Total Variation Distance for Poisson Subset Numbers * †

Larry Goldstein and Gesine Reinert
University of Southern California and Oxford University

August 14, 2006

Abstract

Let n be an integer and $A_0, \ldots, A_k$ random subsets of $\{1, \ldots, n\}$ of fixed sizes $a_0, \ldots, a_k$, respectively chosen independently and uniformly. We provide an explicit and easily computable total variation bound between the distance from the random variable $W = |\cap_{j=0}^k A_j|$, the size of the intersection of the random sets, to a Poisson random variable Z with intensity $\lambda = EW$. In particular, the bound tends to zero when λ converges and $a_j \to \infty$ for all $j = 0, \ldots, k$, showing that W has a asymptotic Poisson distribution in this regime.

1 Introduction

Let n be an integer, take $k+1$ random subsets $A_0, \ldots, A_k$ of $\{1, \ldots, n\}$ with fixed sizes $a_0, \ldots, a_k$, uniformly and independently from all subsets with the given sizes and let W denote the size of their intersection,

$$W = |\bigcap_{j=0}^{k} A_j|. \tag{1}$$

*AMS 2000 subject classifications. 60C05,62E17.
†Key words and phrases: Poisson approximation, Stein's method, size biasing, surprisology

The random variable W arises in a variety of contexts, signal detection and MELK, or Multi-Epitope Ligand Cartography, to name two. MELK, for instance, is a method for direct functional linkage analysis. In a sample of cell tissue, the intensity of a fluorescent marker is measured. The marker is attached to monoclonal antibodies binding specifically to various protein epitopes. A microscopic image of the sample is taken, and the intensity of fluorescence is measured for each pixel. Biologically relevant interactions between specified proteins are expected to show up in a surprisingly large, or surprisingly small, number of pixels at which these concentrations exceed some appropriately chosen threshold values simultaneously. This procedure for determining function has been coined, therefore, subset surprisology. For the choice of suitable thresholds, a distributional approximation of the count W gives information on how surprised one should be, or not be, to observe certain thresholds exceeded. The surprisology methodology was implicitly proposed in Schubert [9].

A Poisson approximation to W, and an asymptotic result, was proved in Dress *et al.* [5] using the inclusion-exclusion principle and some intricate calculations involving the complex form of the exact distribution of W. Using Poisson approximation, one replaces the exact but difficult to compute probabilities $P(W = a)$ by $P(Z = a)$, where Z has a Poisson distribution with the same mean, λ, as W. In particular, with λ given by the simple form (4), one approximates $P(W = a)$ by

$$P(Z = a) = e^{-\lambda}\frac{\lambda^a}{a!}, \quad a = 0, 1, \ldots,. \tag{2}$$

Proving an asymptotic result using the exact distribution of W requires manipulation of the so called Poisson numbers $A_{n|a_0,\ldots,a_k}(a)$, the number of ways subsets $A_j, j = 0, \ldots, k$ of $\{1, \ldots, n\}$, with sizes $a_0, \ldots, a_k$, respectively, can have an intersection of size a. The technique in [5] which justifies the Poisson approximation asymptotically through calculations involving the exact distribution does not give a bound on the error one incurs when approximating W by the Poisson. Here we provide an error bound which holds for all finite n, from which the asymptotic results of [5] follow. We also avoid working directly with the cumbersome form of $A_{n|a_0,\ldots,a_k}(a)$.

Stein's method and its variants provide powerful tools to bound the distance between a random variable W and a given target distribution. In particular, the Chen-Stein method provides explicit error bounds when approximating a random variable W by Z, having a Poisson distribution with parameter $\lambda = EW$. The method was first developed by Stein [10] for normal approximations; more details can be found in [11]. Chen [4] used Stein's ideas in the Poisson context. In Arratia *et al.* [1] Chen's approach was taken with the aim of providing easily computable bounds; Arratia *et al.* [2] provides a friendly introduction. The book [3] by Barbour, Holst and Janson gives an extensive overall treatment of the Poisson approximation method, of which a part is brought to bear here on the problem of interest. The result is not only a limiting Poisson asymptotic for W, but also a bound for the approximation error as measured in total variation distance, which can be explicitly computed as a function of the known parameters of the problem. The Chen-Stein method avoids ofttimes tedious direct calculations involving the exact distribution of W by relying instead on relatively simple coupling constructions, described in detail below for the case at hand.

It is well known that the Poisson distribution is a good model for counting the number of occurrences of rare events in an experiment with many trials (see Feller [6]), and W of (1) is seen to fit this mold by writing the size of the intersection as the sum of n indicators,

$$W = \sum_{\alpha=1}^{n} X_\alpha \quad \text{where} \quad X_\alpha = \mathbf{1}(\alpha \in \cap_{j=0}^{k} A_j). \tag{3}$$

That is, though it might be rare for an α to belong to all the sets, the experiment is being repeated many times if n is large. It is for this reason that we should be able to approximate W by a Poisson random variable Z of the same mean, λ. Since the sets $A_j, j = 0, \ldots, k$ are chosen independently, X_α is the product of independent indicators,

$$X_\alpha = \prod_{j=0}^{k} \mathbf{1}(\alpha \in A_j), \quad \text{and we have} \quad EX_\alpha = \prod_{j=0}^{k} \frac{a_j}{n}.$$

Hence $\lambda = EW = \sum_{\alpha=1}^{n} EX_\alpha$ is given by

$$\lambda = n \prod_{j=0}^{k} \frac{a_j}{n} = n^{-k} \prod_{j=0}^{k} a_j. \tag{4}$$

Though the indicators which comprise the sum W are not independent, the Chen-Stein method handles many forms of dependence through its use of coupling, and characterizing equations. In particular, a random variable Z has the Poisson distribution with parameter $\lambda > 0$ if and only if

$$EZf(Z) = \lambda Ef(Z+1) \tag{5}$$

for all functions f for which the above expectations exists. The characterizing equation (5) is reminiscent of the size bias operation. In particular, for any non-negative random variable X with finite mean $\mu > 0$, we say that X^* has the X-size biased distribution if, for all functions f for which $EXf(X)$ exists,

$$EXf(X) = \mu Ef(X^*). \tag{6}$$

The size bias distribution X^* is that of X weighted by the size of X itself, for which reason it appears in sampling; for instance, it is twice as likely to call a residence with two phone lines than a residence with only one, when dialing numbers at random. For $X \in \{0, 1, \ldots, \}$ with finite mean we can write the distribution of X^* explicitly as

$$P(X^* = a) = \frac{aP(X = a)}{EX}, \quad a = 0, 1 \ldots, .$$

We note in particular that$P(X^* = 0) = 0$. Hence, to size bias a random variable $X \in \{0, 1\}$ with $P(X = 1) > 0$, we must have $X^* = 1$. For discussion of the relation between size biasing and Stein's method, and further constructions such as the one we will apply below, see Goldstein and Rinott [8] and Goldstein and Reinert [7].

We can restate the characterizing equation (5) for the Poisson in terms of size biasing, using equation (6), as follows: Z has a Poisson distribution if and only if $Z^* = Z + 1$, in distribution, or equivalently, if and only if $Z^* - 1$ and Z have the same distribution. Therefore, it makes sense that $W^* - 1$ is close to W if and only if the distribution of W is close to Poisson.

One way to study the proximity of $W^* - 1$ to W is by coupling W and W^* together on the same space. When

$$W = \sum_{\alpha=1}^{n} X_\alpha$$

is the sum of independent, non-negative variables X_α with mean $EX_\alpha = p_\alpha$, a coupling of W to W^* can be accomplished easily as follows. First, choose an index α with probability proportional to p_α. Then replace X_α by a variable X_α^α having the X_α size biased distribution, and which is independent of $X_\beta, \beta \neq \alpha$ (see, for example, [8].) For a sum of dependent variables, the procedure is nearly the same. Choose an index α and replace X_α by X_α^α as before. Then, adjust the remaining variables for $\beta \neq \alpha$ to yield variables X_β^α with the conditional distribution of the original variables X_β, conditioned on the X_α variable taking on its newly chosen value X_α^α. That is, with

$$W^\alpha = \sum_{\beta=1}^{n} X_\beta^\alpha, \quad \alpha = 1, 2, \ldots, n, \tag{7}$$

one constructs the size bias variable W^* by setting it equal to W^α with probability proportional to p_α. This construction often leads to a tractable situation, even in the presence of dependence, which allows computation of distances, or bounds on distances, between W and $W^* - 1$, which then can be translated into distances between W and Z, its Poisson approximant.

Distance between the distributions of non-negative integer valued random variables X and Y can be measured by the total variation distance, defined by

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{a=0}^{\infty} |P(X = a) - P(Y = a)|, \tag{8}$$

or by using its alternate forms expressed in terms of functions f, and 'worst case' sets B,

$$d_{TV}(X, Y) = \sup_{0 \leq f \leq 1} |Ef(X) - Ef(Y)| = \sup_{B \subset \{0,1,\ldots\}} |P(X \in B) - P(Y \in B)|.$$

Hence, focusing on the last formulation, when using the distribution of Y to approximate the probability that X lies in any set, one cannot make an error greater than $d_{TV}(X, Y)$.

The intuition that W is close to Poisson when it is close to $W^* - 1$ is made rigorous in the following result, which combines Theorem 1.B, p.11, of [3], with Lemma 1.1.1, p.7 *ibid.* and adapts it to the structure in our special case.

Theorem 1.1 *Let $W = \sum_{\alpha=1}^n X_\alpha$ be the sum of indicator variables X_α, and W^* be defined on the same space as W with the W size biased distribution,* *constructed by choosing W^α as in (7) with probability proportional to EX_α. Then*

$$\begin{aligned} d_{TV}(W,Z) &\le \left(1-e^{-\lambda}\right) E|W-(W^*-1)| \\ &\le \min(1,\lambda) E|W-(W^*-1)|. \end{aligned}$$

We now apply the construction given above to couple W of (1) to a size biased W^*, and then use Theorem 1.1 to obtain a total variation bound between W and Z.

2 Total Variation Bound

Use of a coupling construction, along with Theorem 1.1, yields the following result.

Theorem 2.1 *Let W be given by (1) and let Z be a Poisson random variable with mean $\lambda = EW$, given in (4). Then*

$$\begin{aligned} d_{TV}(W,Z) &\le \left(1-e^{-\lambda}\right)\lambda\left\{1-\left(1+\frac{1}{n-1}\right)^k \prod_{j=0}^{k}\left(1-\frac{1}{a_j}\right)\right\} && (9) \\ &\le \min(\lambda,\lambda^2)\left\{1-\left(1+\frac{1}{n-1}\right)^k \prod_{j=0}^{k}\left(1-\frac{1}{a_j}\right)\right\} && (10) \\ &\le c_{\underline{a}} \min(\lambda,\lambda^2)\sum_{j=0}^{k}\frac{1}{a_j}, && (11) \end{aligned}$$

where

$$c_{\underline{a}} = -\underline{a}\log\left(1-\frac{1}{\underline{a}}\right) \quad \text{and} \quad \underline{a} = \min_{0\le j\le k} a_j.$$

Proof: A coupling of W, given in its summation form (3), to a variable W^* with the W size biased distribution can be achieved as follows. First, select an index α with probability proportional to EX_α; since EX_α is the same for all α, we simply pick α with uniform probability. Next, we need to size bias X_α and adjust the remaining variables to have their original distribution, conditioned on X_α taking on its newly chosen value.

Since X_α is either zero or one, size biasing it amounts to setting it equal to one to give $X_\alpha^\alpha = 1$. In terms of a new configuration, α needs to be contained in all the sets. If $\alpha \in A_j$ already, we set $A_j^\alpha = A_j$. If $\alpha \notin A_j$, we select an element of A_j, say β_j, uniformly with probability $1/a_j$, remove it from A_j and replace it by α. That is, we set

$$A_j^\alpha = \begin{cases} A_j & \alpha \in A_j \\ (A_j \setminus \{\beta_j\}) \cup \{\alpha\} & \alpha \notin A_j. \end{cases}$$

Now, for the remaining variables, let

$$X_\beta^\alpha = \prod_{j=0}^{k} \mathbf{1}(\beta \in A_j^\alpha),$$

and W^α be given by (7). Since $X_\alpha^\alpha = 1$ by construction, we have

$$W^\alpha = 1 + \sum_{\beta \neq \alpha} X_\beta^\alpha. \tag{12}$$

By (3) and (12),

$$W - (W^\alpha - 1) = X_\alpha + \sum_{\beta \neq \alpha} (X_\beta - X_\beta^\alpha).$$

By construction, for $\beta \neq \alpha$, we have $X_\beta \geq X_\beta^\alpha$ since β might have been removed from one of the sets $A_j, j = 0, \ldots, k$ to make room for α. Therefore $W - (W^\alpha - 1) \geq 0$, and

$$\begin{aligned} E|W - (W^\alpha - 1)| &= EX_\alpha + \sum_{\beta \neq \alpha} E(X_\beta - X_\beta^\alpha) \\ &= p_\alpha + \sum_{\beta \neq \alpha} p_\beta - \sum_{\beta \neq \alpha} EX_\beta^\alpha \\ &= \lambda - \sum_{\beta \neq \alpha} EX_\beta^\alpha. \end{aligned}$$

Using independence to calculate EX_β^α we find

$$EX_\beta^\alpha = \prod_{j=0}^{k} E\mathbf{1}(\beta \in A_j^\alpha) = \prod_{j=0}^{k} \frac{a_j - 1}{n - 1}.$$

The last equality holds because A_j^α has a_j elements, but $\alpha \in A_j^\alpha$ always, while the other $a_j - 1$ members of A_j^α appear uniformly from the remaining $n - 1$ elements.

Hence

$$
\begin{aligned}
E|W-(W^\alpha-1)| &= \lambda - \sum_{\beta\neq\alpha}\prod_{j=0}^{k}\frac{a_j-1}{n-1} \\
&= \lambda-(n-1)\prod_{j=0}^{k}\frac{a_j-1}{n-1} \qquad (13)\\
&= \lambda\left(1-\frac{(n-1)}{n}\frac{\prod_{j=0}^{k}\frac{a_j-1}{n-1}}{\prod_{j=0}^{k}\frac{a_j}{n}}\right) \\
&= \lambda\left\{1-\left(1+\frac{1}{n-1}\right)^k\prod_{j=0}^{k}\left(1-\frac{1}{a_j}\right)\right\}. \qquad (14)
\end{aligned}
$$

Since the last expression is constant in α, averaging over α with probabilities proportional to $EX_\alpha/EW=1/n$ to give W^* shows $E|W-(W^*-1)|$ is given by (14). Now Theorem 1.1 yields (9).

To achieve (10), use $1-e^{-\lambda}\leq\min(1,\lambda)$. Next, note that (14) is bounded above by

$$\lambda\left\{1-\prod_{j=0}^{k}\left(1-\frac{1}{a_j}\right)\right\}.$$

Since

$$1-x\geq e^{-c_{\underline{a}}x} \quad \text{for all } 0\leq x\leq 1/\underline{a},$$

we have

$$1-\prod_{j=0}^{k}\left(1-\frac{1}{a_j}\right)\leq 1-\exp\left(-c_{\underline{a}}\sum_{j=0}^{k}\frac{1}{a_j}\right)\leq c_{\underline{a}}\sum_{j=0}^{k}\frac{1}{a_j},$$

yielding (11).

Remark 2.1 *For $\underline{a}=2$ (only), $c_{\underline{a}}\leq 1.39$, while if $\underline{a}=10$, then $c_{\underline{a}}\leq 1.0537$, which is already very close to the asymptotic, minimal value of*

$$\lim_{\underline{a}\to\infty}-\underline{a}\log\left(1-\frac{1}{\underline{a}}\right)=1.$$

Theorem 2.1 allows us to recover Theorem 4.1 and Corollary 4.4 of Dress *et al.* [5] as an immediate consequence of the bound (11) and Remark 2.1.

Corollary 2.1 *Suppose the sizes* $a_{0,n}, \ldots, a_{k,n}$ *of the subsets* $A_0, \ldots, A_k$ *of* $\{1, \ldots, n\}$ *tend to infinity as* $n \to \infty$. *Then, if* $\lambda = \lambda_n$ *given in (4) satisfies*

$$\limsup_{n \to \infty} \lambda_n < \infty,$$

$d_{TV}(W, Z) \to 0$ *as* $n \to \infty$. *Hence, if* λ_n *converges to* λ_∞, *say, then* W *converges in distribution to a Poisson variable with parameter* λ_∞.

Remark 2.2 *If* $a_j = 1$ *for some* j *then* $A_j^\alpha = \{\alpha\}$, *and hence* $X_\beta^\alpha = 0$ *for all* $\beta \neq \alpha$. *In this case, since* $W \leq \underline{a}$ *always,* $W \in \{0, 1\}$ *and a Poisson approximation will not be valid. We note that in this case bound (9) becomes* $(1 - e^{-\lambda})\lambda$, *which may be greater than one.*

Remark 2.3 *It is not necessary in Corollary 2.1 for* $\limsup_{n \to \infty} \lambda_n < \infty$ *as* $n \to \infty$ *for* $d_{TV}(W, Z) \to 0$. *It suffices that* $\underline{a} \geq 2$ *and* $\lambda_n \sum_{j=0}^{k} 1/a_{j,n} \to 0$.

Remark 2.4 *Recalling (4) and letting*

$$\lambda_n = n^{-k} \prod_{j=0}^{k} a_j \quad \textit{and} \quad \mu_n = (n-1)^{-k} \prod_{j=0}^{k} (a_j - 1)$$

be the intensity for the given problem, and a reduced one where the individual n *is removed from* $\{1, \ldots, n\}$ *and the sizes of all sets* A_j *are reduced by one, in view of (13), Theorem (2.1) can be stated compactly as*

$$d_{TV}(W, Z) \leq (1 - e^{-\lambda_n})(\lambda_n - \mu_n).$$

Remark 2.5 *The bound (9) is valid for all values of* k, *including* $k > n$, *and, for a given sequence* $a_0, a_1, \ldots$, *decreases as* k *increases. But even for small* k, *such as* $k = 1$, *the Poisson approximation may be valid. In particular, if* a_0 *and* a_1 *increase to infinity at a rate such that*

$$\lim_{n \to \infty} \frac{a_0}{n^{1/2}} = c_0 \quad \textit{and} \quad \lim_{n \to \infty} \frac{a_1}{n^{1/2}} = c_1,$$

we have

$$\lambda \to c_0 c_1 \quad \textit{and} \quad \sum_j \frac{1}{a_j} = O(\frac{1}{\sqrt{n}}) \to 0,$$

which imply, by Corollary 2.1, that W *has an asymptotic Poisson distribution with parameter* $c_0 c_1$.

Remark 2.6 *The framework here, and the Poisson approximation, could be further generalized. With some additional effort, we could consider the situation where A_j is a set chosen with a non-uniform distribution from all subsets of size a_j of $\{1, \ldots, n\}$. The size bias construction generalizes in a straightforward way.*

In addition, one could derive a Poisson point process approximation for the point process

$$\Xi = \sum_{\alpha=1}^{n} \delta_\alpha X_\alpha;$$

here δ_α denotes point mass at α. To this purpose one may employ the analog of Theorem 1.1 for point processes, namely Theorem 10.B, p.211, in [3], yielding a point process version of our Theorem 2.1. As Ξ contains information not only on the number of points (pixels) that are present in all the sets, but tells us also which ones these pixels are, such a result could make even better use of spatial information in MELK type data.

3 Numerical Comparison

Below is an extract of the table in [5], comparing the actual values $P(W = a)$ to their Poisson approximants $P(Z = a)$, for various choices of the parameters of the problem, and a. We have augmented the table by including the bound (9) on the total variation distance between W and Z, denoted by Bound, and, for $n \leq 10000$, the actual total variation distance, denoted by TV.

n	$k+1$	a_0	a_1	a_2	a	λ	$P(Z=a)$	$P(W=a)$	TV	Bound
1000	3	100	100	100	3	1	0.0613	0.0604	0.0078	0.0175
100	2	10	10		0	1	0.3679	0.3305	0.0577	0.1149
10000	2	100	100		3	1	0.0613	0.0607	0.0055	0.0125
1000000	2	1000	1000		3	1	0.0613	0.0613		0.0012
1000	2	60	50		5	3	0.1008	0.1022	0.0280	0.1008
1000	2	100	30		0	3	0.0498	0.0403	0.0336	0.1198

The probabilities $P(W=a)$ are calculated using the exact formula in [5],

$$P(W=a)=\frac{A_{n|a_0,\ldots,a_k}(a)}{\prod_{j=0}^{k}\binom{n}{a_j}}$$

where

$$A_{n|a_0,\ldots,a_k}(a)=\binom{n}{a}\sum_{a'=a}^{n}\binom{n-a}{a'-a}(-1)^{a'-a}\prod_{j=0}^{k}\binom{n-a'}{a_i-a'}. \qquad (15)$$

As (15) is complex, it is clear, therefore, why having a good and easily computable approximation such as (2) can be of value. Moreover, as the total variation distance TV in (8) is a sum over all differences $|P(W=a)-P(Z=a)|$, each of which involves the exact probabilities, calculating TV for the larger values of n can become rather cumbersome. Nevertheless, the upper bound, Bound (9), can be easily calculated in all cases.

The total variation distance (8), and hence any upper bound to it, may naturally exceed the difference between the true and approximated probabilities at any particular value. We note, however, that the total variation and its bound are nevertheless of a similar order of magnitude as the observed differences, and, moreover, that the bound on TV is not far from the actual value. Hence if the set sizes $a_j, j=0,\ldots,k$, are such that the upper bound is small, being easily computable, it can be usefully incorporated to obtain conservative and non-approximate error estimates when approximating W by Z.

4 Acknowledgement

The authors gratefully acknowledge Mike Waterman, for bringing this problem to their attention, and Ester Samuel-Cahn for improvements on an earlier version of the manuscript.

References

[1] Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximation: the Chen-Stein method. *Annals of Probability.* **17**, 9–25.

[2] Arratia, R., Goldstein, L., and Gordon, L. (1990). Poisson approximation and the Chen-Stein method. *Statistical Science*, **5**, 403–434.

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

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

[5] Dress, A., Lokot,T., Pustyl'nilov, L.D. and Schubert, W. (2004). Poisson numbers and Poisson distribution in subset surprisology. *Annals of Combinatorics*, **8**, 473-485.

[6] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications.* Vol.1, third edition, Wiley, New York.

[7] Goldstein, L., and Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *Journal of Theoretical Probability* **18**, 185–208.

[8] Goldstein, L., and Rinott, Y. (1996). On multivariate normal approximations by Stein's method and size bias couplings. *Journal of Applied Probability* **33**, 1–17.

[9] Schubert, W. (2003). Topological proteomics, toponomics, MELK-technology. In *Proteomics of Microorganisms: Fundamental Aspects and Application. Advances in Biochemical Engineering/Biotechnology*, **83**, M. vHecker, S.Müllner, eds. Springer Verlag, Berlin, Heidelberg, New York, 189–209.

[10] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583-602, Univ. California Press, Berkeley.

[11] Stein, C. (1986). *Approximate Computation of Expectations.* IMS, Hayward, CA.