Anders Hald

Nicholas Bernoulli's Theorem

Preprint
January
1983
2

Institute of Mathematical Statistics
University of Copenhagen

Anders Hald

NICHOLAS BERNOULLI'S THEOREM

Preprint 1983 No. 2

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

January 1983

Summary.

In 1713 Nicholas Bernoulli derived a much improved version of James Bernoulli's theorem. The significance of this contribution has been overlooked. Nicholas Bernoulli's theorem is the "missing link" between James Bernoulli's theorem and de Moivre's normal approximation to the binomial.

Key words.

History of probability theory. Approximations to the binomial. James and Nicholas Bernoulli. De Moivre.

Introduction.

Occasioned by a discussion of the variations in the ratio of male to female births in London Nicholas Bernoulli wanted to test the hypothesis that the probability of the birth of a male is to the probability of the birth of a female as 18 to 17. He therefore needed an approximation to the tail probability of the binomial better than the one provided by James Bernoulli's theorem. He succeeded in finding a much improved version of the theorem and gave his result in a letter of January 23, 1713 to Montmort (1713, pp. 388-394). In the same letter he informed Montmort that the Ars Conjectandi was in the press at Basel.

Nicholas Bernoulli's result has been overlooked, perhaps because two otherwise reliable witnesses, viz. de Moivre and Todhunter, did not grasp the significance of his result. De Moivre (1730, pp. 96-99) gave a precise account of both theorems with the original examples but without proofs. He does not compare his own result with Nicholas Bernoulli's. Three years later de Moivre (1733; 1738, p. 235; 1756, p. 243) wrote about the results of James and Nicholas Bernoulli that "what they have done is not so much an Approximation as the determining of very wide limits, within which they demonstrated that the Sum of the Terms was contained". This is true for James Bernoulli but as we shall see not for Nicholas Bernoulli. Todhunter (1865, p. 131) does not give Nicholas Bernoulli's result. He writes " His investigation involves a general demonstration of the theorem of his uncle James called Bernoulli's theorem" and furthermore "The whole investigation bears some resemblance to that of James Bernoulli and may have been suggested by it ...". The last statement is certainly true as also certified by Nicholas Bernoulli in his letter.

We shall give a summary and a comparison of the two proofs and show that Nicholas Bernoulli's result is the "missing link" between the result of James Bernoulli and de Moivre's derivation of the normal distribution as an approximation to the binomial.

We shall use the following definitions and notations essentially due to James Bernoulli. Consider a trial with $t = r+s$ equally possible outcomes, and let r be the number of favourable outcomes, so that $p = r/t$ equals the probability of success. Consider $n = kt$ independent trials, let x be the number of successes and denote the probability of a deviation from np of at most d by $P_d = \Pr\{ |x-np| \leqq d \}$. The numbers r,s,k and d are supposed to be positive integers. Notice that $np = kr$ and that

$$P_k = \Pr\{ |x-np| \leqq k \} = \Pr\{ |(x/n) - p| \leqq t^{-1}\}.$$

For given p we may chose t as large as we like.

Since $(p+q)^n = (r+s)^n t^{-n}$ we need only consider the terms of $(r+s)^n$. We shall set

$$(r+s)^n = \sum_0^n \binom{n}{x} r^x s^{n-x} = \sum_{-kr}^{ks} f_i ,$$

say, where

$$f_i = \binom{kr+ks}{kr+i} r^{kr+i} s^{ks-i} , \quad i = -kr, -kr+1, \ldots, ks.$$

As f_{-i}, $i = 0,1,\ldots,kr$, is obtained from f_i, $i = 0,1,\ldots,ks$, by interchange of r and s, results proved for f_i, $i \geqq 0$, also hold for f_{-i}, $i \geqq 0$.

Comparing the central part of the series with the tails we shall prove that

$$P_k = f_0 + \sum_1^k f_i + \sum_1^k f_{-i} > c\{ \sum_{k+1}^{ks} f_i + \sum_{k+1}^{kr} f_{-i} \} = c(1-P_k) \quad (1)$$

for any given $c > 0$ and k sufficiently large. (If $r = 1$ or $s = 1$ the corresponding sum on the right side is empty and should be put equal to zero.) To prove (1) it is sufficient to prove that

$$\sum_1^k f_i \geqq c \sum_{k+1}^{ks} f_i \qquad \text{for } k \geqq k(r,s,c). \qquad (2)$$

From

$$f_i/f_{i+1} = \frac{(kr+i+1)s}{(ks-i)r} > 1, \quad i = 0,1,\ldots, ks-1, \tag{3}$$

it follows that

(a) f_i is a decreasing function of i for $i \geqq 0$,

(b) $f_0 = \max\{f_i\}$,

(c) f_i/f_{i+1} is an increasing function of i for $i \geqq 0$,

(d) $f_0/f_k < f_i/f_{k+i}$ for $i \geqq 1$.

These results are due to James Bernoulli and are common for the two proofs.

James Bernoulli's Theorem (1713).

For any given $c > 0$ we have

$$P_k/(1-P_k) > c \quad \text{or} \quad P_k > c/(c+1) \quad \text{for} \quad k \geq k(r,s,c) \vee k(s,r,c),$$

where $k(r,s,c)$ is the smallest positive integer satisfying

$$k(r,s,c) \geqq \frac{\log c + \log(s-1)}{\log(r+1) - \log r}\left(1 + \frac{s}{r+1}\right) - \frac{s}{r+1} . \tag{4}$$

James Bernoulli's proof. As $P_k > c(1-P_k)$ is identical with (1) we need only prove (2). Using property (a) above we get the rather crude upper bound

$$\sum_{k+1}^{ks} f_i < (s-1) \sum_{1}^{k} f_{k+i} \quad , \quad s \geqq 2, \tag{5}$$

and combining this with property (d) we obtain

$$\sum_{1}^{k} f_i \Big/ \sum_{k+1}^{ks} f_i > \sum_{1}^{k} f_i/\{(s-1) \sum_{1}^{k} f_{k+i}\} > (f_0/f_k)/(s-1). \tag{6}$$

Writing

$$f_0/f_k = \frac{rs+s}{rs-r+(r/k)} \; \frac{rs+s-(s/k)}{rs-r+(2r/k)} \cdots \; \frac{rs+(s/k)}{rs} \tag{7}$$

it will be seen that the k factors lie between $(rs+s)/(rs-r)$ and 1 and that $(rs+s)/(rs-r) > (r+1)/r > 1$ so that by suitable choice of k the *m*th factor ($1 \leqq m \leqq k$) becomes equal to $(r+1)/r$, i.e.

$$\frac{rs+s-(m-1)(s/k)}{rs-r+(mr/k)} = \frac{r+1}{r} ,$$

which gives

$$k = m(1 + \frac{s}{r+1}) - \frac{s}{r+1} .$$

Hence, for this value of k the ratio f_0/f_k consists of m factors greater than or equal to $(r+1)/r$ and $(k-m)$ factors greater than 1 so that m may be found from the inequality

$$f_0/f_k \geqq ((r+1)/r)^m \geqq c(s-1). \qquad (8)$$

Solving for m we obtain

$$m \geqq m^* \geqq \{\log(c(s-1)\}/\{\log((r+1)/r)\}$$

where m^* denotes the smallest positive integer satisfying the inequality. (We have not included this "rounding up" in the formulation of the theorem.) From m^* we then get $k(r,s,c)$ which completes the proof.

<u>Remarks.</u> James Bernoulli's main objective was to prove that x/n converges in probability to p for $n \to \infty$. We may, however, also consider his theorem as giving a lower bound to P_k for given n. As $k = n/t$ we just have to solve the equation $k = k(c) = k(r,s,c) \vee k(s,r,c)$ with respect to c which gives $c(k) = c(r,s,k) \wedge c(s,r,k)$, where $c(r,s,k)$ is obtained from (4), i.e.

$$\ln c(r,s,k) = \frac{k(r+1) + s}{r+1+s} \ln \frac{r+1}{r} - \ln(s-1). \qquad (9)$$

It follows that $P_k > c(k)/(c(k)+1)$ and also that

$$1 - P_k \leqq e^{-\alpha n+\beta} , \quad \alpha > 0,$$

so that Bernoulli's proof implies that the tail probability tends exponentially to zero.

Nicholas Bernoulli's theorem (1713).

$$P_d/(1-P_d) > \min\{(f_0/f_d),\ (f_0/f_{-d})\} - 1 \qquad (10)$$

or

$$P_d > 1 - \max\{(f_d/f_0),(f_{-d}/f_0)\}, \qquad (11)$$

where

$$f_0/f_d \simeq \left(\frac{kr+d}{ks-d+1}\ \frac{kr+1}{kr}\ \frac{ks}{kr}\right)^{d/2} \qquad (12)$$

and f_0/f_{-d} is obtained from (12) by interchange of r and s.

Nicholas Bernoulli's proof. As above we need only consider the right tail because the results for the left is obtained by interchanging r and s. Let us define consecutive sums of length d of terms of the binomial by

$$S_\mu = \sum_1^d f_{\mu d+i}\ ,\ \mu = 0,1,\ \dots\ ,\ [ks/d]\ .$$

From the properties of f_i it follows that

$$S_\mu/S_{\mu+1} > f_{\mu d+1}/f_{\mu d+d+1} > f_0/f_d$$

so that

$$S_\mu < S_0\ (f_d/f_0)^\mu,\quad \mu \geqq 1.$$

Hence we have

$$S_0/(S_1 + \dots + S_{[ks/d]}) > (f_0/f_d) - 1\ .$$

(This is Nicholas Bernoulli's sharpening of (6).) Noting that P_d equals S_0 plus the corresponding sum to the left plus f_0 we have proved (10) and thus also (11).

To evaluate f_0/f_d we consider

$$f_0/f_d = \frac{(kr+d)(kr+d-1)\ \dots\ (kr+1)}{(ks-d+1)(ks-d+2)\ \dots\ (ks)}\left(\frac{ks}{kr}\right)^d . \qquad (13)$$

Bernoulli notes that for large values of n, i.e. of kr and ks, in relation to d, the first d ratios do not differ much and he replaces each ratio by the geometric mean of the largest and the smallest ratio so that

$$f_0/f_d \simeq \left(\frac{kr+d}{ks-d+1}\ \frac{kr+1}{ks}\left(\frac{ks}{kr}\right)^2\right)^{d/2},$$

which leads to (12). (This is Nicholas Bernoulli's sharpening of (8). Actually he uses logarithms and replaces the sum of the d logarithms by d/2 times the sum of the largest and the smallest.)

Nicholas Bernoulli's example. Taking n = 14,000 as the yearly number of births in London and $p = 18/35$ we get $k = 14000/35 = 400$, $kr = 7200$ and $ks = 6800$. To find a lower bound to P_d for $d = 163$ we compute

$$\ln(f_0/f_d) \simeq \frac{163}{2} \ln\left\{ \frac{7363}{6638} \frac{7201}{7200} \frac{6800}{7200} \right\} = 3.8009$$

so that $f_0/f_d \simeq 44.74$. Similarly, Bernoulli finds $f_0/f_{-d} \simeq 44.58$ and thus $P_d > 0.9776$.

Disregarding the requirement that r and s should be integers and using James Bernoulli's theorem for $k = 163$ we find $t = n/k = 85.89$, $r = 44.17$, which leads to $c(k) = 0.15647$ by means of (9) and thus $P_k > 0.1353$. This clearly demonstrates the great improvement obtained by Nicholas Bernoulli.

Using the normal distribution as approximation, as derived by de Moivre (1733), we get $d/\sqrt{npq} = 163/59.14 = 2.76$, which gives $P_d = 0.9942$.

From Nicholas Bernoulli to de Moivre.

From (12) and the example above it will be seen that Bernoulli's approximation to f_0/f_d is easy to compute and perhaps for this reason Nicholas Bernoulli did not go a small (but decisive) step further. Writing np and nq, respectively, for kr and ks we get from (12)

$$\ln(f_0/f_d) \simeq \tfrac{1}{2}d \ln\left\{(1 + \frac{d}{np})(1 + \frac{1}{np})(1 - \frac{d-1}{nq})^{-1}\right\}$$
$$= (d^2/2npq)(1 + (q-p)/d + \dots \quad),$$

which converges to $d^2/2npq$ if only d is of the order of $\sqrt{n}$ and $n \to \infty$. Hence, Bernoulli's result leads easily to $f_0/f_d \simeq \exp\{d^2/2npq\}$, a result that was first found by de Moivre (1733).

In Bernoulli's example we get $d^2/2npq = 3.7987$, which leads to $f_0/f_d \simeq 44.64$ and $P_d > 0.9776$ as above.

At the time it was well-known how to approximate a sum by an integral. The natural continuation of the considerations above would thus be the approximation

$$P_d = t^{-n} f_0 \sum_{-d}^{d} f_i/f_0 \simeq t^{-n} f_0 \int_{-d}^{d} e^{-x^2/2npq}\, dx ,$$

The remaining problem is to evaluate the middle term of the binomial, $t^{-n} f_0$.

De Moivre began his investigations for $p = \frac{1}{2}$ and $n = 2m$, say. From (13) it follows that

$$f_0/f_d = \frac{m+d}{m} \prod_{i=1}^{d-1} \frac{m+i}{m-i}, \qquad d \lneq m.$$

Taking logarithms it will be seen that we have to evaluate sums of the form

$$\sum_{i=1}^{d} \ln(1 + (i/m)) = \sum_{\mu=1}^{\infty} (-1)^{\mu-1} m^{-\mu} \sum_{i=1}^{d-1} i^{\mu} ,$$

which de Moivre did by means of James Bernoulli's formula for the sum of integral powers of integers. De Moivre (1730, pp.128-129) gave his result as

$$\ln(f_0/f_d) \simeq (m+d-\tfrac{1}{2})\ln(m+d-1) + (m-d+\tfrac{1}{2})\ln(m-d+1)$$
$$-2m \ln m + \ln((m+d)/m) , \qquad (14)$$

and at the time he did not go any further. It will be seen that de Moivre's result was fully within the reach of the Bernoullis because he used only the expansion of $\ln(1+x)$ and J. Bernoulli's summation formula. De Moivre's approximation is not essentially better than Nicholas Bernoulli's for large values of n.

De Moivre (1730, pp. 173-174) also found the approximation $t^{-n} f_0 \simeq 2/\sqrt{2\pi n}$ using the same methods as indicated above supplemented by Wallis infinite product for $\pi/2$.

It was first in his 1733-paper that de Moivre remarked that expansion of the logarithms in (14) leads to $\ln(f_0/f_d) \simeq 2d^2/n$ and in the general case to $d^2/2npq$. Having thus derived the normal density function as approximation to the binomial de Moivre found P_d by integrating the density.

References.

Bernoulli, J. (1713). Ars Conjectandi. Basle.

Bernoulli, N. (1713). See Montmort (1713, pp.388-398).

Montmort, P. R. de (1713). Essay d'Analyse sur les Jeux de Hazard, 2nd. ed. Paris.

Moivre, A. de (1730). Miscellanea Analytica. London.

Moivre, A. de (1733). Approximatio ad Summam Terminorum Binomii $(a+b)^n$ in Seriem expansi. London. Printed for private circulation.

Moivre, A. de (1738). The Doctrine of Chances. 2nd. ed. London.

Moivre, A. de (1756). The Doctrine of Chances. 3rd ed. London.

Todhunter, I. (1865). History of the Mathematical Theory of Probability. London.

PREPRINTS 1982

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

No. 2 Andersson, Steen A., Brøns, Hans K. and Jensen, Søren Tolver: Distribution of Eigenvalues in Multivariate Statistical Analysis.

No. 3 Tjur, Tue: Variance Component Models in Orthogonal Designs.

No. 4 Jacobsen, Martin: Maximum-Likelihood Estimation in the Multiplicative Intensity Model.

No. 5 Leadbetter, M.R.: Extremes and Local Dependence in Stationary Sequences.

No. 6 Henningsen, Inge and Liestøl, Knut: A Model of Neurons with Pacemaker Behaviour Recieving Strong Synaptic Input.

No. 7 Asmussen, Søren and Edwards, David: Collapsibility and Response Variables in Contingency Tables.

No. 8 Hald, A. and Johansen, S.: On de Moivre's Recursion Formulae for Duration of Play.

No. 9 Asmussen, Søren: Approximations for the Probability of Ruin Within Finite Time.

PREPRINTS 1983

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jacobsen, Martin: Two Operational Characterizations of Cooptional Times.

No. 2 Hald, Anders: Nicholas Bernoulli's Theorem.