A Stopped Negative Binomial Distribution

Michelle DeVeaux, Michael J. Kane*, Daniel Zelterman

Department of Biostatistics
School of Epidemiology and Public Health
Yale University, New Haven, CT

Abstract

We introduce a discrete distribution suggested by curtailed sampling rules common in early-stage clinical trials. We derive the distribution of the smallest number of independent and identically distributed Bernoulli trials needed to observe either s successes or t failures. This report provides a closed-form expression for the mass function and illustrates limiting approximations.

Keywords: discrete distribution, curtailed sampling

1. Introduction and Motivation

Consider a prototypical early phase, single-arm clinical trial in which 12 patients are enrolled and treated. The binomial probability of success is $p = 0.045$ for each patient under the null hypothesis that treatment is not effective. If two or more patients out of these 12 respond to the treatment then we reject this null hypothesis and the treatment is deemed successful at significance level of 0.1. If fewer than two respond then the null is not rejected and the treatment is judged as ineffective.

☆This research was supported by grants R01CA131301, R01CA157749, R01CA148996, R01CA168733, and PC50CA196530 awarded by the National Cancer Institute along with support from the Yale Comprehensive Cancer Center and the Yale Center for Outcomes Research. We would also like to thank Rick Landin at Ignyta Inc. for his suggestions.

Email addresses: michelle.deveaux@yale.edu (Michelle DeVeaux), michael.kane@yale.edu (Michael J. Kane*), daniel.zelterman@yale.edu (Daniel Zelterman)

* Corresponding author

If all 12 patients are enrolled at once, as in the classic design, then the sample size is 12. However, in most clinical trials the patients are enrolled sequentially, often with one patient's outcome realized before the next one enters the trial. In the present example, observing two successful patients allows us to reach one endpoint so the sample required could be as small as two. Similarly 11 observed treatment failures also ends the trial. This sampling mechanism, in which the experiment ends as soon as a predefined endpoint is reached, is called *curtailed sampling*. Under curtailed sampling the range of the sample size for this trial is between two and 12.

Let us assume each patient outcome can be modeled as an independent, identically distributed Bernoulli(p) random variable. The trial is realized as a sequence of these random variables stopping when either a specified number of successes or failures has been reached.

A hypothetical sample path is illustrated in Fig. 1. The vertical axis denotes the number of successful outcomes. The horizontal axis counts the number of patients enrolled. The horizontal and vertical boundaries represent endpoints for the trial. In this case two successes were reached after enrolling 10 patients: one in the fourth outcome and on in the eleventh. Since the success boundary is reached, we say that the trial succeeds, or more correctly, the treatment succeeds.

More generally the distribution for number of trial enrollments is shown in

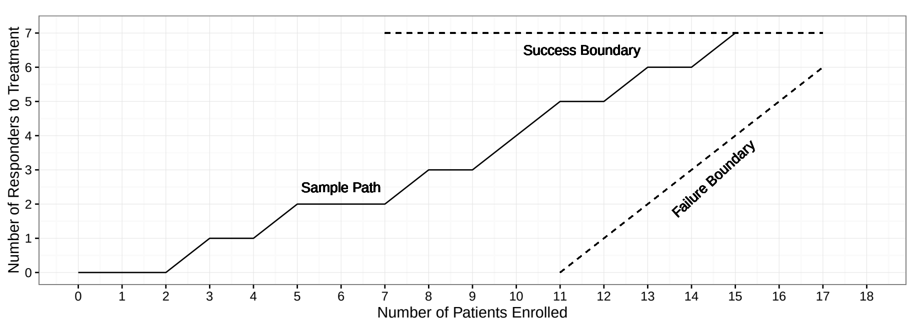
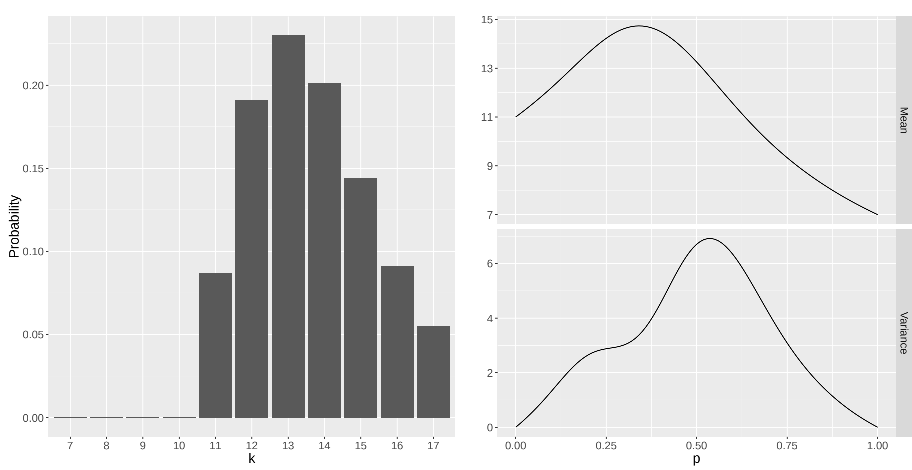


Figure 1: A hypothetical realization of the trial.

Fig. **??** (a). The probability mass is concentrated at the 11^{th} and 12^{th} step since p is small and for most realizations reach 11 non-responders before reaching 2 responders.

Fig. 2 shows the expected value and variance for the number of trial enrollments varying p between zero an 1. When p is small the probability of an individual response is small, the treatment is more likely to fail, and the number of enrollees is large. When p is large the treatment is more likely to succeed and the number of enrolees is small. Since $p = 1$ and $p = 0$ represent the deterministic the corresponding variance is zero. The variance is right-skewed because the support of the distribution corresponding to successful treatment is larger than that of failure.

In the rest of this work we derive the distribution of the number of enrollees needed to observe either s successes or t failures. We refer to this distribution as the Stopped Negative Binomial (SNB). Some of its characteristics are summarized in Tab. 1. This paper derives these results and explores properties of



(a) enrollment distribution

(b) effect of varying p

Figure 2: The distribution of a trial that stops after seven patients respond or 11 patients do not. (a) shows the distribution when $p = 0.2$. (b) shows the result of varying p between zero and one.

Table 1: Characteristics of the Stopped Negative Binomial Distribution

Parameters	p the success probability ($0 \leq p \leq 1$; $q = 1 - p$) s the number of successes before stopping ($s = 1, 2, ...$) t the number of failures before stopping ($t = 1, 2, ...$)
Support	$\min(s, t) \leq k \leq s + t - 1$
$\mathbb{P}[Y = k]$	$\binom{k-1}{s-1} p^s (1-p)^{k-s} + \binom{k-1}{t-1} (1-p)^s p^{k-t}$
$\mathbb{P}[Y \leq k]$	$2 - \mathcal{I}_q(k+1, s) - \mathcal{I}_p(k+1, t)$ where $\mathcal{I}$ is the regularized incomplete beta function.
Mean	$\frac{s}{p}\mathcal{I}_p(s,t) + \frac{p^{s-1}q^{t-1}}{B(s,t)} + \frac{t}{q}\mathcal{I}_q(t,s) + \frac{p^{t-1}q^{s-1}}{B(s,t)}$ where B is the beta function
MGF	$\left(\frac{pe^x}{1-qe^x}\right)^s \mathcal{I}_{1-qe^x}(s,t) + \left(\frac{qe^x}{1-pe^x}\right)^t \mathcal{I}_{1-pe^x}(t,s)$
Compound Distribution	$\binom{k-1}{s-1} \frac{B(\alpha+s, k-s+\beta)}{B(\alpha,\beta)} + \binom{k-1}{t-1} \frac{B(\alpha+k-t, t+\beta)}{B(\alpha,\beta)}$ for Beta(α, β).

the distribution. Section 2 derives the distribution function based on a defined Bernoulli process and gives some basic properties. Section 3 shows how the distribution is related to other standard distributions and connects the SNB tail probability to the Binomial tail probability. Section 4 derives the moment generating function. Section 5 derives the compound distribution using a Beta distribution for p.

2. Probability Mass Function

Let $b_1, b_2, \ldots$ denote a sequence of independent, identically distributed, Bernoulli random variables with $\mathbb{P}[b_i = 1] = p$ and $\mathbb{P}[b_i = 0] = 1 - p$, for probability parameter $0 \leq p \leq 1$. In the clinical trial setting $b_i = 1$ corresponds to a successful patient outcome following treatment. Let s and t be positive integers. Define the SNB random variable Y as the smallest integer value such that $\{b_1, \ldots, b_Y\}$ contains *either* s successes *or* t failures. That is, the SNB distribution of Y is the smallest integer such that either $\sum_i^Y b_i = s$ or $\sum_i^Y 1 - b_i = t$.

The distribution of Y has support on integer values in the range

$$\min(s,t) \leq \; Y \; \leq s + t - 1.$$

The probability mass function is

$$\mathbb{P}[Y = k] = S(k, p, s) \; I_{\{s \leq k\}} + S(k, 1 - p, t) \; I_{\{t \leq k\}} \tag{1}$$

where $I_{\{f\}}$ is the $i$ndicator function, taking the value of one if f is true and zero otherwise, and

$$S(k, p, s) = \binom{k-1}{s-1} p^s (1 - p)^{k-s} \tag{2}$$

is the negative binomial probability mass.

To prove (1), consider the process $\mathbf{X} = \{X(k) : k = 0, 1, ...\}$ with $X(0) = 0$ and

$$X_{k+1} = X_k + b_{k+1} \; I_{\{k-t<X_k<s\}}.$$

At each step a patient's outcome is measured. In Fig. 1 we consider a graphical illustration of the plot X_k against k. If the outcome of the kth patient is a success then the process advances diagonally in the positive horizontal and vertical direction. If the kth patient fails then the sample path advances in the positive horizontal direction only. The process continues until either $X_k = s$ or $X_k = k - t$.

Proposition 1. *The distribution of the stopping time*

$$Y = \operatorname*{argmin}_k \left[X_k \geq s \cup X_k \leq k - t\right]$$

is given at (1).

Proof. The probability a given realization of $\mathbf{X}$ reaches s at the kth outcome is the probability that, at time $k - 1$, there are $s - 1$ successful outcomes and $k - s$ unsuccessful outcomes multiplied by the probability of the final success at time k. This expression is given in (2). Similarly, the probability a given realization reaches $k-t$ is the probability that, at outcome $k-1$, there are $k-t$ successful outcomes and $t-1$ unsuccessful outcomes multiplied by the probability of a final unsuccessful outcome at time k.

To show that (1) sums to one, define

$$R = \sum_{k=s}^{s+t-1} S(k,p,s) + \sum_{k=t}^{s+t-1} S(k,1-p,t).$$

If we substitute $i = k - s$ in the first summation and $j = k - t$ in the second then R can be written as the cumulative distribution function of two negative binomial distributions:

$$R = \sum_{i=0}^{t-1} \binom{i+s-1}{i} p^s (1-p)^i \; + \sum_{j=0}^{s-1} \binom{j+t-1}{j} p^j (1-p)^t. \tag{3}$$

Let $\mathcal{I}_p(s,t)$ be the *regularized incomplete beta function* [1]. This function satisfies $\mathcal{I}_p(s,t) = 1 - \mathcal{I}_{1-p}(t,s)$ [2]. Then

$$\begin{aligned} R &= \sum_{i=0}^{t-1} \binom{i+s-1}{i} p^s (1-p)^i + \sum_{j=0}^{s-1} \binom{j+t-1}{j} p^j (1-p)^t \\ &= 1 - \mathcal{I}_p(s,t) + 1 - \mathcal{I}_{1-p}(t,s) \\ &= 1. \end{aligned}$$

This completes the proof that (1) is the distribution of the stopping time and is a valid probability mass function. □

Next, we consider an interim analysis of a clinical trial after s' successes and t' failures have been observed for $s' < s$ and $t' < t$.

Corollary 1. *The number of subsequent enrollments needed to reach either endpoint behaves as SNB(p, $s - s'$, $t - t'$).*

Having observed s' successes and t' failures, there are $s - s'$ successes needed to reach the success endpoint and $t - t'$ failures needed to reach the failure endpoint.

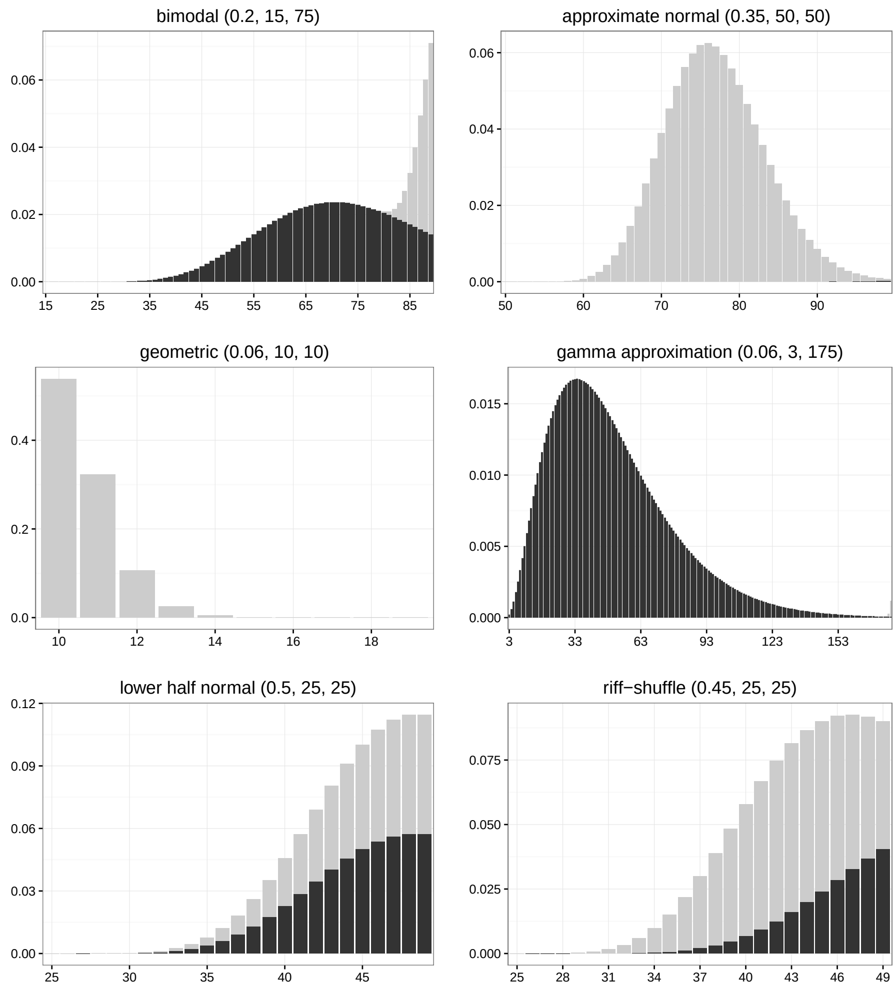


Figure 3: Different shapes of the SNB distribution with parameters (p, s, t), as given. Black indicates mass contributed by reaching s successes before t failures. Grey indicates mass contributed by reaching t failures first.

3. Connections and Approximations by Other Distributions

The SNB is a generalization of the negative binomial distribution. If t is large then $Y - s$ has a negative binomial distribution with

$$\mathbb{P}[Y = s + j] = \binom{s+j-1}{s-1} p^s (1-p)^j$$

for $j = 0, 1, \ldots$. A similar statement can be made when s is large and t is small. As a result, with proper parameter choice, the SNB can mimic other probability distributions in a manner similar to those described in [3] and [4]. Examples are shown in in Fig. 3.

The SNB also generalizes both the minimum (riff-shuffle) and maximum negative binomial distributions up to a translation of the support. For the special case of $s = t$, the distribution of Y is the riff-shuffle, or minimum negative binomial distribution [2]. Similar derivations of the closely-related maximum negative binomial discrete distributions also appear in [5] and [6]. The maximum negative binomial is the smallest number of outcomes necessary to observe at least s successes *and* s failures. The SNB is the number of outcomes to observe *either* s successes *or* t failures.

There is a close connection between the tail probabilities of the SNB and the binomial distributions. The probability of reaching the success endpoint in an SNB(p, s, t) random variable is equal to the probability of at least s successes in a Binomial distributed random variable with size $s + t - 1$ and success probability p. Likewise, the probability of reaching the failure endpoint is equal to the probability of at most $s - 1$ successes in a Binomial distribution. That is, the probability that the number of successes is at least s in the Binomial model is the same that the treatment succeeds (reaches s) in the SNB model.

Proposition 2. *Let Y be distributed as SNB(p, s, t) and let B be distributed Binomial with size $n = s + t - 1$ and success probability p. Then*

$$\mathbb{P}[B \geq s] = \mathbb{P}[\textit{treatment succeeds}]. \tag{4}$$

Proof. The Binomial tail probability is

$$\mathbb{P}[B \geq s] = 1 - \mathcal{I}_{1-p}.$$

The corresponding SNB probability is

$$\mathbb{P}[\text{treatment succeeds}] = \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} p^s (1-p)^{k-s}.$$

Let $i = k - s$. Using the fact that

$$\binom{i+s-1}{s-1} = \binom{i+s-1}{i}$$

the last summation can be rewritten as

$$\begin{aligned} \mathbb{P}[\text{treatment succeeds}] &= \sum_{i=0}^{t-1} \binom{i+s-1}{i} p^s (1-p)^i && (5) \\ &= 1 - \mathcal{I}_{1-p}(t, s) && (6) \end{aligned}$$

completing the proof. □

To illustrate this result, let us return to our initial example where $s = 2$, $t = 11$, and $p = 0.045$. The probability masses in Fig. 4 represented in black are equal in panels a) and b) as are the masses in grey. The probability that s successes are reached in the SNB process is the same as the binomial probability of at least two successes. Likewise, the probability that t failures are reached in the SNB process is the same as the binomial probability of zero or one successes.

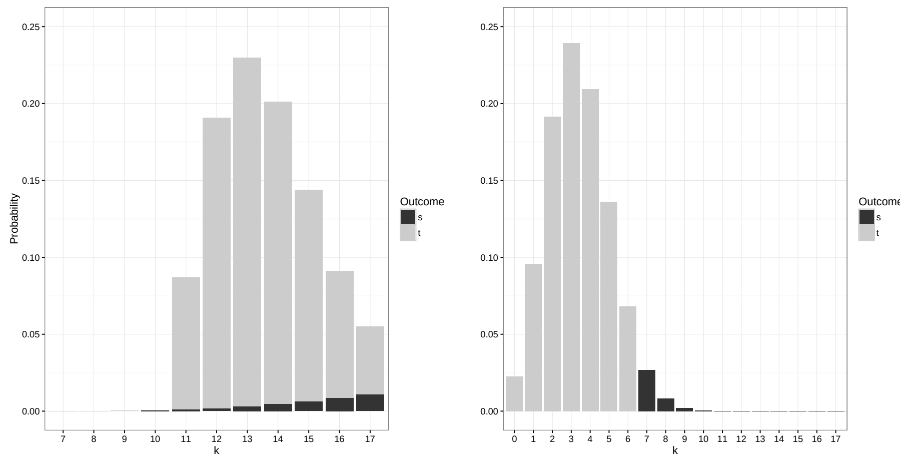
4. The Moment Generating Function

The moment generating function for the SNB is calculated in a manner similar to that of two negative binomial distributions.

Proposition 3. *Let Y be distributed SNB with parameters p, s, and t. Then the moment generating function (MGF) of Y is*

$$\mathbb{E}\ e^{xY} = \left(\frac{pe^x}{1-qe^x}\right)^s \mathcal{I}_{1-qe^x}(s,t) + \left(\frac{qe^x}{1-pe^x}\right)^t \mathcal{I}_{1-pe^x}(t,s) \qquad (7)$$

for $q = 1 - p$ when $x \leq \min\{\log(1/p), \log(1/q)\}$.



(a) SNB distribution

(b) Binomial distribution.

Figure 4: SNB(0.045, 2, 11) with mass contributed from s successes (black) or t failures (grey) along with Bin(0.045, 12) with at least 2 successes (black) or fewer (grey).

Proof. The MGF of the SNB is:

$$\mathbb{E}\, e^{xY} = \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} p^s q^{k-s} e^{kx} + \sum_{k=t}^{s+t-1} \binom{k-1}{t-1} p^{k-t} q^t e^{kx}$$

and can be rewritten as:

$$\mathbb{E}\, e^{xY} = \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} (pe)^{sx} (qe^x)^{k-s} + \sum_{k=t}^{s+t-1} \binom{k-1}{t-1} (qe^x)^t (pe^x)^{k-t}. \qquad (8)$$

The first summation in (8) satisfies

$$\begin{aligned}\sum_{k=s}^{s+t-1} \binom{k-1}{s-1} (pe)^{sx} (qe^x)^{k-s} &= \left(\frac{pe^x}{1-qe^x}\right)^s \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} (qe^x)^{k-s} (1-qe^x)^s \\ &= \left(\frac{pe^x}{1-qe^x}\right)^s \mathcal{I}_{qe^x}(s,t).\end{aligned}$$

Since the incomplete beta function's subscript parameter has support on zero to one, we have $qe^x \leq 1$. This also shows we must restrict $x \leq -\log(q)$. A similar expression can be derived for the second summation in Equation 8 and results in the constraint $x \leq -\log(p)$. □

The SNB's ability to approximate the geometric, normal, gamma, and poisson distributions follow from it generalizing the negative binomial. To recover the MGF of the negative binomial consider the case where $t \to \infty$ in (7). The regularized incomplete beta function in the first term goes to one and zero in the second term. We are left with the MGF of the negative binomial distribution.

When $s = 1$ this is the geometric distribution. The negative binomial can therefore be seen as a sum of i.i.d. geometric distributions. For an appropriately large number of samples the central limit theorem yields a normal approximation.

Drawing connections to the gamma and poisson distributions are more complicated. However a connection to the gamma distribution well-studied problem in the literature (see [7, 8, 4] for examples). A connection to the poisson appears in [9] where it is shown that if the mean of a poisson is proportional to a χ^2_{2k} distribution then the negative binomial is obtained. All of these results work by equating cumulants and then showing that differences between betwen the cumulant generating functions converge to zero.

The lower-half normal distribution can be approximated by setting $s = t$ for appropriately large s and t and $p = 0.5$. In this case the SNB can be viewed as identical, negative binomials approximating a normal and truncated at the median.

5. The Compound Probability Distribution

Let us consider a Bayesian analysis with a Beta distribution on p.

Proposition 4. *The compound distribution of the Stopped Negative Binomial distribution when p is distributed Beta(α, β) is*

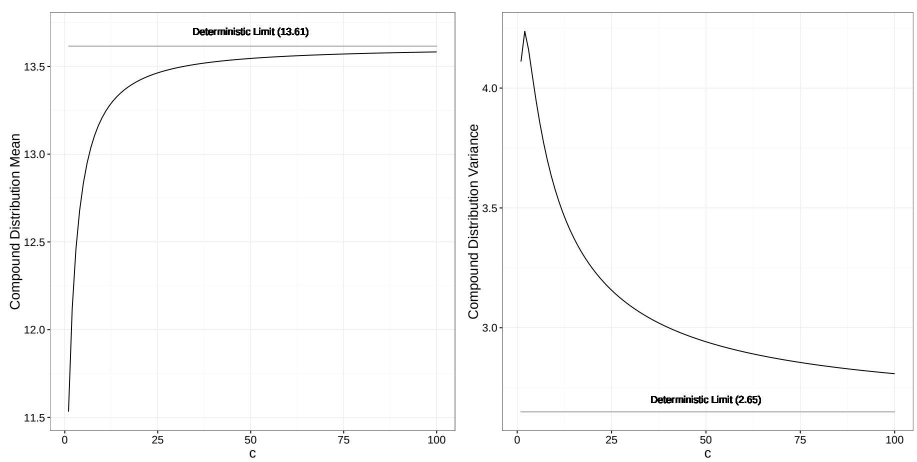
$$\mathbb{P}[Y = k|s, t, \alpha, \beta] = \binom{k-1}{s-1} \frac{B\left(\alpha + s, k - s + \beta\right)}{B(\alpha, \beta)} \; I_{\{s \leq k \leq s+k-1\}} + \binom{k-1}{t-1} \frac{B\left(\alpha + k - t, t + \beta\right)}{B(\alpha, \beta)} \; I_{\{t \leq k \leq s+k-1\}}. \tag{9}$$

Proof. For notational simplicity, assume that k is in the range $\min(s, t) \leq k \leq s + t - 1$. When this is not the case appropriate terms should be removed as dictated by the indicator functions.

$$
\begin{aligned}
f(k|s,t,\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \int_0^1 & \binom{k-1}{s-1} p^{\alpha+s-1}\,(1-p)^{k-s+\beta-1} + \\
& \binom{k-1}{t-1} p^{k-t+\alpha-1}\,(1-p)^{t+\beta-1}\,dp \\
= \frac{1}{B(\alpha,\beta)} \binom{k-1}{s-1} & \int_0^1 p^{\alpha+s-1}\,(1-p)^{k-s+\beta-1}\,dp + \\
& \frac{1}{B(\alpha,\beta)} \binom{k-1}{t-1} \int_0^1 p^{k-t+\alpha-1}\,(1-p)^{t+\beta-1}\,dp.
\end{aligned}
$$

The result in (9) follows by the definition of the Beta function. $\square$

To better understand the compound distribution, let us once again return to our example where $s = 2$, $t = 11$. However, we assume p is a Beta distribution with shape parameters $\alpha = 0.045c$ and $\beta = 0.955c$ for any $c > 0$. This parameterization allows us to examine the relationship between uncertainty in p and the SNB.



(a) mean

(b) variance

Figure 5: The expected size and variance of the compound SNB with $s = 2$, $t = 11$ with $\mathbb{E}\,p = 0.045$ varying the uncertainty in the distribution of p. A larger value of c corresponds to a Beta with smaller variance.

Fig. 5 shows how the expected sample size and variance of the compound

SNB. As the certainty increases the compound distribution sample values converge to corresponding expected values with $p = 0.045$. It may be noted that the compound distribution's mean approaches the deterministic limit slowly. This is because when c is small, the distribution of p is heavily skewed. As c increases this effect is minimized.

Slow convergence in the mean will occur when p takes values close to zero or one. In these cases it may be preferable to select shape parameters based on the median or mode of the beta distribution. These values will vary less in the uncertainty of the distribution.

References

[1] F. W. Olver, NIST handbook of mathematical functions, Cambridge University Press, 2010, Ch. 8, pp. 189–190.

[2] V. Uppuluri, W. Blot, A probability distribution arising in a riff-shuffle, Random Counts in Scientific Work, 1: Random Counts in Models and Structures (1970) 23–46.

[3] D. B. Peizer, J. W. Pratt, A normal approximation for binomial, f, beta, and other common, related tail probabilities, i, Journal of the American Statistical Association 63 (324) (1968) 1416–1456.

[4] D. Best, P. Gipps, An improved gamma approximation to the negative binomial, Technometrics 16 (4) (1974) 621–624.

[5] Z. Zhang, B. A. Burtness, D. Zelterman, The maximum negative binomial distribution, Journal of Statistical Planning and Inference 87 (1) (2000) 1–19.

[6] D. Zelterman, Discrete distributions: applications in the health sciences, John Wiley & Sons, 2005.

[7] J. Ord, Approximations to distribution functions which are hypergeometric series, Biometrika 55 (1) (1968) 243–248.

[8] W. C. Guenther, A simple approximation to the negative binomial (and regular binomial), Technometrics 14 (2) (1972) 385–389.

[9] F. J. Anscombe, Sampling theory of the negative binomial and logarithmic series distributions, Biometrika 37 (3/4) (1950) 358–382.