PascalT

\[ \min_{P \in \Omega} \int \log \left( \frac{dP}{dQ} \right) dP \]

You are here

Beloved Baby Bell is Born

Here are the first seven rows (starting from row 0) of the famous binomial triangle:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1

The numbers in a given row are obtained by adding the two entries of the row above it.

%hide
%r

mpluszs <- function(x, z=1) x+(z/2)*(z+sqrt(z^2+8*x))

P <- function(n=10, col4="red") {
  require(scatterplot3d)
  M <- expand.grid(x=0:n, y=0:n)
  M[M$x+M$y > n,] <- 0
  x <- M$x
  # Code here...
y <- M$y
z <- choose(x+y, y)
cols <- rep("black", length(x))
cols[(x+y)%%2 > 0] <- "white"
cols[(x+y)%%4 == 0] <- col4
s3d <- scatterplot3d(x, y, z, type="h", color=cols, zlab="z = (x+y)
choose y")
s3d$points3d(0:n, 0:n, rep(0, n+1), type="l", col="grey")
x <- 0:(0.5+n-sqrt(1+8*n)/2)
y <- mpluszs(x, z=1)
z <- rep(0, length(x))
s3d$points3d(x, y, z, type="l", lty="dotted", col="green")
x <- 1:n
y <- mpluszs(x, z=-1)
z <- rep(0, length(x))
s3d$points3d(x, y, z, type="l", lty="dotted", col="green")

# For png plotting with sage.

Pplot <- function(n=10, filename="myplot.png", width=480, height=480, ...) {
  png(filename=filename, width=width, height=height, ...)
  P(n)
  .null <- dev.off()
}

pbc <- function(n=10, col1="blue", col2="dark red") {
  x <- 0:n
  p <- choose(n, x)
  plot(x, p, type="h", lwd=2, col=col1, ylab="p")
  points(x, p, pch=16, cex=2, col=col2)
}

ppp <- function(L=c(2, 4, 6, 10, 20), cols=c("orange", "red", "green", "blue", "black")) {
  nn <- L[length(L)]
  x <- ((0:nn)-nn/2)/sqrt(nn/4)
  p <- choose(nn, (0:nn))*sqrt(nn)/2^(nn+1)

  plot(x, p, type="p", lwd=2, col=cols[length(L)], xlab="Standard Units",
       ylab="(prob.) / (Std.Unit)", main="Binomial Coefficients")
  for (j in 1:(length(L)-1)) {
    n <- L[j]
    x <- (0:n - n/2)/sqrt(n/4)
    p <- choose(n, 0:n)*sqrt(n)/2^(n+1)
    points(x, p, pch=16, cex=2, col=cols[j])
  }
}
PPP <- function(L=c(2,4,6,10,20),cols=c("orange","red","green","blue","black"),
       width=480,height=480,...)
{
  png(filename="myplot.png",width=width,height=height)
  ppp(L,cols)
  curve(dnorm,-2,2,add=TRUE)
  .null <- dev.off()
}

Pline <- function(n=10,filename="myplot.png",width=480,height=480,...) {
  png(filename=filename,width=width,height=height,...)
  pbc(n)
  .null <- dev.off()
}

Plot the sixth line of the binomial triangle. It looks "bell-shaped"

(Note: the R code for producing the plots is hidden in the previous cell.

You need to have the 'scatterplot3d' package installed)

%r
Pline(6)
Notice that the entries are symmetrical about the middle $6/2 = 3$ showing the maximum value of 20 for the sixth row.

Now look at the plot of the 10th row. Again the numbers are symmetrical about the middle $10/2 = 5$ and monotonically decreasing from the maximum value of 250 in the center. Notice also that the vertical and the horizontal scales for these plots are very different. The plot for the numbers on the 6th row has 1 vertical unit approximately equal to 5 horizontal units. The plot below for the numbers of the 10th row has 1 vertical unit equal to more than 10 horizontal units.

```r
Pline(10)
```
If we plot all the numbers on the first few rows of the binomial triangle with a common vertical scale, we see clearly one of the hidden secrets of the binomial triangle.

The three plots below show the first 20, 40 and 100 rows of the binomial triangle by displaying the triangle (rotated 120° ccw) on the xy plane and the numbers on the z axis.

All the rows are plotted. The even rows alternate between red and black and the odd rows are shown in white and so they are mostly invisible in the first and second plots.

```r
Pplot(20,width=700)

Loading required package: scatterplot3d
```
```
%r
Pplot(40, width=700, height=650)
```
%r
Pplot(100,width=700,height=650)
Do you see a bell curve getting skinnier and skinnier and running away to infinity along the line \( y = x \)?

In order to find the equation of the bell curve that produces the running away skinny curves in the pictures above, as just location scale deformations of a single curve \( y = \varphi(x) \),

we need to think about the numbers on the binomial triangle.

Consider the number 15 located on the 6th row, 2nd column (both rows and columns starting from 0) of the binomial triangle. This number is denoted by its position on the triangle as \( \binom{6}{2} \).

The construction rule for the binomial triangle says,
In general, \( \binom{n-1}{k-1} + \binom{n-1}{k} \). With the top \( \binom{0}{0} = 1 \)
and the boundaries \( \binom{n}{0} = \binom{n}{n} = 1 \), all the entries in the triangle are then
uniquely specified by the basic construction rule.

If we define \( 0! = 1 \) and \( n! = (n-1)! \cdot n \), then it is easy to check that,
\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]
satisfy the boundary conditions and the basic recurrence rule and thus, provide a formula
for the general entry of the triangle on row \( n \), column \( k \).

**Counting the paths from the top**

The entries in the triangle COUNT the total number of paths from the top to that position.

Consider walking from the top steping down one line at a time either to the right \(+1\) or to the left \(+0\)
until reaching the given \( \binom{n}{k} \) destination. Clearly to arrive to position \( \binom{n}{k} \) we must
follow a path with exactly \( n \) steps (to be able to reach down to the \( n \)th row),
and with exactly \( k \) steps to the right, and thus also exactly \( n - k \) steps to the left (to be able
to reach the \( k \)th column).

It is clear that there are exactly two paths from \( \binom{0}{0} \) to \( \binom{2}{1} \). These are,
\[ LR = +0 + 1 \text{ arriving from the left (i.e. from } \binom{1}{0} \text{) and } RL = +1 + 0 \]
arriving from the right (i.e. from \( \binom{1}{1} \)). It then follows from the basic construction rule,
that \( \binom{n}{k} \) is the total number of paths from the top, since that number is the total number
of paths arriving from the left (i.e. \( \binom{n-1}{k-1} \)) plus the total number of paths arriving
from the right (i.e. \( \binom{n-1}{k} \)).

We therefore conclude that \( \binom{n}{k} \) is the total number of binary sequences of length \( n \) with exactly \( k \) ones (and thus, with exactly \( n - k \) zeros). Furthermore, by specifying a sequence of \( n \) binary digits by the location of its ones (e.g. \( 011000 = \{2,3\} \)) we conclude that there are as many binary sequences of length \( n \) with \( k \) ones as there are subsets with \( k \) elements from a set of \( n \) elements, i.e. exactly \( \binom{n}{k} \).

The probability of \( k \) heads in \( n \) tosses of a fair coin

If we toss a fair coin \( n \) times there are \( 2^n \) possible outcomes corresponding to the total number of binary sequences of length \( n \). If the coin has equal probability for heads and for tails and we assume the tosses independent, then each of the \( 2^n \) outcomes has the same chance \( 2^{-n} \). Thus, by the addition rule for mutually exclusive outcomes, the probability of observing exactly \( k \) heads in the \( n \) tosses is just the proportion of outcomes with exactly \( k \) heads, i.e.,

\[
P(S_n = k) = \frac{\binom{n}{k}}{2^n}
\]

where \( S_n = X_1 + X_2 + \ldots + X_n \) with \( X_j \) taking the value 1 if the \( j \)th toss is heads and the value 0 if tails. Think of \( S_n \) as the sum of \( n \) draws with replacement from a box containing two tickets one ticket with the number 1 the other ticket with the number 0.

We now have a probabilistic interpretation to the entries of the binomial triangle:

\[
\binom{n}{k} = 2^n P(S_n = k)
\]

It follows from this that the sum of the entries of the \( n \)th row of the binomial triangle must be \( 2^n \).
It is also evident from the binomial triangle that the even numbered rows with entries \( \binom{2n}{k} \) are symmetric about the center value \( \binom{2n}{n} \) and this center value is the maximum value of the row.

**The probability of n heads in 2n tosses of a fair coin**

We show the following proposition:

\[
\lim_{n \to \infty} \sqrt{n} P(S_{2n} = n) = \frac{1}{\sqrt{\pi}}
\]

This shows that the more we toss a fair coin the more unlikely it is to get 50% of the tosses to be heads! More over the probability of observing 50% heads decreases to 0 like \( 1/\sqrt{n\pi} \).

**Proof:**

Let \( I_n = \int_0^{\pi/2} \cos^n x \, dx \). Notice that \( 0 < \cos x < 1 \) for \( x \in (0, \pi/2) \) and thus, the sequence \( I_0 = \pi/2, I_1 = 1, \ldots \) is strictly decreasing towards 0. A simple integration by parts shows that, \( nI_n = (n-1)I_{n-2} \) from where it follows that \( I_n/I_{n-2} \to 1 \).

From \( I_n < I_{n-1} < I_{n-2} \) we deduce that \( I_n/I_{n-2} < I_{n-1}/I_{n-2} < 1 \) and therefore \( I_{n-1}/I_{n-2} \to 1 \) also.

From the recursion given by the integration by parts we get,

\[
I_{2n} = \frac{1 \cdot 3 \cdot 5 \ldots (2n-1)}{2 \cdot 4 \ldots (2n)} \frac{\pi}{2}
\]

and

\[
I_{2n-1} = \frac{2 \cdot 4 \ldots (2n-2)}{3 \cdot 5 \ldots (2n-1)}.
\]

Thus,

\[
\frac{I_{2n}}{I_{2n-1}} = \left[ \frac{3 \cdot 5 \ldots (2n-1)}{2 \cdot 4 \ldots (2n)} \right]^2 \cdot (2n) \cdot \frac{\pi}{2}
\]
Multiplying the numerator and the denominator inside the square brackets by $2 \cdot 4 \ldots (2n)$ we obtain,

$$\frac{I_{2n}}{I_{2n-1}} = \left[ \frac{(2n)!}{2^{2n}(n!)^2} \right]^2 \frac{n\pi}{\pi} = \left[ \sqrt{n} \frac{(2n)}{2^{2n}} \right]^2 \pi \to 1$$

from where we deduce that,

$$\lim_{n \to \infty} \sqrt{n} \frac{(2n)}{2^{2n}} = \frac{1}{\sqrt{\pi}}$$

**Stirling's Approximation**

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

The log is strictly increasing. Hence,

$$\int_{k-1}^{k} \log x \, dx < \log k < \int_{k}^{k+1} \log x \, dx$$

adding over $k = 1, 2, \ldots, n$,

$$\int_{0}^{n} \log x \, dx < \log n! < \int_{1}^{n+1} \log x \, dx$$

The antiderivative of $\log x$ is $x \log x - x$ and computing the integrals we obtain,

$$n \log n - n < \log n! < (n + 1) \log (n + 1) - n$$

This suggests to compare $\log n!$ with the in between term, $(n + 1/2) \log n - n$.

We will show that the differences $d_n = \log n! - (n + 1/2) \log n + n$ are decreasing and bounded and therefore they form a convergence sequence. Just notice that,

$$d_n - d_{n+1} = (2n + 1) \frac{1}{2} \log \frac{n+1}{n} - 1 = (2n + 1) \left[ \frac{1}{2n+1} + \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^3} + \ldots \right] - 1$$

the term within the square brackets follows from noticing that,

$$\frac{1}{2} \log \frac{n+1}{n} = \frac{1}{2} \log \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}$$

```
t = var('t')
assume(abs(t)<1)
taylor(log(1+t),t,0,10)
-\frac{1}{10} t^{10} + \frac{1}{9} t^9 - \frac{1}{8} t^8 + \frac{1}{7} t^7 - \frac{1}{6} t^6 + \frac{1}{5} t^5 - \frac{1}{4} t^4 + \frac{1}{3} t^3 - \frac{1}{2} t^2 + t
```

```
taylor((1/2)*log((1+t)/(1-t)),t,0,10)
\frac{1}{9} t^9 + \frac{1}{7} t^7 + \frac{1}{5} t^5 + \frac{1}{3} t^3 + t
```
Hence,
\[
d_n - d_{n+1} = \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \ldots
\]
which shows that,
\[
0 < d_n - d_{n+1} < \frac{r}{3(1-r)} = \frac{1}{12n(n+1)} < \frac{1}{12n^2}
\]
where we have compared the series with the geometric series with \( r = 1/(2n+1)^2 \).

Adding over \( n = 1, 2, \ldots \) we get,
\[
0 < 1 - d_\infty < \frac{\pi^2}{72}
\]
and then,
\[
\lim_{n \to \infty} d_n > 1 - \frac{\pi^2}{72} \approx 0.8629
\]
The sequence \( \{d_n\} \) is thus strictly decreasing and bounded and therefore it must converge to some limit \( C > 0 \).

We conclude that,
\[
\log n! - (n + 1/2) \log n + n - C \to 0
\]
or equivalently,
\[
n! \sim e^C n^{n+1/2} e^{-n}
\]
meaning that the ratio of the two sides converges to 1 as \( n \) increases.

The positive constant \( e^C \) is obtained by using our previous result
\[
\lim_{n \to \infty} \sqrt{n} \frac{(2n)!}{2^{2n} (n!)^2} = \frac{1}{\sqrt{\pi}}
\]
Using (the found above) Stirling's approximation (with the constant \( e^C \)) to replace
the factorials, we get,
\[
\lim_{n \to \infty} \frac{n^{1/2} e^C (2n)^{2n+1/2} e^{-2n}}{2^{2n} (e^C n^{n+1/2} e^{-n})^2} = \frac{2}{e^C} = \frac{1}{\sqrt{\pi}}
\]
From where we deduce that \( e^C = \sqrt{2\pi} \) obtaining the standard Stirling's approximation,
\[
n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}
\]

\[
\text{var('C')}
\]
\[
sol = \text{solve}(2/\exp(C) == 1/\sqrt{\pi}, C)
\]
The sum of \(2n\) draws from a 01 box is expected to be half the number of draws give or take a typical fluctuation (a standard deviation) of the order of \(\sqrt{n}/2\). When we look at the chances of heads in terms of typical fluctuations from \(n\) we discover the bell curve.

We show that for any real value \(x\),

\[
\lim_{n \to \infty} \frac{n}{2} P\left( S_{2n} = n + \frac{x}{\sqrt{n}} \right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

where we take \(n + x \sqrt{n}/2\) to be its closest integer when the expression is not an integer.

If we let \(S^*_{2n}\) to denote \(S_{2n}\) in standard units (i.e. standard deviations from the expected \(n\)) then the above is,

\[
\lim_{n \to \infty} \frac{P(S^*_{2n} = x)}{\sqrt{2/n}} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

Notice that this shows that the bell curve is a density of probability: Probability per unit length. The observed heads \(S_{2n}\) is always an integer from \(\{0, 1, 2, \ldots, 2n\}\). When \(S_{2n} = k\) we have \(S^*_{2n} = (k - n)/\sqrt{n}/2\) in standard units, and when \(S_{2n} = k + 1\) (the next possible value) gives \(S^*_{2n} = (k + 1 - n)/\sqrt{n}/2\). Thus, the size of the step in standard units is \(\sqrt{2/n}\). The above statement says that the probability of observing the number of heads that is \(x\) standard deviations from the expected \(n\) divided by the step of size \(\sqrt{2/n}\) approaches the bell curve.

Proof:

\[
\sqrt{\frac{n}{2}} P\left( S_{2n} = n + x \sqrt{n}/2 \right) = \frac{n^{1/2} \left( \frac{2n}{n+x\sqrt{n}/2} \right)}{2^{2n+1/2}} = \frac{n^{1/2} (2n)!}{2^{2n+1/2} (n + x \sqrt{n}/2)! (n - x \sqrt{n}/2)!}
\]

Now use Stirling's approximation for the factorials to show, that the above is,

\[
\sim \frac{1}{\sqrt{2\pi}} \frac{n^{2n+1}}{(n + x \sqrt{n}/2)^{n+x\sqrt{n}/2+1/2} (n - x \sqrt{n}/2)^{n-x\sqrt{n}/2+1/2}}
\]

Noticing that the exponent of \(n\) in the numerator decomposes as
$$2n + 1 = (n + x\sqrt{n/2} + 1/2) + (n - x\sqrt{n/2} + 1/2)$$ we obtain,

$$\sim \frac{1}{\sqrt{2\pi}} \frac{1}{(1 + x/\sqrt{2n})^{n + x\sqrt{n/2} + 1/2}} \frac{1}{(1 - x/\sqrt{2n})^{n - x\sqrt{n/2} + 1/2}}$$

Erasing the $1/2$ from the exponents (the terms go to 1) and combining the terms raised to the power $n$ we get,

$$\sim \frac{1}{\sqrt{2\pi}} \frac{1}{(1 - x^2/2n)^n(1 + x/\sqrt{2n})^{\sqrt{n}(x/\sqrt{2})}(1 - x/\sqrt{2n})^{-\sqrt{n}(x/\sqrt{2})}}$$

$$\sim \frac{1}{\sqrt{2\pi}} \frac{1}{e^{-x^2/2} e^{x^2/2}} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

```
pp = plot(exp(-x^2/2)/sqrt(2*pi), (x,-4,4))
pp.show(figsize=(6,3))
```

Now look how the rows of the binomial triangle are getting closer and closer to the bell curve when we plot them with Standard Units on the horizontal and Probability Density Units (i.e. probability per std. unit) on the vertical.

```
%r
PPP(L=c(2,4,6,10,20), cols=c("orange","red","green","blue","black"))
```
The picture shows the central values of \( \binom{n}{k} \) when \( n \) is 2, 4, 6, 10 and 20. Different colors correspond to different values for \( n \). At \( x = 0 \) we can see the colors sorted by increasing \( n \) approaching the peak of the bell curve from below.
So YES! when using the correct scales of standard units for the horizontal and probability density units for the vertical we realize that the binomial coefficients approach THE BELL CURVE,

\[ \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \]

Thanks:

Nothing here is original. I have borrowed freely from Herb Robbins, from Feller vol I, from Timothy Gower's blog.

Many thanks also to the creators of sage, TeX, R, and the open source online community.

Carlos Rodriguez <carlos@math.albany.edu>