International Journal of Pure and Applied Mathematics
Volume 81 No. 1 2012, 157-164
ISSN: 1311-8080 (printed version)
url: http://www.ijpam.eu

ijpam.eu

BINOMIAL APPROXIMATION OF NON-ISOLATED VERTICES IN A RANDOM GRAPH

K. Teerapabolarn

Department of Mathematics
Faculty of Science
Burapha University
Chonburi 20131, THAILAND

Abstract: This paper uses Stein's method to give lower and upper bounds on the error in approximating the probability of non-isolated vertices in a random graph $\mathbb{G}(n, \wp)$ by the binomial probability of $q^n = [1-(1-\wp)^{n-1}]^n$.

AMS Subject Classification: 62E17, 05C80, 60C05
Key Words: binomial approximation, isolated vertex, random graph, Stein's method

1. Introduction and Main Result

The study of random graphs has a long history and many useful applications. A systematic study of random graphs began with the influential work of Erdös and Rényi in the 1950s and 1960s [4, 5, 6]. The theory has developed into one of the mainstays of modern discrete mathematics, and has produced a prodigious number of results, many of which are highly ingenious, describing statistical properties of graphs, such as the distribution of component sizes, the existence and the size of a giant component, and the typical vertex-vertex distances.

In this study, we focus on approximating the probability of non-isolated vertices in a random graph, and, for studying this approximation, we first need the following definitions and notations. Let $\mathbb{G}(n, \wp)$ be a graph on n labelled vertices $\{1, 2, \ldots, n\}$ where each possible edge $\{i, j\}$ is present randomly and

Received: August 14, 2012

independently with the probability $\wp \in (0,1)$. Let E_{ij} be independent edge indicator such that $E_{ij} = 1$ if edge $\{i,j\} \in \mathbb{G}(n,\wp)$ and $E_{ij} = 0$ otherwise, we then have $\mathbb{P}(E_{ij} = 1) = \wp$. For each i, suppose that

$$Y_i = \begin{cases} 1 & \text{if vertex } i \text{ is an isolated vertex in } \mathbb{G}(n,\wp), \\ 0 & \text{otherwise,} \end{cases}$$

where a vertex i is an isolated vertex in $\mathbb{G}(n,\wp)$ if the number of edges incident with it is 0. Let $X = \sum_{i=1}^{n} Y_i$, then X is the number of isolated vertices in $\mathbb{G}(n,\wp)$ and, for each i, we obtain the probability

$$\begin{aligned} \mathbb{P}(Y_i = 1) &= \prod_{i=1, i\neq j}^{n} \mathbb{P}(E_{ij} = 0) \\ &= (1-\wp)^{n-1} \end{aligned} \tag{1.1}$$

and the expectation of X is

$$\mathbb{E}(X) = n(1-\wp)^{n-1}. \tag{1.2}$$

Since, for $i \neq j$, $\mathbb{E}(Y_iY_j) = (1-p)^{2(n-2)+1} \neq (1-p)^{2(n-1)} = \mathbb{E}(Y_i)\mathbb{E}(Y_j)$, it indicates that the random variables $Y_1, ..., Y_n$ are dependent, however, this dependence rather weak. We observe that there are non-isolated vertices in $\mathbb{G}(n,\wp)$ if and only if $X = 0$, and our interest of this study is approximating the probability of non-isolated vertices in $\mathbb{G}(n,\wp)$, $\mathbb{P}(X = 0)$.

In 2004, Teerapabolarn et al. [11] used the Stein-Chen method to give a bound of the error in the Poisson approximation of non-isolated vertices in $\mathbb{G}(n,\wp)$ as follows:

$$\left|\mathbb{P}(X=0) - e^{-\lambda}\right| \leq (\lambda + e^{-\lambda} - 1)\left[\frac{(n-2)\wp + 1}{n(1-\wp)}\right], \tag{1.3}$$

where $\lambda = n(1-\wp)^{n-1}$.

As mentioned before, all Y_i are weakly dependent, it is probably to approximate the distribution of X by the binomial distribution with parameters n and $p = (1-\wp)^{n-1}$ as well. In this paper, we are interested to approximate the probability of non-isolated vertices in $\mathbb{G}(n,\wp)$, $\mathbb{P}(X = 0)$, by the binomial probability via Stein's method. The following theorem is our main result.

Theorem 1.1. *For* $p = (1-\wp)^{n-1}$, $q = 1-p$ *and* $n \geq 2$,

$$0 \leq \mathbb{P}(X=0) - q^n \leq \frac{[np - (1-q^n)]\wp}{1-\wp}. \tag{1.4}$$

Remark. Since $\lambda = np$ and $1 - q^n > 1 - e^{-\lambda}$, we have $\lambda + e^{-\lambda} - 1 = np - (1-e^{-\lambda}) > np - (1-q^n)$. Thus, the result (1.4) is certainly better than the result (1.3) when $\wp \leq 0.5$, and the lower and upper bounds in (1.4) are also more narrow than the lower and upper bounds in (1.3).

Example. Let $n = 100$ and $\wp = 0.1$, we have $np = 0.00295127$, $q^n = 0.99705304$ and lower and upper bounds of the error between the probability of non-isolated vertices in a random graph $\mathbb{G}(100, 0.1)$, $\mathbb{P}(X = 0)$, and the binomial probability, q^n, are

$$0.99705304 \leq \mathbb{P}(X=0) \leq 0.99705352.$$

For the Poisson approximation in (1.3), an error bound in this case is as follows:

$$0.99705256 \leq \mathbb{P}(X=0) \leq 0.99705360.$$

In this case, it is evident that the lower and upper bounds of the probability approximation of $\mathbb{P}(X = 0)$ by the binomial probability give more narrow than by the Poisson probability.

The following table, we show some representative Poisson and binomial estimates of the probability $\mathbb{P}(X = 0)$ in (1.3) and (1.4) together with their lower and upper bounds.

The sample values in Table 1.1 point out that both Poisson and binomial estimates are more accurate when $n\wp$ is large, or the bounds in (1.3) and (1.4) are more narrow whenever $n\wp$ is large. In addition, the lower and upper bounds of both estimates are not different when $n\wp$ is also large.

2. Proof of Main Result

We will prove our main result by Stein's method for the binomial approximation. In 1972, Stein [8] introduced a powerful and general method for bounding the error in the normal approximation for dependent random variables. This method was first developed and applied in the setting of the Poisson approximation by Chen [3] in 1975, which is usually called the Stein-Chen method. In 1986, Stein [9] applied his method to the binomial case. The Stein's equation

n	$\wp$	Lower	$e^{-\lambda}$	Upper	Lower	q^n	Upper
10	0.1	0.00000000	0.02077085	0.59976599	0.00744089	0.00744089	0.32762375
10	0.3	0.63323022	0.66795393	0.70267765	0.66238784	0.66238784	0.69064095
10	0.5	0.98046875	0.98065825	0.98084775	0.98063952	0.98063952	0.98081029
10	0.7	0.99980315	0.99980319	0.99980323	0.99980319	0.99980319	0.99980323
20	0.1	0.00000000	0.06709113	0.34223695	0.05488667	0.05488667	0.25006334
20	0.3	0.97734222	0.97746012	0.97757802	0.97744741	0.97744741	0.97755249
20	0.5	0.99996185	0.99996185	0.99996185	0.99996185	0.99996185	0.99996185
30	0.1	0.15101454	0.24340255	0.33579057	0.23518099	0.23518099	0.30720539
30	0.3	0.99903429	0.99903449	0.99903470	0.99903448	0.99903448	0.99903467
30	0.5	0.99999994	0.99999994	0.99999994	0.99999994	0.99999994	0.99999994
40	0.1	0.49505886	0.51844148	0.54182409	0.51562147	0.51562147	0.53479365
40	0.3	0.99996362	0.99996362	0.99996362	0.99996362	0.99996362	0.99996362
50	0.1	0.74620859	0.75102162	0.75583464	0.75040382	0.75040382	0.75448434
50	0.3	0.99999872	0.99999872	0.99999872	0.99999872	0.99999872	0.99999872

Table 1: Sample values of Poisson and binomial estimates of $\mathbb{P}(X = 0)$ with their lower and upper bounds

for the binomial distribution with parameters $n \geq 1$ and $p = (1 - q) \in (0, 1)$, which, given h, is defined by

$$(n - x)pg(x + 1) - qxg(x) = h(x) - \mathcal{B}_{n,p}(h), \tag{2.1}$$

where $\mathcal{B}_{n,p}(h) = \sum_{k=0}^{n} h(k)\binom{n}{k}p^k q^{n-k}$ and g and h are bounded real-valued functions defined on $\{0, 1, ..., n\}$.

For $A \subseteq \{0, 1, ..., n\}$, let $h_A : \{0, 1, ..., n\} \to \mathbb{R}$ be defined by

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{2.2}$$

By following Barbour et al. [2], let $g_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ satisfy (2.1), where $g_A(0) = g_A(1)$ and $g_A(x) = g_A(n)$ for $x \geq n$.

For the case $A = \{0\}$, the solution $g = g_{\{0\}}$ of (2.1) can be written as

$$g(x) = \frac{\mathcal{B}_{n,p}(1 - h_{C_{x-1}})}{x\binom{n}{x}p^x q^{-x+1}}, \tag{2.3}$$

$1 \leq x \leq n$, where $C_x = \{0, ..., x\}$, and it follows from [2] that g is positive and decreasing in $1 \leq x \leq n$.

Let $\Delta g(x) = g(x+1) - g(x)$, then we have $\Delta g(x) \leq 0$ for $x \geq 1$, and we also have the following lemma.

Lemma 2.1. *Let g and Δg be defined as above. Then, for any $j, k \in \mathbb{N}$ and $n \geq 2$,*

$$|g(j) - g(k)| \leq \sup_{x \geq 1} |\Delta g(x)||j - k| \tag{2.4}$$

and for $x \geq 1$,

$$|\Delta g(x)| \leq \frac{np - (1 - q^n)}{(n-1)np^2}. \tag{2.5}$$

Proof. The first inequality follows Lemma 2.1 (1) of Teerapabolarn et al. [11]. Next we will show (2.5) holds. Since $|\Delta g(x)| = 0$ for $x \geq n$, it suffices to show that (2.5) holds for $1 \leq x \leq n-1$. By (2.3), it can be seen that

$$\begin{aligned}
|\Delta g(x)| &= g(x) - g(x+1) \\
&= \frac{\sum_{k=x}^{n} \binom{n}{k} p^k q^{n-k}}{x\binom{n}{x} p^x q^{-x+1}} - \frac{\sum_{k=x+1}^{n} \binom{n}{k} p^k q^{n-k}}{(x+1)\binom{n}{x+1} p^{x+1} q^{-x}} \\
&= \frac{q^{n-1}}{x\binom{n}{x} p^x q^{n-x}} \left\{ \sum_{k=x}^{n} \binom{n}{k} p^k q^{n-k} - \sum_{k=x}^{n-1} \binom{n}{k} p^k q^{n-k} \frac{x(n-k)}{(k+1)(n-x)} \right\} \\
&= \frac{q^{n-1}}{x} \sum_{k=x}^{n} \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{x} p^x q^{n-x}} \left\{ 1 - \frac{x(n-k)}{(k+1)(n-x)} \right\} \\
&= \frac{q^{n-1}}{x} \left\{ \frac{1}{x+1} + \frac{p(x+1)}{p(x)} \left[1 - \frac{x(n-x-1)}{(x+2)(n-x)} \right] + \cdots + \frac{p(n-1)}{p(x)} \right. \\
&\qquad \left. \times \left[1 - \frac{x}{n(n-x)} \right] + \frac{p(n)}{p(x)} \right\} \quad \left(p(i) = \binom{n}{i} p^i q^{n-i} \text{ for } 1 \leq i \leq n \right) \\
&= \frac{q^{n-1}}{x} \left\{ \frac{1}{x+1} + \frac{n-x}{x+1} \frac{p}{q} \left[1 - \frac{x(n-x-1)}{(x+2)(n-x)} \right] + \cdots + \frac{n-x}{x+1} \frac{n-x-1}{x+2} \right. \\
&\qquad \times \cdots \times \frac{2}{n-1} \left(\frac{p}{q} \right)^{n-1-x} \left[1 - \frac{x}{n(n-x)} \right] \\
&\qquad \left. + \frac{n-x}{x+1} \frac{n-x-1}{x+2} \cdots \frac{1}{n} \left(\frac{p}{q} \right)^{n-x} \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{q^{n-1}}{x}\left\{\frac{1}{x+1}+\frac{1}{x+1}\frac{p}{q}\left[\frac{2n-x}{x+2}\right]+\cdots+\frac{1}{x+1}\frac{n-x-1}{x+2}\cdots\frac{2}{n-1}\right.\\
&\quad\left.\times\left(\frac{p}{q}\right)^{n-1-x}\left[\frac{n(n-x)-x}{n}\right]+\frac{n-x}{x+1}\frac{n-(x+1)}{x+2}\cdots\frac{1}{n}\left(\frac{p}{q}\right)^{n-x}\right\}\\
&\le q^{n-1}\left\{\frac{1}{2}+\frac{1}{2}\frac{p}{q}\left[\frac{2n-1}{3}\right]+\cdots+\frac{1}{2}\frac{n-2}{3}\cdots\frac{2}{n-1}\left(\frac{p}{q}\right)^{n-2}\left[\frac{n(n-1)-1}{n}\right]\right.\\
&\quad\left.+\frac{n-1}{2}\frac{n-2}{3}\cdots\frac{1}{n}\left(\frac{p}{q}\right)^{n-1}\right\}\\
&= q^{n-1}\sum_{k=1}^{n}\frac{\binom{n}{k}p^kq^{n-k}}{npq^{n-1}}\left\{1-\frac{n-k}{(k+1)(n-1)}\right\}\\
&= \frac{np-(1-q^n)}{(n-1)np^2}.
\end{aligned}
$$

Hence, (2.5) holds. $\square$

Proof of Theorem 1.1. Substituting $h = h_{\{0\}}$ in (2.1) yields

$$
\begin{aligned}
\mathbb{P}(X=0)-q^n &= \mathbb{E}[(n-X)pg(X+1)-qXg(X)]\\
&= \mathbb{E}[npg(X+1)-Xg(X)]+p\mathbb{E}\{X[g(X)-g(X+1)]\}, \quad (2.6)
\end{aligned}
$$

where g is defined in (2.3).
Let $X_i = X - Y_i$, then, by the same argument detailed as in the proof of Theorem 1.1 in [11], we have

$$
\begin{aligned}
\mathbb{E}[npg(X+1)-Xg(X)] &= \sum_{i=1}^{n}\mathbb{E}[pg(X_i+Y_i+1)-Y_ig(X_i+Y_i)]\\
&= \sum_{i=1}^{n}\left\{p^2\mathbb{E}[g(X_i+2)|Y_i=1]+pq\mathbb{E}[g(X_i+1)|Y_i=0]\right\}\\
&\quad -\sum_{i=1}^{n}p\mathbb{E}\left[g(X_i+1)|Y_i=1\right]\\
&= \sum_{i=1}^{n}p^2\left\{\mathbb{E}[g(X_i+2)|Y_i=1]-\mathbb{E}[g(X_i+1)|Y_i=1]\right\}\\
&\quad +\sum_{i=1}^{n}pq\left\{\mathbb{E}[g(X_i+1)|Y_i=0]-\mathbb{E}\left[g(X_i+1)|Y_i=1\right]\right\}
\end{aligned}
$$

and a similar argument gives

$$\mathbb{E}\{X[g(X) - g(X+1)]\} = \sum_{i=1}^{n} p\{\mathbb{E}[g(X_i+1)|Y_i = 1] - \mathbb{E}[g(X_i+2)|Y_i = 1]\}.$$

Therefore, by replacing these arguments in (2.6), it becomes

$$\begin{aligned}\mathbb{P}(X = 0) - q^n &= \sum_{i=1}^{n} pq\left\{\mathbb{E}[g(X_i+1)|Y_i = 0] - \mathbb{E}\left[g(X_i+1)|Y_i = 1\right]\right\}\\ &= \sum_{i=1}^{n} p\left\{\mathbb{E}[g(X_i+1)] - \mathbb{E}\left[g(X_i+1)|Y_i = 1\right]\right\} \qquad (2.7)\\ &= \sum_{i=1}^{n} p\mathbb{E}\left[g(X - Y_i + 1) - g(X_i^* + 1)\right], \qquad (2.8)\end{aligned}$$

where (2.7) is obtained from Soon [7] on pp. 708 and X_i^* in (2.8) has the same distribution as $X - Y_i$ conditional on $Y_i = 1$, that is, $X_i^* \sim (X - Yi)|Y_i = 1$. By using the fact in Barbour [1] that $X_i^* \geq X - Y_i$ for $1 \leq i \leq n$, thus $g(X - Y_i + 1) - g(X_i^* + 1) \geq 0$, this implies that $\mathbb{P}(X = 0) - q^n \geq 0$. Hence, by (2.8), we have

$$\begin{aligned}0 \leq \mathbb{P}(X = 0) - q^n &\leq \sum_{i=1}^{n} \sup_{x \geq 1} |\Delta g(x)| p\mathbb{E}(X_i^* - X + Y_i) \text{ (by (2.4))}\\ &\leq \frac{np - (1-q^n)}{(n-1)np^2} \sum_{i=1}^{n} p\mathbb{E}(X_i^* - X + Y_i) \text{ (by (2.5)).} \qquad (2.9)\end{aligned}$$

Using the same argument detailed as in the proof of Theorem 1.2 (ii) in [10], we can obtain

$$\begin{aligned}\sum_{i=1}^{n} p\mathbb{E}(X_i^* - X + Y_i) &= \mathbb{E}(X^2) - np(np+1) + np(1-\wp)^{n-1}\\ &= (np)^2(1-\wp)^{-1} - np(1-\wp)^{n-2} - (np)^2 + np(1-\wp)^{n-1}\\ &= \frac{(np)^2(n-1)\wp}{n(1-\wp)}. \qquad (2.10)\end{aligned}$$

Hence, by (2.9) and (2.10), the theorem is obtained. □

References

[1] A.D. Barbour, Stochastic processes: Theory and methods, *Handbook of Statistics* **19**, Elsevier Science (2001), 79-115.

[2] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Studies in probability 2, Clarendon Press, Oxford (1992).

[3] L.H.Y. Chen, Poisson approximation for dependent trials, *Annals of Probability*, **3** (1975), 534-545.

[4] P. Erdös, A. Rényi, On random graphs, *Publications Mathematicae*, **6** (1959), 290-297.

[5] P. Erdös, A. Rényi, On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5** (1960), 17-61.

[6] P. Erdös, A. Rényi, On the strength of connectedness of a random graph, *Acta Mathematica Scientia Hungary*, **12** (1961), 261-267.

[7] S.Y.T. Soon, Binomial approximation for dependent indicators, *Statist. Sinica*, **6** (1996), 703-714.

[8] C.M. Stein, A bound for the error in normal approximation to the distribution of a sum of dependent random variables, *Proc.Sixth Berkeley Sympos. Math. Statist. Probab.*, **3** (1972), 583-602.

[9] C.M. Stein, *Approximate Computation of Expectations*, IMS, Hayward California (1986).

[10] K. Teerapabolarn, A non-uniform bound on Poisson approximation for sums of Bernoulli random variables with small mean, *Thai J. Math.*, **4** (2006), 179-196.

[11] K. Teerapabolarn, K. Neammanee, S. Chongcharoen, Approximation of the probability of non-isolated vertices in random graph, *Annual Meeting in Applied Statistics 2004*, National Institute of Development Administration (2004), 9-18.