World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:6, No:11, 2012

A Note on Negative Hypergeometric Distribution and Its Approximation

S. B. Mansuri

***Abstract*—**In this paper, at first we explain about negative hypergeometric distribution and its properties. Then we use the w-function and the Stein identity to give a result on the poisson approximation to the negative hypergeometric distribution in terms of the total variation distance between the negative hypergeometric and poisson distributions and its upper bound.

***Keywords*—**Negative hypergeometric distribution, Poisson distribution, Poisson approximation, Stein-Chen identity, w-function.

I. Introduction

Let a box contain S items of which are defective and R are non defective. Items are inspected at random (one at a time) without replacement, from box until the number of non defective items reaches a fixed number r.

Let X be the number of defective items the sample, then X has a negative hypergeometric distribution and denoted by $NH(R,S,r)$. Its probability function can be expressed as;

$$p_X(k)=\frac{\binom{r+k-1}{r}\binom{R-r+S-k}{S-k}}{\binom{R+S}{S}}\qquad k=0,1,\dots,S \tag{1}$$

Where $R,S\in\mathbb{N}$ and $r\in\{1,2,\dots,R\}$.

Now, we show the mean and variance of X are $\mu=\frac{rS}{R+1}$ and $\sigma^2=\frac{rS(R+S+1)(R-r+1)}{(R+1)^2(R+2)}$, respectively.

Proof

$$\mu=E(X)=\sum_{k=0}^{S}k\frac{\binom{r+k-1}{r}\binom{R-r+S-k}{S-k}}{\binom{R+S}{S}}$$

$$=\frac{1}{\binom{R+S}{S}}\sum_{k=0}^{S}\frac{(r+k-1)!}{(k-1)!\,(r-1)!}\binom{R-r+S-k}{S-k}$$

$$=\frac{1}{\binom{R+S}{S}}\sum_{k=0}^{S-1}\frac{(r+k)!}{k!\,(r-1)!}\binom{R-r+S-k-1}{S-k-1}$$

$$=\frac{r}{\binom{R+S}{S}}\sum_{k=0}^{S-1}\binom{r+k}{k}\binom{R-r+S-k-1}{S-k-1}$$

Note that, we have

$$\sum_{j=0}^{k}\binom{a+k-j-1}{k-j}\binom{b+j-1}{j}=\binom{a+b+k-1}{k}$$

Then

$$\mu=\frac{r}{\binom{R+S}{S}}\binom{R+S}{S-1}=\frac{rS}{R+1}$$

For variance we obtain $E(X(X-1))$ then,

$$E\big(X(X-1)\big)=\sum_{k=0}^{S}k(k-1)\frac{\binom{r+k-1}{r}\binom{R-r+S-k}{S-k}}{\binom{R+S}{S}}$$

$$=\frac{r(r+1)}{\binom{R+S}{S}}\sum_{k=0}^{S-2}\binom{r+k+1}{k}\binom{R-r+S-k-2}{S-k-2}$$

$$=\frac{r(r+1)}{\binom{R+S}{S}}\binom{R+S}{S-2}=\frac{rS(r+1)(S-1)}{(R+1)(R+2)}$$

Then

$$\sigma^2=\frac{rS(R+S+1)(R-r+1)}{(R+1)^2(R+2)} \tag{2}$$

Suppose that S and R tend to ∞ in such a way that $\frac{S}{R+1}\to\theta$ $(0<\theta<1)$, then the negative hypergeometric distribution converges to the negative binomial distribution with parameters r and $\frac{\theta}{1+\theta}$. Similarly this distribution may converge to the binomial or poisson or normal distribution if the conditions on their parameters are appropriate.

It should be noted that if $\frac{r}{R+1}$ is not be small and S is sufficiently large, then $NH(R,S,r)$ can also approximated by the normal distribution with mean $\frac{Sr}{R+1}$ and variance $\frac{Sr(R-r+1)}{(R+1)^2}$. In this case, a bound on the normal approximation can be derived by using the same method in [3].

In this paper, we use the w-function associated with the random variable X together with the Stein-Chen identity to give an upper bound for the total variation distance between the negative hypergeometric and poisson distributions.

II. Useful Definition and Propositions

A. Let X be a non-negative integer-valued random variable with distribution F and let P_λ denote the poisson distribution with mean λ. The total variation distance between two distribution defined by:

$$d_{TV}(F,P_\lambda)=\sup_A|F(A)-P_\lambda(A)| \tag{3}$$

Where A runs over subset of non-negative integers. To obtain an upper bound for the total variation distance in terms of the w-function, we apply the Stein-chen identity (see [2]) according to which for every positive constant, every subset A of non-negative integers and some function $g=g_{\lambda,A}$,

$$F(A)-P_\lambda(A)=E\big(\lambda g(X+1)-Xg(X)\big) \tag{4}$$

B. Mansouri is with the Department of Science, Payame Noor University, 19395-4697, Tehran, Iran (phone:+989131831016, e-mail: S_mansouri@pnu.ac.ir).

The explicit formula for the function g can be found e.g in [2], but what we really need are the following estimates valid uniformly for all A:

$$\sup_k|g(k)| \ll \min\left(1, \lambda^{-1/2}\right)$$
$$|\Delta g| = \sup_k|\Delta g(k)| \ll \lambda^{-1}\left(1-e^{-\lambda}\right) \quad (5)$$

where $\Delta g(k) = g(k+1) - g(k)$ (see [1]).

B. Let a non-negative integer-valued random variable X with distribution $F = \{p(k), k = 0,1,2,\dots\}$ have mean μ and variance σ^2. Define a function w associated with the random variable X by the relation

C. $\sigma^2 w(k)p(k) = \sum_{i=0}^{k}(\mu - i)p(i), k = 0,1,2,\dots \quad (6)$

Immediately from the above we have

$$w(0) = \frac{\mu}{\sigma^2}$$
$$w(k+1) = \frac{p(k)}{p(k+1)}w(k) + \frac{\mu-(k+1)}{\sigma^2} \qquad k = 0,1,2,\dots \quad (7)$$

And

$$w(k) \gg 0, k = 0,1,2,\dots \quad (8)$$

Proposition1. If a non-negative integer-valued random variable X with distribution $p(k) > 0$, for all k in support of X and $0 < \sigma^2 = \mathrm{Var}(X) < \infty$, then;

$$\mathrm{Cov}\big(X, g(X)\big) = \sigma^2 E(w(X)\Delta g(X)) \quad (9)$$

For any function $g: \mathbb{N}\cup\{0\} \to \mathbb{R}$ for which $E(w(X)\Delta g(X)) < \infty$ .By taking $g(x) = x$, we have $E\big(w(X)\big) = 1$ (see [4]).

Proposition2. (Reference [6])Let w(X) be the w-function associated with the negative hypergeometric random variable, then;

$$w(k) = \frac{(r+k)(S-k)}{(R+1)\sigma^2} \quad (10)$$

Where $\sigma^2 = \frac{rS(R+S+1)(R-r+1)}{(R+1)^2(R+2)}$.

Proof

Following (7), we have

$$w(k) = \frac{p(k-1)}{p(k)}w(k-1) + \frac{\mu-k}{\sigma^2}$$
$$= \frac{\mu}{\sigma^2} + \frac{p(k-1)}{p(k)}w(k-1) - \frac{k}{\sigma^2} \quad (11)$$

With replacing (1) in (11) we have

$$w(k) = \frac{rS}{(R+1)\sigma^2} + \frac{k(R-r+S-k+1)}{(r+k-1)(S-k+1)}w(k-1) - \frac{k}{\sigma^2}$$
$$k = 1,2,\dots,S$$

And as we told before $w(0) = \frac{rS}{(R+1)\sigma^2}$.

We will show that (10) holds for every $k \in \{1,2,\dots,S\}$.

Equation (11) holds for $k = 1$ i.e

$$w(1) = \frac{(r+1)(S-1)}{(R+1)\sigma^2}$$

We assume that (11) holds for $k = i - 1$, then we will prove that holds for $k = i$.

By mathematical induction, (11) holds for every $k \in \{1,2,\dots,S\}$.

III. Poisson Approximation

We will prove our main result by using the w-function associated with the negative hypergeometric random variable X and the Stein-Chen identity.

For the Stein-Chen identity, using definition 1, its applied for every positive constant λ, and every subset A of $g = g_A: \mathbb{N}\cup\{0\} \to \mathbb{R}$, yield

$$NH(R,S,r)\{A\} - Po(\lambda)\{A\} = E(\lambda g(X+1) - Xg(X)) \quad (12)$$

For any subset A of $\mathbb{N}\cup\{0\}$, Barbour et al in [2] proved that:

$$\sup_{A,k}|\Delta g(k)| = \sup_{A,k}|g(k+1) - g(k)| \ll \lambda^{-1}(1-e^{-\lambda}) \quad (13)$$

The following theorem gives a result of the poisson approximation to the negative hypergeometric distribution.

*Theorem***.** Let X be negative hypergeometric random variable, $\lambda = \frac{rS}{R+1}$ and $r \gg S - 1$, then for $A \subseteq \mathbb{N}\cup\{0\}$

$$d_{TV}(NH(R,S,r), Po(\lambda)) \leq \left(1-e^{-\lambda}\right)\frac{(R+1)(r+1) - S(R-r+1)}{(R+1)(R+2)} \quad (14)$$

Proof

From (12) it follows that

$$|NH(R,S,r)\{A\} - Po(\lambda)\{A\}| = |E(\lambda g(X+1) - Xg(X))|$$
$$= |E(\lambda g(X+1)) - Cov(X, g(X)) - \mu E(g(X))|$$
$$= \left|\lambda E\big(\Delta g(X)\big) - Cov(X, g(X))\right|$$
$$= \left|\lambda E\big(\Delta g(X)\big) - \sigma^2 E(w(X)\Delta g(X))\right| \qquad by\ (9)$$
$$\ll E\left|\big(\lambda - \sigma^2 w(X)\big)\Delta g(X)\right|$$
$$\ll \sup_{x\gg 1}|\Delta g(x)|\, E|\lambda - \sigma^2 w(X)|$$
$$\ll \lambda^{-1}(1-e^{-\lambda})E|\lambda - \sigma^2 w(X)| \qquad by\ (13)$$

Then

$$|NH(R,S,r)\{A\} - Po(\lambda)\{A\}| \ll \lambda^{-1}(1-e^{-\lambda})E|\lambda - \sigma^2 w(X)| \quad (15)$$

Now we show that $\lambda - \sigma^2 w(X) \gg 0$. As by proposition 2,

$$\lambda - \sigma^2 w(X) = \frac{rS}{R+1} - \sigma^2\frac{(r+k)(S-k)}{(R+1)\sigma^2}$$
$$= \frac{rS}{R+1} - \frac{(r+k)(S-k)}{(R+1)}$$
$$= \frac{k(k-S+r)}{R+1} \gg 0$$

Thus

$$E|\lambda - \sigma^2 w(X)| = E\big(\lambda - \sigma^2 w(X)\big)$$
$$= \lambda - \sigma^2 E\big(w(X)\big)$$
$$= \lambda - \sigma^2$$
$$= \lambda\frac{(R+1)(r+1) - S(R-r+1)}{(R+1)(R+2)}$$

Then we have

$$d_{TV}(NH(R,S,r), Po(\lambda)) \leq \left(1-e^{-\lambda}\right)\frac{(R+1)(r+1) - S(R-r+1)}{(R+1)(R+2)}$$

If $r = S - 1$ then

$$d_{TV}\big(NH(R,S,r),Po(\lambda)\big) \le \big(1-e^{-\lambda}\big)\frac{(R+1)(r+1)-(r+1)(R-r+1)}{(R+1)(R+2)}$$

Thus

$$d_{TV}(NH(R,S,r),Po(\lambda)) \le \big(1-e^{-\lambda}\big)\frac{r(r+1)}{(R+1)(R+2)} < \frac{r}{R}$$

REFERENCES

[1] A.D. Barbour, G.K. Eagleson, Poisson Approximation for Some Statistics Based on Exchangeable Trials. Adv. Appl. Probab. 15 (1983), 360-385.

[2] A.D. Barbour, L.Hoslt, S. Janson, Poisson Approximation. Oxford Studies in Probability 2, Clarendon Press, Oxford, 1992.

[3] Boonta, K. Neammanee, Bonds on Random Infinite Urn Model. Bull, Malays. Math.Sci.Soc., 30(2007), 121- 128.

[4] T. Cacoullos, V. Papathanasiou, Characterization of Distributions by Variance Bounds. Statist. Probab. Lett., 7 (1989), 351-356.

[5] M. Majsnerowska, A Note on Poisson Approximation by w-function. Appl. Math. 25, 3 (1998), 387-392.

[6] K. Teerapabolarn, On the Poisson Approximation to the Negative Hypergeometric Distribution, Int. Math. Fourm, (2000), 1-8.