

Chapter 7

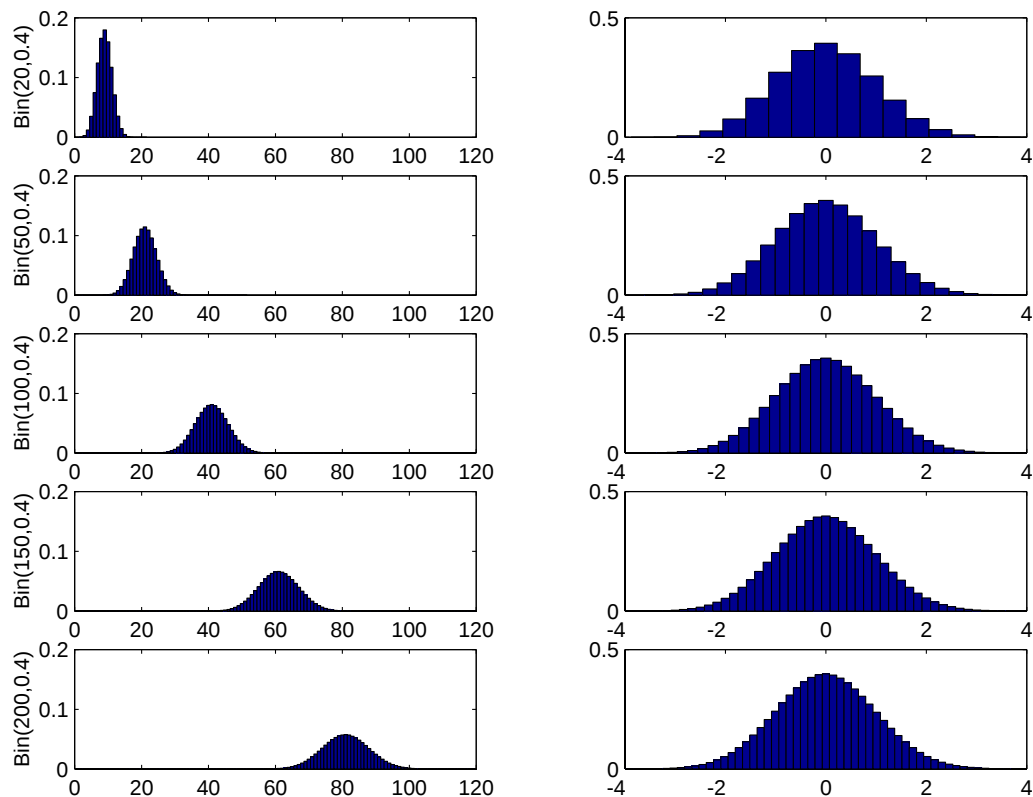
Normal distribution

This Chapter will explain how to approximate sums of Binomial probabilities,

$$b(n, p, k) = \mathbb{P}\{\text{Bin}(n, p) = k\} \quad \text{for } k = 0, 1, \dots, n,$$

by means of integrals of normal density functions.

The following pictures show two series of barplots for the $\text{Bin}(n, 0.4)$, with $n = 20, 50, 100, 150, 200$.



For the five plots on the left (the “raw barplots”), the bar of height $b(n, 0.4, k)$ and width 1 is centered at k , for $k = 0, 1, \dots, n$. As predicted by the Tchebychev inequality, the distributions cluster around the expected values, $n \times 0.4$, and they have a spread proportional to the standard deviation $\sigma_n = \sqrt{n \times 0.4 \times 0.6}$ of the $\text{Bin}(n, 0.4)$ distribution. Actually you may not be able to see the part about the standard deviation. To make the effect clearer, for

the corresponding plots on the right (the “rescaled and recentered barplots”), I have rescaled the bars by the standard deviation and recentered them at the expected value: the bar with height $\sigma_n \times b(n, 0.4, k)$ and width $1/\sigma_n$ is centered at the point $(k - n \times 0.4)/\sigma_n$.

Notice how the plots on the right settle down to a symmetric ‘bell-shaped’ curve. They illustrate an approximation due to de Moivre (1733):

$$<7.1> \quad \mathbb{P}\{\text{Bin}(n, p) \geq np + x\sqrt{np(1-p)}\} \approx \int_x^\infty \frac{\exp(-t^2/2)}{\sqrt{2\pi}} dt.$$

De Moivre, of course, expressed the result in a different way. (See pages 243-259 of the third edition of his *Doctrine of Chances*.)

From Problem Sheet 5 you know that the $b(n, p, k)$ probabilities achieve their maximum in k at a value $k_{\max}$ close to np , the expected value of the $\text{Bin}(n, p)$ distribution. Moreover, the probabilities increase as k increases to $k_{\max}$, and decrease thereafter. It therefore makes sense that a symmetric ‘bell-shaped’ approximating curve should be centered near $k_{\max}$ for the raw barplots, and near 0 for the centered barplots. See Problem Sheet 6 for a partial explanation for why the standard deviation is the appropriate scaling factor.

Put another way, de Moivre’s result asserts that if X_n has a $\text{Bin}(n, p)$ distribution then the standardized random variable $(X - np)/\sqrt{np(1-p)}$ is well approximated by the density function $\phi(t) = \exp(-t^2/2)/\sqrt{2\pi}$, in the sense that tail probabilities for the standardized Binomial are well approximated by tail probabilities for the distribution with density ϕ .

<7.2> **Definition.** A random variable is said to have a STANDARD NORMAL DISTRIBUTION if it has a continuous distribution with density

$$\phi(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \quad \text{for } -\infty < x < \infty$$

The standard normal is denoted by $N(0,1)$. □

Notice how $(X - np)/\sqrt{np(1-p)}$ has been STANDARDIZED to have a zero expected value and a variance of one. Equivalently, we could rescale the standard normal to give it an expected value of np and a variance of npq , and use that as the approximation. As you will see from the next Example, De Moivre’s approximation can also be interpreted as:

If X has a $\text{Bin}(n, p)$ distribution then it is approximately $N(np, np(1-p))$ distributed, in the sense of approximate equalities of tail probabilities.

<7.3> **Example.** Let Z have a standard normal distribution, Define the random variable $Y = \mu + \sigma Z$, where μ and $\sigma > 0$ are constants. Find

- (i) the distribution of Y
- (ii) the expected value of Y
- (iii) the variance of Y

The random variable Y has a continuous distribution. For small $\delta > 0$,

$$\begin{aligned} \mathbb{P}\{y \leq Y \leq y + \delta\} &= \mathbb{P}\{(y - \mu)/\sigma \leq Z \leq (y - \mu)/\sigma + \delta/\sigma\} \\ &\approx \frac{\delta}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right), \end{aligned}$$

where $\phi(\cdot)$ denotes the density function for Z . The distribution of Y has density function

$$\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

which is called the $N(\mu, \sigma^2)$ density. (This method for calculating a density for a function of a random variable works in more general settings, not just for standard normals.)

For the expected value and variance, note that $\mathbb{E}Y = \mu + \sigma\mathbb{E}Z$ and $\text{var}(\mu + \sigma Z) = \sigma^2\text{var}(Z) = \sigma^2$. It therefore suffices if we calculate the expected value and variance for the

•standardized

standard normal. (If we worked directly with the $N(\mu, \sigma^2)$ density, a change of variables would bring the calculations back to the standard normal case.)

$$\mathbb{E}Z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp(-x^2/2) dx = 0 \quad \text{by antisymmetry.}$$

For the variance use integration by parts:

$$\begin{aligned} \mathbb{E}Z^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2/2) dx \\ &= \left[\frac{-x}{\sqrt{2\pi}} \exp(-x^2/2) \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-x^2/2) dx \\ &= 0 + 1 \end{aligned}$$

Thus $\text{var}(Z) = 1$.

The $N(\mu, \sigma^2)$ distribution has expected value $\mu + (\sigma \times 0) = \mu$ and variance $\sigma^2 \text{var}(Z) = \sigma^2$. The expected value and variance are the two parameters that specify the distribution. In particular, for $\mu = 0$ and $\sigma^2 = 1$ we recover $N(0, 1)$, the standard normal distribution. $\square$

The de Moivre approximation: one way to derive it

The representation described in Chapter 6 expresses the Binomial tail probability as an incomplete beta integral:

$$\mathbb{P}\{\text{Bin}(n, p) \geq k\} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt$$

Apply Stirling's approximation (Appendix B) to the factorials, and replace the logarithm of the integrand $t^{k-1}(1-t)^{n-k}$ by a Taylor expansion around its maximizing value t_0 , to approximate the beta integral by

$$A \int_0^p \exp(-B(t-t_0)^2) dt$$

for constants A and B depending on n , p and k . With a change of variable, the integral takes on a form close to the right-hand side of <7.1>. For details see Appendix C.

Normal approximation—the grimy details

How does one actually perform a normal approximation? Back in the olden days, one would interpolate from tables found in most statistics texts. For example, if X has a $\text{Bin}(100, 1/2)$ distribution,

$$\mathbb{P}\{45 \leq X \leq 55\} = \mathbb{P}\left\{\frac{45-50}{5} \leq \frac{X-50}{5} \leq \frac{55-50}{5}\right\} \approx \mathbb{P}\{-1 \leq Z \leq +1\}$$

where Z has a standard normal distribution. From the tables one finds,

$$\mathbb{P}\{Z \leq 1\} \approx .8413.$$

By symmetry,

$$\mathbb{P}\{Z \geq -1\} \approx .8413.$$

so that

$$\mathbb{P}\{Z \leq -1\} \approx 1 - .8413.$$

By subtraction,

$$\mathbb{P}\{-1 \leq Z \leq +1\} = \mathbb{P}\{Z \leq 1\} - \mathbb{P}\{Z \leq -1\} \approx .6826$$

That is, by the normal approximation,

$$\mathbb{P}\{45 \leq X \leq 55\} \approx .68$$

More concretely, there is about a 68% chance that 100 tosses of a fair coin will give somewhere between 45 and 55 heads.

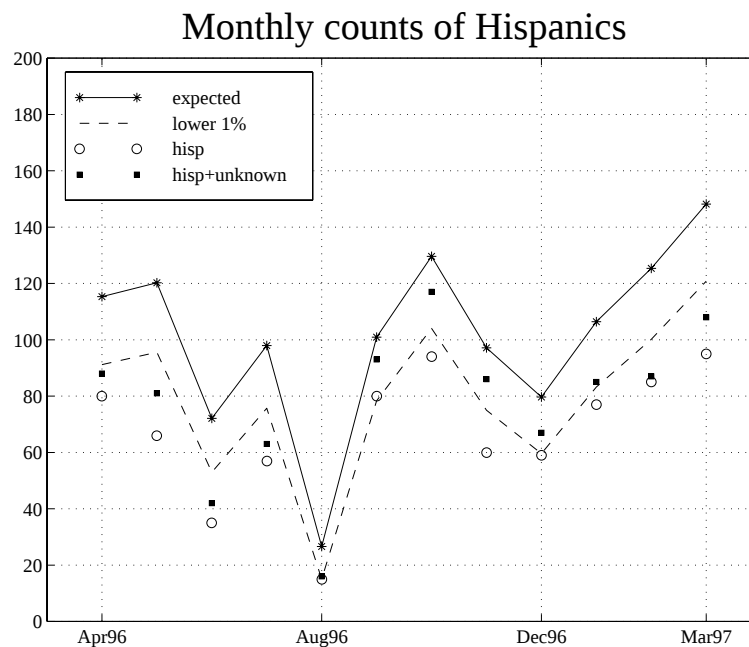
It is possible to be more careful about the atoms of probability at 45 and 55 to improve the approximation, but the refinement is usually not vital.

These days, many computer packages will calculate areas under the normal density curve directly. However one must be careful to read the fine print about exactly which curve and which area is used.

<7.4> **Example.** In response to a questionnaire filled out by prospective jurors at courthouses in the Hartford-New Britain Judicial District, persons were supposed to identify themselves as Hispanic or non-Hispanic. The figures were collected each month:

	Hispanic	Non-Hispanic	unknown	total
Apr 96	80	1659	8	1747
May 96	66	1741	15	1822
Jun 96	35	1050	7	1092
Jul 96	57	1421	6	1484
Aug 96	15	387	1	403
Sep 96	80	1436	13	1529
Oct 96	94	1847	23	1964
Nov 96	60	1386	26	1472
Dec 96	59	1140	8	1207
Jan 97	77	1527	8	1612
Feb 97	85	1812	2	1899
Mar 97	95	2137	13	2245

According to the 1990 Census, the over-18 Hispanic population of the District made up about 6.6% of the total over-18 population (and the proportion was known to have increased since the Census). The solid line in the next Figure shows the expected numbers of Hispanics for each month, calculated under the assumption that the samples were taken from a large population containing 6.6% Hispanics.



The expected counts are larger than the observed counts in every month. Could the discrepancy be due to random fluctuations?

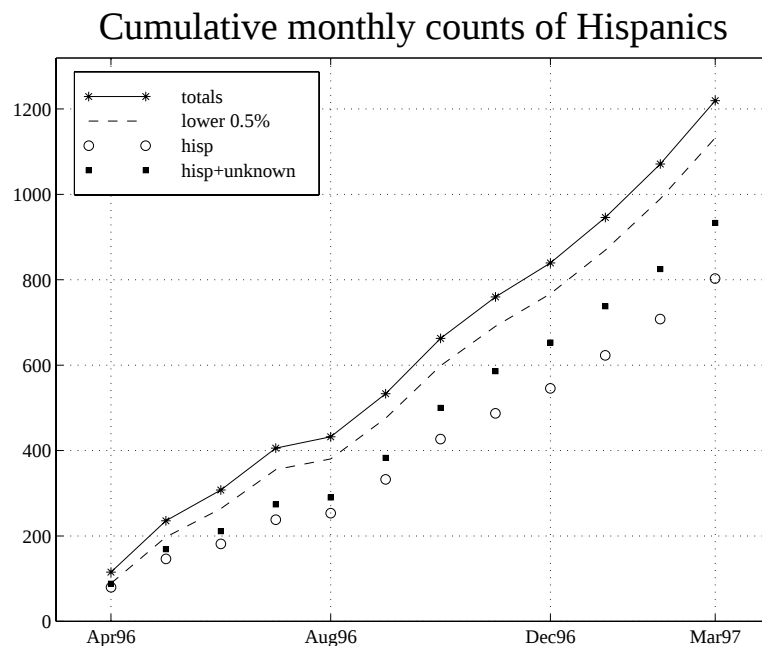
Consider a month in which a total of n questionnaires were collected. Under the model for random sampling from a population containing a fraction $p = 0.066$, the number of Hispanics in the sample should be distributed as $\text{Bin}(n, p)$, with expected value np and variance $np(1 - p)$, which is approximately $N(np, np(1 - p))$. The observed count should be larger than

$$\text{lower } 1\% = np - 2.326\sqrt{np(1 - p)}$$

with probability approximately 0.99, because $\mathbb{P}\{N(0, 1) \geq -2.236\} = 0.99$ (approximately).

The lower 1% bound is plotted as a dashed line in the first Figure. Even if all the unknowns are counted as Hispanic, for half the months the resulting counts fall below the lower 1% values. One could hardly accept the explanation that the low observed counts were merely due to random fluctuations.

The discrepancy is even more striking if a similar calculation is made for the cumulative numbers of Hispanics. (That is, for May 96, add the counts for April and May, to get an observed 146 Hispanics out of 3569 questionnaires; and so on.)



For the cumulative plot, the lower boundary corresponds to the values that the normal approximations would exceed with probability 0.995.

Be careful with the interpretation of the 0.99 or 0.995 probabilities. They refer to each of a sequence comparisons between an observed count and an expected value calculated from a model. There is no assertion that the observed counts should all, simultaneously, lie above the boundary. In fact, if we assume independence between the monthly samples, the probability that all the counts should lie above the lower 1% boundary is approximately $0.99^{12} \approx 0.89$. For the second Figure, the cumulative counts are dependent, which would complicate the calculation for simultaneous exceedance of the lower bound. $\square$

The central limit theorem

The normal approximation to the binomial is just one example of a general phenomenon corresponding to the mathematical result known as the CENTRAL LIMIT THEOREM. Roughly

stated, the theorem asserts:

If X can be written as a sum of a large number of relatively small, independent random variables, then it has approximately a $N(\mu, \sigma^2)$ distribution, where $\mu = \mathbb{E}X$ and $\sigma^2 = \text{var}(X)$. Equivalently, the standardized variable $(X - \mu)/\sigma$ has approximately a standard normal distribution.

To make mathematical sense of this assertion I would need precisely stated mathematical assumptions, which would take us on a detour into material covered more carefully in Statistics 600. (In other words, you wouldn't want to know about it for Statistics 241.)

The normal distribution has many agreeable properties that make it easy to work with. Many statistical procedures have been developed under normality assumptions, with occasional obeisance toward the central limit theorem. Modern theory has been much concerned with possible harmful effects of unwarranted assumptions such as normality. The modern fix often substitutes huge amounts of computing for neat, closed-form, analytic expressions; but normality still lurks behind some of the modern data analytic tools.

<7.5> **Example.** The boxplot provides a convenient way of summarizing data (such as grades in Statistics 241). The method is:

(i) arrange the data in increasing order

(ii) find the split points

LQ = lower quartile: 25% of the data smaller than LQ

M = median: 50% of the data smaller than M

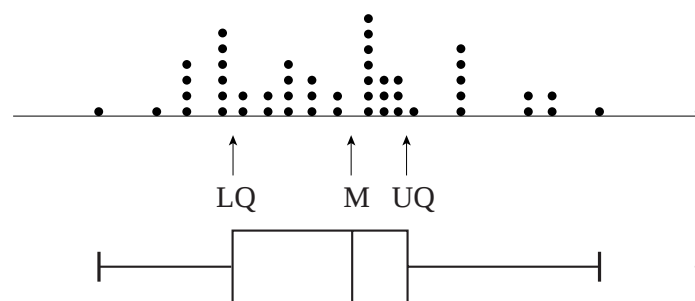
UQ = upper quartile: 75% of the data smaller than UQ

(iii) calculate IQR (= inter-quartile range) = UQ–LQ

(iv) draw a box with ends at LQ and UQ, and a dot or a line at M

(v) draw whiskers out to $UQ + 1.5 \times IQR$ and $LQ - 1.5 \times IQR$, but then trim them back to the most extreme data point in those ranges

(vi) draw dots for each individual data point outside the box and whiskers (There are various ways to deal with cases where the number of observations is not a multiple of four, or where there are ties, or . . .)



Where does the $1.5 \times IQR$ come from? Consider n independent observations from a $N(\mu, \sigma^2)$ distribution. The proportion of observations smaller than any fixed x should be approximately equal to $\mathbb{P}\{W \leq x\}$, where W has a $N(\mu, \sigma^2)$ distribution. From normal tables (or a computer),

$$\mathbb{P}\{W \leq \mu + .675\sigma\} \approx .75$$

$$\mathbb{P}\{W \leq \mu - .675\sigma\} \approx .25$$

and, of course,

$$\mathbb{P}\{W \leq \mu\} = .5$$

For the sample we should expect

$$LQ \approx \mu - .675\sigma$$

$$UQ \approx \mu + .675\sigma$$

$$M \approx \mu$$

and consequently,

$$IQR \approx 1.35\sigma$$

Check that $0.675 + (1.5 \times 1.35) = 2.70$. Before trimming, the whiskers should approximately reach to the ends of the range $\mu \pm 2.70\sigma$. From computer (or tables),

$$\mathbb{P}\{W \leq \mu - 2.70\sigma\} = \mathbb{P}\{W \geq \mu + 2.70\sigma\} = .003$$

Only about 0.6% of the sample should be out beyond the whiskers. $\square$

APPENDIX A

Where does the $\sqrt{2\pi}$ in the normal density come from? Why is it that

$$\int_{-\infty}^{\infty} \exp(-x^2/2) dx = \sqrt{2\pi}?$$

If we write C for the integral on the left-hand side, then

$$C^2 = \iint \exp(-(x^2 + y^2)/2) dx dy,$$

the double integral running over the whole plane. We can even write C^2 as a triple integral, by using the fact that

$$\int_0^{\infty} 1\{r \leq z\} e^{-z} dz = e^{-r} \quad \text{for } r > 0.$$

The indicator function effectively cuts the range of integration to $[r, \infty)$.

Replace r by $(x^2 + y^2)/2$, then substitute into the double integral to get

$$C^2 = \iint \left(\int_0^{\infty} 1\{x^2 + y^2 \leq 2z\} dz \right) dx dy = \int_0^{\infty} \left(\iint 1\{x^2 + y^2 \leq 2z\} dx dy \right) dz.$$

With the change in order of integration, the double integral is now calculating the area of a circle centered at the origin and with radius $\sqrt{2z}$. The triple integral reduces to

$$\int_0^{\infty} \pi (\sqrt{2z})^2 e^{-z} dz = \int_0^{\infty} \pi 2z e^{-z} dz = 2\pi$$

That is, $C = \sqrt{2\pi}$, as asserted.

APPENDIX B: STIRLING'S FORMULA

For positive integers n , the formula asserts that

$$<7.6> \quad n! \approx \sqrt{2\pi n} n^{n+1/2} \exp(-n),$$

in the sense that the ratio of both sides tends to 1 as n goes to infinity.

As the first step towards a proof, write

$$\log n! = \log 1 + \log 2 + \dots + \log n$$

as a sum of integrals of indicator functions:

$$\log n! = \sum_{i=1}^n \int_1^n 1\{1 \leq x < i\} \frac{1}{x} dx = \int_1^n \sum_{i=1}^n 1\{1 \leq x < i\} \frac{1}{x} dx$$

The sum of indicator functions counts the number of integers in the range $1, 2, \dots, n$ that are greater than x . It equals $n - \lfloor x \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x . The difference $\psi(x) = x - \lfloor x \rfloor$ lies in the range $[0, 1)$; it gives the fractional part of x .

The integral representing $\log(n!)$ is equal

$$\int_1^n \frac{n - \lfloor x \rfloor}{x} dx = \int_1^n \frac{n - x + \psi(x)}{x} dx = n \log n - n + \int_1^n \frac{\psi(x)}{x} dx$$

The last integral diverges, because the contribution from the interval $[i, i+1)$ equals

$$\int_i^{i+1} \frac{x - i}{x} dx = \int_0^1 \frac{t}{t+i} dt \approx \frac{1}{2i}.$$

For the approximation I have treated the $t+i$ in the denominator as approximately equal to i and then noted that $\int_0^1 t dt = 1/2$. The sum of the contributions increases like $\frac{1}{2} \log n$.

It seems we have to subtract off an extra

$$\frac{1}{2} \log n = \frac{1}{2} \int_1^n \frac{1}{x} dx$$

to keep the remainder term under control. Splitting the integral into contributions from intervals $[i, i+1)$, we then get

$$<7.7> \quad \log(n!) - (n + 1/2) \log n - n = \sum_{i=1}^n \int_0^1 \frac{t - 1/2}{t+i} dt$$

With the subtraction of the $1/2$ we will get some cancellation between the negative contribution for $0 \leq t \leq 1/2$ and the positive contribution for $1/2 < t \leq 1$.

Make the change of variable $s = 1/2 - t$ for the integral over $[0, 1/2]$, and the change of variable $s = t - 1/2$ over $(1/2, 1]$.

$$\begin{aligned} \int_0^1 \frac{t - 1/2}{t+i} dt &= \int_0^{1/2} \frac{-s}{i + 1/2 - s} ds + \int_0^{1/2} \frac{s}{i + 1/2 + s} ds \\ &= -2 \int_0^{1/2} \frac{s^2}{(i + 1/2)^2 - s^2} ds \end{aligned}$$

The last expression is bounded in absolute value by i^{-2} . The sum of the integrals forms a convergent series. That is, for some constant c ,

$$\int_1^n \frac{\psi(x) - 1/2}{x} dx \rightarrow c \quad \text{as } n \rightarrow \infty$$

Equivalently, from $<7.7>$,

$$\frac{n!}{n^{n+1/2} e^{-n}} \rightarrow e^c \quad \text{as } n \rightarrow \infty$$

This result is equivalent to formula $<7.6>$, except for the identification of $C = e^c$ as the constant $\sqrt{2\pi}$. See the Remark at the end of Appendix C for a way of deriving the value of the constant.

For an alternative derivation of Stirling's formula, see Feller I, pages 52–53.

APPENDIX C

I will explain one method for proving the approximation $<7.1>$, based on the beta-integral representation. To avoid a lot of ugly factors, I will replace k by $k+1$ and n by $n+1$. It has no great effect on the approximation. The de Moivre approximation works with k of the form $np + x\sigma$, with $\sigma = \sqrt{npq}$, for $q = 1 - p$.

Start from the equality

$$\mathbb{P}\{\text{Bin}(n+1, p) \geq k+1\} = \frac{(n+1)!}{k!(n-k)!} \int_0^p t^k (1-t)^{n-k} dt$$

The integrand equals $\exp(H(t))$, where

$$H(t) = k \log t + (n - k) \log(1 - t)$$

By Calculus,

$$H'(t) = \frac{k}{t} - \frac{n - k}{1 - t} \quad \text{and} \quad H''(t) = -\frac{k}{t^2} - \frac{n - k}{(1 - t)^2}$$

The maximum occurs at the t_0 where $H'(t_0) = 0$, that is, at $t_0 = k/n \approx p$. The second derivative at t_0 is

$$H''(t_0) = -\frac{n^3}{k(n - k)} \approx -\frac{n}{pq} = -\frac{n^2}{\sigma^2} \quad \text{for } k \approx np$$

By Taylor's theorem,

$$H(t) \approx H(t_0) - \frac{n^2}{2\sigma^2}(t - t_0)^2 \quad \text{near } t_0.$$

Thus

$$\mathbb{P}\{\text{Bin}(n + 1, p) \geq k + 1\} \approx \frac{(n + 1)n!}{k!(n - k)!} \int_0^p \exp\left(H(t_0) - \frac{n^2}{2\sigma^2}(t - t_0)^2\right) dt$$

Put Stirling's formula,

$$n! \approx Cn^{n+1/2}e^{-n} \quad \text{where } C = \sqrt{2\pi},$$

to work on the factorials sitting in front of the integral.

$$\begin{aligned} \frac{(n + 1)n!}{k!(n - k)!} \exp(H(t_0)) &\approx \frac{nCn^{n+1/2}e^{-n}}{Ck^{k+1/2}e^{-k}C(n - k)^{n-k+1/2}e^{-n+k}} \left(\frac{k}{n}\right)^k \left(\frac{n - k}{n}\right)^{n-k} \\ &= \frac{n^{3/2}}{Ck^{1/2}(n - k)^{1/2}} \end{aligned}$$

For k near np , the factor $k^{1/2}(n - k)^{1/2}$ is close to $(n^2 pq)^{1/2} = \sqrt{n}\sigma$. Thus

$$\mathbb{P}\{\text{Bin}(n + 1, p) \geq k + 1\} \approx \frac{n}{C\sigma} \int_0^p \exp\left(-\frac{n^2(t - t_0)^2}{2\sigma^2}\right) dt \quad \text{for } k \approx np.$$

Make the change of variable $s = n(t - t_0)/\sigma$. The lower terminal becomes $-nt_0/\sigma \approx -np/\sqrt{npq} \approx -\infty$, if n is large. The upper terminal becomes $n(p - t_0)/\sigma$, which equals $-x$ if $k = np + x\sigma$. The change of variable also soaks up the n/σ outside the integral, leaving

$$\mathbb{P}\{\text{Bin}(n + 1, p) \geq k + 1\} \approx \frac{1}{C} \int_{-\infty}^{-x} \exp\left(-\frac{s^2}{2}\right) ds \quad \text{for } k = np + x\sigma.$$

The final integral is not quite in the form asserted by <7.1>, but symmetry of the integrand lets us replace the integral from $-\infty$ to $-x$ by the integral from x to ∞ ; and then we have de Moivre's approximation.

Stirling's formula gives the value of the constant C as $\sqrt{2\pi}$. We could also have deduced the value of C by observing that the Binomial probability on the left-hand side tends to 1 when $x \rightarrow -\infty$, while the right-hand side tends to

$$\frac{1}{C} \int_{-\infty}^{\infty} \exp\left(-\frac{s^2}{2}\right) ds = \frac{\sqrt{2\pi}}{C}.$$

De Moivre himself (page 244 of the *Doctrine of Chances*) was not originally aware of the correct value for the constant. Referring to his calculation of the ratio of the maximum term in the expansion of $(1 + 1)^n$ to the sum, 2^n , he wrote:

When I first began that inquiry, I contented myself to determine at large the Value of B, which was done by the addition of some Terms of the above-written Series; but as I perceived that it converged but slowly, and seeing at the same time that what I had done answered my purpose tolerably well, I desisted from proceeding further till my

worthy and learned Friend Mr. James Stirling, who had applied himself after me to that inquiry, found that the Quantity B did denote the Square-root of the Circumference of a Circle whose Radius is Unity, so that if that Circumference be called c , the Ratio of the middle Term to the Sum of all the Terms will be expressed by $\frac{2}{\sqrt{nc}}$.

Maybe you agree with De Moivre, who called the approximation “the hardest Problem that can be proposed on the Subject of Chance”. My treatment has made little use of two and a half centuries of probability theory. The approximation is usually presented by methods that lead to simple proofs, but at the cost of disguising the reasons for why it works. Look at any standard text to see what I mean.



A. De Moivre, *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*, 3rd edition (1756), pages 242–243. (Photographic reprint of final edition by Chelsea Publishing Company, 1967.)

Corollary.

From this it follows, that if after taking a great number of Experiments, it should be perceived that the happenings and failings have been nearly in a certain proportion, such as of 2 to 1, it may safely be concluded that the Probabilities of happening or failing at any one time assigned will be very near in that proportion, and that the greater the number of Experiments has been, so much nearer the Truth will the conjectures be that are derived from them.

But suppose it should be said, that notwithstanding the reasonableness of building Conjectures upon Observations, still considering the great Power of Chance, Events might at long run fall out in a different proportion from the real Bent which they have to happen one way or the other; and that supposing for Instance that an Event might as easily happen as not happen, whether after three thousand Experiments it may not be possible it should have happened two thousand times and failed a thousand; and that therefore the Odds against so great a variation from Equality should be assigned, whereby the Mind would be the better disposed in the Conclusions derived from the Experiments.

In answer to this, I'll take the liberty to say, that this is the hardest Problem that can be proposed on the Subject of Chance, for which reason I have reserved it for the last, but I hope to be forgiven if my Solution is not fitted to the capacity of all Readers; however I shall derive from it some Conclusions that may be of use to every body: in order thereto, I shall here translate a Paper of mine which was printed November 12, 1733, and communicated to some Friends, but never yet made public, reserving to myself the right of enlarging my own Thoughts, as occasion shall require.

Novemb. 12, 1733

A Method of approximating the Sum of the Terms of the Binomial $\overline{a+b}^n$ expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments.

Altho' the Solution of problems of Chance often requires that several Terms of the Binomial $\overline{a+b}^n$ be added together, nevertheless in very high Powers the thing appears so laborious, and of so great difficulty, that few people have undertaken that Task; for besides James and Nicolas Bernouilli, two great Mathematicians, I know of no body that has attempted it; in which, tho' they have shown very great skill, and have the praise that is due to their Industry, yet some things were further required; for what they have done is not so much an Approximation as the determining very wide limits, within which they demonstrated that the Sum of the Terms was contained.