Le Cam's Inequality and Poisson Approximations

J. Michael Steele

1. INTRODUCTION. For the sum S_n of n independent, non-identically distributed Bernoulli random variables X_i with $P(X_i = 1) = p_i$, Le Cam **[20]** established the remarkable inequality.

$$\sum_{k=0}^{\infty} \left| P(S_n = k) - e^{-\lambda}\lambda^k/k! \right| < 2 \sum_{i=1}^{n} p_i^2, \tag{1.1}$$

where $\lambda = p_1 + p_2 + \cdots + p_n$.

Naturally, this inequality contains the classical Poisson limit law (just set $p_i = \lambda/n$ and note that the right side simplifies to $2\lambda^2/n$), but it also achieves a great deal more. In particular, Le Cam's inequality identifies the sum of the squares of the p_i as a quantity governing the quality of the Poisson approximation.

Le Cam's inequality also seems to be one of those facts that repeatedly calls to be proved—and improved. Almost before the ink was dry on Le Cam's 1960 paper, an elementary proof was given by Hodges and Le Cam **[18]**. This proof was followed by numerous generalizations and refinements including contributions by Kerstan **[19]**, Franken **[15]**, Vervatt **[30]**, Galambos **[17]**, Freedman **[16]**, Serfling **[24]**, and Chen **[11, 12]**. In fact, for raw simplicity it is hard to find a better proof of Le Cam's inequality than that given in the survey of Serfling **[25]**.

One purpose of this note is to provide a proof of Le Cam's inequality using some basic facts from matrix analysis. This proof is simple, but simplicity is not its *raison d'etre*. It also serves as a concrete introduction to the semi-group method for approximation of probability distributions. This method was used in Le Cam **[20]**, and it has been used again most recently by Deheuvels and Pfeifer **[13]** to provide impressively precise results.

The semi-group method is elegant and powerful, but it faces tough competition, especially from the coupling method and the Chen-Stein method. The literature of these methods is reviewed, and it is shown how they also lead to proofs of Le Cam's inequality.

2. MATRIX PROOF OF LE CAM'S INEQUALITY. If one is charged with the task of producing matrices that might help in understanding the distribution of the sum of n independent non-identically distributed Bernoulli random variables, a little time and thought is likely to lead to n matrices P_i like the $N \times N$ matrix

$$P_i = \begin{pmatrix} 1-p_i & p_i & 0 & \cdots & 0 & 0 \\ 0 & 1-p_i & p_i & \cdots & 0 & 0 \\ 0 & 0 & 1-p_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-p_i & p_i \\ 0 & 0 & 0 & \cdots & 0 & 1-p_i \end{pmatrix}. \tag{2.1}$$

This is almost a Markov transition matrix, except of course the last row of P_i does not sum to 1. A benefit of this choice of the P_i is that they can be written as

$$P_i = (1 - p_i)I + p_i R,$$

where I is the $N \times N$ identity matrix and R is the $N \times N$ matrix with $1's$ on the first superdiagonal and $0's$ elsewhere. Since each of the $P_i's$ is just a linear combination of I and R, any pair of the P_i commute, and because the matrices P_i are so much like Markov transition matrices, their analysis is still reminiscent of the elementary theory of Markov chains.

In fact, by the usual considerations that attend the multiplication of Markov matrices, you can quickly convince yourself that for $n < N$ the top row of the n-fold matrix product $P_1 P_2 P_3 \cdots P_n$ is given by $(P(S_n = 0), P(S_n = 1), P(S_n = 2), \ldots, P(S_n = n), 0, 0, \ldots, 0)$, i.e. the first $n + 1$ elements of the top row of $P_1 P_2 \ldots P_n$ correspond precisely to the Bernoulli sum probabilities that we wish to estimate. Also at this point, it may be good to be reminded that N is arbitrary except for the constraint $n < N$, so the padded 0's can go on as far as we like.

So far, we have found a matrix that helps us understand the Bernoulli sum probabilities $P(S_n = k)$, and now we would like to find a matrix that is intimately connected with the Poisson distribution. Given some past experience with calculating matrix functions using the Jordan normal form, one can easily find candidates, but knowledge of Jordan forms is not required. One just needs to compute the exponential of P_i, or, better yet, compute the exponential of a simpler matrix closely connected with P_i.

When we write $P_i = I + Q_i$, we see Q_i has the pleasing form,

$$\begin{pmatrix} -p_i & p_i & 0 & \cdots & 0 & 0 \\ 0 & -p_i & p_i & \cdots & 0 & 0 \\ 0 & 0 & -p_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -p_i & p_i \\ 0 & 0 & 0 & \cdots & 0 & -p_i \end{pmatrix} = -p_i I + p_i R, \tag{2.2}$$

and the Poisson distribution emerges clearly when we compute $\exp Q_i$:

$$\sum_{k=0}^{\infty} Q_i^k / k! = \begin{pmatrix} e^{-p_i} & p_i e^{-p_i} & & \cdots & & e^{-p_i} p_i^{N-1}/(N-1)! \\ 0 & e^{-p_i} & p_i e^{-p_i} & \cdots & & \\ 0 & 0 & e^{-p_i} & \cdots & & \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & e^{-p_i} & p_i e^{-p_i} \\ 0 & 0 & 0 & \cdots & 0 & e^{-p_i} \end{pmatrix}$$

$$= \sum_{r=0}^{N-1} p_i^r e^{-p_i} R^r / r!. \tag{2.3}$$

Note that the Q_i commute, so $\exp(Q_i)\exp(Q_j) = \exp(Q_i + Q_j)$, and, in detail,

$$\prod_{i=1}^{n} \exp(Q_i) = \exp\left(\sum_{i=1}^{n} Q_i \right) = \exp(-\lambda I + \lambda R), \tag{2.4}$$

where $\lambda = p_1 + p_2 + \ldots + p_n$.

The essence of the method is now fully revealed, and we see the proof of Le Cam's inequality boils down to comparing the top rows of $\prod_{i=1}^{n} \exp Q_i$ and $\prod_{i=1}^{n} P_i$. This can be achieved most systematically by introducing matrix norms.

If $A = (a_{ij})$ is any matrix, we set

$$\|A\| = \max_i \sum_j |a_{ij}|. \tag{2.5}$$

This recipe provides a *bona fide* matrix norm, and, in particular, one can easily check the relations $\|AB\| \leq \|A\|\,\|B\|$, $\|A + B\| \leq \|A\| + \|B\|$, and $\|cA\| = c\|A\|$ for $c > 0$.

It is also easy to use the explicit formulas for P_i, Q_i and $\exp Q_i$ to compute their norms: $\|P_i\| = 1$, $\|Q_i\| = 2p_i$, and $\|\exp Q_i\| \leq 1$. When we compare $\prod P_i$ and $\prod \exp Q_i$ using the norm defined by (2.5) we see that the top row attains the maximum, so we have the basic relation,

$$\sum_{k=1}^{N-1} \left|P(S_n = k) - e^{-\lambda}\lambda^k/k!\right| = \left\|\prod_{i=1}^{n} P_i - \prod_{i=1}^{n} \exp Q_i\right\|. \tag{2.6}$$

Next, it is easy to check that

$$P_1 \cdots P_n - \exp Q_1 \cdots \exp Q_n = (P_1 - \exp Q_1)(P_2 \cdots P_n) - (\exp Q_1)(\exp Q_2 \cdots \exp Q_n - P_2 \cdots P_n). \tag{2.7}$$

This identity virtually completes the proof. We just take norms, use the facts that $\|P_2 \cdots P_n\| \leq 1$ and $\|\exp Q_1\| \leq 1$, then repeat the process on the remaining $(n - 1)$-fold product to obtain

$$\begin{aligned}\|P_1 \cdots P_n - \exp Q_1 \cdots \exp Q_n\| &\leq \|P_1 - \exp Q_1\| \\ &\quad + \|\exp Q_2 \cdots \exp Q_n - P_2 \cdots P_n\| \\ &\leq \sum_{i=1}^{n} \|P_i - \exp Q_i\|. \end{aligned} \tag{2.8}$$

How should we bound $\|P_i - \exp Q_i\|$? Since $\exp Q_i$ is defined by the expansion for e^x, we naturally look to Taylor's formula, but we should be careful enough to consider a finite expansion with a remainder term. For any smooth function f we have

$$f(1) = f(0) + f'(0) + \int_0^1 (1 - u) f''(u)\, du, \tag{2.9}$$

so, if we let $f(t) = e^{tQ}$, the derivative calculations $f(0) = I$, $f'(0) = Q$ and $f''(u) = Q^2 e^{uQ}$ yield

$$e^Q = I + Q + \int_0^1 (1 - u) Q^2 e^{uQ}\, du. \tag{2.10}$$

Even for functions of matrices, integrals are just the limits of sums, so taking norms inside an integral only makes it larger, i.e. for any $g(u, v)$ we have $\|\int g(u, Q)du\| \leq \int \|g(u, Q)\|\, du$. Also, just as we computed e^{Q_i} explicitly in order to bound its norm, we can compute e^{uQ_i} explicitly to find $\|e^{uQ_i}\| \leq 1$. Applying these observations to the Taylor representation (2.10), we find

$$\|P_i - e^{Q_i}\| \leq \left\|\int_0^1 (1 - u) Q_i^2 e^{uQ_i}\, du\right\| \leq \tfrac{1}{2}\|Q_i^2\|. \tag{2.11}$$

This is just the tool needed to bound the right side of (2.8). Stringing together the identity (2.6) with inequalities (2.8) and (2.11), we find

$$\sum_{k=0}^{\infty} \left| P(S_n = k) - e^{-\lambda}\lambda^k/k! \right| \le \tfrac{1}{2} \sum_{i=1}^{n} \|Q_i^2\| \le 2 \sum_{i=1}^{n} p_i^2, \tag{2.12}$$

where the last inequality depended on $\|Q_i^2\| \le \|Q_i\|^2$ and our earlier explicit calculation that $\|Q_i\| = 2p_i$. Since $n < N$ is our only restriction on N, we can let $N \to \infty$ to obtain Le Cam's inequality.

3. THE SEMI-GROUP METHOD HAS VIGOROUS COMPETITORS. Deheuvels and Pfeifer **[13]** provide a version of Le Cam's inequality that—in an asymptotic sense—has a solid claim on being the last word:

$$\sum_{k=0}^{\infty} \left| P(S_n = k) - e^{-\lambda}\lambda^k/k! \right| \sim \sqrt{2/\pi e} \left\{ \sum_{i=1}^{n} p_i^2 \right\} \Big/ \left\{ \sum_{i=1}^{n} p_i \right\} \tag{3.1}$$

provided $\sum_{i=1}^{\infty} p_i \to \infty$ *and* $\max(p_1, p_2, \dots, p_n) \to 0$ as $n \to \infty$. The essential ideas behind the proof of (3.1) have been seen in Section 2 in a basic form: one obtains an interpretation of the Bernoulli sum probabilities, introduces a semi-group (like $\exp(tQ)$), finds ways to bound an approximation (like $e^Q - I - Q$), and deals with the difference of two n-fold products. Variations on this pattern are visible in Le Cam **[20]**, Shur **[26]**, and one even can see similar steps in Feller's exposition of Trotter's proof of Lindeberg's central limit theorem.

Through the explicit matrix exponentiation calculations used here, the semi-group method can be seen to be friendly as well as useful. Continued exploration of the method is likely to lead to deep and interesting results, but the semi-group method should not be oversold. There are competitors with considerable power.

The characteristic function method also has a role in Poisson approximations. In particular, the characteristic function method has been used by Rusenko **[23]** to obtain rates of approximation results for repeated samples taken without replacement, by Presman **[22]** to obtain refinements of Le Cam's inequality, and by Yakshyavicius **[32]** to provide an inequality like Le Cam's that is pertinent to classes of discrete distributions other than Bernoulli sums. Despite this litany, applications of the characteristic function method are infrequent in Poisson approximation, and it probably does not rank among the big three: the semi-group method, coupling, and the Chen-Stein method. Enough has been said about the first of these, and it is important to provide some sense of the promise inherent in the other two.

4. THE COUPLING METHOD. Because of its simplicity, the coupling method deserves to be reviewed first. For any two random variables X and Y, we begin by defining their *variation distance* by

$$d(X,Y) = \sup_{A} |P(X \in A) - P(Y \in A)|. \tag{4.1}$$

For random variables that take values in $\mathbb{Z}^+$, the metric $d(\cdot,\cdot)$ has an easily proved alternative expression (cf. Serfling **[25]**, p. 569) that reveals its relevance to Le Cam's inequality:

$$d(X,Y) = \tfrac{1}{2} \sum_{k=0}^{\infty} |P(X = k) - P(Y = k)|. \tag{4.2}$$

The coupling method is based on the simple observation that for random variables

X and Y defined on the same probability space, one has

$$d(X,Y) \leq P(X \neq Y). \tag{4.3}$$

A second observation that helps the coupling method work well with sums is that for $S_n = \sum_{i=1}^n X_i$ and $S_n^* = \sum_{i=1}^n Y_i$ one has

$$d(S_n, S_n^*) \leq \sum_{i=1}^n d(X_i, Y_i). \tag{4.4}$$

From (4.2), (4.3), and (4.4) we see that a good plan for proving Le Cam's inequality consists of building n bivariate couples $Z_i = (X_i, Y_i)$ such that the Z_i are independent, X_i is Bernoulli with parameter p_i, Y_i is Poisson with parameter $\lambda_i = p_i$, and $P(X_i \neq Y_i)$ is as small as possible. This plan has been successfully pursued in Hodges and Le Cam **[18]**, Freedman **[16]**, Serfling **[24]**, Brown **[10]**, Ahmad **[1]**, and Wang **[31]**. In fact, the couplings of Serfling, Brown, and Wang all satisfy

$$d(X_i, Y_i) = P(X_i \neq Y_i) = p_i(1 - e^{-p_i}), \tag{4.5}$$

from which Le Cam's inequality (1.1) follows easily. As it happens, there is no difficulty in constructing variables that satisfy (4.5)—just think how to simulate X_i and Y_i simultaneously using a single uniformly distributed random variable.

5. THE CHEN-STEIN METHOD. The Chen-Stein method may be the most powerful method for obtaining Poisson approximations, and it is often as easy to use as the coupling or semi-group methods, even though it may be more subtle conceptually. If one does not stop for motivation, one can say that the Chen-Stein method is based on the fact that for each $\lambda > 0$ and $A \subset \mathbb{Z}^+$ there is a function $x = x_{\lambda, A}: \mathbb{Z}^+ \to \mathbb{R}$ such that for any non-negative integer-valued random variable T one has the identity:

$$E\{\lambda x(T+1) - Tx(T)\} = P(T \in A) - \sum_{k \in A} e^{-\lambda}\lambda^k/k!. \tag{5.1}$$

Actually, the left-hand side of (5.1) is a natural quantity to consider in the context of Poisson approximation, since by summation by parts one can check that $E\lambda f(T+1) = ETf(T)$ for any f, provided T is Poisson with parameter λ. The identity (5.1) was first developed by Chen **[11]**, and some of its mystery can be removed by studying an analogous identity used by Stein **[27]** in the context of normal approximations. While it is a good exercise to solve (5.1) for x, all one really needs to know about x is that it is bounded and changes slowly. In particular, Barbour and Eagleson **[5]** sharpened earlier bounds of Chen **[12]** and showed that for all A and $\lambda > 0$:

$$|x| \leq \min(1, 4\lambda^{-\frac{1}{2}}) \tag{5.2}$$

and

$$\Delta x = \sup_{m \geq 0} |x(m+1) - x(m)| \leq \lambda^{-1}(1 - e^{-\lambda}). \tag{5.3}$$

From these bounds it is easy to prove—and even sharpen—Le Cam's basic inequality (1.1). If we write $W = S_n$, $W_j = S_n - X_j$, $\lambda = p_1 + p_2 + \cdots + p_n$, and $q_j = 1 - p_j$, we can follow Chen **[12]** and obtain a second identity that together with (5.1) gives one virtually complete information about the Poisson approximation. We evaluate the left side of (5.1) as follows:

$$\begin{aligned}
E\{\lambda x(W+1) - Wx(W)\} &= \sum_{j=1}^{n} E\{p_j x(W+1) - X_j x(W)\} \\
&= \sum_{j=1}^{n} p_j E\{x(W+1) - x(W_j+1)\} \\
&= \sum_{j=1}^{n} p_j \{p_j Ex(W_j+2) \\
&\qquad + q_j Ex(W_j+1) - Ex(W_j+1)\} \\
&= \sum_{j=1}^{n} p_j^2 E\{x(W_j+2) - x(W_j+1)\}. \qquad (5.4)
\end{aligned}$$

From the Chen-Stein identity (5.1) and the Barbour-Eagleson bound on Δx, we see that (5.4) gives

$$\sup_A \left| P(S_n \in A) - e^{-\lambda} \sum_{k \in A} \lambda^k / k! \right| \le \lambda^{-1}(1 - e^{-\lambda}) \sum_{j=1}^{n} p_j^2. \qquad (5.5)$$

Since $\lambda^{-1}(1 - e^{-\lambda}) \le 1$, the identity (4.2) shows that inequality (5.5) is sharper than (1.1). Obviously, the Chen-Stein method is very powerful, though it is only now beginning to be well understood. A richer understanding of the method can be obtained by studying Arratia, Goldstein, and Gordon **[2]**, Barbour **[3]**, Barbour and Eagleson **[5, 6, 7]**, Barbour and Hall **[8]**, Barbour **[4]**, and, of course, Stein **[28]**. A definitive study of Stein's method and its application to Poisson approximation has recently been given in the volume by Barbour, Holst, and Janson **[9]**.

6. CONCLUSION. Le Cam's inequality provides information on the quality of the Poisson approximation, but it also serves as a talisman that is able to charm concrete insights from general techniques. This survey relied on that second service to illustrate the semi-group method, coupling, and the Chen-Stein method. In the course of these illustrations, it has also been possible to survey most of the work on Poisson approximation since the review of Serfling **[25]**, except for the cascade of work coming from the more refined developments of the Chen-Stein method that are dealt with in detail in the monograph of Barbour, Holst, and Janson **[9]**.

REFERENCES

1. Ahmad, I. A. (1985). On the Poisson approximation of multinomial probabilities, *Statist. Probab. Lett.*, 55–56.
2. Arratia, R., Goldstein, L., and Gordon, L. (1990). Two moments suffice for Poisson approximations: the Chen-Stein method *Statistical Science*, 5, 403–434.
3. Barbour, A. D. (1982). Poisson convergence and random graphs, *Math. Proc. Camb. Phil. Soc.*, 92, 349–359.
4. Barbour, A. D. (1987). Asymptotic expansions in the Poisson limit theorem, *Ann. Probab.*, 15, 748–766.
5. Barbour, A. D. and Eagleson, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials, *Adv. Appl. Probab.*, 15, 585–600.
6. Barbour, A. D. and Eagleson, G. K. (1984). On the rate of Poisson convergence, *Math. Proc. Camb. Phil. Soc.*, 95, 473–480.
7. Barbour, A. D. and Eagleson, G. K. (1986). Random association of symmetric arrays, *Stoch. Anal. Appl.*, 4(3), 239–281.

8. Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence, *Math. Proc. Camb. Phil. Soc.*, 95, 473–480.
9. Barbour, A. D., Holst, L., and Janson, S. (1992) *Poisson Approximation*, Oxford University Press, New York, NY.
10. Brown, T. C. (1984). Poisson approximation and the definitions of the Poisson process, *this MONTHLY*, 91, 116–123.
11. Chen, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions, *Ann. Probab.*, 2, 178–180.
12. Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.*, 3, 534–545.
13. Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation, *Ann. Probab.*, 14, 663–676.
14. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Volume II, John Wiley and Sons: New York, NY.
15. Franken, P. (1964). Approximation des Verteilungen von Summen unabhängiger nichtnegativer ganzzahler Zufallsgrössen durch Poissonsche Verteilungen, *Math. Nachr.*, 23, 303–340.
16. Freedman, D. (1974). The Poisson approximation for dependent events, *Ann. Probab.*, 2, 256–269.
17. Galambos, J. (1973). A general Poisson limit theorem of probability theory, *Duke J. Math.*, 40, 581–586.
18. Hodges, S. L. and Le Cam, L. (1960). The Poisson approximation to the binomial distribution, *Ann. Math. Statist.*, 31, 737–740.
19. Kerstan, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam, *Z. Wahrsch. Verw. Gebiete*, 2, 173–179.
20. Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.*, 10, 1181–1197.
21. Le Cam, L. (1963). On the distribution of sums of independent random variables, *Bernoulli, Bayes, Laplace: Proceeding of an International Research Seminar* (J. Neyman and L. M. Le Cam, eds.), Springer: New York, NY, 179–202.
22. Presman, E. L. (1985). Approximation in variation of the distribution of a sum of independent Bernoulli variables with a Poisson law, *Theory Probab. Appl.*, 30, 417–422.
23. Rusenko, N. (1985). On the approximation of the distribution of some statistics by the Poisson laws, *Lith. Math. J.*, 25(4), 363–370.
24. Serfling, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.*, 3, 726–731.
25. Serfling, R. J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials, *SIAM Rev.*, 20, 567–579.
26. Shur, M. G. (1984). The Poisson theorem and Markov chains, *Theory Probab. Appl.*, 29(1), 124–126.
27. Stein, C. (1970). 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.* Vol. 2, 583–602, University of California Press.
28. Stein, C. (1986). *Approximate Computation of Expectations*, Inst. Math. Statist. Lecture Notes—Monograph Series Vol. 7, Hayward, California.
29. Trotter, H. F. (1959). An elementary proof of the central limit theorem, *Arch. Math.*, 10, 226–234.
30. Vervaat, W. (1969). Upper bounds for the distance in total variation between the binomial and the Poisson distribution, *Statist. Neerlandica*, 23, 79–86.
31. Wang, Y. H. (1986). Coupling methods in Poisson approximation, *Canadian J. Statist.*, 14, 69–74.
32. Yakshyavicius, S. V. (1986). Poisson approximation of the binomial, Pascal, negative binomial and geometric distributions (in Russian), *Litovsk. Mat. Sb.*, 26(1), 165–184.

Department of Statistics
Wharton School
University of Pennsylvania
Philadelphia, PA 19104
steele@wharton.upenn.edu

Answer to Who Was the Author:
(p. 14)
Emmy Noether.