Rochester Institute of Technology

RIT Digital Institutional Repository

2-2017

The Conditional Poisson Process and the Erlang and Negative Binomial Distributions

Anurag Agarwal
Rochester Institute of Technology

Peter Bajorski
Rochester Institute of Technology

David L. Farnsworth
Rochester Institute of Technology

James E. Marengo
Rochester Institute of Technology

Wei Qian
Rochester Institute of Technology

Recommended Citation

The Conditional Poisson Process and the Erlang and Negative Binomial Distributions

Anurag Agarwal, Peter Bajorski, David L. Farnsworth, James E. Marengo, Wei Qian

School of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York, USA
Email: anurag.agarwal@rit.edu, pxbeqa@rit.edu, dlfsma@rit.edu, jemsma@rit.edu, wxqsma@rit.edu

Abstract

It is a well known fact that for the hierarchical model of a Poisson random variable Y whose mean has an Erlang distribution, the unconditional distribution of Y is negative binomial. However, the proofs in the literature [1] [2] provide no intuitive understanding as to why this result should be true. It is the purpose of this manuscript to give a new proof of this result which provides such an understanding. The memoryless property of the exponential distribution allows one to conclude that the events in two independent Poisson processes may be regarded as Bernoulli trials, and this fact is used to achieve the research purpose. Another goal of this manuscript is to give another proof of this last fact which does not rely on the memoryless property.

Keywords

Conditional Distribution, Hierarchical Model, Mixture Distribution, Poisson Process, Stochastic Process

1. Introduction

There is much current interest in compounding or mixing distributions and their applications. Indeed, the early history of statistics was greatly concerned with the problem [3]. The work by Greenwood and Yule [4] in the more modern era has been followed up with new results and extensive applications [5], including many based on the Poisson distribution because of its centrality in statistical analysis and probability modeling [6] [7] [8] [9]. The present derivations supply new insights into the structure of this type of modeling by revealing how compounded Poisson variables produce a negative binomial distribution.

There is a relatively simple fact about mixture distributions which says that

if the mean of a conditional Poisson random variable has an Erlang distribution, then the unconditional distribution of this variable is negative binomial. In particular,

Theorem 1. *Let* $m \in N$ *and* $\theta \in R^+$. *Suppose that the random variable* Λ *has the Erlang distribution with probability density function*

$$f(\lambda) = \theta e^{-\theta\lambda} \frac{(\theta\lambda)^{m-1}}{(m-1)!} \quad \text{for} \quad \lambda > 0 \tag{1}$$

and that, given that $\Lambda = \lambda$, $N(t)$ *has the Poisson distribution with probability mass function*

$$p(k|\lambda) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}. \tag{2}$$

Then, the unconditional distribution of $N(t)$ *is that of the number of failures before the* m^{th} *success in Bernoulli trials with success probability* $p = \dfrac{\theta}{\theta + t}$. *That is,*

$$P(N(t) = k) = \binom{k+m-1}{k} \left(\frac{\theta}{\theta+t}\right)^m \left(\frac{t}{\theta+t}\right)^k. \tag{3}$$

The proof involves an application of the law of total probability which conditions on the value of Λ. That is, an integration of the product of (1) and (2) yields (3). This result appears in many different settings ([10], pp. 194-195; [11], p. 191; [2], pp. 332-333; [12], p. 84).

Could this result have been guessed? This proof provides no intuitive understanding as to why it is true. One purpose of this paper is to give another proof of Theorem 1 which provides such an understanding. In that proof, two independent Poisson processes are carefully chosen and the memoryless property of their exponentially distributed interarrival times is used to conclude that the events in these processes may be regarded as Bernoulli trials. In Section 3, without using the memoryless property, another proof that these events are Bernoulli trials has been provided.

2. Proof of Theorem 1

This section contains an alternative proof of Theorem 1 which can facilitate one's intuitive understanding of this result. The proof uses properties of the Poisson process and exponential distribution to obtain (3).

Fix $t \geq 0$, let $\{N_1(u), u \geq 0\}$ be a Poisson process with rate θ, and let $\{N_2(u), u \geq 0\}$ be an independent Poisson process with *rate* t. One may think of the events in the first process as "successes" and those in the second as "failures". Using the well-known facts [2] that the interarrival times in the first (second) process are independent exponentially distributed random variables with rate θ (rate t) and the memoryless property of the exponential distribution ([2], pp. 150, 159; [10], p. 102) one may regard these successes and failures as being Bernoulli trials. That is, the trials are independent, and the

probability of success is the same on each trial. Intuitively, the process probabilistically restarts itself at any point in time. Specifically, suppose an event has just occurred in one of the two processes. Then, regardless of the amount of time that has elapsed since the last event in the other process, the distribution of the amount of time remaining until the next event occurs in the other process is exponential with the rate for that process. Hence, independently of what has occurred up to that point of time, the probability that the next event is a success is the probability that an exponential random variable with rate θ is less than an independent exponential random variable with rate t, and this probability is easily seen to be $\dfrac{\theta}{\theta + t}$ ([12], p. 287). A proof (which does not make direct reference to the memoryless property) that the events in the two processes constitute Bernoulli trials is given in Section 3.

Since the sum of m independent exponential random variables each having rate θ has the Erlang distribution in (1), one may think of Λ as being the time of occurrence of the mth event in the process $\{N_1(u), u \geq 0\}$. That is, Λ is the occurrence time of the mth success ([2], p. 150). Given that $\Lambda = \lambda$, the conditional distributions of both $N(t)$ and $N_2(\lambda)$ are the same Poisson distribution with mean λt. By conditioning on Λ, the unconditional distribution of $N(t)$ is the same as the unconditional distribution of $N_2(\Lambda)$. The proof now follows by observing that $N_2(\Lambda)$ is the number of failures before the time of the mth success.

3. Proof That the Trials Are Bernoulli

This section contains a proof, which does not depend on the memoryless property, that the events in two independent Poisson processes may be regarded as Bernoulli trials.

Theorem 2. *Consider two independent Poisson processes with respective rates λ_1 and λ_2 in which the events that occur in either process are called trials and are referred to as successes or failures according as they come from the first or second process. Then the trials are independent and the probability of success is $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ on each trial. That is, these trials are Bernoulli trials.*

Before proving Theorem 2, the following two lemmas are needed.

Lemma 1. *For nonnegative integers m and n,*

$$\sum_{k=0}^{n} \binom{m+k}{m} = \binom{m+n+1}{m+1}.$$

Proof. The number of possible choices of $m+1$ distinct numbers from the set $\{1, 2, 3, \cdots, m+n+1\}$ is $\binom{m+n+1}{m+1}$. By conditioning on the value of the largest number chosen, one can see that this number of choices is also given by $\sum_{k=0}^{n} \binom{m+k}{m}$. $\qquad \square$

Lemma 2. *Using the terminology in Theorem 2, let $E_{n,k}$ be the event that*

there are exactly k *successes among the first* n *trials. Then*

$$P\left(E_{n,k} \text{ and success on trial } n+1\right) = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{k+1}\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.$$

Proof. Let $T_{i,j}$ be the time between the $(i-1)$ st event and the i th event (*i.e.*, the i th interarrival time) in the j th process. Also, let S_j and F_j be the respective times until the j th success and j th failure, so that

$$S_j = \sum_{i=1}^{j} T_{i,1} \quad \text{and} \quad F_j = \sum_{i=1}^{j} T_{i,2}.$$

Then,

$$\begin{aligned}
P\left(E_{n,k} \text{ and success on trial } n+1\right) &= P\left(F_{n-k} < S_{k+1} < F_{n-k+1}\right) \\
&= P\left(F_{n-k} < S_k + T_{k+1,1} < F_{n-k} + T_{n-k+1,2}\right).
\end{aligned}$$

By conditioning on the independent random variables S_k and F_{n-k}, and using the fact that the interarrival times in a Poisson process are independent random variables, it follows that the last probability is

$$\int_0^\infty \int_0^\infty P\left(v < u + T_{k+1,1} < v + T_{n-k+1,2}\right) f_{S_k}(u) f_{F_{n-k}}(v)\,\mathrm{d}u\mathrm{d}v, \tag{4}$$

where f_{S_k} and $f_{F_{n-k}}$ are the pdf's of S_k and F_{n-k}, respectively. Now, use the fact that $T_{k+1,1}$ and $T_{n-k+1,2}$ are independent and have respective exponential distributions with failure rates λ_1 and λ_2. If $u \le v$, it follows that

$$P\left(v < u + t < v + T_{n-k+1,2}\right) = \begin{cases} 0 & \text{if } t \le v - u \\ e^{-\lambda_2(t+u-v)} & \text{if } t > v - u. \end{cases}$$

By conditioning on the value of $T_{k+1,1}$, it can be concluded that

$$\begin{aligned}
P\left(v < u + T_{k+1,1} < v + T_{n-k+1,2}\right) &= \int_{v-u}^\infty e^{-\lambda_2(t+u-v)} \lambda_1 e^{-\lambda_1 t}\,\mathrm{d}t \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-\lambda_1(v-u)}.
\end{aligned} \tag{5}$$

Similarly, if $u > v$,

$$P\left(v < u + t < v + T_{n-k+1,2}\right) = e^{-\lambda_2(t+u-v)},$$

and hence

$$\begin{aligned}
P\left(v < u + T_{k+1,1} < v + T_{n-k+1,2}\right) &= \int_0^\infty e^{-\lambda_2(t+u-v)} \lambda_1 e^{-\lambda_1 t}\,\mathrm{d}t \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-\lambda_2(u-v)}.
\end{aligned} \tag{6}$$

Using the fact that S_k and F_{n-k} have Erlang distributions with respective shape parameters k and $n-k$ and respective scale parameters λ_1 and λ_2, substituting (5) and (6) into (4) leads to

$$\begin{aligned}
&P\left(E_{n,k} \text{ and success on trial } n+1\right) \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}\left(\int_0^\infty \int_u^\infty e^{-\lambda_1(v-u)} \frac{(\lambda_1 u)^{k-1}}{(k-1)!} \lambda_1 e^{-\lambda_1 u} \frac{(\lambda_2 v)^{n-k-1}}{(n-k-1)!} \lambda_2 e^{-\lambda_2 v}\,\mathrm{d}v\mathrm{d}u \right. \\
&\quad \left. + \int_0^\infty \int_v^\infty e^{-\lambda_2(u-v)} \frac{(\lambda_1 u)^{k-1}}{(k-1)!} \lambda_1 e^{-\lambda_1 u} \frac{(\lambda_2 v)^{n-k-1}}{(n-k-1)!} \lambda_2 e^{-\lambda_2 v}\,\mathrm{d}u\mathrm{d}v \right).
\end{aligned} \tag{7}$$

To evaluate the first double integral in (7), one can use the fact that the waiting time until the $(n-k)^{th}$ event in a Poisson process with rate $\lambda_1 + \lambda_2$ exceeds u if and only if the number of events in this process that occur by time u is at most $n-k-1$, and that this number has a Poisson distribution with mean $(\lambda_1 + \lambda_2)u$. Hence the first integral in (7) is

$$
\begin{aligned}
&\frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^{n-k}} \int_0^\infty \frac{u^{k-1}}{(k-1)!} \int_u^\infty \frac{((\lambda_1+\lambda_2)u)^{n-k-1}}{(n-k-1)!} (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)v} \mathrm{d}v \mathrm{d}u \\
&= \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^{n-k}} \int_0^\infty \frac{u^{k-1}}{(k-1)!} \sum_{j=0}^{n-k-1} e^{-(\lambda_1+\lambda_2)u} \frac{((\lambda_1+\lambda_2)u)^j}{j!} \mathrm{d}u \\
&= \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \sum_{j=0}^{n-k-1} \binom{k-1+j}{k-1} \int_0^\infty \frac{((\lambda_1+\lambda_2)u)^{j+k-1}}{(j+k-1)!} (\lambda_1 + \lambda_2) e^{-(\lambda_1+\lambda_2)u} \mathrm{d}u \qquad (8) \\
&= \sum_{j=0}^{n-k-1} \binom{k-1+j}{k-1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \\
&= \binom{n-1}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k},
\end{aligned}
$$

where the penultimate equation follows from the fact that the integral in the preceding expression is one, and the last equation follows by an application of Lemma 1. By interchanging λ_1 with λ_2, k with $n-k$, and once again applying Lemma 1, it can be concluded in a similar manner that the value of the second double integral in (7) is

$$
\begin{aligned}
&\sum_{j=0}^{k-1} \binom{n-k-1+j}{n-k-1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \\
&= \binom{n-1}{k-1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.
\end{aligned}
\qquad (9)
$$

From (7), (8) and (9), it now follows that

$$
\begin{aligned}
&P\left(E_{n,k} \text{ and success on trial } n+1\right) \\
&= \left[\binom{n-1}{k} + \binom{n-1}{k-1}\right] \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{k+1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \\
&= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{k+1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}.
\end{aligned}
$$

The case $n=0$ is easy and left to the reader. The argument just presented assumes that $0 < k < n$. The cases $k=0$ and $k=n$ are simpler and are also left to the reader. The proof of Lemma 2 is complete. □

Proof of Theorem 2. It will now be shown by induction on m that for $m \geq 2$, the first m trials are independent and that the probability of success on each of these trials is $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$.

First, suppose $m = 2$. Set $n = k = 0$ in Lemma 2 to conclude that

$$
P(\text{success on the 1st trial}) = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \qquad (10)
$$

and hence $P(\text{failure on the 1st trial}) = 1 - \dfrac{\lambda_1}{\lambda_1 + \lambda_2} = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$.

Set $n = 1$ and $k = 1$ in Lemma 2 to see that

$$P(\text{success on each of the first two trials}) = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2, \tag{11}$$

and set $n = 1$ and $k = 0$ to obtain

$$P(\text{failure on the 1st trial and success on the 2nd trial}) = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}.$$

Adding the last two probabilities gives

$$P(\text{success on the 2nd trial}) = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 + \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{12}$$

The independence of the first two trials now follows from (10), (11) and (12), and consequently Theorem 2 is true for $m = 2$.

Suppose that Theorem 2 is true for some $m \geq 2$. The first m trials are therefore Bernoulli trials, so that the number of successes has a binomial distribution. Specifically,

$$P(E_{m,k}) = \binom{m}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{m-k}. \tag{13}$$

From Lemma 2, we can condition on the number of successes in the first m trials to see that

$$
\begin{aligned}
P(\text{success on trial } m+1) &= \sum_{k=0}^{m} P(E_{m,k} \text{ and success on trial } m+1) \\
&= \sum_{k=0}^{m} \binom{m}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{k+1}\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{m-k} \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\end{aligned} \tag{14}
$$

It follows that on each of the first $m+1$ trials, the probability of success is $\dfrac{\lambda_1}{\lambda_1 + \lambda_2}$. Furthermore, it follows from Lemma 2, (13), and (14) that for $k = 0, 1, \cdots, m$

$$P(E_{m,k} \text{ and success on trial } m+1) = P(E_{m,k}) P(\text{success on trial } m+1),$$

from which one may conclude that the first $m+1$ trials are independent. The proof of Theorem 2 is now complete.

$\square$

4. Conclusion

In this paper, a new proof has been provided for the fact that, in the hierarchical model of a Poisson random variable Y whose mean has an Erlang distribution, the unconditional distribution of Y is negative binomial. A new proof that the events in two independent Poisson processes may be regarded as Bernoulli trials

has also been provided. The distinguishing feature of this proof is that it does not make use of the memoryless property of the exponential distribution.

References

[1] Karlin, S. and Taylor, H.M. (1975) A First Course in Stochastic Processes. 2nd Edition, Academic Press, San Deigo, CA.

[2] Ross, S.M. (2014) Introduction to Probability Models. 11th Edition, Academic Press, San Diego, CA.

[3] Stigler, S.M. (1986) The History of Statistics: The Measurement of Uncertainty before 1900. Belknap Press of Harvard University Press, Cambridge, MA.

[4] Greenwood, M. and Yule, G.U. (1920) An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. *Journal of the Royal Statistical Society A*, **83**, 255-279. https://doi.org/10.2307/2341080

[5] Sundt, B. and Vernic, R. (2009) Recursions for Convolutions and Compound Distributions with Insurance Applications. Springer-Verlag, Berlin Heidelberg.

[6] Albrecht, P. (1982) On Some Statistical Methods Connected with the Mixed Poisson Process. *Scandinavian Actuarial Journal*, No. 1, 1-14. https://doi.org/10.1080/03461238.1982.10405427

[7] Antzoulakos, D. and Chadjiconstantinidis, S. (2004) On Mixed and Compound Mixed Poisson Distributions. *Scandinavian Actuarial Journal*, No. 3, 161-188. https://doi.org/10.1080/03461230110106525

[8] Nadarajah, S. and Kotz, S. (2006) Compound Mixed Poisson Distributions I. *Scandinavian Actuarial Journal*, No. 3, 141-162. https://doi.org/10.1080/03461230600783384

[9] Nadarajah, S. and Kotz, S. (2006) Compound Mixed Poisson Distributions II. *Scandinavian Actuarial Journal*, No. 3, 163-181. https://doi.org/10.1080/03461230600715253

[10] Casella, G. and Berger, R.L. (1990) Statistical Inference. Wadsworth & Brooks/Cole, Pacific Grove, CA.

[11] Hogg, R., McKean, J.W. and Craig, A.T. (2005) Introduction to Mathematical Statistics. 6th Edition, Pearson, Upper Saddle River, NJ.

[12] Taylor, H.M. and Karlin, S. (1998) An Introduction to Stochastic Modeling. 3rd Edition, Academic Press, San Deigo, CA.