E and its complement \bar{E} are mutually exclusive and collectively exhaustive (i.e. $E \cup \bar{E} = S_S$) and using the third axiom, the following relations may be inferred:

$$\begin{aligned} P(S_S) = P(E \cup \bar{E}) = P(E) + P(\bar{E}) = 1 &\Rightarrow P(\bar{E}) = 1 - P(E). \\ P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2) &\end{aligned} \quad (3)$$

Figure 1 presents the relationships between various types of events.

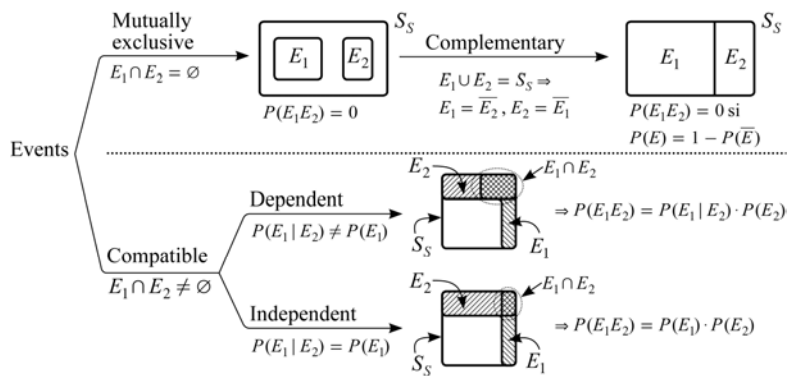


Fig. 1 – Relationships between events.

Sometimes the occurrence of one event may depend on the occurrence of another event (statistically dependent events) and this usually leads to difficulties in evaluation of the joint probability. Therefore, the probability of occurrence of an event E_1 is different if another event e.g. E_2 occurred, $P(E_1 | E_2) \neq P(E_1)$. The notation $P(E_1 | E_2)$ stands for the conditional probability or the probability of occurrence of one event given that the other has occurred. The probability of the concurrent occurrence of two statistically dependent $P(E_1 E_2)$ is:

$$P(E_1 E_2) = P(E_1 | E_2) \cdot P(E_2). \quad (4)$$

Figure 2 illustrates the concept of conditional probability. The same experiment will be presented later on for statistically independent events in order to provide a convenient comparison between the two. In Fig. 2, we consider a continuous random experiment $n \rightarrow \infty$ which generates several events E_j :

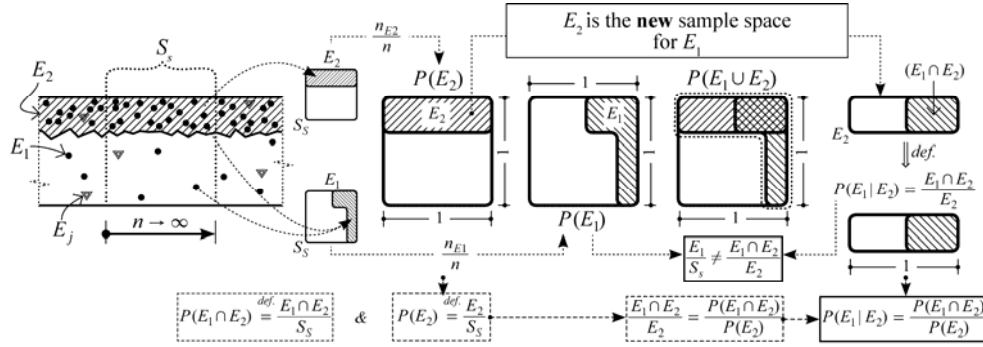


Fig. 2 – Conditional probabilities.

In some cases the occurrence of one event may not depend on the occurrence of the other, these events are called statistically independent events. The probability of event E_1 does not depend on the occurrence of E_2 and thus the probability of E_1 remains unaffected – $P(E_1 | E_2) = P(E_1)$. The joint probability (intersection) of two independent events $P(E_1 E_2)$ is:

$$P(E_1 E_2) = P(E_1) \cdot P(E_2). \quad (5)$$

Same random experiment is run in order to clarify this expression and the relation between the compatible events E_1 and E_2 is tracked throughout the process.

Once the relevant relations of probabilities were briefly introduced, is time to apply the previous concepts to earthquake type events. The probability distributions presented hereinafter are based on Poisson process, which satisfies the following requirements:

1. The events are statistically independent (there is no conditioning between events). The probability of occurrences in any disjoint interval (usually 1 year) of earthquake events is independent from each other.
2. Only two outcomes are possible in one trial (occurrence or nonoccurrence of a seismic event). This assumption basically indicates that the probability of occurrence of a seismic event within one interval is $p \leq 1$.

Next, we track the number of earthquake events which may occur during the sequence of 1, 2... n intervals (trials). The sample space consists of current events as well as all previous recorded ones during preceding trials. It is apparent that earthquakes cannot happen simultaneously but successively, yet as the sample space increases it may be possible acquiring several additional seismic events after i -trials ($i = 1 \dots n$), which are further available in the same time within the sample space. Otherwise stated, the events produced by each interval are compatible and independent and complies to a Poisson process if the chosen interval is small enough to get at most one event per interval.

This breakdown approach presumes an overall examination of probabilities after each additional disjoint interval appended to original sample space.

Figure 3 summarizes this technique, where E_i represents the success (earthquake occurrence on interval i) and p denotes the probability of success of event E_i on i -interval (the probability p shall remain unchanged throughout entire experiment according to the Poisson process assumptions). The probability p is obtained using equation (1), where n_E represents the number of successes out of all n -trials.

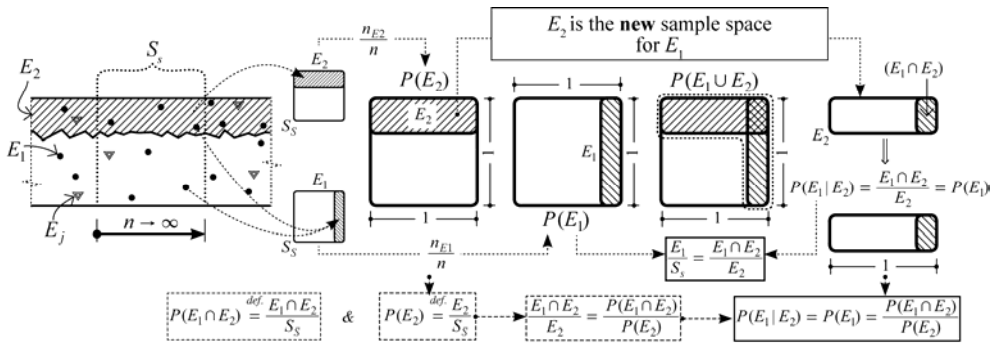


Fig. 3 – Independent probabilities.

The binomial and geometric distributions clearly arise observing Fig. 4. Where the geometric distribution deals with the first occurrence time of an event or in other words it provides the probability of required trials to get one success. It is important to notice that the geometric distribution requires nonoccurrence of all

events until $(i-1)$ trial and then the success only at time i . This sequence of probabilities represents indeed a valid distribution (see next section).

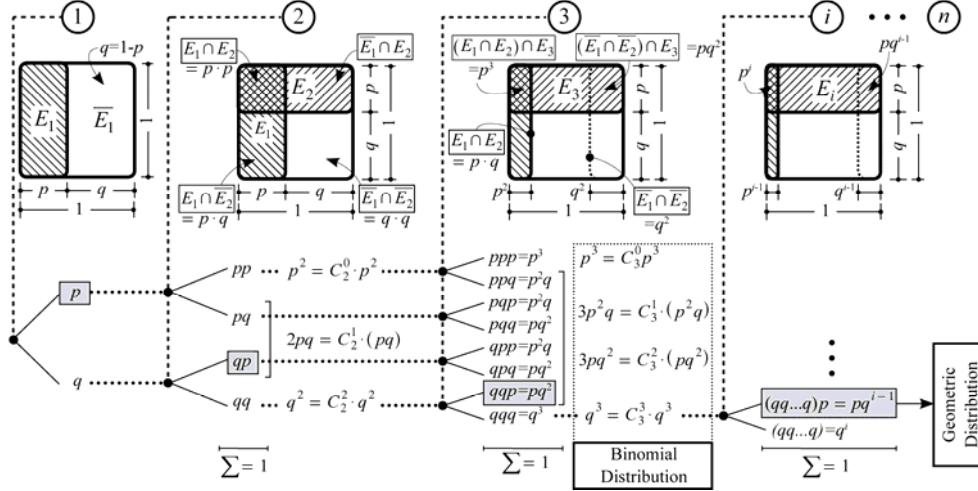


Fig. 4 – Probabilities of statistically independent events throughout a Poisson process.

Only the mean of geometric distribution (interval between two successes) is relevant to the structural engineer as it represents the return period. It is obvious observing Fig. 6 that the largest probability (mode) corresponds to the first trial, which does not mean that the 1-year interval is the most frequent return period but indicates that is more likely getting a success during the first trial than at moment i (as the experiment assumes $(i-1)$ failures until first i success).

The binomial distribution addresses all possible combinations of k successes out of exactly n -trials. It provides the probability of occurrence for k earthquakes out of n years provided that Poisson process conditions are met. The binomial distribution yields the overall picture of probabilities for any combination of k -earthquakes out of exactly n -years, while the geometric distribution runs through one by one all n -intervals tracking the probability of occurrence for a single success at that particular instant.

3. EXPONENTIAL DISTRIBUTION: LIMITING CASE OF GEOMETRIC DISTRIBUTION

The exponential distribution can be derived as a limiting case of the geometric distribution. The first occurrence time of an event is of great interest in engineering thus a continuous form of the geometric distribution is convenient for application to PSHA. The exponential distribution is also appealing to practitioners as its expression is simple and easy to handle. Both distributions deal with time between events in a Poisson process.

The probability density function (PDF) of a random variable Y is given by the next relation (with reference to Fig. 4). The random variable Y is assigned $y_n = 1, 2, \dots, n, \dots$, then one could write:

$$f_Y(y) = P(Y = n) = (1 - p)^{n-1} \cdot p. \quad (6)$$

First, we should prove that this expression represents indeed a valid PDF, thus we verify if the sum of this geometric series is in fact 1. Denoting q the probability of failure, then $q = 1 - p$, hence the sum is:

$$\begin{aligned} \sum_{n=1}^{\infty} q^{n-1} (1-q) &= (1-q) \sum_{n=1}^{\infty} q^{n-1} = \sum_{n=1}^{\infty} q^{n-1} - q \sum_{n=1}^{\infty} q^{n-1} = \\ &= \left(q^0 + \sum_{n=2}^{\infty} q^{n-1} \right) - \sum_{n=1}^{\infty} q^n = 1 + \sum_{n=2}^{\infty} q^{n-1} - \sum_{n=2}^{\infty} q^{n-1} = 1. \end{aligned} \quad (7)$$

The mean of geometric mean is:

$$\begin{aligned} \mu_Y &= \sum_n y_n \cdot f_Y(y_n) = \sum_{n=1}^{\infty} y_n \cdot [q^{n-1} (1-q)] = 1 \cdot (1-q) + 2 \cdot q \cdot (1-q) + \dots \\ q \cdot \mu_Y &= q \cdot (1-q) + 2 \cdot q^2 \cdot (1-q) + 3 \cdot q^3 \cdot (1-q) + \dots \end{aligned} \quad (8)$$

Subtracting the two above expressions one could obtain:

$$\mu_Y - q \cdot \mu_Y = \sum_{n=1}^{\infty} (1-q) \cdot q^{n-1} = 1 \Rightarrow \mu_Y = \frac{1}{1-q} = \frac{1}{p}. \quad (9)$$

The cumulative distribution function (CDF) is obtained using the same method:

$$\begin{aligned} P(Y \leq n) = \text{CDF} &= \sum_n q^{n-1} (1-q) = (1-q) + (1-q) \cdot q + \dots + (1-q) \cdot q^{n-1} \\ q \cdot \text{CDF} &= (1-q) \cdot q + (1-q) \cdot q^2 + (1-q) \cdot q^3 + \dots + (1-q) \cdot q^n. \end{aligned} \quad (10)$$

Subtracting the two above expressions, we get:

$$\begin{aligned} \text{CDF} - q \cdot \text{CDF} &= (1-q) - (1-q) \cdot q^n \Rightarrow (1-q) \cdot \text{CDF} = (1-q) \cdot (1 - q^n) \\ &\Rightarrow \text{CDF} = 1 - q^n = 1 - (1-p)^n. \end{aligned} \quad (11)$$

Therefore, the cumulative distribution function is

$$F_Y(y) = 1 - (1 - p)^n. \quad (12)$$

In order to illustrate the equivalence between the geometric and exponential distributions, the entire interval t is divided into smaller and smaller sub-intervals by means of a parameter r . Fig. 5 depicts the reasoning involved in exponential distribution determination out of geometric distribution.

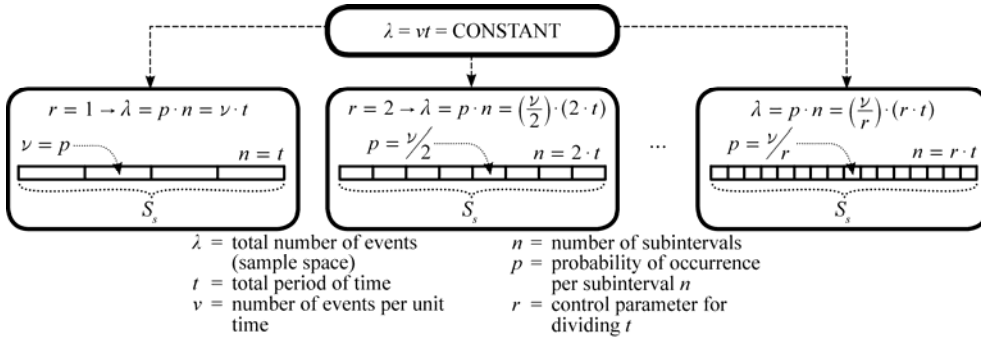


Fig. 5 – Determination of exponential distribution.

Replacing the following quantities $p = v/r$ and $n = r \cdot t$ into the geometric CDF equation, one could obtain:

$$\begin{aligned}
 1 - (1 - p)^n &= 1 - \left(1 - \frac{v}{r}\right)^{t \cdot r} \xrightarrow[r \rightarrow \infty]{r \rightarrow \infty} \lim_{r \rightarrow \infty} \left\{ 1 - \left[\left(1 - \frac{v}{r}\right)^r \right]^t \right\} \\
 &= 1 - (e^{-v})^t = 1 - e^{-vt},
 \end{aligned} \quad (13)$$

where the following identity was applied: $e^{-v} = \lim_{r \rightarrow \infty} \left(1 - \frac{v}{r}\right)^r$.

The probability density function of exponential distribution is acquired using the definition $F_T(t) = \int_0^t f_T(t) dt \Rightarrow f_T(t) = \frac{dF_T(t)}{dt}$ [3].

$$PDF: \quad P(T=t) = f_T(t) = \frac{dF_T(t)}{dt} = D_t \left(1 - e^{-vt}\right) = ve^{-vt}. \quad (14)$$

The mean of exponential distribution is given by:

$$\mu_{\text{geometric}} = \frac{1}{p} \longleftrightarrow \mu_{\text{exponential}} = \frac{1}{\nu} \tag{15}$$

It is emphasized once again that the time between two successes rendered by random variable T represents the random variable in case of geometric distribution while the number of k successes over the entire period t stands for the random variable of binomial distribution. Fig. 6 presents a comparison between geometric and exponential distribution using various probability of success p .

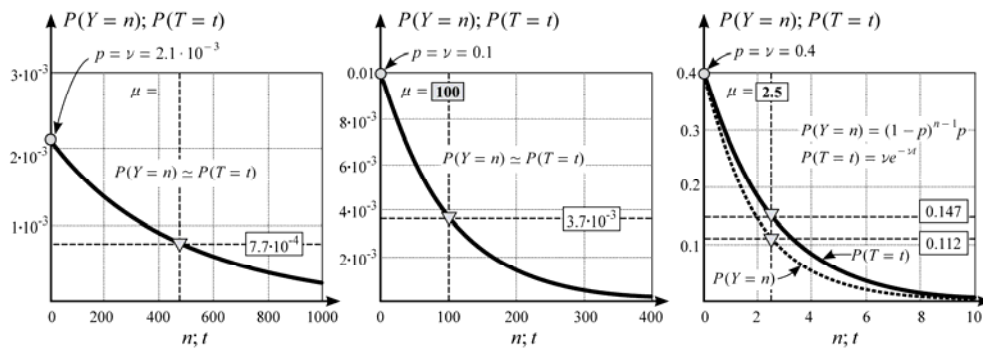


Fig. 6 – Comparison between geometric and exponential distribution.

The two distributions are almost equivalent for small values of ν and gets divergent as ν increases.

4. POISSON DISTRIBUTION: LIMITING CASE OF BINOMIAL DISTRIBUTION

The previous constrain employed to geometric distribution determination is now removed, and instead of tracking probabilities for $(i-1)$ failures before first i success we are now looking for k successes out of n trials. The success may now occur in any of the n trials. The framework related to the Poisson process is maintained.

The binomial distribution provides the probability of occurrence of the number of k successes (earthquakes) in a sequence of n independent trials (years). The experiment handles only binomial type events (such as yes/no), hence the seismic magnitude of the event has no relevance but only the occurrence or nonoccurrence of an earthquake. The method may be extended later on to earthquakes with magnitude greater than a specific threshold or within a particular range but preserving the binomial character of the event.

The binomial distribution is built on a sample space S_S consisting of different sets each of each encloses n trials with k successes, hence each set correspond to

one sample point S_p . Counting all possible combinations of a given k earthquakes that may be initiated in any of the n years, and observing that one sub-interval (1 year) may contain one earthquake at most, then the following well-known relation may be used [4]:

$$C_n^k = \frac{n!}{(n-k)!k!} \tag{16}$$

Fig. 7 shows a schematic of the possible combinations for a given number of k earthquakes, where λ is the mean total number of earthquakes which may occur during n years. The probability of each combination is constant $p^k q^{(n-k)}$.

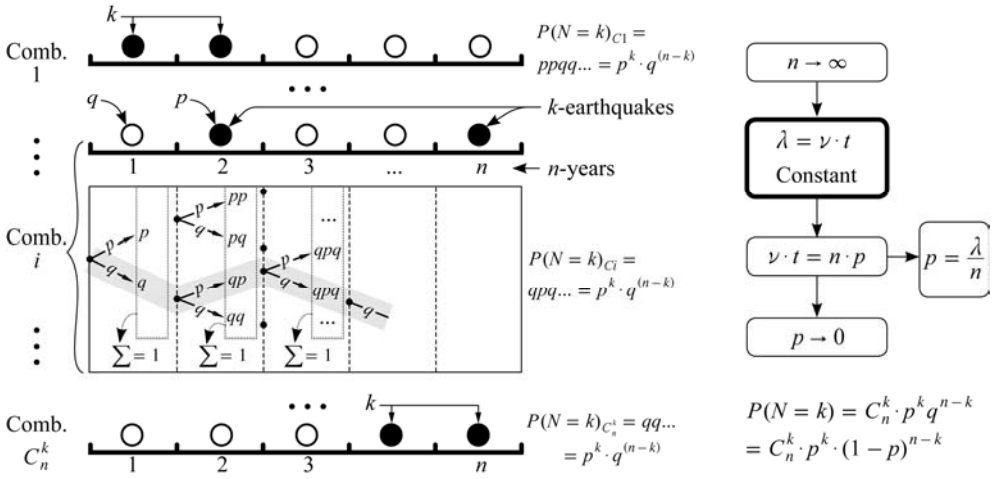


Fig. 7 – Possible combinations for a given number of k earthquakes. Determination of the Poisson distribution.

The probability density function of binomial distribution is given by the following expression:

$$P(N = k \text{ earthquakes in a sequence of } n \text{ years}) = C_n^k p^k (1 - p)^{n-k} \tag{17}$$

There is a justified reason for this notation $\lambda = np = \nu t$ as the mean of binomial distribution is in fact the product np . This is rather apparent as the mean frequency of earthquake occurrence per year is p (or ν) and thus in n years we may get np earthquakes. The probability density function is obtained by variation the parameter $k = 0 \dots n$, which further yields individual scenarios with $k = 0 \dots n$ earthquakes.

It is important to notice the conceptual distinction between ν which stands for the mean success rate associated to any subinterval and thus it may hypothetically yield values even larger than 1 (which in turn means that it cannot be a probability), and p on the other side, representing the probability of success for a small enough subinterval, chosen so that p being less than 1 (see Fig. 12 for the relationship between p and ν). However, both values indicate the same fact: what is the likelihood for an earthquake occurrence within that particular subinterval. Both values are calculated in the same way by means of the ratio of total number of earthquakes λ to the total number of trials (n or t):

$$p = \frac{\lambda}{n} \quad \text{or} \quad \nu = \frac{\lambda}{t}. \quad (18)$$

Observing Fig. 7, p gets smaller as n increases as long as λ remains unchanged, hence in order to show that the Poisson distribution is a limiting case of binomial distribution the following proof could be written:

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{n!}{(n-k)!k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ \xrightarrow{n \rightarrow \infty} & 1 \cdot \left(\frac{\lambda^k}{k!}\right) \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \frac{(\nu t)^k}{k!} \cdot e^{-(\nu t)}, \end{aligned} \quad (19)$$

where the following limits were applied:

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \frac{n \cdot (n-1) \dots (n-(k-1))}{n^k} \cdot \frac{(n-k)!}{(n-k)!} \\ \xrightarrow{n \rightarrow \infty} \lim_{n \rightarrow \infty} & \left[1 \cdot \left(1 - \frac{1}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \right] = (1 \cdot 1 \dots 1)_{k\text{-terms}} = 1, \end{aligned} \quad (20)$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= y \xrightarrow{\ln} \ln y = \lim_{n \rightarrow \infty} \ln \left[\left(1 - \frac{\lambda}{n}\right)^n \right] = \lim_{n \rightarrow \infty} \left[n \cdot \ln \left(1 - \frac{\lambda}{n}\right) \right] \\ &= -\lambda \lim_{n \rightarrow \infty} \left[\frac{1}{1 - \frac{\lambda}{n}} \right] = -\lambda \Rightarrow \ln y = -\lambda \Rightarrow y = e^{-\lambda}, \end{aligned} \quad (21)$$

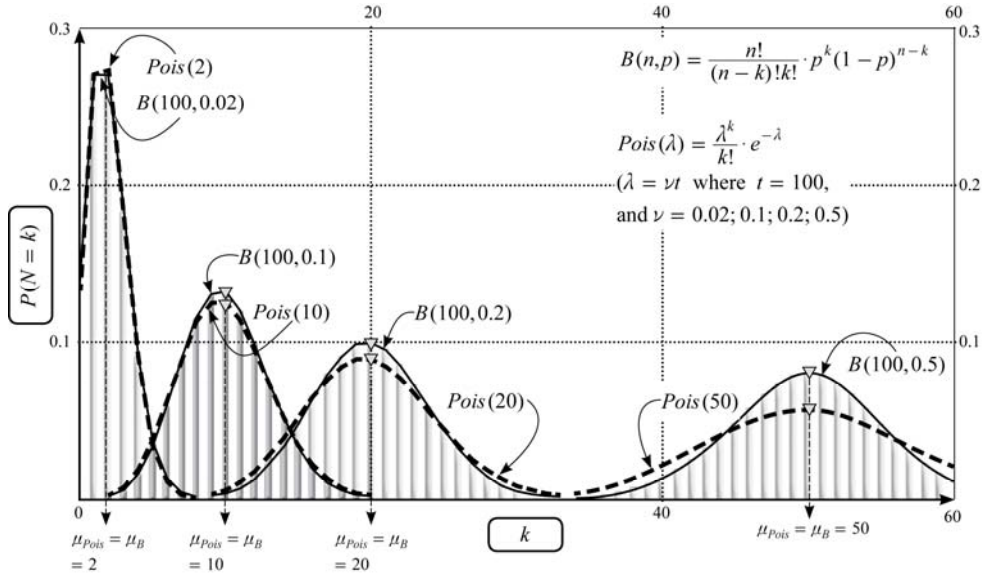


Fig. 8 – Comparison between binomial and Poisson distribution for various $\lambda = 2, 10, 20, 50$.

The Poisson distribution is usually employed instead of its binomial counterpart mainly due to convenience of handling a simpler expression, even if this approximation leads to less accurate results once the entire period t is divided into fewer n subintervals.

5. RELATIONSHIP BETWEEN POISSON AND EXPONENTIAL DISTRIBUTION

The Poisson distribution can be viewed as arising from the binomial distribution or from the exponential density. Therefore, in order to conclude the discussion is necessary to show its connection with the latter.

Fig. 9 summarizes the deduction of relationship between Poisson and exponential distribution, which basically indicates that waiting time distribution of a Poisson process is exponential.

The $k = 0$ corresponds to the probability associated to nonoccurrence of any event within $0 \dots t$, which means that if there is no success along the entire period t , then clearly the time between two successes is greater than t :

$$P(k=0) = P(\text{time between two successes} > t) = P(T > t). \quad (22)$$

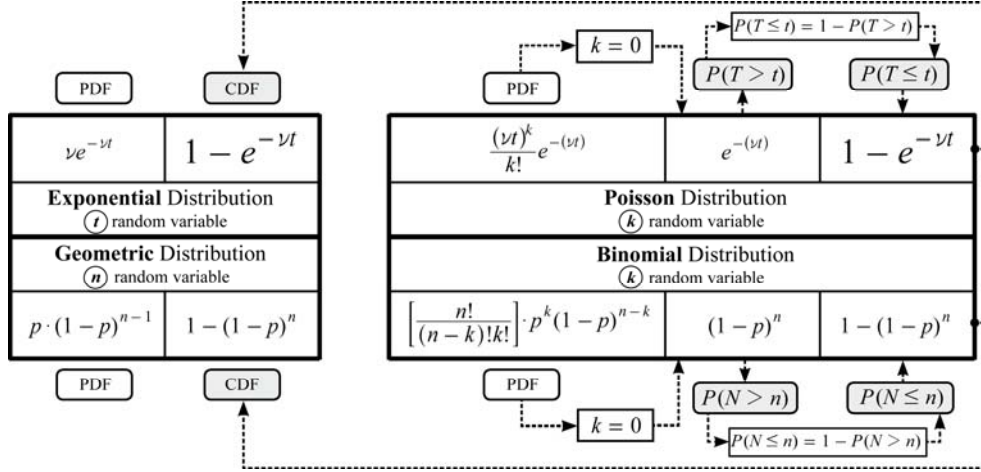


Fig. 9 – Relationship between Poisson and exponential distribution.

The complementary probability $P(T \leq t) = 1 - P(T > t)$ based on equation (3), indicates the likelihood of time between two successes being less than t , or in other words the odds of getting at least one event within $0 \dots t$. It can easily be noticed that the probability $P(T \leq t)$ obtained from Poisson distribution is in fact the exponential cumulative distribution function $1 - e^{-\nu t}$.

Fig. 10 illustrates the connection between Poisson PDF and exponential CDF. No integration has been involved as the Poisson distribution is discrete (the PDF is valid only for $k \in \mathbb{N}$) same as geometric and binomial. On the contrary, the exponential distribution is a continuous density function on the entire $t \in \mathbb{R}^+$.

All calculations were done maintaining unchanged the mean success rate ν while the distributions behavior was traced only by varying the period t .

The relationship between Poisson and exponential distribution is given by the following expression:

$$\overbrace{P(k > 0)}^{\text{Poisson Distribution}} \text{ for a given } t = \overbrace{P(T \leq t)}^{\text{Exponential Distribution}} = 1 - e^{-\nu t}. \quad (23)$$

This relation indicates that the probability of occurrence for one event within a given period t is the same thing as determining the likelihood that a time elapsed between two successive occurrences is at most t .

Relation (23) may be also employed to determination of return period T_r [5] given the exposure time $t = T_{exp}$ and probability $P(k > 0)$ of having at least one

success if $t = T_{exp}$. The inverse of success rate ν is the return period T_r according to equation (15).

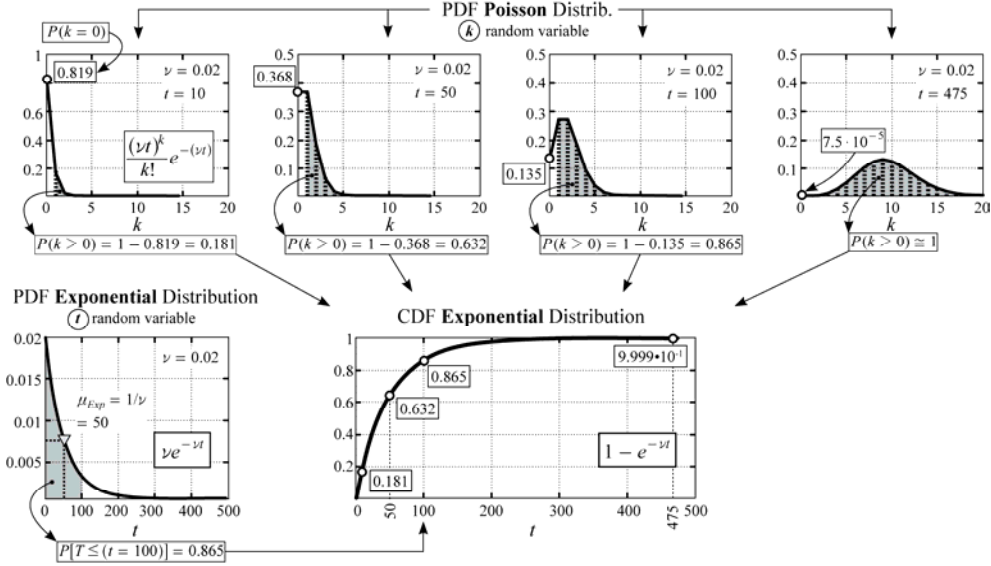


Fig. 10 – The connection between Poisson PDF and exponential CDF.

$$\begin{aligned}
 & \text{Poisson} \\
 & \text{Distribution} \\
 & \overbrace{P(k > 0)}^{\text{for given } t=T_{exp}} = P(k > 0 | T_{exp}) = 1 - e^{-\nu T_{exp}} \Rightarrow e^{-\nu T_{exp}} = 1 - P(k > 0 | T_{exp}) \\
 & T_r = \frac{1}{\nu} = \frac{-T_{exp}}{\ln[1 - P(k > 0 | T_{exp})]}. \quad (24)
 \end{aligned}$$

Assuming that the period is measured in years then for small values of mean success rate, the annual probability $P(k > 0 | T_{exp} = 1)$ could be approximated by the mean rate of success/occurrence (in this case annual frequency) ν . Therefore, the annual probability and the mean annual rate are used virtually interchangeably in common PSHA [6].

This comparison is self evident as the exponential distribution is the continuous version of geometric distribution. Therefore, the annual probability $P(k > 0 | T_{exp} = 1)$ represents in fact exactly the probability of occurrence of an event within a subinterval (in this case 1 year) denoted hereinbefore with p . This probability equals the mean rate of success ν if $n \rightarrow \infty$ (Fig. 5).

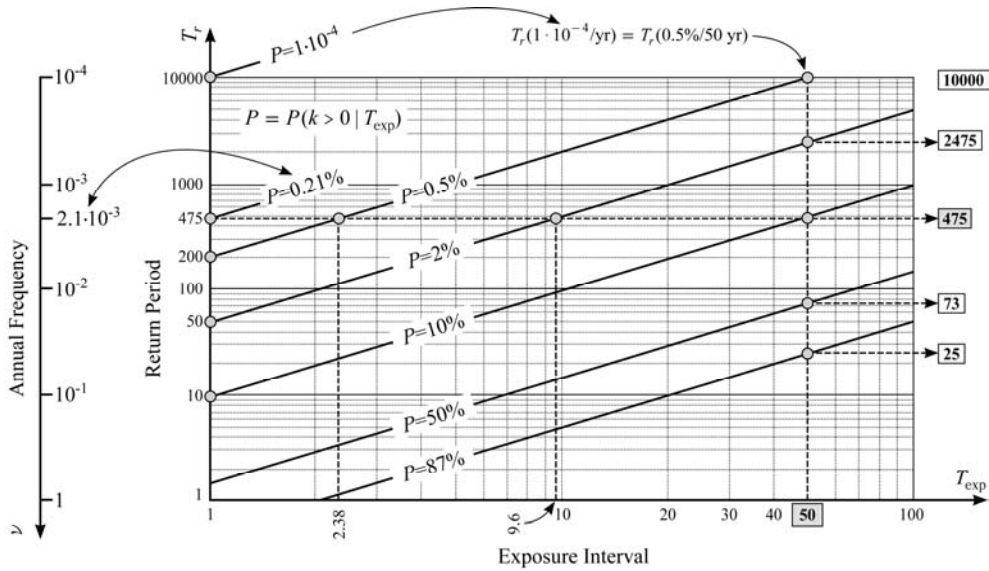


Fig. 11 – Key values for the return period T_r .

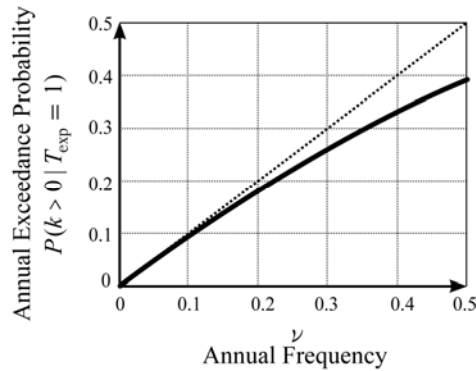


Fig. 12 – Comparison between annual exceedance probability and annual frequency.

REFERENCES

1. ELISHAKOFF, I., *Probabilistic Theory of Structures*, Second Edition, Mineola, Dover Publications, New Work, 1999.
2. HALDAR, A., MAHADEVAN, S., *Probability, Reliability and Statistical Methods in Engineering Design*, John Wiley & Sons, New York, 2000.
3. LUNGU, D., GHIOCEL, D., *Metode Probabilistice in Calculul Constructiilor*, Edit. Tehnică, Bucharest, 1982.
4. CHONG, C.C., MENG, K.K., *Principles and Techniques in Combinatorics*, World Scientific Publishing, Singapore, 1992.

5. GUMBEL, E.J., *Statistics of Extremes*, Columbia University Press, Dover Publications inc., New York, 2004, 1958.
6. SENIOR SEISMIC STANDARD COMMITTEE (SSHAC), *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, Electric Power Research Institute (NUREG/CR-6372), Washington, 1997.
7. HWEI, P.H., *Probability, Random Variables and Random Processes*, McGraw-Hill, New Jersey, 1997.

Received on February 18, 2012