

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 17 patients are enrolled and treated. The binomial probability a patient responds to treatment is $p = 0.2$ under the null hypothesis that the treatment is not
5 effective. If seven or more patients out of these 17 respond to the treatment then we reject the hypothesis and the treatment is deemed successful at significance level of 0.1. If fewer than seven respond then the null is not rejected and the treatment is deemed 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 17 patients are enrolled at once, as in the classic design, then the sample size is 17. However, in most clinical trials the patients are enrolled sequentially with one patient's outcome realized before the next one enters the trial. In the present example, observing seven successful patients is one endpoint and the number of enrollees required could be as small as seven. Similarly 11 observed treatment failures also ends the trial. This sampling mechanism, in which the experiment ends as soon as any predefined endpoint is reached, is called *curtailed sampling*. Under curtailed sampling the range of the sample size for this trial is between seven and 17.

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 Bernoulli samples that stops when either a specified number of responders or non-responders have 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 a seventh response was reached on the 15th enrollment. Since the success boundary is reached, we say the treatment succeeds.

More generally the distribution of the number of trial enrollments is shown in Fig. 2 (a). There is relatively little probability mass for values of seven through 10 since p is small and the treatment succeeding quickly is unlikely. Fig. 2

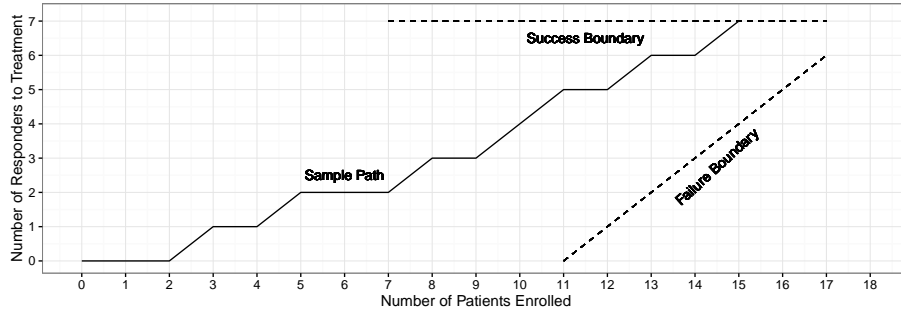


Figure 1: A hypothetical realization of the trial.

(b) shows the expected value and variance for the number of trial enrollments varying p between zero and one. When p is small the probability of an individual response is small, the treatment is more likely to fail shortly after the 11th enrollment. When p is large the treatment is more likely to succeed and the number of enrollees approaches seven from above. Since $p = 1$ and $p = 0$ represent the deterministic case the corresponding variance is zero. The “bump” around $p = 0.25$ results from inequality in the endpoints.

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 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

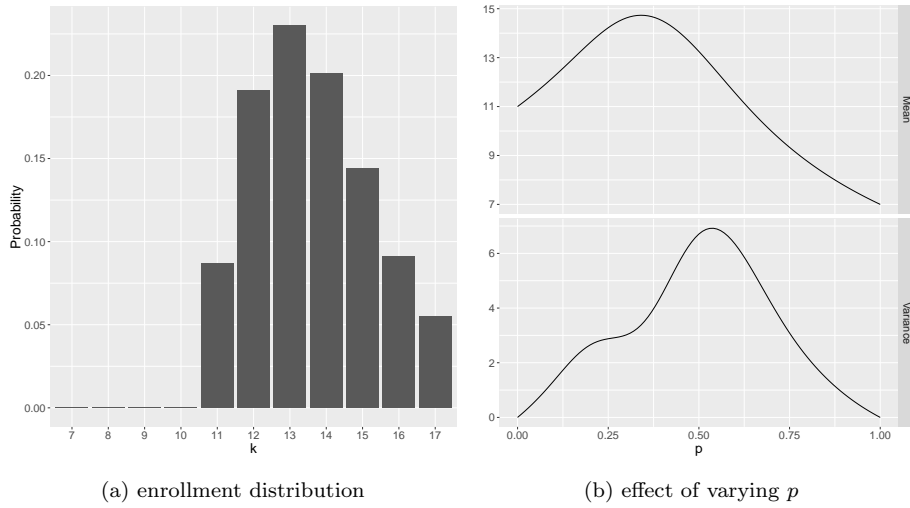


Figure 2: The distribution of a trial that stops after seven patients respond to treatment or 11 patients do not. (a) shows the distribution when $p = 0.2$. (b) shows the mean and variance of the distribution varying p between zero and one.

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 ($s = 1, 2, \dots$) t the number of failures before stopping ($t = 1, 2, \dots$)
Support	$\min(s, t) \leq k \leq s + t - 1$
$\mathbb{P}[Y = k]$	$\binom{k-1}{s-1} p^s (1-p)^{k-s} + \binom{k-1}{t-1} (1-p)^t p^{k-t}$
$\mathbb{P}[Y \leq k]$	$2 - \mathcal{I}_q(k+1, s) - \mathcal{I}_p(k+1, t)$ 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).

45 generating function. Section 5 derives the compound distribution when p has a beta distribution.

2. Probability Mass Function

Let b_1, b_2, \dots 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
50 probability parameter $0 \leq p \leq 1$. In the clinical trial setting $b_i = 1$ corresponds to a patient responding to treatment. Let s and t be positive integers. Define the SNB random variable Y as the smallest integer value such that $\{b_1, \dots, b_Y\}$ contains *either* s responders *or* t non-responders. 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

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

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

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

55 is the negative binomial probability mass.

To prove (1), consider the process $\mathbf{X} = \{X(k) : k = 0, 1, \dots\}$ 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 k th patient responds to treatment then the process advances diagonally in the positive horizontal and vertical direction. If the k th patient does not respond then the sample path
60 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*

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

is given at (1).

Proof. The probability a given realization of \mathbf{X} reaches s at the k th outcome is the probability that, at time $k-1$, there are $s-1$ successful outcomes and $k-s$
65 unsuccessful outcomes multiplied by the probability of a 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:

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

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

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

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

Next, we consider an interim analysis of a clinical trial after s' patients respond to treatment and t' do not for $s' < s$ and $t' < t$.

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

Having observed s' responders and t' non-responders, there are $s - s'$ more responders needed to reach the success endpoint and $t - t'$ more non-responders needed to reach the failure endpoint.

3. Connections and Approximations to Other Distributions

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

$$\mathbb{P}[Y = s + j \mid t \text{ is large}] \rightarrow \binom{s+j-1}{s-1} p^s (1-p)^j$$

80 for $j = 0, 1, \dots$. 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

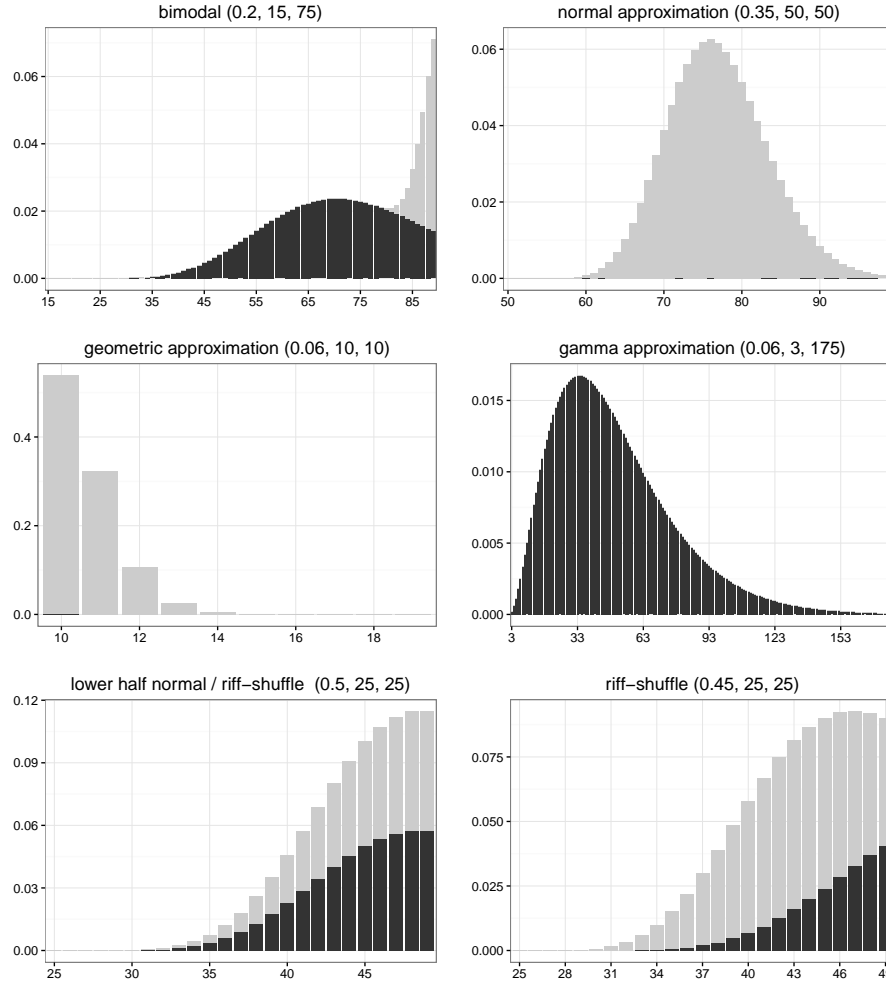


Figure 3: Different shapes of the SNB distribution with parameters (p, s, t) , as given. Black indicates mass contributed by reaching s responders before t non-responders. Grey indicates mass contributed by reaching t non-responders first.

distributions in a manner similar to those described in [3] and [4]. Examples are shown in in Fig. 3.

The SNB is the number of enrollment outcomes to observe *either* s responders 85 *or* t non-responders. The SNB 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 exactly the riff-shuffle, or minimum negative binomial distribution [2]. The maximum negative binomial [5, 6] is the smallest number of outcomes necessary to observe at least s responders 90 *and* s non-responders and is equivalent to a translated version of the riff-shuffle.

There is an equivalence between the probability of reaching an endpoint in the SNB model and the tail probability of the binomial distribution. That is, the probability that the number of responders is at least s in the binomial model 95 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 response probability p . Then*

$$\mathbb{P}[B \geq s] = \mathbb{P}[\text{treatment succeeds}]. \quad (4)$$

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

$$\mathbb{P}[\text{treatment succeeds}] = \sum_{i=0}^{t-1} \binom{i+s-1}{i} p^s (1-p)^i \quad (5)$$

$$= 1 - \mathcal{I}_{1-p}(t, s) \quad (6)$$

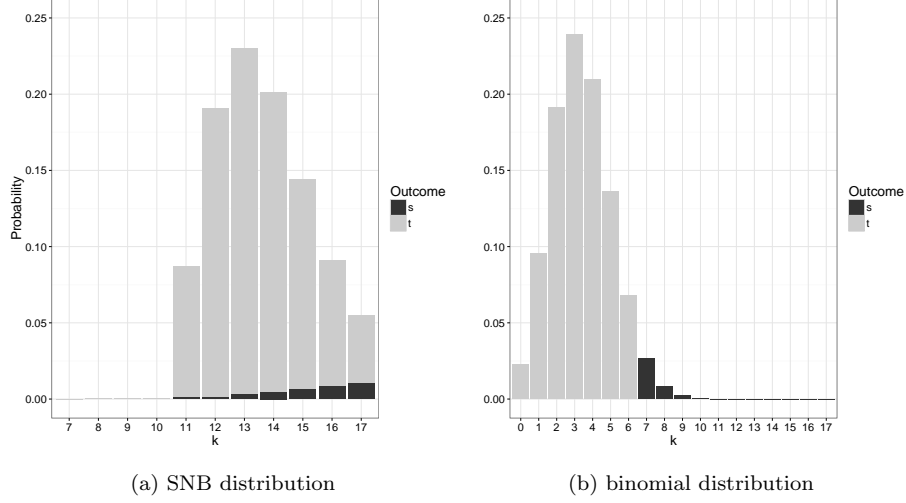


Figure 4: SNB(0.2, 7, 11) with mass contributed from s responders (black) or t non-responders (grey) along with Binomial(0.2, 17) with at least 2 responders (black) or fewer (grey).

completing the proof. □

To illustrate this result, let us return to our initial example where $s = 7$, $t = 11$, and $p = 0.2$. 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 responders are reached in the SNB process is the same as the binomial probability of at least seven responders. Likewise, the probability that t non-responders are reached in the SNB process is the same as the binomial probability of zero through six responders.

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*

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

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

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:

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

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
110 similar expression can be derived for the second summation in Equation 8 and results in the constraint $x \leq -\log(p)$. \square

The SNB's ability to approximate the geometric, normal, gamma, and poisson distributions follow from it generalizing the negative binomial distribution. To recover the MGF of the negative binomial consider the case where $t \rightarrow \infty$ in
115 (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 $s = 1$ (and t is still large) the SNB's MGF is the same as that of the geometric distribution. The negative binomial can therefore be seen as a sum
120 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

125 in [9] where it is shown that if the mean of a poisson is proportional to a χ_{2k}^2
distribution then the negative binomial is obtained. Both of these approxima-
tions work by equating cumulants and then showing that differences between
between the cumulant generating functions converge to zero.

The lower-half normal distribution can be approximated by setting $s = t$ for
130 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 the SNB distribution where p has a beta distribution.

Proposition 4. *The compound distribution of the SNB distribution when p is distributed $Beta(\alpha, \beta)$ is*

$$\begin{aligned} \mathbb{P}[Y = k|s, t, \alpha, \beta] &= \binom{k-1}{s-1} \frac{B(\alpha + s, k - s + \beta)}{B(\alpha, \beta)} I_{\{s \leq k \leq s+k-1\}} + \\ &\quad \binom{k-1}{t-1} \frac{B(\alpha + k - t, t + \beta)}{B(\alpha, \beta)} I_{\{t \leq k \leq s+k-1\}}. \end{aligned} \quad (9)$$

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.

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

135 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 $s = 7$, $t = 11$. However, we assume p is a beta distribution

with shape parameters $\alpha = 0.2c$ and $\beta = 0.8c$ for any $c > 0$. This parameterization allows us to examine the relationship between uncertainty in p and the SNB.

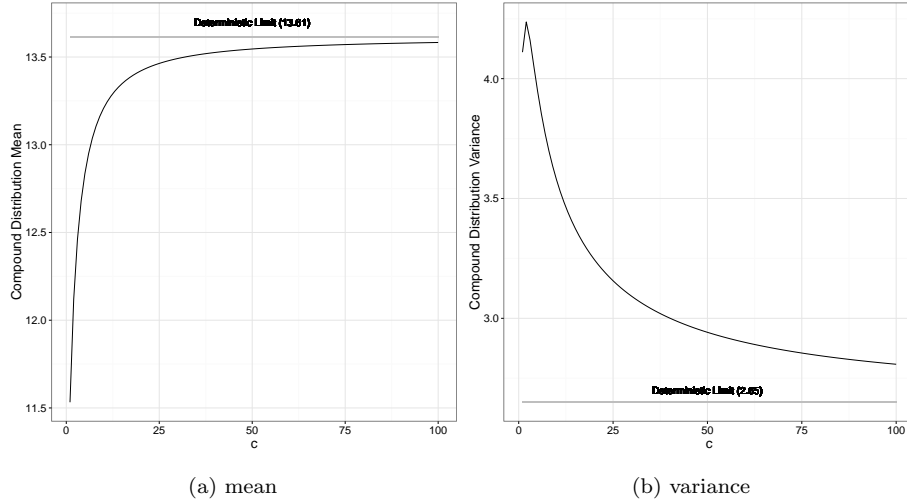


Figure 5: The mean and variance of the compound SNB with $s = 7$, $t = 11$ with $\mathbb{E} p = 0.2$ varying the uncertainty in the distribution of p . A larger value of c corresponds to a beta with smaller variance.

Fig. 5 shows how the mean and variance of the compound distribution is related to the SNB's mean and variance. As c increases the compound distribution sample statistics converge to corresponding statistics for the SNB with $p = 0.2$. When the distribution of p is centered close to zero or one the mean estimate approach the deterministic limit relatively slowly in the uncertainty parameter c . Because of this, we note that for estimation it may be preferable to select shape parameters based on the median or mode of the beta distribution since they are less sensitive to distributional uncertainty.

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.
- 155 [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.
- 160 [5] Z. Zhang, B. A. Burtneiss, 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.
- 165 [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.
- 170