Clinical Trial Design Using A Stopped Negative Binomial Distribution

February 23, 2017

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.

1 Introduction and Motivation

Consider a prototypical early phase, single-arm clinical trial in which 17 patients are enrolled and treated. Suppose the Bernoulli probability of a patient responding to treatment is $p = 0.2$ under the null hypothesis that the

treatment is not effective. If seven or more patients out of these 17 respond to the treatment then we reject this hypothesis and the treatment is deemed successful at a significance level of 0.1. If fewer than seven respond then the null hypothesis is not rejected and the treatment is deemed ineffective. The trial is realized as a sequence of Bernoulli(p) samples stopping when either the specified number of responders or non-responders is reached.

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 over time. In the present example, observing seven successful patients ends the trial and so 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 seven through 17.

A hypothetical sample path is illustrated in Fig. 3. The vertical axis

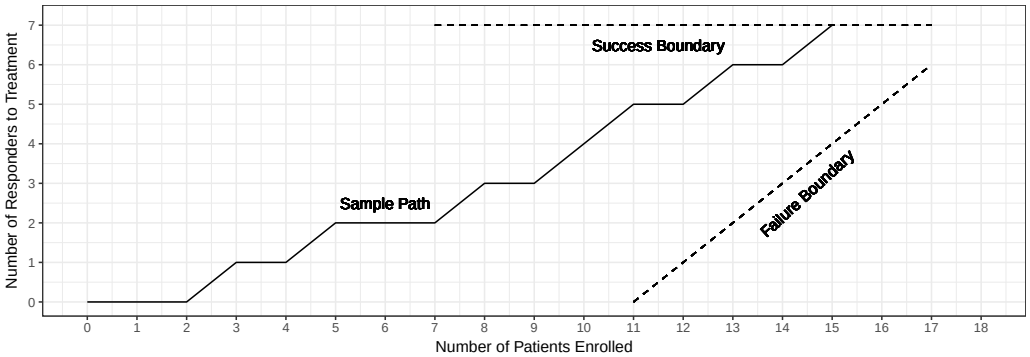


Figure 1: A hypothetical realization of the trial.

denotes the number of successful outcomes. The horizontal axis counts the number of patients enrolled. The horizontal and tilted vertical boundaries represent endpoints for the trial. In this case, a seventh response was reached on the 15^{th} enrollment. Since the success boundary is reached, we say the treatment succeeds.

The distribution of the number of trial enrollments is shown in Fig. 3. There is relatively little probability mass for values of sample size equal to seven through 10 since p is small and it is unlikely the treatment will succeed quickly. Fig. 3 shows the expected value and variance for the number of trial enrollments varying p between zero and one. When p is small then the treatment is more likely to fail shortly after the 11^{th} enrollment. When p is large then the treatment is more likely to succeed and the number of enrollees approaches seven from above.

When $p = 0$ or 1 the processes are deterministic and variance is zero. Values between zero and one change the mixing proportions of the two endpoints. The saddle around $p = 0.25$ results from inequality in the size of the support of the two 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). This paper derives this distribution and explores its properties. 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 stan-

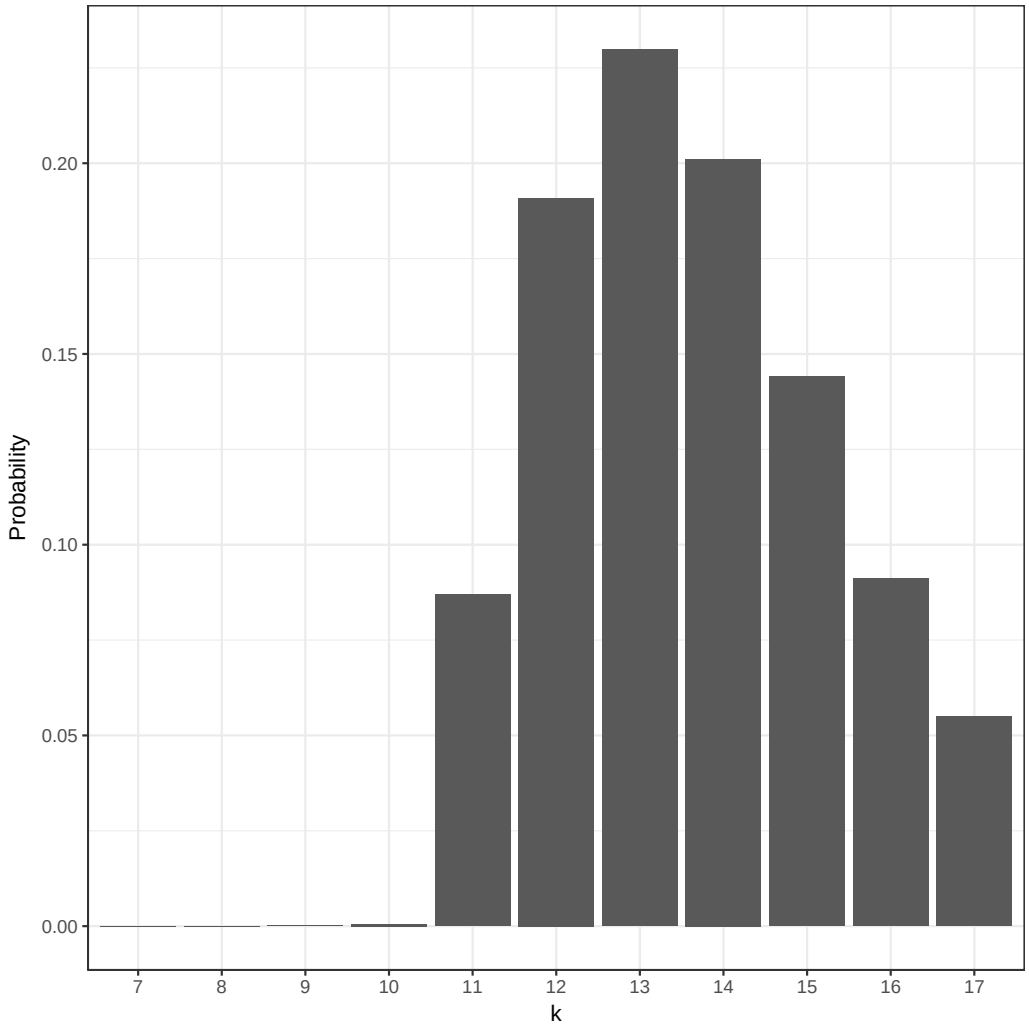


Figure 2: The distribution of the sample size in a trial that stops after seven patients respond to treatment or 11 do not when $p = 0.2$.

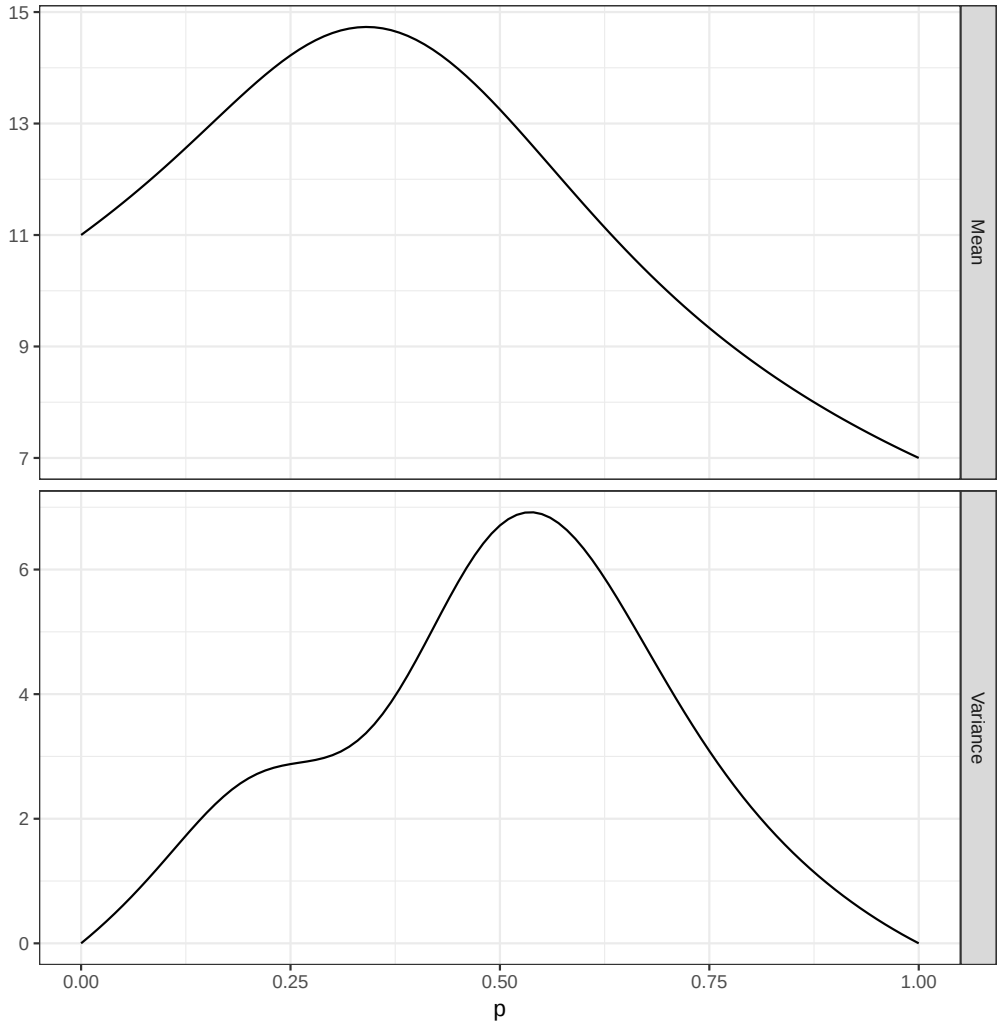


Figure 3: The effect of varying p between zero and one when the trial stops after seven patients respond or 11 do not.

dard distributions and connects the SNB tail probability to the binomial tail probability. Section 4 derives the moment generating function. Section 5 derives the predictive and posterior distributions when p has a beta prior distribution.

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 patient responding to treatment.

Let s and t be positive integers. Define the SNB random variable Y as the smallest integer value such that $\{b_1, \ldots, b_Y\}$ 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

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

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

zero otherwise, and

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

is the negative binomial probability mass. The CDF of the negative binomial is commonly expressed in terms of the *regularized incomplete beta function* [7]. Using standard notation, (2) is written as $\mathcal{I}_{1-p}(t,s)$.

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. 3 we consider a graphical illustration of the plot X_k against k. If the outcome of the kth patient responds to treatment then the process advances diagonally in the positive horizontal and vertical direction. If the kth patient does not respond then the sample path advances in the positive horizontal direction only. The process continues until either $X_k = s$ or $X_k = k - t$ corresponding to the success and failure boundaries in Fig. 3, respectively.

Proposition 1. *The distribution of the stopping time*

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

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 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. That is,

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

as claimed in (1).

To show (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 (CDF) of two negative binomial distributions:

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

The regularized incomplete beta function satisfies $\mathcal{I}_p(s,t) = 1 - \mathcal{I}_{1-p}(t,s)$

[10]. 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 (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' patients respond to treatment and t' fail to respond for $s' < s$ and $t' < t$.

Corollary 1. *The number of subsequent enrollments needed to reach either s or t endpoints 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.

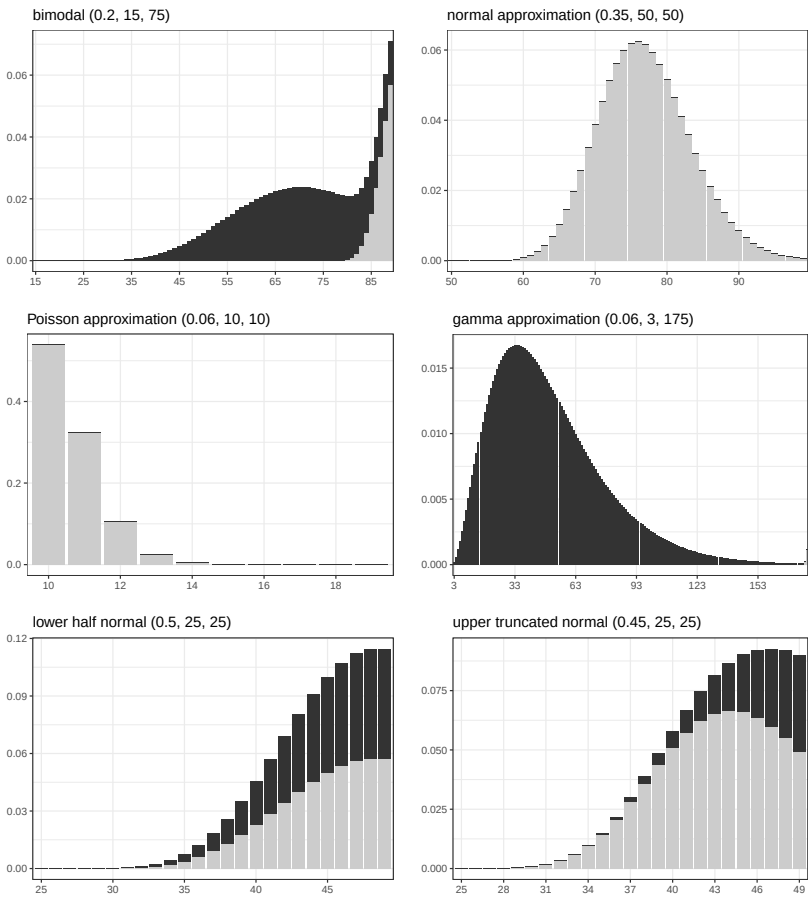


Figure 4: Different shapes of the SNB distribution with parameters (p, s, t), as given. Black areas indicate the mass contributed by reaching s responders before t non-responders. Grey indicates mass contributed by reaching t non-responders first.

3 Connections and Approximations to Other Distributions

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

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

for $j = 0, 1, \ldots$. A similar statement can be made when s is large and t is moderate. As a result, with proper parameter choice, the SNB can mimic other probability distributions in a manner similar to those described in [9] and [3]. Examples are shown in in Fig. 4.

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 the riff-shuffle, or minimum negative binomial distribution [10, 6]. The maximum negative binomial [6, 12, 11] is the smallest number of outcomes necessary to observe at least s responders *and* s non-responders. This is equivalent to a translated version of the riff-shuffle.

There is also an equivalence between the probability of reaching an endpoint in the SNB model and the tail probability of the binomial distribution. Specifically, the probability that the number of responders is at least s in the binomial model is the same as the probability the treatment succeeds

(reaches s) in the SNB model.

Proposition 2. *Let Y be distributed as SNB(p, s, t) and let X_Y correspond to the number of responders at the end of the trial. Let B be distributed binomial with index parameter $n = s+t-1$ and response probability p. Then*

$$\mathbb{P}[B \geq s] = \mathbb{P}[X_Y = s]. \tag{4}$$

Proof. The binomial tail probability is

$$\mathbb{P}[B \geq s] = 1 - \mathcal{I}_{1-p}(s, t)$$

The corresponding success probability is

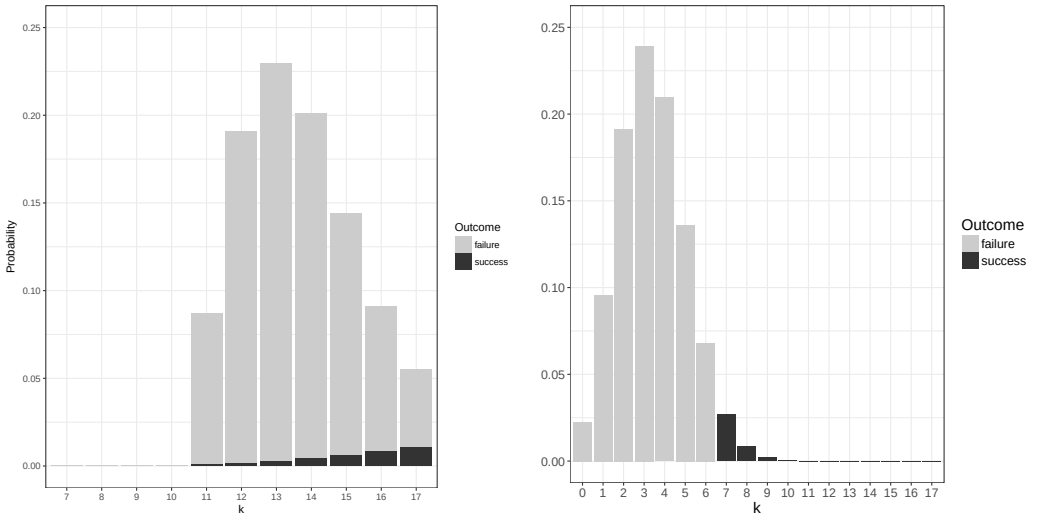
$$\mathbb{P}[X_Y = s] = \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} p^s (1-p)^{k-s}. \tag{5}$$

Let $i = k - s$. Since

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

the summation in (5) can be rewritten as

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



(a) SNB distribution (b) binomial distribution

Figure 5: SNB(0.2, 7, 11) with mass contributed from 7 responders (black) or 11 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. 5 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*

$$\mathbb{E}\ e^{xY} = \left(\frac{pe^x}{1-qe^x}\right)^s \mathcal{I}_{1-qe^x}(s,t) + \left(\frac{qe^x}{1-pe^x}\right)^t \mathcal{I}_{1-pe^x}(t,s) \tag{6}$$

for $q = 1-p$ and is defined for $x < \min\{\log(1/p), \log(1/q)\}$.

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:

$$\mathbb{E}\ e^{xY} = \sum_{k=s}^{s+t-1} \binom{k-1}{s-1} (pe^x)^s (qe^x)^{k-s} + \sum_{k=t}^{s+t-1} \binom{k-1}{t-1} (qe^x)^t (pe^x)^{k-t}. \tag{7}$$

The first summation in (7) satisfies

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

Since the p parameter in $\mathcal{I}_p$ has support on zero to one, we have $0 \le pe^x < 1$.

This implies $x < -\log(p)$. A similar expression can be derived for the second summation in (7) and results in the constraint $x < -\log(1-p)$. □

Depending on the parameter values the SNB can be approximated by the geometric, normal, gamma, and Poisson distributions.

Proposition 4. *The MGF of the SNB converges to that of the negative binomial when either s or t gets large. That is*

$$lim_{t\to\infty}\ \mathbb{E}e^{xY} = \left(\frac{pe^x}{1-qe^x}\right)^s$$

as. The analogous result holds when $s \to \infty$.

Proof. The second incomplete beta function in (6) can be written in terms of a cumulative binomial distribution

$$\mathcal{I}_{1-pe^x}(t,s) = \mathbb{P}\left[B \leq s-1\right]$$

where B is distributed as Binomial$(t-k,\, pe^x)$. From Chebychev's inequality it follows that

$$\mathbb{P}\left[B \leq s-1\right] \leq \frac{(t-k)pe^x(1-pe^x)}{\left(s-(t-k)pe^x\right)^2} \tag{8}$$

As t gets large $\mathcal{I}_{1-pe^x}(t,s)$ tends to zero and $\mathcal{I}_{1-qe^x}(s,t)$ approaches one. The proof follows by realizing

$$0 < \frac{qe^x}{1-pe^x} < 1$$

over the support of x. □

When $s = 1$ and t is large the SNB's MGF is approximately the same as that of 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 distributions, the central limit theorem yields a normal approximation.

Drawing connections to the gamma and Poisson distributions can also be shown. A connection to the gamma distribution well-studied problem in the literature (see [3, 8, 4] for examples). A connection to the Poisson appears in [1] where it is shown that when the mean of a Poisson is proportional to a chi-square distribution with $2k$ degrees of freedom then the negative binomial is obtained. Both of these approximations work by equating cumulants and then showing that differences between 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 binomial distributions approximating a normal and truncated at the median.

5 The Posterior and Predictive Probability Distribution

Let us consider the Bayesian setting where the rate parameter is distributed as Beta(α, β) and denoted P. The posterior distribution of P is proportional

to the likelihood, given by the function

$$f_{P|Y}(p,k,s,t,\alpha,\beta) \propto \frac{\binom{k-1}{s-1}}{B(\alpha,\beta)} p^{\alpha+s-1}(1-p)^{k+\beta-s-1} + \frac{\binom{k-1}{t-1}}{B(\alpha,\beta)} p^{k+\alpha-t-1}(1-p)^{\beta+t-1} \tag{9}$$

where $0 \leq p \leq 1$, $s \leq k \leq s+k-1$, and $t \leq k \leq s+k-1$. This result can be found directly by multiplying the probability mass function in (1) by the density of the Beta distribution.

The predictive distribution of the SNB can be found by integrating p over the interval zero to one and applying the definition of the beta function.

$$\begin{aligned} f_Y(k,s,t,\alpha,\beta) &= \int_0^1 f_P(p|\alpha,\beta) f_{Y|P}(p,k,s,t)dp \\ &= \binom{k-1}{s-1} \frac{B(\alpha+s, k-s+\beta)}{B(\alpha,\beta)} + \binom{k-1}{t-1} \frac{B(\alpha+k-t, t+\beta)}{B(\alpha,\beta)}. \end{aligned} \tag{10}$$

If both α and β are non-negative integers then the predictive distribution is a mixture of hypergeometric distributions.

$$f_Y(k,s,t,\alpha,\beta) = \frac{\binom{k-1}{s-1}\binom{\alpha+\beta}{\alpha}}{\binom{\alpha+\beta+k-1}{\alpha+s}} \frac{\alpha}{\alpha+\beta} \frac{\beta}{k-s+\beta} + \frac{\binom{k-1}{t-1}\binom{\alpha+\beta}{\beta}}{\binom{\alpha+\beta+k-1}{t+\beta}} \frac{\beta}{\alpha+\beta} \frac{\alpha}{k-t+\alpha}$$

The ratio of combinations in the first term can be interpreted as the probability of $s-1$ responders from $k-1$ patients in $\alpha+s$ draws from a population size of $\alpha+\beta+k-1$. This value is multiplied by $\alpha/(\alpha+\beta)$, the expected response rate of the prior. The final term in the product weigths the prior based on

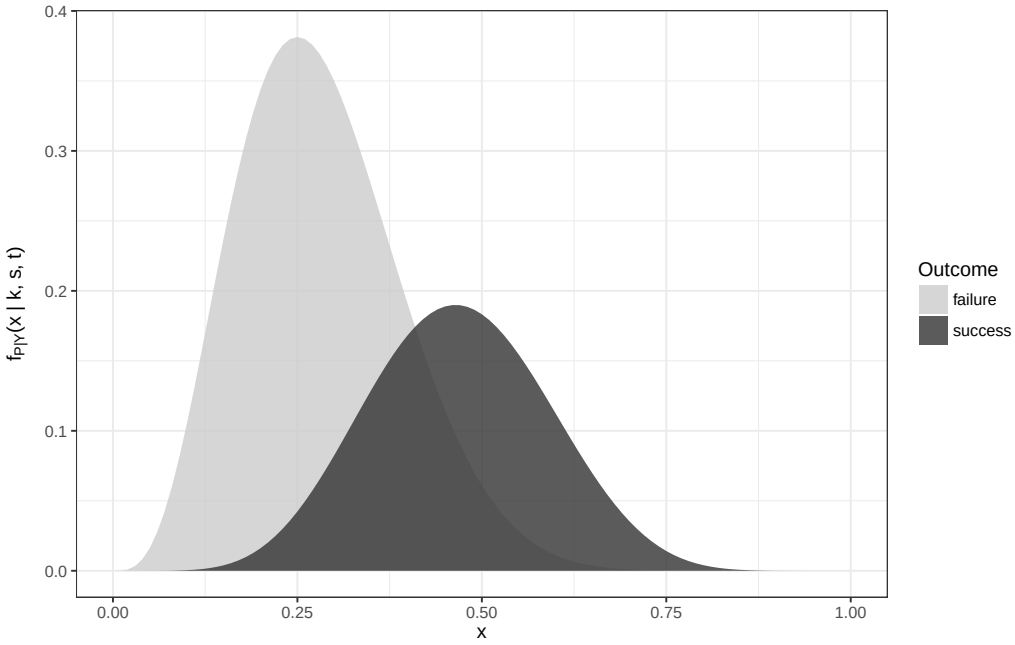


Figure 6: The posterior distribution of the response probability with the Jeffreys prior ($\alpha = \beta = 1/2$) for the trial where $s = 7$, $t = 11$, $k = 15$.

the number of non-responders ($k - s$). Terms in the second summand are interpreted similarly for non-responders.

The ratio of (9) divided by (10) gives the posterior distribution of P. It is a mixture of beta distributions. The mixing parameter depend on the endpoints (s and t), the number of enrollees needed to reach an endpoint (k), and the prior parameters (α and β).

The posterior result above is for the case where the parameters are known and the endpoint is not. That is, we do not which boundary was reached. Fig. 6 shows this distribution based on the parameters in the hypothetical trial assuming the Jeffreys prior [5]. If we include the fact that the trajectory

reaches the endpoint then the second terms in (9) and (10) are both zero and the posterior distribution is Beta(7.5, 8.5). This is proportional to the area labeled "success" in Fig. 6.

6 Powering Trials using Confidence Sets

The previous section shows the posterior distribution P at each of the endpoints in the domain is defined by a Beta distribution. This fact can be used to provide power calculations based on confidence sets with specified minimum response rate. Let $\mathcal{P}_s(s,t)$ be the set of random variables associated success. Then a trial whose minimum response rate is p_{min} at a significance level of at least α_C is one where each distribution is at least p_{min} with probability $1-\alpha_C$. That is,

$$\mathbb{P}\left[X > p_{min}\right] > (1-\alpha_C) \quad \forall X \in \mathcal{P}_s(s,t) \tag{11}$$

Letting $\mathcal{Q}_t(s,t)$ be the set of random variables associated with failure, a trial whose power is at least $1-\beta_C$ has

$$\mathbb{P}\left[X < p_{min}\right] < \beta_C \quad \forall X \in \mathcal{Q}_t(s,t).$$

Proposition 5. *Let P_k be the posterior distribution of reaching the success*

endpoint at the kth enrollment with density

$$f_k(p) = \frac{\binom{k-1}{s-1}}{B(\alpha,\beta)} p^{\alpha+s-1}(1-p)^{k+\beta-s-1}.$$

If $0 \leq \alpha_C \leq 1 - 1/k$ *then*

$$\mathbb{P}\left[P_k \geq \alpha_C\right] \leq \mathbb{P}\left[P_{k-1} \geq \alpha_C\right]. \tag{12}$$

Proof. Consider the ratio of the posterior distributions

$$\begin{aligned}
\frac{f_k(p)}{f_{k+1}(p)} &= \frac{\frac{\binom{k-1}{s-1}}{B(\alpha,\beta)} p^{\alpha+s-1}(1-p)^{k+\beta-s-1}}{\frac{\binom{k}{s-1}}{B(\alpha,\beta)} p^{\alpha+s-1}(1-p)^{k+\beta-s}} \\
&= \frac{p^{\alpha+s-1}(1-p)^{k+\beta-s-1}}{kp^{\alpha+s-1}(1-p)^{k+\beta-s}} \\
&= \frac{1}{k(1-p)}.
\end{aligned}$$

This implies that the density $f_{k+1}(p)$ is greater than $f_k(p)$ for all values of $0 \leq p \leq 1 - 1/k$. This implies the integral of $f_{k+1}(p)$ must also be greater than $f_k(p)$ over a similar interval. The result follows by realizing that this relationship is reversed over the complement of the interval. □

The maximum significance of a trial is $\mathbb{P}\left[P_{s+t-1} \leq \alpha_C\right]$ Proposition 5 guarantees the significance for all other success endpoints will be smaller. A similar calculation can be performed for power to show that the minimum power is $\mathbb{P}\left[Q_{s+t-1} \leq 1 - \beta_C\right]$ where Q_{s+t-1} is the posterior distribution of reaching

a trial failure at the $s + t - 1$ step.

These two results imply a relatively simple procedure for finding the sample size for trials of this type. A minimum response rate, maximum significance, and minimum power are selected. The task is to find the values of s and t that fulfill these requirements given the minimum response. Start with $s = t = 1$. If the maximum significance criterion is not met then increase s by one. If the minimum power criterion is not met then increase t by one. Repeat the procedure until both criteria have been met. The sample size is then determined by the calculation $s + t - 1$.

7 A Bayesian Sequential Probability Ratio Test

8 The Approximate Distribution for Two Linear Boundaries

Consider an alternative visualization of the sample path counting the number of responders and non-responders in the vertical and horizontal direction respectively, as in Figure 7. This alternative formulation can be rotated and scaled to form a random walk on the integers. A functional central limit can be then applied when the endpoints are sufficiently far from the origin and the process converges to a Brownian motion. The hitting time of the success

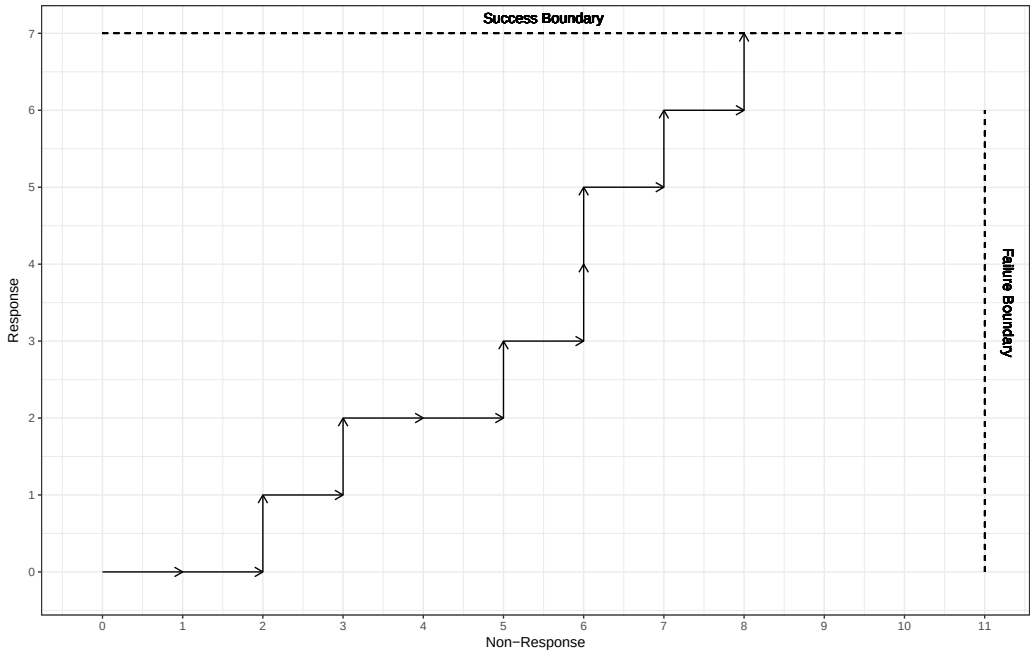
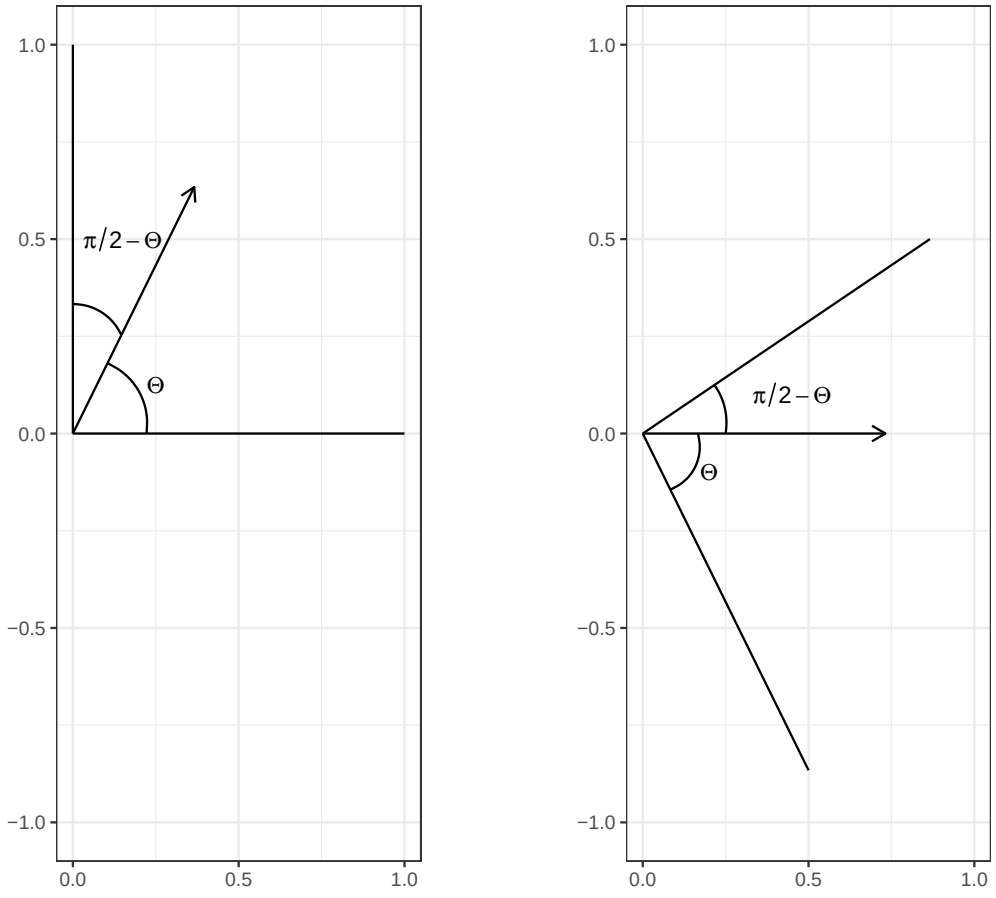


Figure 7: An alternative visualization of the hypothetical trial.

and failure endpoints is under this reparameterization is a mixture of normal distributions.

For a response rate p, the expected direction of each step in the sample path is $\tan^{-1}(p)$ radians from the vertical direction. If the graph is rotated $\pi/2 - \Theta$ radians then the expected value of the process is zero at any time. Let $\Theta = tan^{-1}(p)$. Then, after rotation, the process moves $\sin\Theta$ units in the horizontal direction and $\cos\Theta$ units in the positive vertical direction for each response. Each non-response corresponds to $\cos(\Theta)$ units in the horizontal direction and $\sin\Theta$ units in the negative vertical direction. The rotation is shown pictorially in Figure 8.

After the rotation, a response step is normalized by $\sin(\pi/2 - \Theta)$ in the



(a) Before rotation

(b) After rotation

Figure 8: The direction of a single step in the binomial process before (a) and after rotation (b).

vertical direction and $\cos(\pi/2 - \Theta)$ in the horizontal direction. Likewise, a non-response step is normalized by $\sin(\Theta)$ and $\cos(\Theta)$ in the vertical and horizontal direction respectively. Each step is a Rademacher random variable. The resulting process forms a random walk on the integers. Let m be the shortest path from the origin to any endpoint, let step $k = mt$ the process, and let R_i be a sample from a Rademacher random variable. Then it is well known that a cádlág version of the rescaled process on the natural filtration converges to a Brownian motion as m gets large

$$t\frac{1}{\sqrt{tm}}\sum_{i=0}^{tm} R_i \to B(t).$$

The density of the first time a Brownian motion hits a sloping line with intercept a and is [2]

$$f_{a,b}(t) = \frac{a}{t^{3/2}}\phi\left(\frac{a+bt}{\sqrt{t}}\right) \tag{13}$$

where $\phi(t)$ is the density of the standard normal distribution at t. Therefore, for success and failure endpoints defined by lines $a + bt$ and $a' + b't$ with support $[s_0 \leq t \leq s_m]$ and $[r_0 \leq t \leq r_m]$ respectively, the density of the SNB approximation is

$$f(t) = \frac{f_{a,b}(t)[s_0 \leq t \leq s_m] + f_{a',b'}(t)[r_0 \leq t \leq r_m]}{\int_{s_0}^{s_m} f_{a,b}(x)dx + \int_{r_0}^{r_m} f_{a',b'}(x)dx}. \tag{14}$$

References

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

[2] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.

[3] DJ Best and PG Gipps. An improved gamma approximation to the negative binomial. *Technometrics*, 16(4):621–624, 1974.

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

[5] Harold Jeffreys. An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 186, pages 453–461. The Royal Society, 1946.

[6] Norman L Johnson, Adrienne W Kemp, and Samuel Kotz. *Univariate discrete distributions*, volume 444. John Wiley & Sons, 2005.

[7] Frank WJ Olver. *NIST handbook of mathematical functions*, chapter 8, pages 189–190. Cambridge University Press, 2010.

[8] JK Ord. Approximations to distribution functions which are hypergeometric series. *Biometrika*, 55(1):243–248, 1968.

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

[10] VRR Uppuluri and WJ Blot. Probability distribution arising in a riff-shuffle. Technical report, Oak Ridge National Lab., Tenn., 1967.

[11] Daniel Zelterman. *Discrete distributions: applications in the health sciences*. John Wiley & Sons, 2005.

[12] Zhongxin Zhang, Barbara A Burtness, and Daniel Zelterman. The maximum negative binomial distribution. *Journal of Statistical Planning and Inference*, 87(1):1–19, 2000.