Generation of pseudo-Poisson distributed discrete intervals

Génération d'intervalles discrets avec distribution quasi de Poisson

Bruno CERNUSCHI-FRIAS

Larrea 1065, 1117 CAPITAL FEDERAL, ARGENTINA

Bruno Cernuschi-Frías was born in Montevideo, Uruguay, on April 7, 1952. He received the Ingeniero Electromecánico Orientación Electrónica degree from the Facultad de Ingeniería de la Universidad de Buenos Aires, Argentina, in 1976 and the M. Sc. and the Ph. D. degrees in electrical engineering from Brown University, Providence, RI, in 1983 and 1984 respectively. Since 1984 he has been at FIUBA as a member of the Scientific Research Career of the CONICET, where he is involved in research in computer vision.

Dr. Cernuschi-Frías is a member of ACM, IEEE and AAAS. He has received fellowships from the Organization of American States, Facultad de Ingeniería, and IBM.

Daniel Claudio CASAGLIA

Isaac Newton 1528, 1676 SANTOS LUGARES, ARGENTINA

Daniel Casaglia was born in Buenos Aires, Argentina, on November 23, 1956. He received the Ingeniero Electromecánico Orientación Electrónica degree from the Facultad de Ingeniería de la Universidad de Buenos Aires (FIUBA), in 1981.

In November 1981, he joined the Instituto de Ingeniería Biomédica (FIUBA), where he is now a member of the División de Procesamiento de la Información. He has received fellowships from the CONICET to work in digital processing of electroencephalograms. Since 1983 he is also engaged in post degree teaching.

Mr. Casaglia is a candidate for the Scientific Research Carreer of the CONICET, where he plans to continue his research in information theory and processing.

SUMMARY

A method for generating random discrete intervals with a pseudo-Poisson distribution, is described. The resulting process has the same first and second order statistics as the continuous one before quantization. This is accomplished by using a roundoff coefficient ε and changing the mean value λ of the original process. The method is valid for values of $\lambda T<2$, where T is the quantization interval. The experimental purpose of this method is the simulation of the instants of occurrence of muscle artifacts in an EEG.

KEY WORDS

Quantization, Poisson process, simulation.

RÉSUMÉ

On présente une méthode pour obtenir des intervalles discrets, aléatoires, selon une distribution proche de la loi de Poisson de paramètre λ. On opère en discrétisant des intervalles tirés suivant une loi de Poisson de paramètre λ. Le processus obtenu a les mêmes deux premiers moments que la loi de Poisson que l'on veut approcher. Outre le paramètre λ il dépend d'un « coefficient d'approximation » ε lié à la discrétisation des intervalles. Cette méthode est utilisable sous la condition $\lambda T<2$, T étant le pas de quantification des intervalles. Elle a été mise au point pour permettre la simulation numérique des instants qui, sur un électroencéphalogramme, correspondent à l'apparition de petits mouvements musculaires.

MOTS CLÉS

Quantification, processus de Poisson, simulation.

TABLE DES MATIÈRES

1. Introduction
2. Poisson process
3. Generation of pseudo-Poisson distributed discrete intervals
4. Proposed method
5. Conclusions
References

1. Introduction

The Poisson process is the simplest associated with counting random numbers of events. Its importance is due to the fact that it appears in several specific situations such as: failure analysis [1], queueing theory [2], etc. The interest in simulating this process arises in several cases, as for example, in optical detection [3] where the number of photoelectrons produced on a photosensitive surface has a Poisson distribution. Algorithms for generating these numbers can be found in the literature [4, 5].

In many other cases, it is more important to generate the intervals between Poisson distributed events. For example, in the optimization of damaged device replacement we can estimate the time of failure occurrence. In [1] it is shown that the instants in which shocks producing failures occur, have a Poisson distribution. And also for studying queueing systems, it may be usefull to simulate the arrival times of communications, that are Poisson distributed [2].

Though it is not particularly difficult to generate intervals with a Poisson distribution, a problem arises when those intervals must belong to a discrete space. What we intend is to quantize a Poisson distributed interval that can take any value on $\mathbb{R}^+$. This must be done so that the mean and variance of the quantized process are the same as the first and second moments of the original Poisson process. Now it should be said that, though the distribution of these intervals is determined by only one parameter λ, when quantization is done two parameters ε and λ' are needed to accomplish the desired statistics.

This method was developed for the simulation of the instants of appearence of muscle artifacts in electroencephalograms. These artifacts can be simulated as the impulse response of a second order linear filter. A histogram of their arrival time shows that it may be suitably approximated by an exponential, so the driving process of the filter may be Poisson distributed [6]. For simulation purposes, the time interval between artifact appearences have to be multiples of the sampling interval T, that depends on the data acquisition equipment. This poses the quantization problem of the Poisson distributed intervals.

2. Poisson process

A Poisson process $x(\tau)$ is inherently discrete since it can only take discrete values [7, 8]. Its mathematical description is

$$(1)\qquad \begin{aligned} x(\tau)&=0 \quad \text{for} \quad \tau=0\\ P\{x(\tau)=k\}&=\frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} \quad \text{for} \quad \tau>0,\\ &k\in\mathbb{N}\end{aligned}$$

where $\lambda\tau$ is the mean value of the process in the interval $(0, \tau)$. The number of events in an interval is a random variable, and when two or more of these random variables correspond to nonoverlapping intervals they are statistically independent.

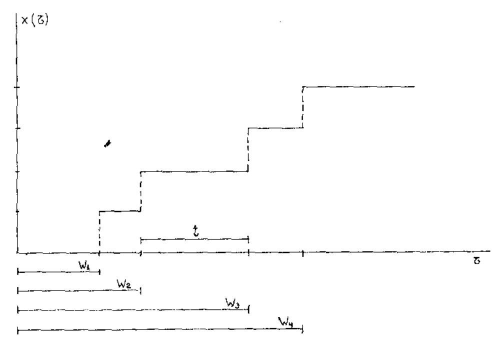

Fig. 1

The waiting time sequence $\{W_n\}$ of the process is shown in Figure 1. There W_n is the interval of time elapsed from the beginning of the observation at $\tau=0$, and the nth event. We define a new random variable t that is the time between consecutive events. The statistics of this random variable t is given by

$$(2)\qquad \begin{cases} F(s)=1-P\{t>s\}=1-P\{x(s)=0\}=1-e^{-\lambda s},\\ \forall s\in\mathbb{R}^+\end{cases}$$

so the distribution and density function are

$$(3)\qquad \begin{cases} F(t)=1-e^{-\lambda t}\\ f(t)=\lambda e^{-\lambda t}\end{cases}$$

and the first two moments

$$(4)\qquad \begin{aligned} E\{t\}&=\frac{1}{\lambda}\\ \sigma^2\{t\}&=\frac{1}{\lambda^2}\end{aligned}$$

3. Generation of pseudo-Poisson distributed discrete intervals

A method for generating Poisson distributed random intervals is described in [9]. To state it briefly, we make

$$(5)\qquad F(t)=1-e^{-\lambda t}=U$$

where U is a random variable with a uniform distribution over [0, 1]. Thus

$$(6)\qquad t=F^{-1}(U)=-\frac{1}{\lambda}\ln(1-U)$$

gives the length of the intervals between consecutive events in a Poisson process. As $(1-U)$ also has a uniform distribution over [0, 1] we can use the equation

$$(7)\qquad t=-\frac{1}{\lambda}\ln(U)$$

As it was mentioned above, these random intervals should be put in terms of a quantization interval T that is a fixed parameter. The quantization of the continuous random variable t gives a new discrete random variable nT. We would like this discrete random variable to have the same first and second order statistics as the continuous one. We say that the discrete random variable nT takes a given value kT if the continuous random variable t falls in the interval $(kT-\varepsilon, (k+1)T-\varepsilon)$, where we define ε as the roundoff coefficient. This coefficient ε is related to T as follows

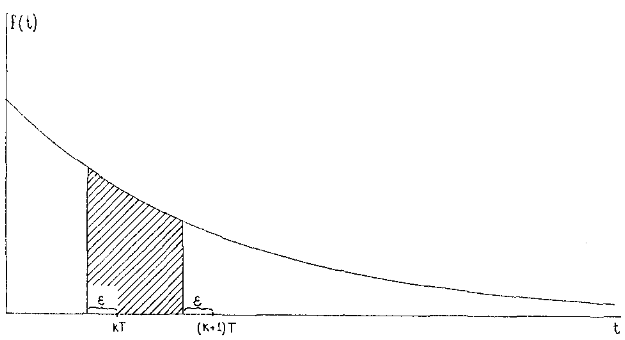

Fig. 2

$$(8)\qquad 0\leqq\varepsilon<T$$

Figure 2 shows that the probability of the value kT is given by the shaded area

$$(9)\qquad P\{nT=kT\}=\int_{kT-\varepsilon}^{(k+1)T-\varepsilon}\lambda e^{-\lambda t}\,dt=e^{\lambda\varepsilon}[1-e^{-\lambda T}]e^{-k\lambda T}$$

and

$$(10)\qquad P\{nT=0\}=\int_0^{T-\varepsilon}\lambda e^{-\lambda t}\,dt=1-e^{-\lambda T}e^{\lambda\varepsilon}$$

The mean value of this discrete variable nT is

$$(11)\qquad \begin{cases} E\{nT\}=\sum_{n=0}^{\infty} nT e^{\lambda\varepsilon}e^{-\lambda nT}[1-e^{-\lambda T}] \\ E\{nT\}=\dfrac{e^{\lambda\varepsilon}Te^{-\lambda T}}{1-e^{-\lambda T}} \end{cases}$$

and the variance

$$(12)\qquad \sigma^2\{nT\}=\frac{T^2e^{\lambda\varepsilon}e^{-\lambda T}}{1-e^{-\lambda T}}+\frac{T^2e^{-2\lambda T}}{(1-e^{-\lambda T})^2}[2e^{\lambda\varepsilon}-e^{2\lambda\varepsilon}]$$

As our purpose is to simulate the random discrete intervals with the same mean and variance as the continuous one, from equations (4), (11) and (12) we get

$$(13)\qquad \begin{cases} \dfrac{e^{\lambda\varepsilon}Te^{-\lambda T}}{1-e^{-\lambda T}}=\dfrac{1}{\lambda} \\ \dfrac{T^2e^{\lambda\varepsilon}e^{-\lambda T}}{1-e^{-\lambda T}}+\dfrac{T^2e^{-2\lambda T}}{(1-e^{-\lambda T})^2}\times[2e^{\lambda\varepsilon}-e^{2\lambda\varepsilon}]=\dfrac{1}{\lambda^2} \end{cases}$$

The problem with this system is that ε is the only free variable and so we cannot solve it for both moments simultaneously.

4. Proposed method

Another degree of freedom should be included. A simple way is to change the parameter λ of the original process to λ' before it is quantized, in order to obtain the desired statistics for the discrete process. By introducing this new parameter λ' instead of the original λ in (11) and (12) we get

$$(14)\qquad \frac{e^{\lambda'\varepsilon}Te^{-\lambda' T}}{1-e^{-\lambda' T}}=\frac{1}{\lambda}$$

$$(15)\qquad \frac{T^2e^{\lambda'\varepsilon}e^{-\lambda' T}}{1-e^{-\lambda' T}}+\frac{T^2e^{-2\lambda' T}}{(1-e^{-\lambda' T})^2}\times[2e^{\lambda'\varepsilon}-e^{2\lambda'\varepsilon}]=\frac{1}{\lambda^2}$$

From (14)

$$(16)\qquad e^{\lambda'\varepsilon}=\frac{1}{\lambda}\,\frac{1-e^{-\lambda' T}}{Te^{-\lambda' T}}$$

and together with (15) we obtain

$$(17)\quad \begin{cases} \mathrm{T}+\dfrac{2\,\mathrm{T}\,e^{-\lambda'\mathrm{T}}}{1-e^{-\lambda'\mathrm{T}}}-\dfrac{1}{\lambda}=\dfrac{1}{\lambda} \\ \dfrac{2\,\mathrm{T}\,e^{-\lambda'\mathrm{T}}}{1-e^{-\lambda'\mathrm{T}}}=\dfrac{2-\lambda\mathrm{T}}{\lambda} \\ \left[2\,\mathrm{T}+\dfrac{2-\lambda\mathrm{T}}{\lambda}\right]e^{-\lambda'\mathrm{T}}=\dfrac{2-\lambda\mathrm{T}}{\lambda} \\ e^{-\lambda'\mathrm{T}}=\dfrac{2-\lambda\mathrm{T}}{2+\lambda\mathrm{T}} \end{cases}$$

The righthand side of this last equation must be positive, so $\lambda\mathrm{T}<2$ appears as a restriction. By taking this into account we have

$$(18)\quad \lambda'=\frac{1}{\mathrm{T}}\ln\frac{2+\lambda\mathrm{T}}{2-\lambda\mathrm{T}}$$

Using (16) and (17) we get

$$(19)\quad \begin{cases} e^{\lambda'\varepsilon}=\dfrac{1}{\lambda}\,\dfrac{1-[(2-\lambda\mathrm{T})/(2+\lambda\mathrm{T})]}{\mathrm{T}[(2-\lambda\mathrm{T})/(2+\lambda\mathrm{T})]}=\dfrac{2}{2-\lambda\mathrm{T}} \\ \varepsilon=\dfrac{1}{\lambda'}\ln\dfrac{2}{2-\lambda\mathrm{T}} \end{cases}$$

Finally using (18) in (19)

$$(20)\quad \varepsilon=\mathrm{T}\frac{\ln[2/(2-\lambda\mathrm{T})]}{\ln[(2+\lambda\mathrm{T})/(2-\lambda\mathrm{T})]}$$

The parameter ε obtained from (20) agrees with the inequality given in (8).

We have mentioned before that the quantization interval T is a fixed parameter. We then write the normalized equations for the proposed method as follows

$$(21)\quad \frac{1}{\mathrm{T}}\mathrm{E}\{n\mathrm{T}\}=\frac{e^{\lambda'\varepsilon}e^{-\lambda'\mathrm{T}}}{1-e^{-\lambda'\mathrm{T}}}$$

$$(22)\quad \frac{1}{\mathrm{T}^2}\sigma^2\{n\mathrm{T}\}=\frac{e^{\lambda'\varepsilon}e^{-\lambda'\mathrm{T}}}{1-e^{-\lambda'\mathrm{T}}}+\frac{e^{-2\lambda'\mathrm{T}}}{(1-e^{-\lambda'\mathrm{T}})^2}[2e^{\lambda'\varepsilon}-e^{2\lambda'\varepsilon}]$$

$$(23)\quad \lambda'\mathrm{T}=\ln\frac{2+\lambda\mathrm{T}}{2-\lambda\mathrm{T}}$$

$$(24)\quad \frac{\varepsilon}{\mathrm{T}}=\frac{\ln[2/(2-\lambda\mathrm{T})]}{\ln[(2+\lambda\mathrm{T})/(2-\lambda\mathrm{T})]}$$

TABLE I

Continuous variable Poisson distribution		Discrete variable pseudo-Poisson distribution					
Normalized mean value $\frac{1}{\lambda\mathrm{T}}$	Normalized variance $\frac{1}{(\lambda\mathrm{T})^2}$	Normalized mean value $\frac{1}{\mathrm{T}}\mathrm{E}\{n\mathrm{T}\}$			*Normalized variance* $\frac{1}{\mathrm{T}^2}\sigma^2\{n\mathrm{T}\}$		
		a	*b*	*c*	*a*	*b*	*c*
0.50001	0.25001	0.15652	0.42547	0.50001	0.18102	0.37764	0.25001
0.55	0.3025	0.19377	0.48096	0.55	0.23132	0.43603	0.3025
1	1	0.58198	0.95952	1	0.92067	1.15568	1
2	4	1.54149	1.97932	2	3.91769	4.16383	4
5	25	4.51666	4.99168	5	24.91683	25.16621	25
10	100	9.50833	9.99583	10	99.91671	100.16655	100
20	400	19.50417	19.99792	20	399.91668	400.16664	400
50	2,500	49.50167	49.99917	50	2,499.91667	2,500.16667	2,500

TABLE II

Continuous variable Poisson distribution		Simulated discrete variable pseudo-Poisson distribution						$\frac{\varepsilon}{\mathrm{T}}$	$\lambda'\mathrm{T}$
Normalized mean value $\frac{1}{\lambda\mathrm{T}}$	Normalized variance $\frac{1}{(\lambda\mathrm{T})^2}$	Normalized mean value $\frac{1}{\mathrm{T}}\mathrm{E}\{n\mathrm{T}\}$			Normalized variance $\frac{1}{\mathrm{T}^2}\sigma^2\{n\mathrm{T}\}$				
		a	*b*	*c*	*a*	*b*	*c*		
0.50001	0.25001	0.1568	0.4256	0.5002	0.175	0.36	0.25	0.9398	11.517529
0.55	0.3025	0.1952	0.481	0.5478	0.219	0.412	0.295	0.7876	3.044523
1	1	0.5824	0.9602	0.9996	0.873	1.107	0.952	0.6309	1.098612
2	4	1.5304	1.9624	1.9856	3.836	4.09	3.923	0.5632	0.510825
5	25	4.514	4.981	4.9896	23.581	23.761	23.657	0.5251	0.20067
10	100	9.495	9.9784	9.983	94.438	94.526	94.356	0.5125	0.100083
20	400	19.4734	19.978	19.98	377.679	377.827	377.65	0.5062	0.05001
50	2,500	49.4408	49.9344	49.936	2,360.899	2,361.94	2,361.75	0.5025	0.02

For each value of λ T equations (23) and (24) give the parameters λ' T and ε/T that are used for generating the discrete random intervals. In this case the discrete variable will have the same first two moments as the continuous one, because that was the purpose of the proposed method. If we rather use the classical methods of truncation or symmetrical roundoff, the normalized mean value and the normalized variance of the discrete random variable may be obtained from equations (21) and (22). Results are shown in Table I where the following three cases are analyzed:

(*a*) Truncation ($\varepsilon/T=0$ and $\lambda' T=\lambda T$).

(*b*) Symmetrical roundoff ($\varepsilon/T=0.5$ and $\lambda' T=\lambda T$).

(*c*) Proposed method.

In order to confirm the theoretical results, a test of the algorithm was run. The algorithm for simulating random discrete intervals with a pseudo-Poisson distribution, was programmed in Fortran and run on a PDP 11/34 with a sample size of 5,000 values. The program computes the normalized mean value and the normalized variance of the simulated discrete intervals, the normalized roundoff coefficient ε/T and the normalized λ' T value. The results are given in Table II.

5. Conclusions

An approach to the generation of pseudo-Poisson distributed discrete intervals was presented. The introduction of the roundoff coefficient ε and the auxiliary λ' value proved to be usefull specially for keeping the variance fit, when $1/\lambda$ T takes the smallest possible values. Besides, for large values of $1/\lambda$ T the normalized roundoff coefficient ε/T tends to 0.5 in accordance with equation (24).

Manuscrit reçu le 6 *janvier* 1986.

REFERENCES

[1] M. Abdel-Hameed and I. N. Shimi, Optimal Replacement of Damaged Devices, *J. Appl. Prob.*, 15, No. 1, March 1978, pp. 153-161.

[2] D. L. Minh, The Discrete-Time Single-Server Queue with Time-Inhomogeneous Compound Poisson Input and General Service Time Distribution, *J. Appl. Prob.*, 15, No. 3, September 1978, pp. 590-601.

[3] S. Karp *et al.*, Communication Theory for the Free-Space Optical Channel, *Proc. IEEE*, 58, No. 10, October 1970, pp. 1611-1626.

[4] R. H. Snow, Algorithm 342, Generator of Random Numbers Satisfying the Poisson Distribution, *Comm. ACM*, 11, (12), 1968, p. 819.

[5] H. E. Schaffer, Algorithm 369, Generator of Random Numbers Satisfying the Poisson Distribution, *Comm. ACM*, 13, (1), 1970, p. 49.

[6] T. L. Johnson *et al.*, Filtering of Muscle Artifacts from the Electroencephalogram, *IEEE Trans. on Biomed. Eng.*, 26, No. 10, October 1979, pp. 556-563.

[7] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

[8] D. L. Snyder, *Random point processes*, Wiley, New York, 1975.

[9] D. Knuth, *The Art of Computing Programming*, II, *Seminumerical Algorithm*, Addison-Wesley, Reading Mass., 1981.