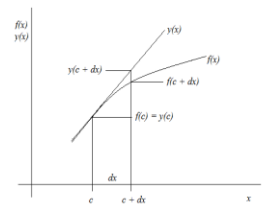
ISSN: 2037-6367

MatematicaMente



Pubblicazione mensile della sezione veronese della MATHESIS – Società Italiana di Scienze Matematiche e Fisiche – Fondata nel 1895 – Autorizzazione del Tribunale di Verona n. 1360 del 15 – 03 – 1999 – I diritti d'autore sono riservati. Direttore: Luciano Corso – Redazione: Alberto Burato, Fabrizio Giugni, Michele Picotti, Sisto Baldo – Via IV Novembre, 11/b – 37126 Verona – tel e fax (045) 8344785 – 338 6416432 – e-mail: lcorso@iol.it – info@mathesisverona.it – Stampa in proprio – Numero 224 – Pubblicato il 05 – 05 – 2017

One, Two, Three … Infinity?

Len Bos [*]

One of the most powerful proof techniques is that of mathematical induction. It allows us, for example, to *verify* that a formula such as $1+2+\cdots+n = n(n+1)/2$ is true after perhaps having guessed it by doing a number of examples.
There is always the evil temptation to skip the details, arguing that if a formula is true for $n = 1, 2, 3$ and even 4, then it must be true for all n. We all know that this is really not the case, and it is easy to make *artificial* counterexamples. What I would like to show you is quite a *natural* geometric example of a formula that is true for $n = 1, 2, 3, 4$ and 5, but is *not* true in general. It was first shown to me by my former colleague Richard Guy (who, by the way, is now over 100 years old and still an active Mathematician!). The problem is this: take n points (in general position) on the perimeter of a circle and connect each one to all the others by a line. How many regions does this network of lines create? Here "in general position" means that connecting lines have simple intersections, i.e., no three connecting lines intersect in the same point. Figures One and Two show the examples of $n = 2$ and $n = 3$.

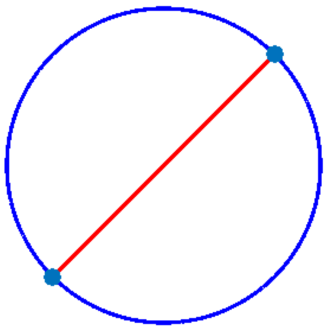


Figure One: $n = 2$; two regions

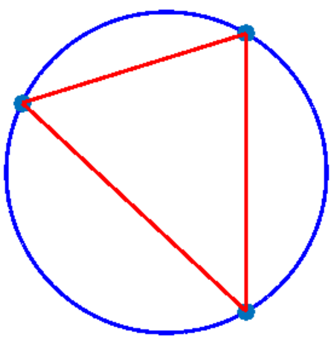


Figure Two: $n = 3$; four regions

Let F_n denote the number of regions (also known as faces) for n points. We have that $F_2 = 2$ and $F_3 = 4$. If there is only one point then there are no connecting lines and the circle is not subdivided. Hence, we may reasonably take $F_1 = 1$. What about higher values of n?

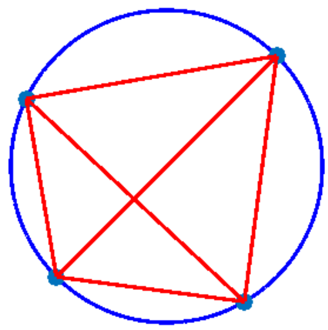


Figure Three: $n = 4$; eight regions

From Figure Three we easily see that $F_4 = 8 = 2^3$ and we might be tempted, at this point, to think that $F_n = 2^{n-1}$. But maybe we should first try the next value, n = 5. The result is in Figure 4.

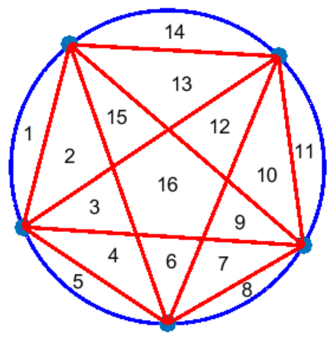


Figure Four: $n = 5$; sixteen regions

As expected $F_5 = 16 = 2^4$ and our conjecture that $F_n = 2^{n-1}$ is looking good. BUT, let's now check $n = 6$. The result is in Figure Five.

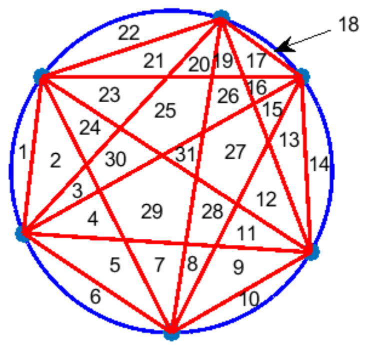


Figure Five: $n = 6$; 31 regions

The value of F_6 is 31, not $2^5 = 32$ as we may have thought!
Thus the formula $F_n = 2^{n-1}$ is *not* true, even though it is true for $n = 1, 2, 3, 4$ and 5 !

Why does this happen? What then is the correct formula for F_n? Let's see. First of all, in the graph that the connecting lines produce, a *vertex* is a point of intersection of two of the lines (plus the original points). An *edge* is the line segment between any two vertices. Let V_n denote the number of vertices and E_n the number of edges in the graph. Note that each vertex, other than one of the original points, is the intersection of exactly two different lines, each connecting a different pair of points. Such intersection points are determined by choosing any 4 of the n points. Hence

$$V_n = \binom{n}{4} + n,$$

the n being the original n points. Further, each vertex lies on two different lines and contributes one edge on each of the two lines. But, also it is the case that e.g., three internal points divide an interval into *four* pieces, i.e., one more than the number of division points. Hence we must add the number of connecting lines, i.e.,

$$E_n = 2(V_n - n) + \binom{n}{2} = 2\binom{n}{4} + \binom{n}{2}.$$

Finally, we may use Euler's famous graph formula that

$$V_n - E_n + f_n = 1.$$

No, f_n is not a typo! It stands for the interior faces, i.e., those not bounded by an arc of the circle connecting two of our points; $F_n = f_n + n$. Hence

$$\begin{aligned} F_n &= n + 1 - V_n + E_n \\ &= n + 1 - \left(n + \binom{n}{4}\right) + (2\binom{n}{4} + \binom{n}{2}) \\ &= 1 + \binom{n}{2} + \binom{n}{4}. \end{aligned}$$

Now a little algebra will reveal that an alternate formula is

$$F_n = \binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \binom{n-1}{3} + \binom{n-1}{4}.$$

In either case we see that F_n is a quartic polynomial in n, and not 2^n!

[*] Professore Ordinario di Analisi Numerica, Università degli Studi di Verona. E-mail: leonardpeter.bos@univr.it

Sulle variabili aleatorie doppie in statistica

(*di Luciano Corso*) Un antico detto latino (di Orazio) dichiara: *Longum iter per praecepta; breve et efficax per exempla*. Seguendo questa visione educativa tentiamo di insegnare a studenti che si imbattono per la prima volta in alcuni concetti di calcolo delle probabilità e di statistica il significato di variabili aleatorie, di probabilità marginali, congiunte e condizionate e di misure, in generale, a esse connesse attraverso un semplice esempio.

Si lancia una moneta ideale n = 4 volte. Dall'esperimento, gli eventi possibili sono quaterne ordinate di testa e croce; in ogni lancio, infatti, o che esce testa, cui assegniamo il valore 1, con probabilità 1/2, o croce (non testa), cui assegniamo il valore 0 con probabilità 1/2.

Spazio campionario Ω e le variabili aleatorie

Lo spazio campionario rappresenta una partizione di eventi riferibili a questo esperimento. La |Ω| è 2^4=16 eventi che presento qui di seguito:

$$\Omega = \{\begin{array}{lllll} 0000 & 0001 & 0011 & 0111 & 1111 \\ & 0010 & 0101 & 1011 & \\ & 0100 & 0110 & 1101 & \\ & 1000 & 1001 & 1110 & \\ & & 1010 & & \\ & & 1100\}. & & \end{array}$$

Gli eventi sono equiprobabili e ciascuno ha probabilità 2^{-4} di verificarsi ogni 4 lanci.

Consideriamo ora le seguenti variabili aleatorie:
X = "*numero di teste in* 4 *lanci di moneta*";
Y ="*numero di cambiamenti di segno in* 4 *lanci di moneta*".
Le due variabili per ogni evento di Ω hanno assegnato un preciso valore; per esempio, la sequenza 0101 assegna a X il valore 2 e a Y il valore 3. I valori e le probabilità associate a esse sono sintetizzati nella tabella 1.

Tabella 1						
(X, Y)	x = 0	x = 1	x = 2	x = 3	x = 4	
y = 0	1 / 16				1 / 16	2 / 16
y = 1		2 / 16	2 / 16	2 / 16		6 / 16
y = 2		2 / 16	2 / 16	2 / 16		6 / 16
y = 3			2 / 16			2 / 16
	1 / 16	4 / 16	6 / 16	4 / 16	1 / 16	1

La tabella 1 è a doppia entrata ed è di facile lettura. In tabella ci sono due tipi di misure: le probabilità marginali dei diversi valori di X e di Y (ultima riga e ultima colonna della tabella) e le probabilità congiunte (interne alla tabella). Per le probabilità marginali prendiamo $Pr(x=2)$ = 6/16, per le probabilità congiunte $Pr(x=3, y=2)$ = 2 / 16 e per le probabilità condizionate $Pr(y=1|\, x=2)$ = 2/6.
Lo studio in congiunzione di queste due variabili aleatorie fa nascere interessanti nozioni concettuali di calcolo delle probabilità e di statistica.

Misure e misure condizionate

Applichiamo alle variabili aleatorie X e Y le più importanti misure statistiche. Calcoliamo medie aritmetiche, varianze e coefficienti di variazione delle due variabili.

$$M(X) = \sum_{j=1}^{5} x_j \cdot p_{.j} = 0 \cdot \frac{1}{16} + \cdots + 4 \cdot \frac{1}{16} = 2$$

$$M(Y) = \sum_{i=1}^{4} y_i \cdot p_{i.} = 0 \cdot \frac{2}{16} + \cdots + 3 \cdot \frac{2}{16} = 1{,}5$$

Dove j è il contatore delle colonne e i quello delle righe.
Posto, per semplicità di scrittura, $M(X) = \bar{x}$ e $M(Y) = \bar{y}$ le varianze, le deviazioni standard e i coefficienti di variazione sono:

$$V(X) = \sum_{j=1}^{5} \left(x_j - \bar{x}\right)^2 \cdot p_{.j} = M(X^2) - \bar{x}^2 = 1 \,;$$

$$V(Y) = \sum_{i=1}^{4} (y_i - \bar{y})^2 \cdot p_{i.} = M(Y^2) - \bar{y}^2 = 0{,}75 \,;$$

$$\sigma(X) = 1 \quad , \quad \sigma(Y) = 0{,}866025 \,;$$

$$v(X) = \frac{\sigma(X)}{\bar{x}} = \frac{1}{2} = 0{,}5 \quad , \quad v(Y) = \frac{\sigma(Y)}{\bar{y}} = \frac{0{,}866025}{1{,}5} = 0{,}5773 \,.$$

I coefficienti di variazione $v(.)$ dimostrano che Y ha una dispersione media relativa maggiore di X. Occhio! Una comparazione della variabilità delle due v.a. basata sulle deviazioni standard sarebbe in questo caso scorretta. Il lettore saprebbe darne una ragione? Calcoliamo anche una media condizionata: $M(y|x = 2) = 2$. Si verifichi il risultato.

Le due variabili sono statisticamente indipendenti? Per verificarlo basta provare che il teorema delle probabilità composte per eventi statisticamente indipendenti in questo caso non viene rispettato. L'enunciato del teorema nel nostro caso è:

$$Pr\left(x_j \cap y_i\right) = Pr\left(x_j\right) \cdot Pr(y_i) \qquad \forall\, i, j \qquad (1)$$

Applicando la relazione (1) al caso (x=3, y=2) si ottiene $Pr(x=3 \cap y=2) = (4/16)\cdot(6/16)$ = 6 / 64 che non corrisponde al valore osservato. Confrontando le probabilità congiunte sperimentali con quelle che si calcolano con la (1) si può notare che le due variabili non sono statisticamente indipendenti; basta che in una sola casella non si verifichi l'uguaglianza tra valore osservato e valore teorico per arrivare a questa conclusione.
A questo punto ci possiamo chiedere quale tipo di dipendenza ci potrebbe essere tra X e Y. Questa fase, che normalmente è molto complicata, si chiama identificazione del modello. Possiamo innanzi tutto verificare se le due variabili sono correlate linearmente attraverso il calcolo del coefficiente di correlazione lineare (*di Bravais-Pearson*):

$$r(x, y) = \frac{cov(x,y)}{\sigma(x)\cdot\sigma(y)} = \frac{M(x\cdot y) - \bar{x}\bar{y}}{\sigma(x)\cdot\sigma(y)} = \frac{0}{1\cdot 0{,}866025} = 0. \qquad (2)$$

L'indice conferma che non esiste una correlazione lineare tra X e Y.