Thinning process algorithms for compound poisson process having nonhomogeneous poisson process (NHPP) intensity functions

S Abdullah1, F Ikhsan2, S Ula1 and Y Rukmayadi1
1 Department of Mechanical Engineering, University of Sultan Ageng Tirtayasa, Jl. Jenderal Sudirman Km 03 Cilegon, Banten, Indonesia 42435
2 Department of Metallurgy Engineering, University of Sultan Ageng Tirtayasa

E-mail: abdullahsyarifayis@gmail.com

Abstract. One stochastic process that is often used to model real phenomena is the compound Poisson process (CPP). CPP is a process in which a component in the process of the events occurred is assumed to be a Poisson process with a certain intensity function (homogeneous or nonhomogeneous). Thinning process algorithm is usually used to generate events that occurred in the Poisson process, but not yet in CPP. This study aims to find out the algorithm to produce CPP which has the function of Poisson nonhomogeneous (NHPP) intensity, where in addition to knowing the process of events that occur, it also takes into account the extent of the consequences of these events. The value of the load caused by the Poisson process is assumed to be a family of i.i.d random variables and the variables are also independent of the Poisson process. The results of this study have obtained the thinning process algorithm and its generalizations for compound Poisson process having nonhomogeneous Poisson process (NHPP) intensity functions. This algorithm is the result of theoretical development and analysis of computational simulations that can be applied in various fields of science such as the analysis of reliability and risk models.

1. Introduction
Reliability models that take into account random failures are usually based on stochastic processes. A Stochastic process is a process that describes an event or phenomenon that is uncertain. This process is useful for modelling phenomena related to probability rules. Stochastic processes can be classified into discrete stochastic processes and stochastic processes with continuous time. One continuous stochastic process is a Poisson process. This process is classified into two, namely homogeneous and nonhomogeneous Poisson process. The homogeneous Poisson process (HPP) can be described as the number of failures during a certain operating period or within a certain range that changes randomly for the next operating period or the next interval has a constant intensity function, while the NHPP has a non-constant intensity function (depending on time). One example of the application of homogeneous and nonhomogeneous Poisson processes in the engineering field can be seen in [1], [2] and [3]. The stochastic process theory allows random evolution modelling that does not only count the number of events that occur or fail at certain times or intervals but also can calculate corrective maintenance costs or the size of the consequences that occur. One stochastic process that is often used for calculating this problem is the compound Poisson process.
The compound Poisson process (CPP) is useful in modelling real phenomena. CPP is a process in which a component in the process of the events occurred is assumed to be a Poisson process with a certain intensity function (homogeneous or nonhomogeneous). One example of the application of this process is in the fields of insurance and finance, physics, demography, seismology, biology (see the references in [4]) and engineering ([5] and [6]). In this application, CPP mostly uses the assumption of an HPP. If an event has a greater chance to occur at a certain time interval compared to another time, then this assumption is not appropriate. Therefore, if an event that occurred is not constant (in the form of a time-dependent function), then we must use a CPP which has an NHPP. Furthermore, if the event is periodic or cyclic, then we must use a compound cyclic Poisson process (CCPP), a process where its Poisson component has a periodic intensity function. The period of this process can be one day, one week, one year, or in the form of another unit. The cyclic Poisson process, the intensity of the previous function is the same as that in the next period. Hence, this process is useful for predicting an event in a subsequent period. The distribution of a CPP is very difficult to find, so the first step is to find its mean and variant. Because this process depends on time, the mean and its variant are functions. The estimation of the mean and variance functions of a CCPP has been investigated by Ruhiyat et al. [7], Mangku et al. [8] and Makhmudah et al. [9]. Furthermore, the study is extended to a CCPP in the presence of linear trends. This process is very appropriate in describing the phenomenon that occurs periodically, but the increase follows the linear trend. The estimation of the mean and variance functions of this process has been investigated by Wibowo et al. [10] and Abdullah et al. [4].

Generating random numbers of events that occur from an HPP can be done by generating the first $T$ time units. While the most famous method for generating an NHPP is thinning process, where this process is a method of acceptance-rejection. This process was introduced by Lewis and Shedler [11]. The thinning process algorithm still does not discuss generating random numbers from a CPP. Thus, the study aims to construct the algorithm for generating CPP that have NHPP intensity functions. This research is very important because in order to see and approach the actual phenomenon, an appropriate random number must be generated.

2. Compound Poisson process having a nonhomogeneous Poisson process intensity functions (CPP-NHPP)

The counting process $\{N(t), t \geq 0\}$, which represents the total number of events that occur by time $T$, is said to be HPP if it has an intensity function in the form of a constant function $\lambda > 0$. Whereas if it has an intensity function in the form of non-constant function $\lambda(t) > 0$, it is said to be NHPP. This means that the HPP does not depend on time because it is a constant function, whereas in NHPP it is a function that depends on the time. The definitions and theorems about HPP and NHPP can be seen in Ross [12].

The expectation and variance of HPP respectively as follows

$$E[N(t)] = \lambda t, \text{Var}[N(t)] = \lambda t. \quad (1)$$

While the expectation and variance of NHPP respectively as follows

$$E[N(t)] = \Lambda(t), \text{Var}[N(t)] = \Lambda(t), \quad (2)$$

with $\Lambda(t) = \int_0^t \lambda(s) \, ds$. The CPP is generalizations of the Poisson process. The definition of a CPP, as follows

**Definition 2.1.** A stochastic process $\{Y(t) \geq 0\}$ is said to be a CPP if it can be represented as

$$Y(t) = \sum_{i=1}^{N(t)} X_i, \ t \geq 0, \quad (3)$$

where $\{N(t), t \geq 0\}$ is a Poisson process, and $\{X_i, i \geq 1\}$ is a family of independent and identically distributed (i.i.d) random variables that is also independent of $\{N(t), t \geq 0\}$ [12].

CPP can be classified into two types based on the component of the Poisson process, which is a CPP that has an HPP intensity function (CPP-HPP) and a CPP that has an NHPP intensity function (CPP-
NHPP). Let $E[X_1] = \mu_1$, $E[X_1^2] = \mu_2$. From equation (1), (3) and with simple proofing, we get the mean and variance function of CPP-HPP, denoted by $\psi_1(t)$ and $V_1(t)$ respectively, given by

$$\psi_1(t) = E[Y(t)] = E[N(t)]E[X_1] = \lambda t \mu_1$$
$$V_1(t) = \text{Var}(Y(t)) = E[N(t)]E[X_1^2] = \lambda t \mu_2.$$  

(4) (5)

While from equation (2), (3) and with simple proofing also obtained the mean and variance function of CPP-NHPP, $Y(t)$, denoted by $\psi_2(t)$ and $V_2(t)$ respectively, given by

$$\psi_2(t) = E[Y(t)] = E[N(t)]E[X_1] = \Lambda(t) \mu_1$$
$$V_2(t) = \text{Var}(Y(t)) = E[N(t)]E[X_1^2] = \Lambda(t) \mu_2.$$ 

(6) (7)

with $\Lambda(t) = \int_0^t \lambda(s) \, ds$. It should be noted because this process depends on time, the mean and its variance are functions of time as well.

The algorithm for generating random numbers from events that occur from an HPP can be done by generating the first $T$ time unit [12] and [13]. The most famous method for generating an NHPP is the thinning process. The intuitive idea behind thinning process is to first find a constant intensity function $\lambda$ which dominates the desired intensity function $\lambda(t)$, next generate from the implied HPP with intensity function $\lambda$, and then reject an appropriate fraction of the generated events so that the desired intensity function $\lambda(t)$ is achieved. The basis of this process can be seen in [11]. Ross [12] and [13] provides modification to the thinning method to mitigate excessive rejection. Furthermore, in this paper, the algorithm and its generalizations will be modified and developed for generating CPP-HPP and CPP-NHPP.

3. Results and Discussion

Previous studies discuss the algorithm for generating random variables only on HPP or NHPP, which is about the events that occur. This research discusses where in addition to knowing the process of events that occur, it also calculates the extent of the consequences of these events, which is by obtaining algorithms about CPP-HPP and CPP-NHPP. So that the previous algorithms will be modified and developed into more complex algorithms.

3.1. Algorithm for generating CPP-HPP

According to the form of the CCP in equation (3), for generating a CCP-HPP has three major steps: (i) generating an HPP (say $Z_t$), (ii) generating a family of i.i.d random variables ($X_i$) and (iii) counting CCP ($Y(T)$). According to the definition of HPP and the definition in equation (3), suppose a counting process $\{N(t), t \geq 0\}$ is defined on a space $\Omega$, such that for each $\omega \in \Omega$, and defined on a probability space $(\Omega, F, P)$. This process has an intensity function $\lambda$ and observed. The function $N(\omega)$ is a realization of the number of events that occur in the interval $(0, t)$, with $N(0) = 0$.

Suppose that for each data point in the observed realization $N(\omega)$ is corresponding with a family of i.i.d random variable $X_i$ that have a $G$ distribution. The algorithm to generate an HPP of $Z_t$ can be seen in [13]. Every time $Z_t$ is obtained, then $Z_t$ bring in a single realization to be able to start generating $X_i$ and giving the value by the distribution $G$. Suppose we already have an algorithm for generating $G$. So $Y(T)$ will also be obtained by calculating the sum of each the value of the load of $X_i$. The steps of the algorithm to generate the CPP-HPP as follows.

| Step | Instructions | Step | Instructions |
|------|--------------|------|--------------|
| 1 :  | $t = 0, I = 0, Y = 0$. | 5 :  | Generate $X$ distributed as $G$. |
| 2 :  | Generate $U$. | 6 :  | Set $I = I + 1, Y = Y + X$. |
| 3 :  | $t = t - \frac{1}{\Lambda} \log(U)$. | 7 :  | Go to step 2. |
| 4 :  | If $t > T$, stop. |      |              |

The steps to generate the CPP-HPP describe in the flowchart in figure 1 as follows.
The meaning of the symbols in the algorithm and the flowchart are: \( t \) represented as time, \( I \) represented a number of events that have occurred by time \( t \), \( Y \) represented as random number, \( \lambda \) represented as intensity function of HPP that is constant function, \( T \) represented as first \( T \) time units and \( X \) represented as random variables that distributed as \( G \) which has i.i.d properties. The final values \( I \) and \( Y \) represent the number of events that occur by time \( T \) and represent the total amount of sequence of \( X_i \), respectively. If the reader wants to show the time of the event, it can be added in Step 6.

3.2. Thinning process algorithm for generating CPP-NHPP type 1

According to the form of the CCP in equation (3), for generating a CCP-NHPP type 1 has three major steps: (i) generating an NHPP (say \( Z_i \)), (ii) generating a family of i.i.d random variables \( (X_i) \) and (iii) counting CCP \( (Y(T)) \). According to the definition of NHPP and the definition in equation (3), suppose a counting process \( \{ N(t), t \geq 0 \} \) is defined on a space \( \Omega \), such that for each \( \omega \in \Omega \), and defined on a probability space \( (\Omega, F, P) \). This process has an intensity function \( \lambda(t) \) and observed. The function \( N(\omega) \) is a realization of the number of events that occur in the interval \( (0,t] \), such that \( N(0) = 0 \).

Suppose that for each data point in the observed realization \( N(\omega) \) is corresponding with a family of i.i.d random variable \( X_i \) that have a \( G \) distribution. The algorithm to generate a NHPP of \( Z_i \) type 1 can be seen in [13]. Remember that to generate NHPP, we must choose the value \( \lambda \) so that \( \lambda(t) \leq \lambda \) for all \( t \leq T \) and with probability \( \lambda(t)/\lambda \). Every time \( Z_i \) is obtained, then \( Z_i \) bring in a single realization to be able to start generating \( X_i \) and giving the value by the distribution \( G \). Suppose we already have an algorithm for generating \( G \). So \( Y(T) \) will also be obtained by calculating the sum of each the value of the load of \( X_i \). The following preposition is the basis of the method for generating CPP-NHPP with a thinning process.

**Preposition 1.** Consider form of CCP-NHPP in equation (3). Suppose \( Y(T) \) is representing CCP, \( X'_1, X'_2, ..., X'_n \) are representing a family of i.i.d random variables that is independent of \( \{ N(t), t \geq 0 \} \), and \( T'_1, T'_2, ..., T'_n \) are random variables representing event times from the NHPP with intensity function, \( \lambda'(t) \) for all \( t \in [0, t_0] \). Let \( \lambda(t) \) be an intensity functions such that \( 0 \leq \lambda(t) \leq \lambda'(t) \) for all \( t \in [0, t_0] \). If the \( i \)th event time \( T_i' \) is independently deleted with probability \( 1 - \lambda(t)/\lambda'(t) \) for \( i = 1, 2, ..., n \), then the remaining event time of NHPP with the intensity function \( \lambda(t) \) in the interval \( (0, t_0] \), bring in an observed realization of \( X'_j \) for \( j = 1, 2, ..., n \) and \( Y(T) \) can also be calculated at that interval.

The steps to generate CPP-NHPP type 1 are described in the flowchart in figure 2 as follows
The steps of the algorithm to generate the CPP-NHPP type 1 as follows

| Step | Instructions | Step | Instructions |
|------|--------------|------|--------------|
| 1    | $t = 0, I = 0, Y = 0$. | 6    | If $U_2 \leq \lambda(t)/\lambda$. |
| 2    | Generate $U_1$. | 7    | Set $I = I + 1$. |
| 3    | Set $t = t - \frac{1}{\lambda} \log(U_1)$. | 8    | Generate $X$ distributed as $G$. |
| 4    | If $t > T$, stop. | 9    | Set $Y = Y + X$. |
| 5    | Generate $U_2$. | 10   | Go to step 2. |

The meaning of the symbols in the flowchart and the algorithm are: $t$ represented as time, $I$ represented as a number of events that have occurred by time $t$, $Y$ represented as CPP-NHPP, $U_1$ and $U_2$ represented as random numbers, $\lambda(t)$ represented as intensity function of NHPP that is non-constant function with $\lambda(t) \leq \lambda$, $T$ represented as first $T$ time units and $X$ represented as random variables that distributed as $G$ which has i.i.d properties. The final values $I$ and $Y$ represent the number of events that occur by time $T$ and represent the total amount of sequence of $X_i$, respectively. If the reader wants to show the time of the event, it can be added in Step 7-9.

3.3. **Thinning process algorithm for generating CPP-NHPP type 2**

In this section, the algorithm in type 1 is modified and developed by dividing the interval into subintervals. After being divided, procedures are used such as type 1. This procedure makes sense because it has fewer numbers of rejected events when divided into several subintervals when $\lambda(t)$ approaches $\lambda$ along with the interval. That is, determine appropriate values $k, 0 = t_0 < t_1 < t_2 < \cdots < t_k < t_{k+1} = T, \lambda_1, \lambda_2, \ldots, \lambda_k, \lambda_{k+1}$ such that

$$\lambda(s) \leq \lambda_i \text{ if } t_{i-1} \leq s < t_i, i = 1, 2, \ldots, k, k+1. \quad (8)$$

According to the form of the CCP in equation (3), for generating a CCP-NHPP type 2 has four major steps: (i) break the interval into subinterval, (ii) generating an NHPP (say $Z_i$), (iii) generating a family of i.i.d random variables ($X_i$) and (vi) counting CCP ($Y(T)$). Suppose that for each data point in the observed realization $N(\omega)$ is corresponding with a family of i.i.d random variable $X_i$ that have a $G$ distribution. The algorithm to generate an NHPP of $Z_i$ can be seen in [13]. Remember that to generate NHPP, we must choose the value $\lambda_i$ so that $\lambda(s) \leq \lambda_i$ for all $s \in (t_{i-1}, t_i)$ and with the probability $\lambda(t)/\lambda_i$. Because exponential has no memory, use $\lambda_i[Z - (t_i - t)]/\lambda_{i+1}$ for the next exponential form with the rate $\lambda_{i+1}$. Every time $Z_i$ is obtained, then $Z_i$ bring in a single realization to be able to start
generating $X_i$ and giving the value by the distribution $G$. Suppose we already have an algorithm for generating $G$. So $Y(T)$ will also be obtained by calculating the sum of each the value of the load of $X_i$. The steps to generate the CPP-NHPP type 2 describe in the flowchart in figure 3 as follows.

![Flowchart of algorithm for generating CPP-NHPP type 2](image)

**Figure 3.** Flowchart of algorithm for generating CPP-NHPP type 2.

The steps of the algorithm to generate the CPP-NHPP type 2 as follows

| Step | Instructions | Step | Instructions |
|------|--------------|------|--------------|
| 1    | $t = 0, J = 1, I = 0, Y = 0$. | 9    | Generate $X$ distributed as $G$ |
| 2    | Generate $U_1$. | 10   | Set $Y = Y + X$. |
| 3    | Set $Z = -\frac{1}{\lambda_j} \log(U_1)$. | 11   | Go to Step 2. |
| 4    | If $t + Z > t_j$, go to Step 12. | 12   | If $J = k + 1$, stop. |
| 5    | Set $t = t + Z$. | 13   | Set $Z = \frac{(2-t_j+\epsilon)\lambda_j}{\lambda_j + 1}$, |
| 6    | Generate $U_2$. | 14   | Set $t = t_j$. |
| 7    | If $U_2 \leq \frac{\lambda(t)}{\lambda_j}$. | 15   | Set $J = J + 1$. |
| 8    | Set $I = I + 1$. | 16   | Go to Step 4. |

The meaning of the symbols in the flowchart and the algorithm are: $t$ represented as present time, $J$ represented as present interval, $I$ represented a number of events that have occurred by time $t$, $Y$ represented as CPP-NHPP, $U_1$ and $U_2$ represented as random number, $\lambda(t)$ represented as intensity function of NHPP that is non-constant function with $\lambda(t) \leq \lambda$ and $X$ represented as random variables that distributed as $G$ which has i.i.d properties. The final values $I$ and $Y$ represent the number of events that occur by time $T$ and represent the total amount of sequence of $X_i$, respectively. If the reader wants to show the time of the event, it can be added in Step 8-10.

The algorithms above are used by assuming that the random variable $X_i$ is a family of i.i.d random variables, such as exponential distribution, Gamma distribution or others. So that the algorithms can be modified and developed. The first algorithm (CPP-HPP) can be used to solve the equation (4) and (5). Whereas the second and third algorithms (CPP-NHPP Type 1 and 2) can be used to solve the equations (6) and (7). These algorithms can also be used to simulate and estimate the mean and variance functions of a CCP when we have the estimator. Furthermore, these algorithms can also be used in maintenance planning production, predicting the cost of corrective maintenance, the reliability model analysis, determining the price of premiums for policies, risk analysis for company products, determining reserve benefits for companies and in applied various fields of science as mentioned earlier.
4. Conclusions
The results of this study have obtained algorithms and generalizations for compound Poisson processes. The first algorithm is used to generate compound Poisson process having homogeneous Poisson process intensity functions (CPP-HPP), where it is a constant function. The second algorithm, the thinning process algorithm (CPP-NHPP Type 1), is used to generate compound Poisson process having nonhomogeneous Poisson process intensity functions, where it is a non-constant function. The second algorithm is modified and developed by breaking the interval into subintervals. From these results, the third algorithm (CPP-NHPP Type 2) is obtained.

Acknowledgments
The authors would like to thank referees for remarks and suggestions for improving the quality of this paper, and also thank the editor for improving the appearance of this paper.

References
[1] Pahlajani CD, Poulakakis I, Herbert G and Tanner HG 2014. Networked decision making for Poisson processes: Application to nuclear detection. *IEEE Transaction on Automatic Control*. 59(1):193-198. doi.org/10.1109/TAC.2013.2267399.
[2] Cha JH 2013. On a stochastic failure model under random shocks. International Conference on Mathematical Modelling in Physical Sciences IOP Publishing. *Journal of Physics: Conference Series*. 410 (2013) 012108. doi:10.1088/1742-6596/410/1/012108.
[3] Soltanali H, Rohani A, Tabazizadeh M, Abbaspour-Fard MH and Parida A 2019. Operational reliability evaluation-based maintenance planning for automotive production line. *Quality Technology & Quantitative Management. Taylor and Francis Group*. doi:10.1080/16843703.2019.1567664.
[4] Abdullah S, Mangku IW and Siswadi 2017. Estimating the variance function of a compound cyclic Poisson process in the presence of linear trend. *Far East Journal of Mathematical Sciences*. 102(3):559-572. doi.org/10.17654/MS102030559.
[5] Andrzejczak K, Młyńczak M and Selech J 2018. Poisson-distributed failures in the predicting of the cost of corrective maintenance. *Eksploatacja i Niezawodnosc–Maintenance and Reliability*. 20(4):602–609. doi.org/10.17531/ein.2018.4.11.
[6] Grabski F 2018. Nonhomogeneous stochastic processes connected to Poisson process. *Scientific Journal of Polish Naval Academy*. 213(2):1-15. doi:10.2478/sjpna-2018-0009.
[7] Ruhiyat, Mangku IW and Purnaba IGP 2013. Consistent estimation of the mean function of a compound Poisson process. *Far East Journal of Mathematical Sciences*. 77(2):183-194.
[8] Mangku IW, Ruhiyat and Purnaba IGP 2013. Statistical properties of an estimator for the mean function of a compound cyclic Poisson process. *Far East Journal of Mathematical Sciences*. 82(2):227-237.
[9] Makhmudah FI, Mangku IW and Sumarno H 2016. Estimating the variance function of a compound cyclic Poisson process. *Far East Journal of Mathematical Sciences*. 100(6):911-922. doi.org/10.17654/MS100060911.
[10] Wibowo BA, Mangku IW and Siswadi 2017. Statistical properties of an estimator for the mean function of a compound cyclic Poisson process in the presence of linear trend. *Arab Journal of Mathematical Science*. 23(2):173-185. doi:10.1016/j.ajmsc.2016.08.004.
[11] Lewis PAW and Shedler GS 1979. Simulation Poisson process by thinning. Naval Postgraduate School Monterey [Thesis]. California (US).
[12] Ross SM 2010. *Introduction to Probability Models*. 9th Ed. Florida (US): Academic Press.
[13] Ross SM 2012. *Simulation*. 5th Ed. Amsterdam (NL): Academic Press.