An Interpretation of the Mixture of Poisson Distributions with a Gamma Distributed Parameter

Suneung Ahn
Hanyang University
Department of Industrial Engineering
Ansan, Korea

Abstract

Used as a mixing distribution for an unknown Poisson parameter, the gamma distribution leads to the negative binomial distribution. The hyperparameters of the gamma distribution have their own meanings according to what the Poisson parameter represents. Different sources in the randomness of the Poisson parameter give different interpretations of the negative binomial distribution.

Keywords : negative binomial distribution, mixture, Poisson distribution, conjugacy, gamma distribution

1. INTRODUCTION

The Poisson distribution arises naturally in the study of data taking the form of counts. For a homogeneous Poisson process \( N(t), t \geq 0 \) having rate \( \lambda \), the number of events in any interval of length \( t \) is Poisson distributed with mean \( \lambda t \). That is, for all \( s, t \geq 0 \)

\[
\Pr\{N(t+s)-N(s)=y\}=e^{-\lambda t} \left(\frac{\lambda t}{y!}\right)^{y}, \quad y=0,1,\ldots
\]

Note that the expected value of \( N(t) \) is \( \lambda t \), or \( \mathbb{E}[N(t)]=\lambda t \), which explains why \( \lambda \) is called the rate of the process. Denote the distribution (1) by \( \text{Poisson}(y|\lambda t) \).

We will further assume that the parameter \( \lambda t \) is a random variable with prior distribution \( p(\lambda t) \). Specially, we consider the case that the random parameter \( \lambda t \) is distributed as the gamma distribution (or the Erlangian distribution) with the shape and scale parameters \( \alpha \) positive integer and \( \beta > 0 \) respectively, which is of the form

\[
p(\lambda t) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\beta(\lambda t)}
\]

and is denoted by \( \text{gamma}(\lambda t|\alpha, \beta) \). The parameters \( \alpha \) and \( \beta \) characterize the gamma distribution and are often called “hyperparameters” to avoid confusion. Since the gamma distribution is a conjugate family for the Poisson likelihood [3] and Bayes’ theorem states that the posterior distribution is related to the prior and likelihood distributions according to

\[
p(\lambda t|y) = \frac{p(y|\lambda t)}{p(y)} p(\lambda t), \quad y=0,1,\ldots
\]

the posterior distribution of \( \lambda t \) conditional on \( y \) is given by \( \text{gamma}(\alpha+y, \beta+1) \). The unconditional marginal distribution \( p(y) \) can be obtained according to

\[
p(y) = \int p(y, \lambda t) d(\lambda t)
\]

where the integration is taken over the admissible range of \( \lambda t \). That is, by marginalizing the joint distribution of \( y \) and \( \lambda t \), \( p(y, \lambda t) = p(y|\lambda t) p(\lambda t) \) with respect to \( \lambda t \), we can find the distribution \( p(y) \);

\[
p(y) = \int p(y|\lambda t) p(\lambda t) d(\lambda t)
\]

\[
= \int_0^\infty \left( e^{-\lambda t} \left(\frac{\lambda t}{y!}\right)^{y} \right) \frac{\beta^\alpha}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\beta(\lambda t)} d(\lambda t)
\]

\[
= \left(\frac{\alpha+y-1}{\beta+1}\right)^\alpha \left(\frac{\beta}{\beta+1}\right)^{\alpha-1} \frac{1}{\Gamma(\alpha)} \frac{1}{\beta+1}^y,
\]

which is known as the negative binomial distribution with parameters \( \alpha \) and \( \beta/(\beta+1) \). Denote (5) by \( \text{neg-bin}(y|\alpha, \beta) \). Note that the expected value of the random variable \( \alpha+y \) in (5) is \( \alpha \) times the mean \( (\beta+1)/\beta \) of a geometric
random variable, that is, \( E(\alpha + y) = \alpha (\beta + 1) / \beta \). For \( \alpha \) constant, \( E[y] = \alpha / \beta \). The derivation of (5) shows that the negative binomial distribution is a mixture of Poisson \( (y | \lambda t) \) with random parameter, \( \lambda t \), that follows gamma \( (\lambda t | \alpha, \beta) \) \[2]\:

\[
\text{neg - bin}(y | \alpha, \beta) = \int \text{Poisson}(y | \lambda t) \text{gamma}(\lambda t | \alpha, \beta) d(\lambda t). \tag{6}
\]

We now consider the assumption that the parameter \( \lambda t \) follows gamma \( (\lambda t | \alpha, \beta) \). Since the random mean rate \( \lambda t \) in (1) corresponds to the random mean number of events occurred during the time interval \( t \) when \( \lambda \) is the rate per unit time interval, we have three cases in allowing for the parameter \( \lambda t \) to be random: First, \( \lambda t \) is the random mean number of events occurred with the random rate per unit time interval, \( \lambda \), during a fixed time interval \( t \). Secondly, \( \lambda t \) is the random mean number of events occurred with a fixed rate per unit time interval, \( \lambda \), during the random time interval \( t \). Thirdly, \( \lambda t \) is the random mean number of events occurred with the random rate per unit time interval, \( \lambda \), during the random time interval \( t \). As far as the last case is concerned, the interpretation of the unconditional marginal distribution (5) is simple. That is, when the occurrences of events follow Poisson \( (y | \lambda t) \) and further the random mean rate \( \lambda t \) follows gamma \( (\lambda t | \alpha, \beta) \), the distribution of the number of such events is Neg - bin \( (y | \alpha, \beta) \). However, by allowing one of the two variables \( \lambda \) and \( t \) to be random and the other to be fixed, we can have different interpretations of the unconditional marginal distribution \( p(y) \). We now consider the first two cases.

2. A MIXTURE OF POISSON DISTRIBUTION HAVING GAMMA DISTRIBUTED RANDOM PARAMETERS WITH A FIXED TIME INTERVAL

When the Poisson random parameter \( \lambda t \) is gamma \( (\lambda t | \alpha, \beta) \) distributed and \( t \) is a fixed time interval, by the change of variable method, we can see that the random rate per unit time interval \( \lambda \) follows gamma \( (\lambda | \alpha, \beta t) \). By the Bayes’ formula (3) and the same method as in the derivation of (5), the posterior distribution of \( \lambda \) given the data point \( y \) becomes gamma \( (\alpha + y, \beta t + t) \) and the unconditional marginal distribution of \( y \) becomes

\[
p(y) = \left( \frac{\alpha + y - 1}{\alpha - 1} \right) \left( \frac{1}{\beta t + t} \right)^{\alpha} \left( \frac{1}{\beta t + t} \right)^{\alpha - 1}. \tag{7}
\]

Note that (7) becomes (5) when \( t = 1 \).

Now we interpret the possible meanings of the parameters \( \alpha \) and \( \beta \). Because the prior mean rate of \( \lambda t \) in the Poisson \( (y | \lambda t) \) likelihood (1) is \( E(\lambda t) = \alpha / \beta \), or equivalently \( E(\lambda) = \alpha / \beta t \), the shape parameter \( \alpha \) can be interpreted as a total count in a prior test of \( \beta \) observations (or \( \beta t \) time units). On the other hand, when we compare the form of the Poisson \( (y | \lambda t) \) likelihood (1)

\[
p(y | \lambda t) \propto (\lambda t)^{y} e^{-\lambda t} \tag{8}
\]

with the form of the gamma \( (\lambda t | \alpha, \beta) \)

\[
p(\lambda t | \alpha, \beta) \propto (\lambda t)^{\alpha - 1} e^{-\beta \lambda t},
\]

or

\[
p(\lambda) \propto \lambda^{\alpha - 1} e^{-\beta(\alpha \lambda), \tag{9}
\]

we can say that the prior distribution (9) is, in some sense, equivalent to a total count of \( \alpha - 1 \) in prior \( \beta t \) units. However, choosing (9) for \( \lambda \) in (8) is still intended for using \( \alpha / \beta t \) as the mean of the random parameter \( \lambda \) rather than \( (\alpha - 1) / \beta t \). Note that if \( \alpha / \beta t \) converges to a constant, then \( \alpha / \beta t \) and \( (\alpha - 1) / \beta t \) are approximately same for \( \beta t \) large. It is further noted that the shape parameter \( \alpha \) is a quantity of no unit, whereas the scale parameter \( \beta t \) has units of scale, in this case, time, associated with it.

In the posterior distribution of \( \lambda \) given \( y \) which is gamma \( (\alpha + y, \beta t + t) \), the parameter \( (\alpha + y) \) is referred to as the combined number of events, whereas \( (\beta t + t) \) is the combined total observed time units. The effect of gamma \( (\alpha, \beta) \) prior distribution on \( \lambda t \) is to increase the observed number of events \( y \) by \( \alpha \) and to increase the observed total time units \( t \) by \( \beta t \). This is a clear interpretation of the effect of the prior distribution in the analysis. The unconditional marginal distribution of \( y \) is given by (7) with mean and variance given by \( \alpha / \beta \) and \( \alpha (\beta + 1) / \beta^2 \), respectively.
The dominant factor in selecting a prior model for \( \lambda t \) in (1) is that the selected model represent the analyst’s knowledge and experience regarding \( \lambda t \). That is, the prior distribution should reflect the analyst’s prior belief about \( \lambda t \). The flexibility present in the gamma distribution (2) through the choices of \( \alpha \) and \( \beta \) allows the analyst to select the model that best expresses the current state of knowledge about \( \lambda t \).

### 3. A MIXTURE OF POISSON DISTRIBUTION HAVING GAMMA DISTRIBUTED RANDOM PARAMETERS WITH A FIXED RATE PER UNIT TIME INTERVAL

When the Poisson random parameter \( \lambda t \) is given by \( \gamma(\lambda|\alpha, \beta) \) distributed and \( \lambda \) is a fixed rate per unit time interval, by the change of variable method, we can see that the random parameter \( t \) follows \( \gamma(t|\alpha, \beta, \lambda) \). By the Bayes’ formula (3) and the same method as in the derivation of (5), the posterior distribution of \( t \) given the data point \( y \) becomes \( \gamma(t|\alpha + y, \beta \lambda + \lambda) \) and the unconditional marginal distribution of \( y \) becomes

\[
\begin{align*}
p(y) &= \left(\frac{\alpha + y - 1}{\alpha - 1}\right) \left(\frac{\beta \lambda}{\beta \lambda + \lambda}\right)^\alpha \left(\frac{\lambda}{\beta \lambda + \lambda}\right)^{y-\alpha} \cdot (\alpha > 0, \beta > 0) \\
&= \int \gamma(\lambda|\alpha, \beta) \gamma(t|\alpha + y, \beta \lambda + \lambda) \ dt.
\end{align*}
\]

(10)

Note that (10) becomes (5) when \( \lambda = 1 \).

Consider a homogeneous Poisson process with a fixed parameter \( \lambda \) and let \( N(t) \) be the total number of “events” in the time interval \([0, t]\) wherein \( t \) is a nonnegative random time interval. In the case of \( \lambda = 1 \), Engel and Zijlstra [1] showed the fact that \( N(t) = y \) has a negative binomial distribution with parameters \( \alpha \) and \( \beta \) for \( t \) if and only if \( t \) is gamma distributed with parameters \( \alpha > 0 \) and \( \beta > 0 \); That is,

\[
\text{neg-bin}(y|\alpha, \beta) = \int \gamma(t|\alpha + y, \beta \lambda + \lambda) \ dt.
\]

For any \( \lambda \) we have the following relation:

\[
\text{neg-bin}(y|\alpha, \beta) = \int \gamma(\lambda|\alpha, \beta) \gamma(t|\alpha + y, \beta \lambda + \lambda) \ dt(\lambda t).
\]

or equivalently,

\[
\text{neg-bin}(y|\alpha, \beta, \lambda) = \int \gamma(\lambda|\alpha, \beta) \gamma(t|\alpha + y, \beta \lambda + \lambda) \ dt.
\]

(11)

The negative-binomial distribution as the mixture of (11) is now interpreted by considering two independent Poisson processes with respective rates of \( \lambda \) and \( \beta \lambda \). The \( \text{Poisson}(y|\lambda t) \) in (11) represents the probability that exactly \( y \) independent events, each of which has the exponential rate \( \lambda \), occur during a certain time interval \([0, t]\). The \( \gamma(t|\alpha, \beta, \lambda) \) in (11) plays a role of weighting factor for reflecting the effect of the length of \( t \) in determining the probability that exactly \( y \) independent events with the exponential rate \( \lambda \) occur during a certain time interval \([0, t]\).

This implies that the distribution \( \gamma(t|\alpha, \beta, \lambda) \) generates the time interval \( t \) which varies. We can now interpret the right-hand-side of (11) as the probability that exactly \( y \) independent events with the exponential rate \( \lambda \) occur during the time interval \([0, t]\) for all possible values of \( \lambda t \) wherein \( t \) is generated from \( \gamma(\lambda|\alpha, \beta) \). Furthermore, the value of \( t \) is known as the sum of \( \alpha \) independent and identically distributed exponential random variables having mean \( 1/(\beta \lambda) \), or having the exponential rate \( \beta \lambda \). Therefore, the integration of \( \text{Poisson}(y|\lambda t) \) over all range of \( \lambda t \) where \( t \) follows \( \gamma(t|\alpha, \beta, \lambda) \) is equivalent to the probability that exactly \( y \) independent events with the exponential rate \( \lambda \) occur until the \( \alpha \) th independent event with the exponential rate \( \beta \lambda \) occurs, which implies the negative binomial distribution.

The above negative binomial distribution can be derived due to the fact that a homogeneous Poisson process has the properties that the process from any point on is independent of all that has previously occurred, and also has the same distribution as the original process. In other words, the process has no memory, and hence exponential interarrival times are to be expected [4]. The probability that exactly \( y \) events occur in one Poisson process with rate \( \lambda \) until the \( \alpha \) th event occur in a second and independent Poisson process with rate \( \beta \lambda \) can be derived as follows: Let \( \{N_1(t), t \geq 0\} \) and \( \{N_2(t), t \geq 0\} \) be two independent Poisson process having respective rates \( \lambda \) and \( \beta \lambda \). Also, let \( S_1^y \) denote the time of \( y \) th event of the first process, and \( S_2^\alpha \) the time of the \( \alpha \) th event of the second process. We seek
Let us consider the special case \( y = \alpha = 1 \). Since \( S'_1 \), the time of the first event of the \( N_1(t) \) process, and \( S''_1 \), the time of the first event of the \( N_2(t) \) process, are both exponentially distributed random variables with respective means \( 1/\lambda \) and \( 1/(\beta \lambda) \), it follows that

\[
P[S'_1 < S''_1] = \frac{\lambda}{\beta \lambda + \lambda}. \tag{12}\]

This is the probability that one exponential random variable is smaller than another. Let us now consider the special case \( 1 \leq \alpha < \beta \). From Equation (12), with probability \( \lambda/(\beta \lambda + \lambda) \), now given that the initial event is from the \( N_1(t) \) process, the next thing that must occur for \( S''_1 \) to be less than \( S'_1 \) is for the second event also to be an event of the \( N_1(t) \) process. However, when the first event occurs both processes start all over again (by the memoryless property of Poisson processes) and hence this conditional probability is also \( \lambda/(\beta \lambda + \lambda) \), and hence the desired probability is given by

\[
P[S'_1 < S''_1] = \left(\frac{\lambda}{\beta \lambda + \lambda}\right)^2. \tag{13}\]

In fact this reasoning shows that each event occurs is going to be an event of the \( N_1(t) \) process with probability \( \lambda/(\beta \lambda + \lambda) \) and an event of the \( N_2(t) \) process with probability \( \beta \lambda/(\beta \lambda + \lambda) \), independent of all that has previously occurred. In other words, the probability that the \( N_1(t) \) process reaches \( y \) until the \( N_2(t) \) process reaches \( \alpha \) is just the probability that equals the product of the probability of obtaining exactly \( \alpha - 1 \) events (from the \( N_2(t) \) process) in the first \( \alpha + y - 1 \) trials (\( \alpha - 1 \) events from \( N_1(t) \) process and \( y \) events from \( N_2(t) \) process) and the probability of an event on the \( \alpha \) th trial (from the \( N_2(t) \) process). Thus, we have (10).

4. DISCUSSION

The variances of the negative binomial distributions (5), (7), and (10) are always greater than their corresponding means, in contrast to the Poisson, whose variance is always equal to its mean. The negative binomial distribution is a two-parameter family that allows the mean and variance to be fitted separately, with variance at least as great as the mean. This is the reason that the negative binomial can be used as a robust alternative to the Poisson distribution. In the limit as \( \beta \to \infty \) with \( \alpha/\beta \) remaining constant, the underlying gamma distribution approaches a spike, and the negative binomial distribution approaches the Poisson distribution.

Here, the probability models include an uncertain but essential parameter governing a system’s behavior. Under the modeling procedure, the so-called “prior” distribution expresses our initial opinion about the parameter. And we keep updating its form according to the data available from the real observations. In this way the prior becomes a “posterior” distribution of the parameter. In the next application, this posterior distribution plays a role of the prior distribution in the current analysis. In this way, the Bayesian probabilistic modeling procedure includes a learning process.

REFERENCES


