JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, March 2011, Vol. 56, No. 2
doi: 10.1111/j.1556-4029.2010.01638.x
Available online at: onlinelibrary.wiley.com

TECHNICAL NOTE

CRIMINALISTICS

Jos J. A. M. Weusten,[1] *Ph.D.*

Representative Drug Sampling: Sample Size Calculations Revisited

ABSTRACT: Calculating the required number of samples to be tested from a consignment of pills suspected of containing drugs can be performed from a Bayesian perspective. Procedures in literature are based on the outstanding work of Aitken. However, in the mathematical treatment of the problem, the limitedness of the consignment is not systematically used. The current Technical Note addresses this problem. A suitable prior distribution for the number of positive pills is derived, being a betabinomial distribution with the consignment size as one of the parameters. A hypergeometric likelihood is used, as sampling generally proceeds without replacement. The betabinomial posterior distribution is mathematically identical to the predictive distribution as reported elsewhere. The currently used large consignment approximation can be derived from the betabinomial posterior, but the quality is not optimal when compared to the exact betabinomial-based results. A new approximation is derived, with better properties, as illustrated in some examples.

KEYWORDS: forensic science, representative drug sampling, sample size calculation, finite consignment, hypergeometric distribution, betabinomial distribution

A consignment of pills is obtained, suspected of containing drugs. The size of the consignment is limited; though, it can be quite large. A sample is drawn consisting of a given number of pills, and they are tested for the presence of drugs. Given the observed number of drug containing pills ("positives") in the sample, what can be said about the number of drug containing pills in the total consignment?

Sampling can be performed from a Bayesian perspective. There is some prior belief concerning the expected fraction of drug containing pills. In the literature (e.g., the guidelines of the ENFSI [1] or the United Nations [2], both based on the work of Aitken [3,4]), this concept is worked out according to the following lines.

Consider a large consignment of which the fraction positives is to be assessed. A beta distribution can be used as prior distribution for this fraction, and a binomial distribution for the likelihood given the number of pills tested. This leads to a beta distribution as posterior distribution. In case of small consignments, inferences on the number of positives in the remainder of the consignment are then based on constructing the predictive distribution for the number of positives. The final result is a betabinomial distribution. Nordgaard (5) gives an interesting strategy to set the prior distribution parameters based on experience with drug seizures.

There are some problems with this approach. The first problem is the use of the binomial distribution for the observed number of positives. This would be suitable for infinite consignment sizes, or for sampling with replacement. However, main interest is in consignments of limited size, and sampling without replacement is more obvious, so a hypergeometric likelihood would be more suitable for the problem at hand. The second is the problem that the prior chosen concerns the fraction of positives, viewed as a continuous variable. Also here the limitedness of the size of the consignment is ignored. It would be better to define a prior distribution for the number of positive pills given the actual size of the consignment and to combine this with the hypergeometric likelihood. This approach was also used by Coulson et al. (6), using a user-defined discrete prior distribution for the number of positives in combination with a hypergeometric likelihood.

The current document aims at deriving an exact equation for the posterior distribution of the number of positives in a consignment using a suitable discrete prior and a hypergeometric likelihood. Such prior is found in the betabinomial distribution. This prior is derived by viewing the consignment at hand as a small random sample from all the pills that are produced in a given laboratory and by combining a beta distribution with a binomial distribution. The posterior distribution for the number of positives follows by combining the betabinomial prior with the hypergeometric distribution for the likelihood. From the resulting posterior for finite consignment sizes, a large consignment approximation is derived.

Mathematical Treatment

In this section, the mathematics required to derive the posterior distribution for the number of positives is described.

Symbols and Definitions

The following symbols are used:

N_{tot}	Total number of pills in the consignment
N_{pos}	The number of pills with drugs (positives) in the total consignment
N_s	Number of pills tested (sample size)
x	Number of positive pills in the sample
a, b	Parameters of the prior beta distribution for the fraction
B(v, w)	beta function, $B(v, w) = \Gamma(v)\,\Gamma(w)/\Gamma(v + w)$

[1]DSM Resolve, The Maastricht Forensic Institute, PO Box 18, 6160 MD Geleen, The Netherlands.

Received 6 Oct. 2009; and in revised form 11 Jan. 2010; accepted 30 Jan. 2010.

The Prior Distribution for N_{pos}

The prior is constructed in two steps as described in Appendix 1. The consignment at hand is viewed as a small sample from some super-population, with a certain fraction of positives. A suitable prior for this fraction would be a beta distribution with parameters a and b. Given the size of the consignment, N_{tot}, the number of positives is a random variable with a binomial distribution. Combining these leads to the following prior:

$$p(N_{pos}) = \frac{B(a + N_{pos},\, N_{tot} + b - N_{pos})}{B(a,\, b)} \binom{N_{tot}}{N_{pos}} \tag{1}$$

This probability distribution is known as the betabinomial distribution with parameters N_{tot}, a, and b.

The betabinomial prior distribution of Eq. 1 includes a noninformative prior; on choosing $a = b = 1$ Eq. 1 reduces to $p(N_{pos}) = 1/(N_{tot} + 1)$, which is a uniform discrete distribution over the range of integers in the interval $[0, N_{tot}]$.

The Likelihood Function of N_{pos} *Given the Experimental Result*

Given the number of positive pills in the total consignment, the number of positives in the sample follows a hypergeometric distribution according to

$$p(x|N_{pos}) = \frac{\binom{N_{pos}}{x}\binom{N_{tot} - N_{pos}}{N_s - x}}{\binom{N_{tot}}{N_s}} \tag{2}$$

The Posterior Distribution for N_{pos}

The posterior distribution follows from combining the prior (Eq. 1) and the likelihood (Eq. 2), according to

$$p(N_{pos}|x) \propto p(x|N_{pos})p(N_{pos})$$

According to the details described in Appendix 2, it now follows

$$\begin{aligned} &p(N_{pos}|x) = \frac{B(N_{pos} + a,\, N_{tot} - N_{pos} + b)}{B(x + a,\, N_s + b - x)} \binom{N_{tot} - N_s}{N_{pos} - x} \\ &N_{pos} - x \sim \text{BetaBinom}(N_{tot} - N_s,\, x + a,\, N_s + b - x) \\ &x \le N_{pos} \le N_{tot} - N_s + x \end{aligned} \tag{3}$$

This is precisely in line with the results reported elsewhere (1–4), although derived in another way (see Introduction). As $N_{pos} - x$ reflects the number of positive pills in the part of the consignment not subjected to testing, and $N_{tot} - N_s$ is the size of this remainder, the equation corresponds with the interpretation of the predictive distribution.

Approximation for Large Consignments

In the literature, a beta distribution is used in case the consignment is large. It is worthwhile to study if a large consignment approximation can be derived from Eq. 3. For this approximation, a useful expression is needed for $\Gamma(s + t)/\Gamma(s)$ for s approaching infinity and fixed t. In Appendix 3, it is shown that in approximation

$$\Gamma(s + t) \approx s^t \Gamma(s) \text{ for large } s \tag{4}$$

This can be applied to Eq. 3, after expansion of all beta functions and binomial terms to gamma functions. Written in the proportionality notation to retain only those terms that do contain N_{pos}, it follows from Eq. 3:

$$p(N_{pos}) \propto \frac{\Gamma(N_{pos} + a)}{\Gamma(N_{pos} - x + 1)} \frac{\Gamma(N_{tot} - N_{pos} + b)}{\Gamma(N_{tot} - N_{pos} - N_s + x + 1)} \tag{5}$$

Applying Eq. 4 yields in approximation:

$$\begin{aligned} p(N_{pos}) &\propto \frac{\Gamma(N_{pos})N_{pos}^a}{\Gamma(N_{pos})N_{pos}^{-x+1}} \frac{\Gamma(N_{tot} - N_{pos})(N_{tot} - N_{pos})^b}{\Gamma(N_{tot} - N_{pos})(N_{tot} - N_{pos})^{-N_s+x+1}} \\ &= N_{pos}^{a+x-1}(N_{tot} - N_{pos})^{b+N_s-x-1} \\ &\propto \left(\frac{N_{pos}}{N_{tot}}\right)^{a+x-1}\left(1 - \frac{N_{pos}}{N_{tot}}\right)^{b+N_s-x-1} \end{aligned}$$

For large consignments, this approaches a beta distribution according to

$$\frac{N_{pos}}{N_{tot}} \sim Beta(a + x,\, b + N_s - x) \tag{6}$$

This is the result for large consignments as reported elsewhere (1–4), though now derived via a different strategy: instead of deriving Eq. 3 from Eq. 6, as was carried out by Aitken (3,4), now Eq. 6 is derived as large consignment approximation of Eq. 3.

This is not the only way in which the betabinomial distribution of Eq. 3 can be approximated by a beta distribution. In fact, the choices are limitless. To illustrate this, introduce two interim parameters μ and ν. From Eq. 5 in combination with Eq. 4, it now follows

$$\begin{aligned} p(N_{pos}) &\propto \frac{\Gamma(N_{pos} + \mu + a - \mu)}{\Gamma(N_{pos} + \mu - x + 1 - \mu)} \\ &\quad \times \frac{\Gamma(N_{tot} - N_{pos} + \nu + b - \nu)}{\Gamma(N_{tot} - N_{pos} + \nu - N_s + x + 1 - \nu)} \\ &\approx \frac{\Gamma(N_{pos} + \mu)(N_{pos} + \mu)^{a-\mu}}{\Gamma(N_{pos} + \mu)(N_{pos} + \mu)^{-x+1-\mu}} \\ &\quad \times \frac{\Gamma(N_{tot} - N_{pos} + \nu)(N_{tot} - N_{pos} + \nu)^{b-\nu}}{\Gamma(N_{tot} - N_{pos} + \nu)(N_{tot} - N_{pos} + \nu)^{-N_s+x+1-\nu}} \\ &= (N_{pos} + \mu)^{a+x-1}(N_{tot} - N_{pos} + \nu)^{b+N_s-x-1} \propto \\ &\propto \left(\frac{N_{pos} + \mu}{N_{tot} + \mu + \nu}\right)^{a+x-1}\left(1 - \frac{N_{pos} + \mu}{N_{tot} + \mu + \nu}\right)^{b+N_s-x-1} \end{aligned}$$

so it follows in approximation:

$$\frac{N_{pos} + \mu}{N_{tot} + \mu + \nu} \sim Beta(a + x,\, b + N_s - x) \tag{7}$$

Eq. 6 follows from this expression by setting $\mu = \nu = 0$.

A criterion for the optimal choice for μ and ν can be that the mean and variance of N_{pos} as can be derived from Eqs 3 and 7 are identical. As described in Appendix 4, this criterion leads to

$$\frac{N_{pos}+\mu}{N_{tot}\sqrt{\left(1-\frac{N_s}{N_{tot}}\right)\left(1+\frac{a+b}{N_{tot}}\right)}} \sim Beta(a+x,\ b+N_s-x) \quad (8)$$

with $\mu = (a+x)\frac{N_{tot}-N_s}{a+b+N_s}\left(\sqrt{1+\frac{a+b+N_s}{N_{tot}-N_s}}-1\right)-x$

It can be shown (Appendix 4) that $-x \le \mu \le \frac{1}{2}(a-x)$ for any $N_s \le N_{tot}$. Therefore, for large consignment sizes, μ can be ignored compared to N_{pos}, and the denominator of Eq. 8 can be approximated as N_{tot}, so Eq. 8 reduces to the original approximation of Eq. 6.

For the calculation of the posterior probability $\Pr(N_{pos} \ge n)$ for any n, use can be made of a continuity correction, converting Eq. 8 into

$$\Pr(N_{pos} \ge n) = \Pr\left(B > \frac{n-\frac{1}{2}+\mu}{\sqrt{(N_{tot}-N_s)(N_{tot}+a+b)}}\right) \quad (9)$$

with $B \sim Beta(a+x,\ b+N_s-x)$

Sample size calculations aim at selecting the proper value of N_s such that the probability as calculated using Eq. 9 is sufficiently high (typically >0.95 or >0.99) for a given reasonable value of n (typically corresponding to a considerable fraction of the total consignment).

Examples

A few numerical examples are given, based on the tables as published by the ENFSI (1). The tables in the UN guideline (2) are the same. The tables present the required number of samples to be tested (N_s), given the expected number of negatives (pills not containing drugs) in the sample, to ensure with the required confidence that at least a certain fraction of the total consignment of pills contains drugs. This fraction is chosen to be 0.5, 0.7, or 0.9, and the confidence level is set at 0.95 and 0.99. The guidelines provide the sample sizes for the large consignment approximation (Eq. 6), which is

TABLE 1—*Required sample sizes to guarantee with a probability of 95% or 99% that a given seizure contains at least a proportion of* k *drugs, if expected that 0, 1, or 2 negatives are observed.*

				95% Confidence			99% Confidence		
Parameters Prior	Expect. Negs	N_{tot}	Computation Method	k = 0.5	k = 0.7	k = 0.9	k = 0.5	k = 0.7	k = 0.9
a = 1	0	50	Betabinomial (Eq. 3)	4	7	19	5	10	25
b = 1			New approx. (Eq. 9)	4	7	19	5	10	25
		100	Betabinomial (Eq. 3)	4	7	22	6	11	32
			New approx. (Eq. 9)	4	7	22	6	11	32
		500	Betabinomial (Eq. 3)	4	8	27	6	12	41
			New approx. (Eq. 9)	4	8	27	6	12	41
		>50	Guidelines (1,2)	4	8	28	6	12	43
	1	50	Betabinomial (Eq. 3)	6	11	28	8	15	34
			New approx. (Eq. 9)	6	11	28	8	15	34
		100	Betabinomial (Eq. 3)	7	12	35	9	17	45
			New approx. (Eq. 9)	7	12	35	9	17	45
		500	Betabinomial (Eq. 3)	7	13	43	9	18	59
			New approx. (Eq. 9)	7	13	43	9	18	59
		>50	Guidelines (1,2)	7	13	45	10	19	63
	2	50	Betabinomial (Eq. 3)	9	15	36	11	19	41
			New approx. (Eq. 9)	9	15	36	11	19	40
		100	Betabinomial (Eq. 3)	9	17	46	12	21	55
			New approx. (Eq. 9)	9	17	46	12	21	55
		500	Betabinomial (Eq. 3)	10	18	57	13	24	74
			New approx. (Eq. 9)	10	18	57	13	24	74
		>50	Guidelines (1,2)	10	18	60	13	24	80
a = 0.5	0	50	Betabinomial (Eq. 3)	3	5	14	5	8	21
b = 0.5(*)			New approx. (Eq. 9)	3	5	14	5	8	21
		100	Betabinomial (Eq. 3)	3	5	16	5	9	25
			New approx. (Eq. 9)	3	5	16	5	9	25
		500	Betabinomial (Eq. 3)	3	6	18	5	9	30
			New approx. (Eq. 9)	3	6	18	5	9	30
		>50	Guidelines (1,2)	3	6	18	5	10	32
	1	50	Betabinomial (Eq. 3)	6	10	26	8	14	32
			New approx. (Eq. 9)	6	10	26	8	14	32
		100	Betabinomial (Eq. 3)	6	11	31	8	15	41
			New approx. (Eq. 9)	6	11	31	8	15	41
		500	Betabinomial (Eq. 3)	6	12	36	9	16	51
			New approx. (Eq. 9)	6	12	36	9	16	51
		>50	Guidelines (1,2)	6	12	38	9	17	55
	2	50	Betabinomial (Eq. 3)	9	15	35	11	18	40
			New approx. (Eq. 9)	9	15	34	11	18	39
		100	Betabinomial (Eq. 3)	9	16	43	11	20	53
			New approx. (Eq. 9)	9	16	43	11	20	53
		500	Betabinomial (Eq. 3)	9	17	51	12	22	68
			New approx. (Eq. 9)	9	17	51	12	22	68
		>50	Guidelines (1,2)	9	17	54	12	22	73

*The text of the ENFSI guideline (1) indicates a = 0.5, b = 1, but the legend and the numbers suggest $a = b$ = 0.5. This would be in line with the UN guideline (2).

stated to be valid for consignment sizes of 50 or more. In Table 1, the required sample sizes according to the exact betabinomial distribution and according to the approximated equation (including continuity correction) are presented, as function of the total consignment size (of 50, 100, or 500 pills) and some choices. The results show that the original large consignment approximation is rather conservative. Also, it can prescribe sample sizes exceeding 50 (up to 80 in the table), which is clearly impossible for a consignment size of 50, for which the guideline tables should be applicable. The results of the betabinomial and the newly derived approximated beta distribution are well in line with each other. In some cases, the required sample sizes differ by one unit. For example, for $a = b = 0.5$, $N_{tot} = 50$, two expected negatives, 95% confidence, $k = 0.9$, the calculated sample sizes are 35 using the betabinomial distribution and 34 using the approximation. For $N_s = 34$, the exact betabinomial probability equals 0.94984, which is below 0.95, whereas the approximated beta distribution leads to a probability of 0.95078, which is above the threshold. Taking the result with the betabinomial distribution as the correct result, the table reveals that the new approximation (Eq. 9) works very well for consignment sizes as small as 50, whereas the original approximation (Eq. 6) can give rather poor results for consignment sizes up to 100. This illustrates the improved quality of Eq. 9 over Eq. 6.

Discussion

In the current article, the problem of sample size calculations is revisited. In contrast to the earlier reported derivations, now a strategy is described taking the limited size of the consignment and the discrete nature of the random variables into account in all steps. Fortunately, the resulting posterior is the same as is currently in use, and as recommended by organizations like the ENFSI or the UN. Also, strategies to set the values of the a and b parameters of the prior as described by Nordgaard (5) can still be used. The large consignment approximation as derived in the current paper is an improvement compared to the originally reported one. It can easily be implemented in packages like Microsoft Excel. A ready-to-use workbook is available on request. This workbook contains the equations derived in the current Technical Note, as well as the non-Bayesian calculations using the hypergeometric distribution (as described in detail in ENFSI [1] and United Nations [2]).

References

1. ENFSI Guidelines on representative drug sampling. http://www.enfsi.eu/page.php?uid=195 2007 version (accessed on October 6, 2009).
2. United Nations Office on Drugs and Crime Guidelines on representative drug sampling. New York, NY: United Nations, 2009 http://www.unodc.org/unodc/en/scientists/guidelines-on-representative-drug-sampling.html (accessed on October 6, 2009).
3. Aitken CGG. Sampling—how big a sample? J Forensic Sci 1999;44:750–60.
4. Aitken C, Taroni F. Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Chichester: John Wiley & Sons, 2004;179–204.
5. Nordgaard A. Quantifying experience in sample size determination for drug analysis of seized drugs. Law Probab Risk 2005;4:217–25.
6. Coulson SA, Coxon A, Buckleton JS. How many samples from a drug seizure need to be analyzed? J Forensic Sci 2001;46:1456–61.

Additional information and reprint requests:
Jos Weusten, Ph.D.
Statistician
DSM Resolve, The Maastricht Forensic Institute
PO Box 18
6160 MD Geleen
The Netherlands
E-mail: Jos.weusten@DSM.com

Appendix 1: Constructing the Prior

The consignment at hand is viewed as a sample from the super-population of pills manufactured at a given laboratory. Let the true fraction of positives in this super-population be π, being a continuous variable over the range of zero to one. A typical prior for such a number is the beta distribution with parameters a and b, according to

$$p(\pi) = \frac{1}{\mathrm{B}(a,\, b)}\,\pi^{a-1}(1-\pi)^{b-1}$$

Given π, the actual number of N_{pos} in a consignment of size N_{tot} is a binomial random variable with parameters π and N_{tot}.

$$p(\,N_{pos}|\pi\,) = \binom{N_{tot}}{N_{pos}}\pi^{N_{pos}}(1-\pi)^{N_{tot}-N_{pos}}$$

The prior for N_{pos} follows from integration over all π:

$$\begin{aligned} p(N_{pos}) &= \int_{\pi=0}^{1} p(\,N_{pos}|\pi\,)p(\pi)d\pi \\ &= \frac{1}{\mathrm{B}(a,\, b)}\binom{N_{tot}}{N_{pos}}\int_{\pi=0}^{1}\pi^{N_{pos}+a-1}(1-\pi)^{N_{tot}-N_{pos}+b-1}d\pi \end{aligned}$$

The integral reflects the definition of the beta function, hence leading to the betabinomial distribution of Eq. 1.

Appendix 2: Assessing the Posterior

The posterior distribution follows from the prior (Eq. 1) and the likelihood (Eq. 2). Using the proportionality notation and retaining only the N_{pos} containing terms leads to:

$$\begin{aligned} p(N_{pos}|x) &\propto p(x|N_{pos})p(N_{pos}) \\ &\propto \binom{N_{pos}}{x}\binom{N_{tot}-N_{pos}}{N_s-x}\binom{N_{tot}}{N_{pos}}\Gamma(N_{pos}+a)\Gamma(N_{tot}-N_{pos}+b) \end{aligned}$$

Writing out the binomials in terms of factorials leads to

$$\begin{aligned} p(N_{pos}|x) &\propto \frac{\Gamma(N_{pos}+a)\Gamma(N_{tot}-N_{pos}+b)}{(N_{pos}-x)!(N_{tot}-N_{pos}-N_s+x)!} \\ &\propto \frac{\Gamma(N_{pos}+a)\Gamma(N_{tot}+b-N_{pos})}{\Gamma(N_{tot}+a+b)} \\ &\quad\times\frac{(N_{tot}-N_s)!}{(N_{pos}-x)!(N_{tot}-N_{pos}-N_s+x)!} \\ &= \mathrm{B}(N_{pos}+a,\, N_{tot}+b-N_{pos})\binom{N_{tot}-N_s}{N_{pos}-x} \end{aligned}$$

Now, $N_{pos} - x$ is the number of positives in the part of the consignment that is not tested, and $N_{tot} - N_s$ is the size of that part. Let the former be referred to as y, and the latter as N_{rest}, and the current result can be written as

$$\begin{aligned} p(y|x) &\propto \mathrm{B}(x+a+y,\, N_{rest}+N_s+b-x-y)\binom{N_{rest}}{y} \\ &\propto \frac{\mathrm{B}(x+a+y,\, N_{rest}-y+N_s+b-x)}{\mathrm{B}(x+a,\, N_s+b-x)}\binom{N_{rest}}{y} \end{aligned}$$

for $0 \le y \le N_{rest}$. This is a betabinomial distribution with parameters N_{rest}, $x + a$ and $N_s + b - x$. This leads to Eq. 3.

Appendix 3: Approximating Gamma Functions

Approximating $\Gamma(s + t)/\Gamma(s)$ for s approaching infinity and fixed t can be done using Stirling's approximation for the gamma function:

$$\Gamma(s) \approx \sqrt{\frac{2\pi}{s}}\left(\frac{s}{e}\right)^s$$

With this,

$$\frac{\Gamma(s+t)}{\Gamma(s)} \approx \frac{\sqrt{\frac{2\pi}{s+t}}\left(\frac{s+t}{e}\right)^{s+t}}{\sqrt{\frac{2\pi}{s}}\left(\frac{s}{e}\right)^{s}} = \frac{(s+t)^{s+t-\frac{1}{2}}}{s^{s-\frac{1}{2}}e^t} = \frac{s^{s+t-\frac{1}{2}}\left(1+\frac{t}{s}\right)^{s+t-\frac{1}{2}}}{s^{s-\frac{1}{2}}e^t}$$
$$\approx \frac{s^{s+t-\frac{1}{2}}e^t}{s^{s-\frac{1}{2}}e^t} = s^t$$

to be written as Eq. 4 in the text.

Appendix 4: Characterizing the Parameters of the Large Consignment Approximation

A criterion for the optimal choice for μ and v can be that the mean and variance of N_{pos} as can be derived from Eqs 3 and 7 are identical. The requirement of equal expectations can be written as

$$E(N_{pos}) = \frac{(N_{tot} - N_s)(a+x)}{a+b+N_s} + x = \frac{(N_{tot} + \mu + v)(a+x)}{a+b+N_s} - \mu$$

Requiring that the variances are equal leads to

$$Var(N_{pos}) = \frac{(N_{tot} - N_s)(a+x)(b+N_s-x)(N_{tot}+a+b)}{(a+b+N_s)^2(1+a+b+N_s)}$$
$$= \frac{(N_{tot} + \mu + v)^2(a+x)(b+N_s-x)}{(a+b+N_s)^2(1+a+b+N_s)}$$

Solving μ and v requires some straightforward calculus, leading to

$$\begin{cases} \mu = (a+x)\frac{N_{tot}-N_s}{a+b+N_s}\left(\sqrt{1+\frac{a+b+N_s}{N_{tot}-N_s}}-1\right) - x \\ N_{tot} + \mu + v = N_{tot}\sqrt{\left(1-\frac{N_s}{N_{tot}}\right)\left(1+\frac{a+b}{N_{tot}}\right)} \end{cases}$$

This can be substituted in Eq. 7, leading to Eq. 8.

Note that the expression for μ can be written as

$$\mu = (a+x)\xi\left(\sqrt{1+1/\xi}-1\right) - x$$

with $\xi = (N_{tot} - N_s)/(a+b+N_s)$. It follows for positive valued ξ

$$\begin{cases} \lim_{\xi\to 0}\xi\left(\sqrt{1+1/\xi}-1\right) = \lim_{\xi\to 0}\left(\sqrt{\xi^2+\xi}-\xi\right) = 0 \\ \lim_{\xi\to\infty}\xi\left(\sqrt{1+1/\xi}-1\right) = \lim_{\xi\to\infty}\xi\left(\left(1+\frac{1}{2\xi}\right)-1\right) = \frac{1}{2} \\ \frac{\partial}{\partial\xi}\left(\sqrt{\xi^2+\xi}-\xi\right) = \frac{\xi+\frac{1}{2}}{\sqrt{\xi^2+\xi}} - 1 = \frac{\sqrt{\xi^2+\xi+\frac{1}{4}}}{\sqrt{\xi^2+\xi}} - 1 > 0 \end{cases}$$

It now immediately follows $0 \le \xi \le \frac{1}{2}$ hence $-x \le \mu \le \frac{1}{2}(a-x)$ for all possible N_s.