J. Appl. Prob. **53,** 1265–1270 (2016)
*doi:*10.1017/jpr.2016.79
© *Applied Probability Trust* 2016

IMPROVED CHEN–STEIN BOUNDS ON THE PROBABILITY OF A UNION

SHELDON M. ROSS,* *University of Southern California*

Abstract

We improve the Chen–Stein bounds when applied to the probability of a union. When the probability is small, the improvement in the distance from the lower to the upper bound is roughly a factor of 2. Further improvements are determined when the events of the union are either negatively or positively dependent.

Keywords: Chen–Stein Poisson approximation; improved bounds on unions; negative and positive dependencies

2010 Mathematics Subject Classification: Primary 60E15
Secondary 60E99

1. Introduction

Let X_i, $i = 1, \ldots, n$, be Bernoulli random variables with means $\lambda_i = \mathbb{E}[X_i]$, and set $W = \sum_{i=1}^n X_i$. Also, let $\lambda = \sum_{i=1}^n \lambda_i = \mathbb{E}[W]$. For Z being a Poisson random variable with mean λ, the Chen–Stein approach is a way to bound $|\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)|$ for any set of integers A.

Whereas the Chen–Stein approach can be used to bound the error of the Poisson approximation of $\mathbb{P}(W \in A)$ for any set A, many important applications are concerned with bounding $\mathbb{P}(W = 0)$. We mention a few.

- Consider independent trials that each result in any of the outcomes $1, \ldots, n$ with probabilities $p_1, \ldots, p_n$, $\sum_{i=1}^n p_i = 1$, and suppose that there are specified numbers $r_1, \ldots, r_n$.

  In the generalized birthday problem we are interested in M, defined as the number of trials until there have been at least r_i type-i outcomes for some $i = 1, \ldots, n$. (The classical problem has $n = 365$, $p_i \equiv 1/n$, and all $r_i = 2$.) With N_i being the number of type-i outcomes in the first k trials, $i = 1, \ldots, n$, and $X_i = \mathbf{1}\{N_i \geq r_i\}$, we have $\mathbb{P}(M > k) = \mathbb{P}(\sum_{i=1}^n X_i = 0)$.

  The generalized coupon collecting problem concerns N, the number of trials until there have been at least r_i type-i outcomes for every $i = 1, \ldots, n$. (The classical coupon collecting problem has all $r_i = 1$.) With N_i defined as the number of type-i outcomes in the first k trials, $i = 1, \ldots, n$, and $X_i = \mathbf{1}\{N_i < r_i\}$, we have $\mathbb{P}(N \leq k) = \mathbb{P}(\sum_{i=1}^n X_i = 0)$.

- In reliability systems with components $1, \ldots, m$, we often suppose that there are specified subsets of components $C_1, \ldots, C_n$, none of which is a subset of another, such that the system fails if and only if all components in at least one of these subsets fail. Thus, with X_i

Received 22 July 2015; revision received 7 January 2016.
* Postal address: Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email address: smross@usc.edu
This material is based upon work supported by, or in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF-14-1-0166, and by the National Science Foundation under contract/grant number CMMI1233337.

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press

the indicator of the event that all components in C_i fail, we see that $\mathbb{P}(\text{system is failed}) = \mathbb{P}(\sum_{i=1}^{n} X_i > 0) = 1 - \mathbb{P}(\sum_{i=1}^{n} X_i = 0)$.

- With $Y_1, Y_2, \ldots$ independent and identically distributed discrete random variables, we are often interested in N, the first time that the pattern $y_1, y_2, \ldots, y_r$ appears. With X_i defined as the indicator of the event that $Y_{i-k} = y_{r-k}$, $k = 0, \ldots, r-1$, we see that $\mathbb{P}(N > n) = \mathbb{P}(\sum_{i=r}^{n} X_i = 0)$.
- In DNA matching problems we are often interested in the largest common subsequence in the subsequences $W_1, \ldots, W_{r+k}$ and $Y_1, \ldots, Y_{s+k}$. In particular we often want to determine the probability that there would be a common subsequence of length k if the $r + s + 2k$ data values were independent and identically distributed. If we let $X_{i,j}$ equal the indicator of the event that $W_{i+m} = Y_{j+m}, m = 1, \ldots, k$ the probability of such subsequences is $\mathbb{P}(\sum_{i \le r, j \le s} X_{i,j} > 0)$.
- The Erdös–Rényi model for a random graph on a set of n nodes supposes that each of the $\binom{n}{2}$ pair of nodes is, independent of others, an edge with probability p. Suppose that we are interested in the probability that the maximum of the degrees of the n nodes is at least k. Then, with X_i the indicator of the event that the degree of node i is at least k, the desired probability is $\mathbb{P}(\sum_{i=1}^{n} X_i > 0)$.
- If $Y_1, \ldots, Y_n$ has a multivariate normal distribution then, with $X_i = \mathbf{1}\{Y_i > x\}$, we have

$$\mathbb{P}(\max_i Y_i \le x) = \mathbb{P}\left(\sum_{i=1}^{n} X_i = 0\right).$$

We now show how the usual Poisson bounds for $\mathbb{P}(W = 0)$ can be improved. To begin, however, we review some basic Chen–Stein theory.

2. Some relevant Chen–Stein theory

As before, suppose that the X_i are Bernoulli random variables with means $\lambda_i = \mathbb{E}[X_i]$, $i = 1, \ldots, n$; set $W = \sum_{i=1}^{m} X_i$, and let $\lambda = \mathbb{E}[W]$. Also, let Z be a Poisson random variable with mean λ.

For any set of nonnegative integers A, let f_A be as follows:

$$\begin{aligned} f_A(0) &= 0, \\ \lambda f_A(j+1) &= j f_A(j) + \mathbf{1}\{j \in A\} - \mathbb{P}(Z \in A), \qquad j \ge 0. \end{aligned} \tag{1}$$

The following lemma is key to the Chen–Stein approach. (For a proof, see [1].)

Lemma 1. *For all A,*

$$|f_A(j) - f_A(i)| \le \frac{1 - \mathrm{e}^{-\lambda}}{\lambda} |j - i|.$$

It follows from (1) that

$$\lambda f_A(W+1) - W f_A(W) = \mathbf{1}\{W \in A\} - \mathbb{P}(Z \in A).$$

Taking expectations yields

$$\lambda \mathbb{E}[f_A(W+1)] - \mathbb{E}[W f_A(W)] = \mathbb{P}(W \in A) - \mathbb{P}(Z \in A). \tag{2}$$

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press

Now,

$$\mathbb{E}[Wf_A(W)] = \sum_{i=1}^{n} \mathbb{E}[X_i f_A(W)] = \sum_{i=1}^{n} \mathbb{E}[f_A(W) \mid X_i = 1]\lambda_i = \sum_{i=1}^{n} \mathbb{E}[f_A(1+V_i)]\lambda_i, \quad (3)$$

where $V_1, \ldots, V_n$ are any random variables such that $V_i =_{\mathrm{st}} \sum_{j \neq i} X_j \mid X_i = 1$. It follows from (2) and (3) that

$$\sum_i \lambda_i (\mathbb{E}[f_A(W+1)] - \mathbb{E}[f_A(1+V_i)]) = \mathbb{P}(W \in A) - \mathbb{P}(Z \in A).$$

Hence,

$$\mathbb{P}(W \in A) - \mathbb{P}(Z \in A) = \sum_i \lambda_i \mathbb{E}[f_A(W+1) - f_A(1+V_i)], \quad (4)$$

which, from Lemma 1, yields

$$|\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \leq \frac{1-\mathrm{e}^{-\lambda}}{\lambda} \sum_i \lambda_i \mathbb{E}[|W - V_i|], \quad (5)$$

which is the Chen–Stein bound (see [3]).

3. Improvements when bounding $\mathbb{P}(W = 0)$

The function $f_0 \equiv f_{\{0\}}$ is the unique solution of

$$f_0(0) = 0, \qquad \lambda f_0(1) = 1 - \mathrm{e}^{-\lambda}, \qquad \lambda f_0(j+1) = j f_0(j) - \mathrm{e}^{-\lambda}, \quad j \geq 1.$$

Lemma 2. *It holds that*

$$f_0(j) = \int_0^1 \mathrm{e}^{-\lambda t} t^{j-1} \, \mathrm{d}t, \qquad j \geq 1.$$

Proof. As the preceding clearly holds when $j = 1$, we need to show that

$$\int_0^1 \lambda \mathrm{e}^{-\lambda t} t^j \, \mathrm{d}t = j \int_0^1 \mathrm{e}^{-\lambda t} t^{j-1} \, \mathrm{d}t - \mathrm{e}^{-\lambda}, \qquad j \geq 1,$$

which follows upon integrating the left-hand side of the preceding by parts. □

Corollary 1. *It holds that $f_0(j)$ is a nonnegative, decreasing, convex function of j.*

Proof. This immediately follows from Lemma 2 upon using the fact that t^j is, when $0 < t < 1$, a nonnegative, decreasing, convex function of j. □

Corollary 2. *For $i > 0$ and $j > 0$,*

$$|f_0(i) - f_0(j)| \leq |j-i|(f_0(1) - f_0(2)) = \frac{\mathrm{e}^{-\lambda} - (1-\lambda)}{\lambda^2} |j-i|.$$

Proof. For $0 < i < j$, it follows from the fact that f_0 is decreasing and convex that

$$\begin{aligned} 0 &\leq f_0(i) - f_0(j) \\ &= f_0(i) - f_0(i+1) + f_0(i+1) - f_0(i+2) + \cdots - f_0(j) \\ &\leq (j-i)(f_0(1) - f_0(2)), \end{aligned}$$

which proves the result. □

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press

Consequently, if $A = \{0\}$ then going from (4) to (5) to bound $|\mathbb{P}(W = 0) - e^{-\lambda}|$, the constant $f_0(1) = (1 - e^{-\lambda})/\lambda$ can be replaced by the smaller constant $f_0(1) - f_0(2) = (e^{-\lambda} - (1 - \lambda))/\lambda^2$, thus proving the following result.

Proposition 1. *It holds that*

$$|\mathbb{P}(W = 0) - e^{-\lambda}| \le \frac{e^{-\lambda} - (1 - \lambda)}{\lambda^2} \sum_i \lambda_i \mathbb{E}[|W - V_i|],$$

where V_i, $i = 1, \ldots, n$, *are any random variables such that* $V_i =_{\text{st}} \sum_{j \neq i} X_j \mid X_i = 1$.

For instance, when $\lambda = 1$, the preceding bound on $|\mathbb{P}(W = 0) - e^{-\lambda}|$ is smaller than that given by (5) by the factor $e - 1$, and as $\lambda \to 0$, it is smaller by a factor that converges to 2.

Remark. The strategy of attempting to improve the Chen–Stein bounds for $|\mathbb{P}(W = j) - \mathbb{P}(Z = j)|$ by exploiting properties of the functions $f_{\{j\}}$ was mentioned in [1, Section 2.4].

Further improvements on the Chen–Stein bounds on $\mathbb{P}(W = 0) - e^{-\lambda}$ can be obtained in the special cases where knowledge that $X_i = 1$ makes it less likely that $X_j = 1$, $j \neq i$, or when such knowledge makes it more likely that $X_j = 1$, $j \neq i$.

Definition. With $\boldsymbol{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, we say that $X_1, \ldots, X_n$ are *negatively dependent* if

$$\boldsymbol{X}_{-i} \mid X_i = 1 \le_{\text{st}} \boldsymbol{X}_{-i}, \qquad i = 1, \ldots, n, \tag{6}$$

and say that $X_1, \ldots, X_n$ are *positively dependent* if

$$\boldsymbol{X}_{-i} \mid X_i = 1 \ge_{\text{st}} \boldsymbol{X}_{-i}, \qquad i = 1, \ldots, n. \tag{7}$$

Proposition 2. *If* $X_1, \ldots, X_n$ *are negatively dependent then*

$$\mathbb{P}(W = 0) \le \prod_{i=1}^{n} (1 - \lambda_i).$$

If $X_1, \ldots, X_n$ *are positively dependent then*

$$\mathbb{P}(W = 0) \ge \prod_{i=1}^{n} (1 - \lambda_i). \tag{8}$$

Proof. Suppose that $X_1, \ldots, X_n$ are negatively dependent. Using the fact that negative dependence implies that $\mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 1) \ge \mathbb{P}(X_j = 0,\ j \le i)$, we obtain

$$\begin{aligned}
&\mathbb{P}(X_j = 0,\ j \le i) \\
&\quad = \mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 1)\lambda_{i+1} + \mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 0)(1 - \lambda_{i+1}) \\
&\quad \ge \mathbb{P}(X_j = 0,\ j \le i)\lambda_{i+1} + \mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 0)(1 - \lambda_{i+1}).
\end{aligned}$$

Hence, $\mathbb{P}(X_j = 0,\ j \le i) \ge \mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 0)$, implying that

$$\begin{aligned}
\mathbb{P}(X_{i+1} = 0 \mid X_j = 0,\ j \le i) &= \frac{\mathbb{P}(X_{i+1} = 0)\mathbb{P}(X_j = 0,\ j \le i \mid X_{i+1} = 0)}{\mathbb{P}(X_j = 0,\ j \le i)} \\
&\le \mathbb{P}(X_{i+1} = 0).
\end{aligned}$$

The result now follows from the identity $\mathbb{P}(W = 0) = \mathbb{P}(X_1 = 0)\prod_{i=1}^{n-1} \mathbb{P}(X_{i+1} = 0 \mid X_j = 0,\ j \le i)$. With all inequalities being reversed, the proof is similar in the positive dependency case. □

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press

Remarks. (i) The positive-dependent result of Proposition 2 is not new. It follows from [4] that in the positive case $X_1, \ldots, X_n$ are associated and, thus, $\mathbb{P}(X_i = 0,\ i = 1, \ldots, n) \geq \prod_{i=1}^{n} \mathbb{P}(X_i = 0)$.

(ii) The negative-dependent condition (6) holds in the generalized birthday and coupon collecting problems. It also holds in the pattern problem when the pattern has no overlap, in the sense that $(i_j, \ldots, i_r) \neq (i_1, \ldots, i_{r-j+1})$ for any $j > 1$. For instance, to show that (6) holds in the generalized birthday problem, we use a coupling argument as follows. For fixed i, let N_i be binomial with parameters (k, p_i), and let N_i' have the distribution of $N_i \mid N_i \geq r_i$. Because $N_i' \geq_{\mathrm{st}} N_i$ we can couple them so that $N_i' \geq N_i$. Now generate the results of $k - N_i$ independent trials, with each trial resulting in outcome j with probability $p_j/(1 - p_i),\ j \neq i$; let N_j denote the number of type-j outcomes in these $k - N_i$ trials, and let N_j' be the number of type-j outcomes in the first $k - N_i'$ of these $k - N_i$ trials. Because, for $j \neq i$, $N_j \geq N_j'$, we see that $Y_j \equiv \mathbf{1}\{N_j \geq r_j\} \geq \mathbf{1}\{N_j' \geq r_j\} \equiv Y_j',\ j \neq i$. As it is easy to see that $(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$ has the distribution of $\boldsymbol{X}_{-i}$, whereas $(Y_1', \ldots, Y_{i-1}', Y_{i+1}', \ldots, Y_n')$ has the distribution of $\boldsymbol{X}_{-i} \mid X_i = 1$, condition (6) is shown. The coupling arguments for showing that (6) holds in the generalized coupon collecting and in the no-overlap pattern problems are similar.

(iii) The inequality $\mathbb{P}(W = 0) \leq \prod_{i=1}^{n}(1 - \lambda_i)$ was directly proven for the generalized coupon and birthday problems in [6]. Other notions of negative dependency are given in [2].

(iv) The positive dependent condition (7) holds in the reliability system example when the component states are independent. It also holds in the Erdös–Rényi example. The coupling arguments for showing these results are straightforward.

(v) Among other bounds of $\mathbb{P}(W = 0)$ are those of [5] and [8], which exploit dependency graphs for $X_1, \ldots, X_n$. An upper bound for $\mathbb{P}(W = 0)$ is given by the conditional expectation inequality (see Section 3.2 of [7]).

Example 1. Consider the generalized birthday problem when all $r_i = 3$ and $p_i = \frac{1}{365}, i = 1, \ldots, 365$. Thus, we are asking for the probability that among k people there will be a group of size three that all share the same birthday. Because $\lambda_i = \mathbb{P}(\text{binomial}(k, \frac{1}{365}) \geq 3)$, when $k = 88$, we have $\lambda_i \approx 0.001\,896\,6$. Hence, with W the number of days having at least three birthdays, the Poisson approximation is that W is approximately Poisson with mean $\lambda = 365(0.001\,896\,6) = 0.692\,26$, so that the Poisson approximation of the probability that there will be a set of size three having the same birthday is

$$\mathbb{P}(W > 0) \approx 1 - \mathrm{e}^{-0.692\,26} \approx 0.499\,56.$$

To compute $\mathbb{E}[V_i]$, let $B(n, p)$ be a binomial (n, p) random variable. Then

$$\begin{aligned}\mathbb{E}[W] &= \sum_{j=0}^{2} \mathbb{E}[W \mid N_1 = j]\mathbb{P}\left(B\left(88, \frac{1}{365}\right) = j\right) + \mathbb{E}[W \mid N_1 \geq 3]\mathbb{P}\left(B\left(88, \frac{1}{365}\right) \geq 3\right) \\ &= 364 \sum_{j=0}^{2} \mathbb{P}\left(B\left(88 - j, \frac{1}{364}\right) \geq 3\right)\mathbb{P}\left(B\left(88, \frac{1}{365}\right) = j\right) \\ &\quad + (1 + \mathbb{E}[V_1])\mathbb{P}\left(B\left(88, \frac{1}{365}\right) \geq 3\right).\end{aligned}$$

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press

Solving the preceding yields $\mathbb{E}[V_1] = 0.628\,803\,581\,25$. Consequently, the upper bound on the error of the Poisson approximation as given by (5) is $(1 - e^{-\lambda})(\lambda - \mathbb{E}[V_i]) \approx 0.031\,70$, giving the bounds

$$0.467\,86 \leq \mathbb{P}(W > 0) \leq 0.531\,26,$$

whereas the results of this paper show that

$$0.499\,88 \leq \mathbb{P}(W > 0) \leq 0.517\,22.$$

Example 2. Consider the Erdös–Rényi graph model where each pair of nodes is independently an edge with probability p. Let X_i denote the indicator of the event that there are no edges involving node i. Then $\mathbb{P}(W = 0)$ is the probability that there are no isolated nodes. Then, with $q = 1 - p$, $\lambda_i = q^{n-1}$, and $\lambda = nq^{n-1}$, $\mathbb{E}[V_i] = (n-1)q^{n-2}$. Because $X_1, \ldots, X_n$ are positively dependent, $V_i \geq_{\text{st}} W - X_i$, and so coupling such that this holds yields

$$\begin{aligned}\mathbb{E}[|V_i - W|] &\leq \mathbb{E}[|V_i - (W - X_i)| + |X_i|] \\ &= \mathbb{E}[V_i - W + X_i] + \mathbb{E}[X_i] \\ &= (n-1)q^{n-2} - nq^{n-1} + 2q^{n-1}.\end{aligned}$$

Consequently, with $\alpha = (n-1)q^{n-2} - (n-2)q^{n-1}$, the original Chen–Stein bounds are

$$e^{-nq^{n-1}} - (1 - e^{-nq^{n-1}})\alpha \leq \mathbb{P}(W = 0) \leq e^{-nq^{n-1}} + (1 - e^{-nq^{n-1}})\alpha.$$

On the other hand, because $X_1, \ldots, X_n$ are positively dependent the bounds presented in this paper yield

$$(1 - q^{n-1})^n \leq \mathbb{P}(W = 0) \leq e^{-nq^{n-1}} + \left(\frac{e^{-nq^{n-1}} - 1 + nq^{n-1}}{nq^{n-1}}\right)\alpha.$$

For instance, if $n = 10$ and $p = 0.5$ then the original Chen–Stein bounds are $0.980\,28 \leq \mathbb{P}(W = 0) \leq 0.981\,04$, whereas the new bounds yield $0.980\,64 \leq \mathbb{P}(W = 0) \leq 0.980\,85$.

References

[1] Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximations*. Oxford University Press.

[2] Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *Ann. Prob.* **10,** 765–772.

[3] Chen, L. H. Y. (1975). Poisson approximations for dependent trials. *Ann. Prob.* **3,** 534–545.

[4] Cohen, A. and Sackrowitz, H. B. (1994). Association and unbiased tests in statistics. In *Stochastic Orders and Their Applications*, Academic Press, Boston, MA, pp. 251–274.

[5] Janson, S. (1998). New versions of Suen's correlation inequality. *Random Structures Algorithms* **13,** 467–483.

[6] Mallows, C. (1968). An inequality involving multinomial probabilities. *Biometrika* **55,** 422–424.

[7] Ross, S. M. (2002). *Probability Models for Computer Science*. Academic Press, San Diego, CA.

[8] Suen, W.-C. S. (1990). A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures Algorithms* **1,** 231–242.

https://doi.org/10.1017/jpr.2016.79 Published online by Cambridge University Press