Academic Advances of the CTO, Vol. 1, Issue 3, 2017
Original Research Article

A Generalized Geometric Distribution from Vertically Dependent Bernoulli Random Variables

Rachel Traylor*, Ph.D.

Abstract
This paper generalizes the notion of the geometric distribution to allow for dependent Bernoulli trials generated from dependency generators as defined in Traylor and Hathcock's previous work. The generalized geometric distribution describes a random variable X that counts the number of dependent Bernoulli trials until the first success. The main result of the paper is X can count dependent Bernoulli trials from any dependency structure and retain the same distribution. That is, if X counts Bernoulli trials with dependency generated by $\alpha_1 \in \mathcal{C}_\delta$, and Y counts Bernoulli trials with dependency generated by $\alpha_2 \in \mathcal{C}_\delta$, then the distributions of X and Y are the same, namely the generalized geometric distribution. Other characterizations and properties of the generalized geometric distribution are given, including the MGF, mean, variance, skew, and entropy.

Keywords
dependent Bernoulli variables — probability theory–geometric distribution

Contents

*Office of the CTO, Dell EMC

Introduction

The standard geometric distribution counts one of two phenomena:

(1) The count of i.i.d. Bernoulli trials until the first success

(2) The count of i.i.d. Bernoulli trials that resulted in a failure prior to the first success

The latter case is simply a shifted version of the former. However, this distribution, in both forms, has limitations because it requires a sequence of independent and identically distributed Bernoulli trials. Korzeniowski [2] originally defined what is now known as first-kind (FK) dependent Bernoulli random variables, and gave a generalized binomial distribution that allowed for dependence among the Bernoulli trials. Traylor [4] extended the work of Korzeniowski into FK-dependent categorical random variables and derived a generalized multinomial distribution in a similar fashion. Traylor and Hathcock [5] extended the notion of dependence among categorical random variables to include other kinds of dependency besides FK dependence, such as sequential dependence. Their work created a class of *vertical dependency structures* generated by a set of functions

$$\mathscr{C}_\delta = \{\alpha : \mathbb{N}_{\geq 2} \to \mathbb{N} : \alpha(n) < n \text{ and } \forall n \exists j \in \{1,...,n-1\} : \alpha(n) = j\},$$

where the latter property is known as *dependency continuity*. In this paper, we derive a generalized geometric distribution from identically distributed but *dependent* Bernoulli random variables. The main result is that the pdf for the generalized geometric distribution is the same regardless of the dependency structure. That is, for any $\alpha \in \mathscr{C}_\delta$ that generates a sequence of identically distributed but dependent Bernoulli trials, the generalized geometric distribution remains unchanged.

1. Background

The standard geometric distribution is built from a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables with probability of success p and probability of failure $q = 1 - p$. There are two "versions" of the geometric distribution:

(1) A random variable X has a geometric distribution if it counts the number of Bernoulli trials needed to observe the first success.

(2) A random variable $Y = X - 1$ has a geometric distribution if it counts the number of failures in a sequence of Bernoulli trials before the first observed success.

In the first case, X has support $\{1,2,3,...\}$, because we are looking for the first success, which can occur on trial 1, 2, 3, and so forth. In the latter case, Y has support $\{0,1,2,...\}$ because we are counting the number of failures before the first success occurs. That is, if the first success occurs on Trial 1, then there were 0 failures preceding the first success. If the first success occurs on trial 2, then one failure occurred prior to the first success, and thus $Y = 1$. Essentially, Version 2 is a shifted Version 1, because our perspective changes– we do not include the success in the count in Version 2.

For Version 1, the pdf is given by

$$f_X(k) = q^{k-1}p, \quad k = 1,2,3,\ldots \tag{1}$$

For Version 2, (the shifted generalized geometric distribution)the pdf is given by

$$f_Y(k) = q^k p, \quad k = 0,1,2,\ldots \tag{2}$$

The next section derives the generalized geometric distribution for FK-dependent random variables, and then shows that the pdf is the same regardless of dependency structure.

2. Generalized Geometric Distribution

Derivation from FK-Dependent Bernoulli Random Variables

Suppose we have a sequence of FK-dependent Bernoulli Random variables. Recall from [2] and [4] that FK-dependent random variables are weighted toward the outcome of the first variable ε_1. That is, in the Bernoulli case, $P(\varepsilon_1 = 1) = p$ and $P(\varepsilon_1 = 0) = q = 1 - p$. For subsequent variables in the sequence,

$$P(\varepsilon_n = 1|\varepsilon_1 = 1) = p^+ P(\varepsilon_n = 1|\varepsilon_1 = 0) = p^-$$
$$P(\varepsilon_n = 0|\varepsilon_1 = 1) = q^- P(\varepsilon_n = 0|\varepsilon_1 = 0) = q^+$$

for $n \geq 2$, where $q = 1 - p$, $p^+ = p + \delta q$, $p^- = p - \delta p$, $q^- = q - \delta q$, $q^+ = q + \delta p$, and $0 \leq \delta \leq 1$ is the *dependency coefficient*.

We will first give the generalized "Version 1" of the geometric distribution for FK-dependent random variables.

Proposition 1. *Suppose $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots)$ is a FK-dependent sequence of Bernoulli random variables. Let X be the count of such Bernoulli variables needed until the first success. Then X has a generalized geometric distribution with pdf*

$$f_X(k) = \begin{cases} p, & k = 1 \\ q\left(q^+\right)^{k-2} p^-, & k \geq 2 \end{cases} \tag{3}$$

Proof. The probability of the first success occurring on the first trial is the probability that $\varepsilon_1 = 1$, so

$$P(X = 1) = P(\varepsilon_1) = 1 = p.$$

The probability of the first success occurring on the second trial is the probability of the following sequence $\varepsilon = (0,1)$, and thus $P(X = 2) = P(\varepsilon = (0,1)) = qp^-$. In general, for $k \geq 2$,

$$P(X = k) = P(\varepsilon = (0,0,...,0,1))$$

with $k - 1$ failures (or 0s). Since the variables are FK-dependent, the probability of failure after the first failure is q^+, and the probability of success given that $\varepsilon_1 = 0$ is p^- regardless of where in the sequence that success occurs. Therefore, $P(\varepsilon = (0,0,\ldots,0,1) = q\left(q^+\right)^{k-2} p^-$. □

In a similar fashion, suppose we prefer to count the number of failures before the first success occurs. For this generalized "Version 2", we have the following proposition.

Proposition 2. *Suppose $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots)$ is a FK-dependent sequence of Bernoulli random variables. Let $Y = X - 1$ be the count of failures prior to the first success. Then Y has a shifted generalized geometric distribution with pdf*

$$f_Y(k) = \begin{cases} p, & k = 0 \\ q\left(q^+\right)^{k-1} p^-, & k \geq 1 \end{cases} \tag{4}$$

Proof. The proof follows in an identical fashion to the proof of Proposition 1 □

Generalized Geometric Distribution for any Vertical Dependency Structure

Propositions 1 and 2 were derived for FK-dependent random variables, but in fact these random variables X and Y remain distributed according to the generalized geometric distribution and the shifted generalized geometric distribution regardless of the vertical dependency structure specified, as long as the dependency structure was generated from a function in $\mathscr{C}_\delta$.

Theorem 1. *Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots)$ be a vertically dependent sequence of Bernoulli random variables, where the dependency is generated by $\alpha \in \mathscr{C}_\delta$. Let X be the count of such Bernoulli trials needed until the first success, and let $Y = X - 1$ be the count of failures of such Bernoulli trials prior to the first success. Then the pdf of X and Y is identical to those given in Propositions 1 and 2.*

Proof. Let $\alpha \in \mathscr{C}_\delta$ generate the vertically dependent Bernoulli sequence ε. Then $P(X = 1) = P(Y = 0) = P(\varepsilon_1 = 1) = p$. For $k \geq 2$, $P(X = 1) = P(\varepsilon = (0, 0, \ldots, 0, 1))$, where there are $k - 1$ 0s prior to the first 1. Then $P(\varepsilon_k = 1) = p^-$, because $\alpha(k) \in \{1, ..., k-1\}$, and $\varepsilon_i = 0$ for all $i \leq k$. $P(\varepsilon_1 = 0) = q$ for $k \geq 2$. Then, for $2 \leq i \leq k - 1$, $\alpha(i) = j \in \{1, ..., i-1\}$, and $\varepsilon_j = 0 \forall j = 1, ..., i-1$ and $\forall i = 1, ..., k-1$. Thus $P(\varepsilon_i = 0 | \varepsilon_{\alpha(i)} = 0) = q^+$ for all $i = 2, \ldots, k-1$, and therefore $P(X = k) = q\left(q^+\right)^{k-2} p^-$. A similar argument follows for $P(Y = k), k \geq 1$. □

This result is quite powerful, and not one that holds for all generalized distributions constructed from dependent random variables. Given any vertical dependency structure generated from the broad class $\mathscr{C}_\delta$, the count of trials before a success and the count of failures before a success have the same probability distribution. Thus, if this information is desired, no information about the dependency structure other than the membership of its generating function in $\mathscr{C}_\delta$ is necessary. The only information needed to calculate the generalized geometric probabilities for dependent Bernoulli trials is p and δ.

The next section gives some basic properties of the Generalized Geometric Distribution, such as the moment generating function and selected moments.

3. Properties of the Generalized Geometric Distribution

3.1 Moment Generating Function

Fact 1. *The moment generating function of the generalized geometric distribution is*

$$M_X(t) = pe^t + \frac{qp^- e^{2t}}{1 - q^+ e^t} \tag{5}$$

Derivation. $M_X(t) = E\left[e^{tX}\right]$ by definition, so

$$\begin{aligned} M_X(t) &= pe^t + \sum_{k=2}^{\infty} q\left(q^+\right)^{k-2} p^- e^{kt} \\ &= pe^t + \frac{qp^- e^{2t}}{1 - q^+ e^t} \end{aligned}$$

Using the moment generating function, we can give moments of the generalized geometric distribution.

3.2 Mean

Fact 2. *The mean of the generalized geometric distribution is*

$$E[X] = \mu = \frac{1 - \delta p}{p(1 - \delta)}$$

Derivation. Taking the first derivative of (5) and evaluating at $t = 0$ yields the mean after arithmetic simplification. See Section 5 for full derivation.

The effect of dependence can be seen in the plot of $E[X]$ below in Figure 1. For fixed p, when $\delta \to 1$, $E[X] \to \infty$, though the rate changes with p.

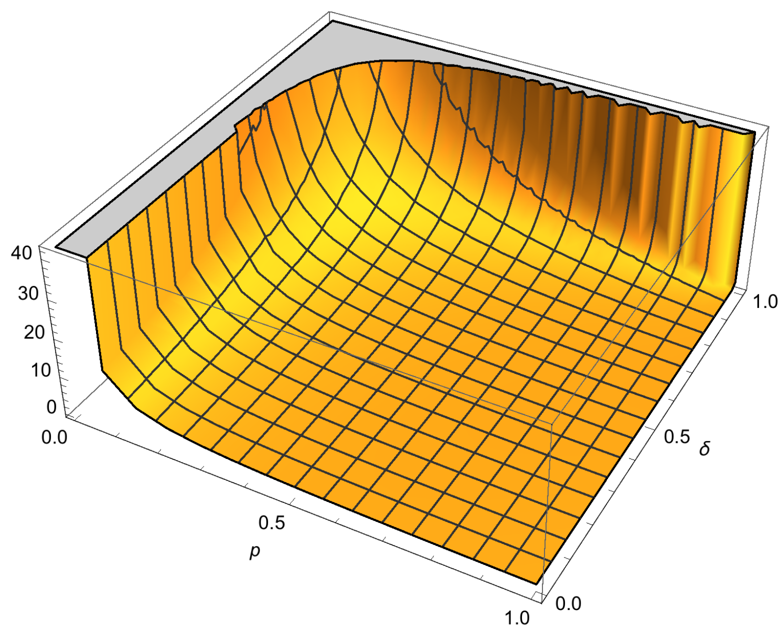


Figure 1. Expectation of the Generalized Geometric Distribution

To explore further, suppose $p = 1/2$, and the Bernoulli trials are thus balanced between success and failure. Figure 2 shows the effect of delta for a fixed p. Notice that the effect of δ on the expected value become more pronounced as $\delta \to 1$. In particular, for $\delta = 1/2$ and $p = 1/2$, $E[X] = 3$, but after this point, an increase of only $1/6$ in δ to $\delta = 2/3$ increased the expected value to 4 trials before a success. To double the expected number of trials before a success again to $E[X] = 8$ requires an increase of δ by only $4/21$ to $6/7$.

A smaller probability of success p will yield an expected value μ that is much more susceptible to effects of dependency δ, and a larger p will yield an expected value more resistant to high dependency δ. Since the geometric distribution is a count of the number of trials needed to obtain the first success, a higher p increases the probability that the first success occurs on the first trial, while a lower p decreases that probability. Therefore, the dependency δ would have a higher effect for lower p, because a longer (dependent) sequence is expected to be generated prior to the first success, which increases the expected number of trials faster than if the Bernoulli trials were independent.

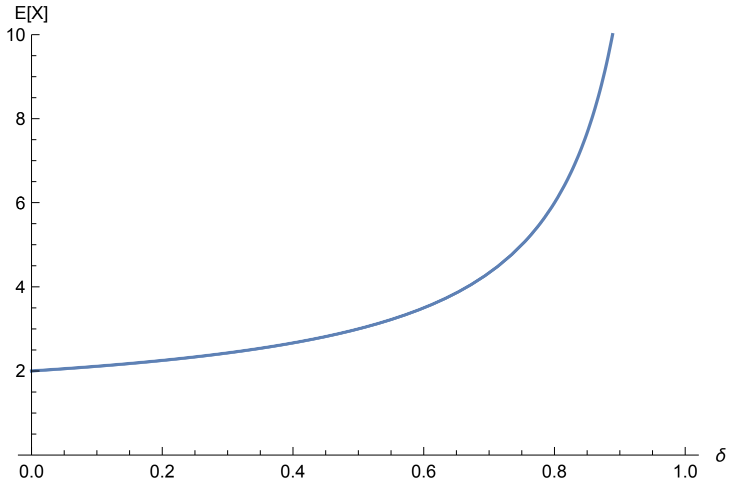


Figure 2. Expected value of generalized geometric distribution with fixed $p = 1/2$.

Remark. *Notice that when $\delta = 0$, the Bernoulli trials are independent. The mean of the generalized geometric distribution when $\delta = 0$ is $E[X] = \frac{1}{p}$, the mean of the standard geometric distribution.*

3.3 Variance

Fact 3. *The variance of the generalized geometric distribution is*

$$\text{Var}(X) = \sigma^2 = \frac{1 - p + \delta p(1-p)}{p^2(1-\delta)^2}$$

Derivation. See Section5.

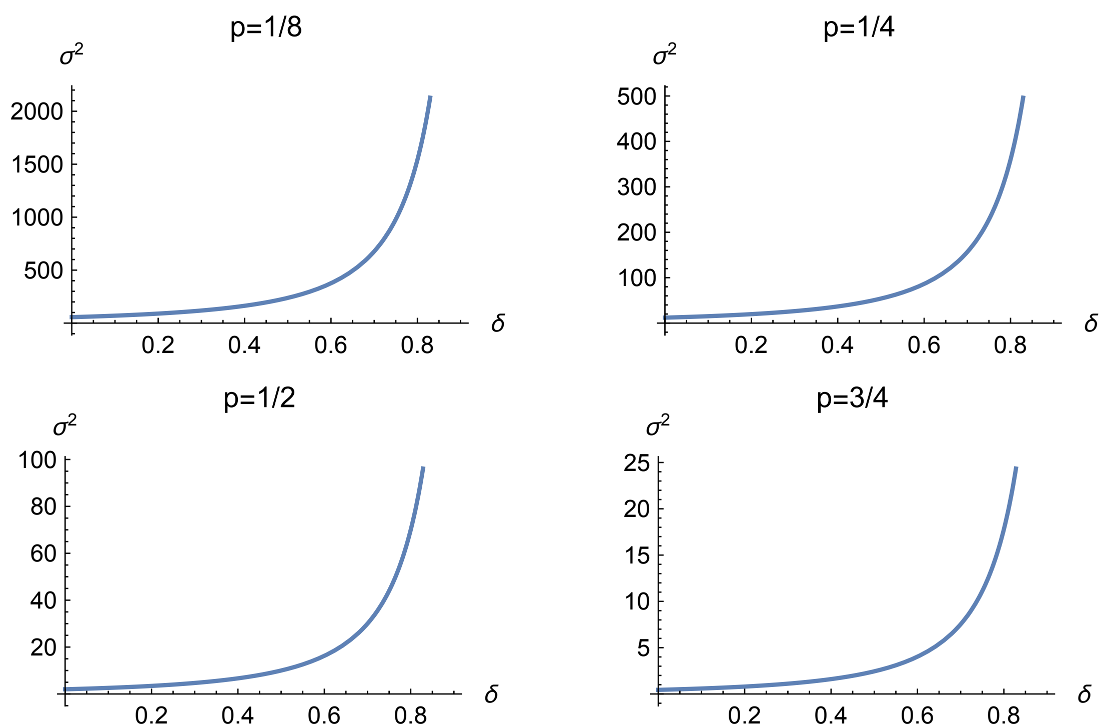


Figure 3. Effect of δ on Var$[X]$ for various values of fixed p.

Figure 3 shows the effect of δ on the variance for different values of p. As with the mean, a smaller p induces a higher effect of δ on the variance of the number of dependent Bernoulli trials before the first success is observed. The shape of all 4 cases is similar, but the scales are vastly different. As p increases, the scale of the variance decreases dramatically.

Remark. Again, note that when $\delta = 0$, the variance of the generalized geometric distribution reduces to that of the standard geometric distribution.

3.4 Skew

Fact 4. *The skew of the generalized geometric distribution is given by*

$$\text{Skew}[X] = \frac{2 - 3p + p^2 + \delta p[q + \delta pq + p(2\delta - 1 - p)]}{(q + \delta pq)^{3/2}}$$

Derivation. See Section 5.

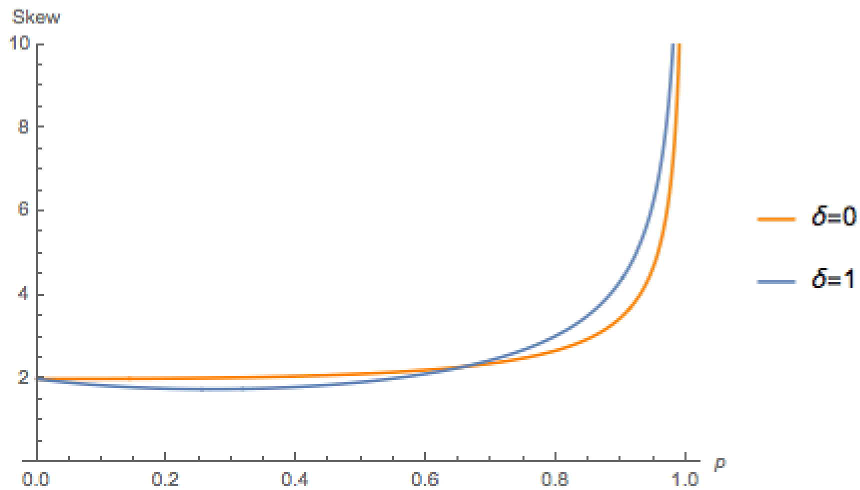


Figure 4. Skew of Generalized Geometric Distribution as a function of p for $\delta = 0$ and $\delta = 1$.

The skew of the generalized geometric distribution gets more complicated in its behavior as a function of both p and δ. Figure 4 shows the skew as a function of p for the two extreme cases: complete independence ($\delta = 0$) and complete dependence $\delta = 1$. From $p = 0$ to $p \approx 0.658$, the skew for the independent geometric distribution is greater than the completely dependent case. For $p \gtrapprox 0.658$, the skew is greater under complete dependence.

3.5 Entropy

The *entropy* of a random variable measures the average information contained in the random variable. It can also be viewed as a measure of how unpredictable or "truly random" the variable is [1] . The definition of entropy, denoted $H(X)$, was coined by Claude Shannon [3] in 1948.

Definition 1 (Entropy). $H(X) := -\sum_i P(x_i) \log_2(P(x_i))$

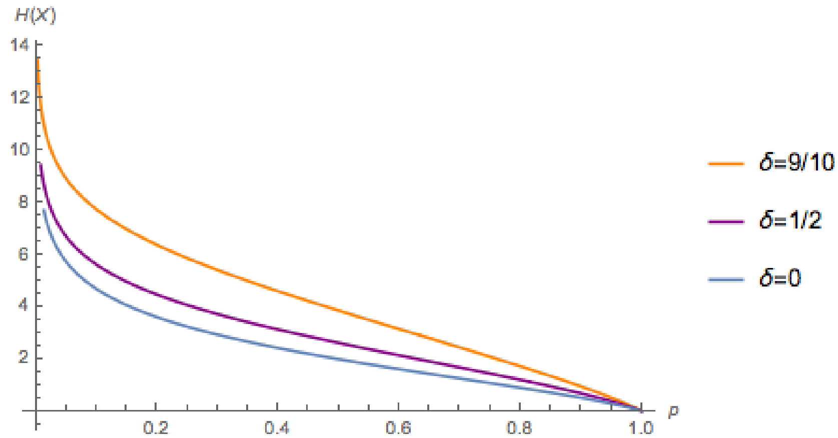
For the standard geometric distribution, the entropy is given by

$$H_{sg}(X) = \frac{-(1-p)\log_2(1-p) - p\log_2(p)}{p}$$

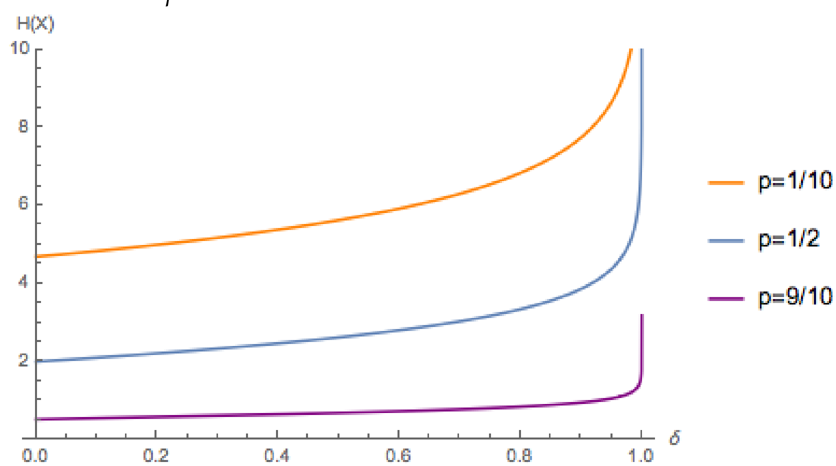
Fact 5. *The entropy for the generalized geometric distribution is*

$$H_{gg}(X) = \frac{-\left[pp^-\log_2(p) + qp^-\log_2(qp^-) + qq^+\log_2(q^+)\right]}{p^-}$$

Derivation. See Section 5.



(a) Entropy of the generalized geometric distribution as a function of p for selected values of δ.



(b) Entropy of the generalized geometric distribution as a function of δ for selected values of p

Figure 5. Entropy from two different perspectives: fixed δ and fixed p

Figure 5a shows $H_{gg}(X)$ as a function of p for fixed values of δ. Notice that while the entropy decreases to 0 for all curves as $p \to 1$, the entropy curve is shifted upward for larger δ. Figure 5b fixes p and looks at entropy as a function of δ. Notice that for smaller p, the entropy is much higher, which aligns with intuition.

4. Conclusion

The standard geometric distribution counts the number of independent Bernoulli trials until the first success. This paper uses the works of Koreniowski, and Traylor and Hathcock [2, 4, 5] on dependent Bernoulli and categorical random variables to develop a generalized geometric distribution built from dependent Bernoulli random variables. The main result of the paper is that the pdf for the generalized geometric distribution is independent of the dependency structure of the Bernoulli random variables that comprise it. That is, regardless of dependency structure, the pdf for the generalized geometric distribution is given by Proposition 1. Various properties and characterizations were given, including the moment generating function, mean, variance, skew, and entropy. The effect of dependency on each property was studied.

5. Appendix

5.1 Derivation of the mean

$$
\begin{aligned}
E[X] &:= \sum_{k=1}^{\infty} kP(k) \\
&= p + \sum_{k=2}^{\infty} k\left(q(q^+)^{k-2}p^-\right) \\
&= p + \frac{qp^-(2-q^+)}{(p^-)^2} \\
&= \frac{pp^- + q(2-q^+)}{p^-} \\
&= \frac{pp^- + q(1+1-q^+)}{p^-} \\
&= \frac{pp^- + q(1+p^-)}{p^-} \\
&= \frac{pp^- + (1-p)(1+p^-)}{p^-} \\
&= \frac{1-p+p^-}{p^-} \\
&= \frac{1-\delta p}{p(1-\delta)}
\end{aligned}
$$

5.2 Derivation of the variance

$\mathrm{Var}[X] = E[X^2] - E[X]^2$, and thus we must first obtain $E[X^2]$.

$$
\begin{aligned}
E[X^2] &:= \sum_{k=1}^{\infty} k^2 P(k) \\
&= p + \sum_{k=2}^{\infty} k^2\left(q(q^+)^{k-2}p^-\right) \\
&= p + \frac{qp^-(4-3q^+ + (q^+)^2))}{(p^-)^3} \\
&= p + \frac{q(4-3q^+ + (q^+)^2)}{(p^-)^2} \\
&= \frac{p(p^-)^2 + q(q^+)^2 + 4q - 3qq^+}{(p^-)^2}
\end{aligned}
$$

We will simplify the numerator. First,

$$
\begin{aligned}
p(p^-)^2 + q(q^+)^2 &= p^3 + q^3 - 2p^3\delta + 2p^2\delta q + p^3\delta^2 + p^2\delta^2 q \\
&= 1 - 3p + 3p^2 + 2p(q^2\delta - p^2\delta) + \delta^2 p^2(p+q) \\
&= 1 - 3p + 3p^2 + 2p\delta(1-2p) + \delta^2 p^2
\end{aligned}
$$

Then

$$
\begin{aligned}
4q - 3qq^+ &= 4(1-p) - 3(1 - 2p + p^2 + \delta p - \delta p^2) \\
&= 1 + 2p - 3p^2 - 3p\delta + 3\delta p^2
\end{aligned}
$$

Combining the two, the numerator is given by

$$\begin{aligned} p(p^-)^2 + q(q^+)^2 + 4q - 3qq^+ &= 1 - 3p + 3p^2 + 2p\delta(1-2p) + \delta^2p^2 + 1 + 2p - 3p^2 - 3p\delta + 3\delta p^2 \\ &= 2 - p - \delta p(1 - \delta p + p) \end{aligned}$$

Thus $E[X^2] = \frac{2-p-\delta p(1-\delta p+p)}{p^2(1-\delta)^2}$
Finally,

$$\begin{aligned} \mathrm{Var}[X] &= \frac{2 - p - \delta p(1 - \delta p + p)}{p^2(1-\delta)^2} - \left(\frac{1-\delta p}{p(1-\delta)}\right)^2 \\ &= \frac{2 - p - \delta p(1 - \delta p + p) - [1 - 2\delta p + \delta^2 p^2]}{p^2(1-\delta)^2} \\ &= \frac{1 - p + \delta pq}{p^2(1-\delta)^2} \end{aligned}$$

5.3 Derivation of Skew

The skew of a random variable is given by

$$\mathrm{Skew}[X] := E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{E[X^3] - 3\mu\mathrm{Var}[X] - \mu^3}{(\mathrm{Var}[X])^{3/2}}$$

First, we derive $E[X^3]$.

$$\begin{aligned} E[X^3] &= \sum_{k=1}^{\infty} k^3 P(k) \\ &= p + \sum_{k=2}^{\infty} k^3 q(q^+)^{k-2} p^- \\ &= \frac{6 - 6p + p^2(1-\delta)^2 = p^3\delta(1-\delta)^2}{p^3(1-\delta)^3} \end{aligned}$$

Then

$$\begin{aligned} \mathrm{Skew}[X] &= \frac{\frac{6-6p+p^2(1-\delta)^2-p^3\delta(1-\delta)^2}{p^3(1-\delta)^3} - 3\left(\frac{1-\delta p}{p(1-\delta)}\right)\left(\frac{1-p+\delta p(1-p)}{p^2(1-\delta)^2}\right) - \left(\frac{1-\delta p}{p(1-\delta)}\right)^3}{\left(\frac{1-p+\delta p(1-p)}{p^2(1-\delta)^2}\right)^{3/2}} \\ &= \frac{6 - 6p + p^2(1-\delta)^2 - p^3\delta(1-\delta)^2 - 3(1-\delta p)(1 - p + \delta p(1-p)) - (1-\delta p)^3}{(1 - p + \delta p(1-p))^{3/2}} \end{aligned}$$

Now, the numerator is expanded to

$$\begin{aligned} 6 - 6p + &p^2(1 - 2\delta + \delta^2) - p^3\delta(1 - 2\delta + \delta^2) - 3(1-\delta p)((1-p) + \delta p(1-p)) + (1 - \delta p + \delta^2 p^2 - \delta^3 p^3) \\ &= 6 - 6p + p^2 - 2\delta p^3 - p^3\delta + 2\delta^2 p^3 - 3[(1-p)(1-\delta^2p^2)] - 1 + \delta p \\ &= 2 - 3p + p^2 - 2\delta p^2 - p^3\delta - \delta^2 p^3 + 3\delta^2 p^2 + \delta p \\ &= 2 - 3p + p^2 + \delta p(q + \delta pq + p(2\delta - 1 - p)) \end{aligned}$$

And thus the skew is as given in Fact 3.

5.4 Derivation of Entropy

Entropy is defined as $H(X) := \sum_{i=1}^{\infty} P(x_i)\log_2(P(x_i))$. For the generalized geometric distribution,

$$\begin{aligned}H(X) &= -\left(p\log_2(p) + \sum_{i=2}^{\infty} qp^{-}(q^{+})^{i-2}\log_2(qp^{-}(q^{+})^{i-2}))\right)\\ &= -\left(p\log_2(p) + qp^{-}\log_2(qp^{-})\sum_{i=2}^{\infty}(q^{+})^{i-2} + qp^{-}\log_2(q^{+})\sum_{i=2}^{\infty}(i-2)(q^{+})^{i-2}\right)\end{aligned}$$

Simplifying the above yields the result.

Acknowledgments

The author is grateful to Jason Hathcock for editorial review and the DPD CTO of Dell EMC and Stephen Manley for continued funding and support.

References

[1] Skewness. https://en.wikipedia.org/wiki/skewness.

[2] KORZENIOWSKI, A. On correlated random graphs. *Journal of Probability and Statistical Science* (2013), 43–58.

[3] SHANNON, C. A mathematical theory of communication. *The Bell System Technical Journal* (1948).

[4] TRAYLOR, R. A generalized multinomial distribution from dependent categorical random variables. *Academic Advances of the CTO* (2017).

[5] TRAYLOR, R., AND HATHCOCK, J. Vertical dependency in sequences of categorical random variables. *Academic Advances of the CTO* (2017).