Almost All About the Compound Gamma-Poisson Probability Model and Applications

Nate Josephs Ben Draves

October 3, 2017

Abstract

The Gamma-Poisson (G-P) probability model is a discrete, two-parameter compound distribution where the mean of a Poisson random variable follows a Gamma distribution. We derive standard properties of the G-P model like its density, first and second moments, moment generating function and characteristic function. We show how special properties arise when at least one parameter is known. We relate G-P to other models and families of distributions. Finally, we show theoretical and applied applications of the G-P model.

1 Introduction

The Gamma-Poisson (G-P) probability model is a compound distribution. A compound distribution, or mixing distribution, is given by its mixture density

$$f(x) = \int g(x|\theta)h(\theta)d\theta \tag{1}$$

where $g(x|\theta)$ is a density function whose parameter $\theta \sim h(\theta)$. Recall that the density of a Poisson random variable (RV) is given by

$$P(N = k) = \frac{e^{-k}\lambda^k}{k!}\mathbb{1}_{k\in\mathbb{N}} \tag{2}$$

and the density of a gamma RV is given by

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathbb{1}_{x\in\mathbb{R}^+} \tag{3}$$

where $\alpha, \beta > 0$ are the shape and rate parameters, respectively. The G-P model is used for a Poisson process where its mean is taken from a gamma distribution. Many properties of the G-P distribution are derived using the laws of total expectation and variance by conditioning over the underlying gamma distribution.

As we will see, the G-P model is intricately related to both the gamma and Poisson models, and also has close relationships with other models such as the negative binomial distribution and the Exponential-Poisson distribution. The compound nature of the G-P model provides more tuning flexibility than a Poisson model and such a framework is especially useful for modeling frequencies.

2 Properties of the Model

2.1 Standard Properties

The G-P probability model is a compound distribution where the mean of a Poisson RV, X, follows a gamma distribution. That is,

$$X \sim \text{Pois}(\Lambda) \quad \text{where} \quad \Lambda \sim \text{Gamma}(\alpha, \beta) \tag{4}$$

Since X has a compound distribution, we derive its density using the definition given by (1). That is, by marginalizing over the distribution of Λ we can define the density of X as follows:

$$\begin{aligned}
P_X(X = n) &= \int_0^\infty P(X = n|\Lambda = \lambda) f_\Lambda(\lambda) d\lambda \\
&= \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \left(\frac{1}{\Gamma(a)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \right) d\lambda \\
&= \frac{1}{n!\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{-\lambda(1+\frac{1}{\beta})} \lambda^{n+\alpha-1} d\lambda \\
&= \frac{\Gamma(n+\alpha)(\frac{\beta}{\beta+1})^{\alpha+n}}{n!\Gamma(\alpha)\beta^\alpha} \int_0^\infty \frac{x^{n+\alpha-1} e^{-\lambda(1+\frac{1}{\beta})}}{\Gamma(\alpha+n)(\frac{\beta}{\beta+1})^{\alpha+n}} d\lambda \\
&= \frac{1}{n!\Gamma(\alpha)\beta^\alpha} \Gamma(n+\alpha) \left(\frac{\beta}{\beta+1} \right)^{n+\alpha} \\
&= \frac{\Gamma(\alpha+n)\beta^n}{n!\Gamma(\alpha)(\beta+1)^{n+\alpha}} \quad \text{for} \quad n = 0, 1, 2 \ldots
\end{aligned} \tag{5}$$

Note that the fourth equality is obtained by recognizing the integrand as the kernel of a Gamma$(n+\alpha, \frac{\beta+1}{\beta})$ density function.

Like most mixture models, the moments of the compound distribution can be related to the moments of the underlying random variables. In our case, we look to express the mean and variance of X in terms of these values for Λ. To leverage this relationship, we use the law of total expectation and law of total variance to find these quantities for X.

$$E[X] = E\left[E[X|\Lambda]\right] = \int_0^\infty E[X|\Lambda] f_\Lambda(\lambda) d\lambda = \int_0^\infty \Lambda f_\Lambda(\lambda) d\lambda = E[\Lambda]$$

Therefore, we see that the mean of Λ directly determines the mean of X. To compute the variance, we again relate X to Λ.

$$Var(X) = E\big[Var(X|\Lambda)\big] + Var\big(E[X|\Lambda]\big) = E\big[\Lambda\big] + Var\big(\Lambda\big)$$

Again, we see that the variance is perfectly determined by the underlying distribution of Λ. Notice that if $Y \sim \text{Pois}(\theta)$ for fixed rate θ, then $Var(Y) = E[Y]$. In our case, however, we see that $Var(X) \geq E[X]$ with equality occurring only when $Var(\Lambda) = 0$, i.e. Λ is constant. This increase in variance is due directly to the variance in our mean Λ. This feature distinguishes the G-P distribution from a univariate Poisson distribution and is quite useful in several applications, which we will discuss further in Section 3.

Using the fact that $\Lambda \sim \text{Gamma}(\alpha, \beta)$ we can write the mean and variance of X as follows

$$E[X] = \alpha\beta \qquad Var(X) = \alpha\beta + \alpha\beta^2 \tag{6}$$

Unfortunately, no closed form exists for the median, which is the solution m to the equation

$$P_X(X \leq m) = \sum_{k=0}^{m} \frac{\Gamma(\alpha+k)\beta^k}{k!\Gamma(\alpha)(\beta+1)^{k+\alpha}} = 1/2$$

Using the methods above, we derive the moment generating function (MGF) and the characteristic function (CF) of a G-P model. As before, we leverage our knowledge of the MGFs and CFs for Poisson and gamma distributions.

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= E[E[e^{tX}|\Lambda]] \\
&= E[e^{\Lambda(e^t-1)}] \\
&= M_\Lambda(e^t-1) \\
&= \frac{1}{(1-\beta t)^\alpha}\Big|_{t=e^t-1} \qquad t<\frac{1}{\beta} \\
&= (1+\beta-\beta e^t)^{-\alpha} \qquad t<\frac{1}{\beta}
\end{aligned}
$$

$$
\begin{aligned}
\varphi_X(t) &= E[e^{itX}] \\
&= E[E[e^{itX}|\Lambda]] \\
&= E[e^{\Lambda(e^{it}-1)}] \\
&= E[e^{\Lambda(\cos(t)+i\sin(t)-1)}] \\
&= E[e^{i\Lambda(\cos(t)/i+\sin(t)-1/i)}] \\
&= \varphi_\Lambda\left(\cos(t)/i+\sin(t)-1/i)\right) \\
&= \frac{1}{(1-\beta it)^\alpha}\Big|_{t=\cos(t)/i+\sin(t)-1/i} \\
&= \frac{1}{(1-\beta i(\cos(t)/i+\sin(t)-1/i))^\alpha} \\
&= (1+\beta-\beta e^{it})^{-\alpha}
\end{aligned}
$$

Using these functions, we can derive higher moments of the G-P model. Namely, we can find the skewness and kurtosis of the G-P distribution. These values are defined as

$$
\begin{aligned}
Skew(X) &= E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[X^3]-3\alpha\beta E[X^2]+2(\alpha\beta)^3}{(\alpha\beta+\alpha\beta^2)^{3/2}} \\
Kurt(X) &= E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{E[X^4]-4E[X^3](\alpha\beta)^2+6E[X^2](\alpha\beta)^2-3(\alpha\beta)^4}{(\alpha\beta+\alpha\beta^2)^2}
\end{aligned}
$$

Using the MGF of X, we find the necessary values in the equations above.

$$
\begin{aligned}
E[X^2] &= (\alpha\beta)^2+\alpha\beta^2+\alpha\beta \\
E[X^3] &= (\alpha\beta)^2\left[\alpha\beta^2+(3\alpha+1)\beta^2+(3\alpha+1)\beta+\beta^2+2\beta+1\right] \\
E[X^4] &= \alpha\beta\big[\alpha^3\beta^3+(6\alpha^2+4\alpha+1)\beta^3+(6\alpha^2+4\alpha+1)\beta^2+(7\alpha+4)\beta^3+(14\alpha+8)\beta^2 \\
&\quad +(7\alpha+4)\beta+\beta^3+3\beta^2+3\beta+1\big]
\end{aligned}
$$

Substituting these values into the equations above yields the skewness and kurtosis of the G-P probability model.

$$
Skew(X) = \frac{1+2\beta}{(\alpha\beta+\alpha\beta^2)^{1/2}} \qquad Kurt(X) = \frac{3\alpha\beta^2+6\beta^2+3\alpha\beta+6\beta+1}{\alpha\beta+\alpha\beta^2} \tag{7}
$$

2.2 Other Properties

We derive more properties of the G-P model when one of the parameters is known. Suppose that the collection $S_1, S_2, \ldots, S_n$ are G-P RVs, where $S_i \sim \text{Pois}(\Theta_i)$ for $\Theta_i \sim \text{Gamma}(\alpha_i, \beta)$. Note that here we are fixing β by requiring that all X_i have identical scale parameter. Next let $S = \sum_{i=1}^n S_i$. Then we claim that $S \sim \text{G-P}\big(\sum_{i=1}^n \alpha_i, \beta\big)$. To prove the claim, we show that the probability mass function of S matches that found in (5). First note that S is the sum of Poisson RVs, so $S \sim \text{Pois}\big(\Theta\big)$ where $\Theta = \sum_{i=1} \Theta_i$. Moreover, using this notation, then Θ is the sum of gamma RVs so $\Theta \sim \text{Gamma}(\alpha, \beta)$ where $\alpha = \sum_i^n \alpha_i$. Now, we have $S \sim \text{Pois}(\Theta)$ where $\Theta \sim \text{Gamma}(\alpha, \beta)$. We recognize S as a compound distribution and compute $f_S(x)$ by marginalizing over the distribution of Θ.

$$
\begin{aligned}
P_S(S=n) &= \int_0^\infty P(S=n|\Lambda=\lambda) f_\Lambda(\lambda) d\lambda \\
&= \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \left(\frac{1}{\Gamma(a)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \right) d\lambda \\
&= \frac{1}{n!\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{-\lambda(1+\frac{1}{\beta})} \lambda^{n+\alpha-1} d\lambda \\
&= \frac{1}{n!\Gamma(\alpha)\beta^\alpha} \Gamma(n+\alpha) \left(\frac{\beta}{\beta+1} \right)^{n+\alpha} \\
&= \frac{\Gamma(\alpha+n)\beta^n}{n!\Gamma(\alpha)(\beta+1)^{n+\alpha}} \quad \text{for} \quad n=0,1,2\ldots
\end{aligned}
$$

As before, the fourth equality follows by relating the integrand with the kernel of a density from a gamma RV with shape parameter $n+\alpha$ and scale parameter $\frac{1+\beta}{\beta}$. Thus we have demonstrated that the sum of G-P RVs has a G-P distribution with shape parameter $\sum_i^n \alpha_i$ and scale parameter β.

This time, fixing α we show that the G-P model is an exponential family. Taking $c(\beta) = \frac{1}{(\beta+1)^\alpha}$, $h(n) = \frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)}$, $w(\beta) = \log(\frac{\beta}{\beta+1})$, and $t(n) = -n$, we see that

$$
\begin{aligned}
P_X(X=n) &= \frac{\Gamma(\alpha+n)\beta^n}{n!\Gamma(\alpha)(\beta+1)^{n+\alpha}} \\
&= \left(\frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)} \right) \left(\frac{1}{(\beta+1)^\alpha} \right) \left(\frac{\beta}{\beta+1} \right)^n \\
&= \left(\frac{1}{(\beta+1)^\alpha} \right) \left(\frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)} \right) \exp\left(-n \log(\frac{\beta}{\beta+1}) \right) \\
&= c(\beta) h(n) \exp\left\{ w(\beta) t(n) \right\} \quad \text{for} \quad n=0,1,2\ldots
\end{aligned}
$$

Since $\alpha, \beta > 0$, it follows that $c(\beta), h(n) > 0$, hence G-P is an exponential family when α is known. This is an important result for estimation and sufficiency.

2.3 Relationship with Other Distributions

In this section, we focus primarily on relating the G-P probability model to similar, if not entirely derivative, probability models. By understanding where the G-P model is situated among other distributional families, we can further infer properties providing more information about this mixture model. Since the G-P probability model is a discrete distribution that is parameterized by the pair (α, β), we consider other discrete distributions, compound or otherwise, of two parameters similar to the structure of the G-P model.

First, recall that if $W \sim \text{Gamma}(1, \beta)$ then $W \sim \text{Exp}(\beta)$. Hence, by fixing $\alpha = 1$, that the G-P reduces to a compound Exponential-Poisson model where the mean of the Poisson RV follows an exponential distribution. Thus, we see that the G-P model generalizes the compound Exponential-Poisson probability model.

Perhaps more interesting, suppose that $\beta = \frac{p}{1-p}$ for $p \in (0,1)$ and $\alpha = r$ for $r \in \mathbb{N}$. Then the density of the G-P model reduces to

$$
\begin{aligned}
P_X(X=n) = \frac{\Gamma(\alpha+n)\beta^n}{n!\Gamma(\alpha)(\beta+1)^{n+\alpha}} &= \frac{\Gamma(r+n)(\frac{p}{1-p})^n}{n!\Gamma(r)(\frac{1}{1-p})^{n+r}} \\
&= \frac{(r-1+n)!}{n!(r-1)!} \left(\frac{p/(1-p)}{1/(1-p)} \right)^n \left(\frac{1}{1/(1-p)} \right)^r \\
&= \binom{r+n-1}{n} p^n (1-p)^r \quad \text{for} \quad n=0,1,2\ldots \qquad (8)
\end{aligned}
$$

We recognize (8) as the density of a negative binomial distribution with parameters (p, r). Notice that all we require for this reduction is that α takes on nonnegative integer values. Due to this relation, the results derived in Section 2.2 also hold for the negative binomial model. That is, the sum of a collection of negative binomial RVs is also a negative binomial RV. Moreover, we see that fixing r reveals the negative binomial is an exponential family. The authors were surprised to find how closely these probability models were related. In most elementary probability courses, the negative binomial model provides a methodology to analyze how many trials or experiments are needed to acquire a certain number of successes where the number of successes are integer valued. Here, we see that the G-P generalizes this framework where the successes are no longer required to be integer valued, but instead can take any positive real value.

Having shown that the G-P generalizes the negative binomial, it follows that the G-P generalizes any RVs that are special cases of the negative binomial. For instance, consider the case that $\alpha = 1$. Then, as discussed above, this distribution is given by the compound Exponential-Poisson probability model. Furthermore, letting $\beta = \frac{p}{p-1}$, we have

$$P_X(X = n) = \frac{\Gamma(\alpha + n)\beta^n}{n!\Gamma(\alpha)(\beta + 1)^{n+\alpha}} = \frac{\Gamma(1 + n)(\frac{p}{1-p})^n}{n!\Gamma(1)(\frac{1}{1-p})^{n+1}} = p^n(1 - p) \quad \text{for} \quad n = 0, 1, 2 \dots \tag{9}$$

which is a geometric density with parameter p. Hence, we see that the G-P model also generalizes the geometric probability model.

3 Applications

3.1 Theoretical Applications

An immediate theoretical application of the G-P model is understanding how it decomposes into several familiar distributions. In this light, the relationship between the G-P model and the negative binomial makes sense since the gamma, Poisson, and negative binomial models are all intertwined. First we will consider the relationship between the Poisson and gamma distributions. Let $X \sim \text{Gamma}(\alpha, 1)$. Then for $\alpha \in \mathbb{N}$ we have

$$P(X \geq x) = \int_x^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy = \frac{1}{\Gamma(\alpha)} \int_x^\infty y^{\alpha-1} e^{-y} dy$$

Using integration by parts, we obtain the recursive formula

$$\frac{1}{\Gamma(\alpha)} \int_x^\infty y^{\alpha-1} e^{-x} dy = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} + \frac{1}{\Gamma(\alpha - 1)} \int_x^\infty y^{\alpha-2} e^{-y} dy$$

By iteratively integrating by parts we have

$$\begin{aligned} P(X \geq x) &= \frac{1}{\Gamma(\alpha)} \int_x^\infty y^{\alpha-1} e^{-y} dy \\ &= \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} + \frac{1}{\Gamma(\alpha - 1)} \int_x^\infty y^{\alpha-2} e^{-y} dy \\ &= \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} + \frac{x^{\alpha-2}e^{-x}}{\Gamma(\alpha - 1)} + \frac{1}{\Gamma(\alpha - 2)} \int_x^\infty y^{\alpha-3} e^{-y} dy \\ &\vdots \\ &= \sum_{k=1}^{\alpha} \frac{x^{k-1}e^{-x}}{\Gamma(k)} = \sum_{k=0}^{\alpha-1} \frac{x^k e^{-x}}{k!} \end{aligned}$$

But notice that this is exactly the density of a Poisson RV with rate parameter x. In fact, we see that this is just the cumulative probability function of a Poisson at $\alpha - 1$. That is if $Y \sim \text{Pois}(x)$ then we have

$$P(X \geq x) = P(Y \leq \alpha - 1) \tag{10}$$

This relation shows that the distribution function of both gamma and Poisson RVs are closely connected.

Next, consider the following relationship between the gamma and negative binomial distributions. If $Y \sim NB(r,p)$, then the RV $pY \sim Gamma(r,1)$ as $p \to 0$. To see this, recall that $M_Y(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$. It follows that

$$M_{pY}(t) = E[e^{pYt}] = E[e^{Y(pt)}] = M_Y(pt) = \left(\frac{p}{1-(1-p)e^{pt}}\right)^r$$

Thus as $p \to 0$

$$M_{pY}(t) = \lim_{p\to 0}\left(\frac{p}{1-(1-p)e^{pt}}\right)^r = \left(\lim_{p\to 0}\frac{p}{1-(1-p)e^{pt}}\right)^r = \left(\lim_{p\to 0}\frac{1}{-te^{pt}+e^{pt}+tpe^{pt}}\right)^r = \left(\frac{1}{1-t}\right)^r$$

Note that the third equality follows from an application of L'Hospital's rule. Hence, $M_{pY}(t)$ is just the MGF of a gamma RV with shape parameter r and rate parameter 1.

Finally, consider the relationship between the Poisson and the negative binomial distributions. Specifically, under certain conditions, we show that the MGF of a negative binomial limits to the MGF of Poisson. Suppose $r \to \infty$, $p \to 1$ and $r(1-p) \to \lambda$. Then

$$\begin{aligned}
M_{NB}(t) &= \left(\frac{p}{1-(1-p)e^t}\right)^r \\
&= \left(\frac{1-(1-p)}{1-(1-p)e^t}\right)^r \\
&= \left(\frac{1-(1-p)e^t+(1-p)e^t-(1-p)}{1-(1-p)e^t}\right)^r \\
&= \left(\frac{1-(1-p)e^t+(1-p)(e^t-1)}{1-(1-p)e^t}\right)^r \\
&= \left(1+\frac{(1-p)(e^t-1)}{1-(1-p)e^t}\right)^r \\
&= \left(1+\frac{1}{r}\frac{r(1-p)(e^t-1)}{1-(1-p)e^t}\right)^r
\end{aligned}$$

Now, letting $r \to \infty$ then $p \to 1$ we have

$$\lim_{\substack{r\to\infty \\ p\to 1}} M_{NB}(t) = \exp\left\{\lim_{r(1-p)\to\lambda} r(1-p)(e^t-1)\right\} = \exp\left\{e^t - 1\right\} \tag{11}$$

Hence, we see that the MGF of a negative binomial RV converges to a MGF of a Poisson RV.

In this section, we have shown the interconnectedness of the gamma, Poisson, and negative binomial models. We demonstrated how under certain conditions, the negative binomial behaves like a Poisson or gamma distribution. The theoretical utility of the G-P model is that is provides a simple, yet elegant way to connect the gamma and Poisson probability models back to the negative binomial. By working in a mixture setting, we generalized the negative binomial parameter space from $\mathbb{N} \times (0,1)$ to $\mathbb{R}^+ \times \mathbb{R}^+$. The advantage of working in this compound probability model setting is that we begin to infer and study *classes* and *generalizations* of RVs which can be applied to more specific cases as needed. In our case, we showed that by understanding and proving interesting features of the G-P model, our results extend far beyond our initial frame of inference to other probability models.

3.2 Applied Applications

There are many applied applications of the G-P model in wide-ranging disciplines. The commonality between these applications is the goal to model frequencies. For example, the G-P model is used in the insurance industry to model claim frequency. Though insurers pool risk during rating and underwriting, there is known segmentation between risk groups. The G-P model provides insurers the ability to compute expected claim frequencies for all policies while flexibly assigning different average claim-frequencies to individual risk groups. This provides a more accurate framework for claims-frequency data that is right-skewed and heavy around zero, which does not follow traditional distributions.

Another application of the G-P model is text classification. In text classification, one interest is in modelling the frequency of given words. A Poisson distribution is appropriate since it models the number of successes, here occurrences of a given word, during some time interval, say a word document. However, if it is known that there are different word-document classes, then a G-P model is preferable since it provides the flexibility to specify the frequency of a given word. For instance, a text classifier would expect "anaphlaxis" to occur more frequently in *New England Journal of Medicine* than in *Rolling Stone*, and this can be accounted for using the G-P framework.

Finally, the G-P model inherits all the applications of the negative binomial distribution when $\alpha \in \mathbb{N}$. The classic examples include modeling the waiting time for the number of successes or failures such as the number of potential jurors that need to be interviewed to fill a jury. This approach lends itself to a technique known as inverse binomial sampling.

4 Conclusion

The G-P model is a compound distribution where the mean of a Poisson random variable follows a gamma distribution. From the definition of a compound distribution, we derived its density. Using the underlying Poisson and gamma distributions, we easily obtained its moments, as well as the MGF and CF. By fixing one of the parameters, we demonstrated that the sum of G-P RVs is G-P, and by fixing the other parameter we saw that the G-P model is an exponential family. We also recognized that by limiting the parameter space, the G-P model is simply the negative binomial, and we explored the relationships between the gamma, Poisson, and negative binomial models. Finally, we covered the applications of the G-P distribution in modeling frequency.

5 References

[1] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, 2002. ISBN: 0534243126.

[2] John D. Cook. *Notes on the Negative Binomial*. 2009. URL: `https://www.johndcook.com/negative_binomial.pdf`.

[3] B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Monographs on Applied Probability and Statistics. Chapman and Hall, 1981. ISBN: 0412224208.

[4] Lawrence Leemis. *Gamma-Poisson distribution*. URL: `http://www.math.wm.edu/~leemis/chart/UDR/PDFs/Gammapoisson.pdf`.

[5] Dan Ma. *The Negative Binomial Distribution*. 2011. URL: `https://probabilityandstats.wordpress.com/tag/poisson-gamma-mixture/`.

[6] Hiroshi Ogura et al. "Gamma-Poisson Distribution Model for Text Categorization". In: *ISRN Aritifical Intelligence* (2013). DOI: `https://www.hindawi.com/journals/isrn/2013/829630/`.

[7] Brian Peacock. *Statistical Distributions*. Wiley series in probability and statistics. Wiley, 2011. ISBN: 9780470390634.