THE POWER OF RANDOMNESS:

TWO HOPEFULLY THOUGHT-PROVOKING NUMERICAL EXPERIMENTS

J. M. Sanz-Serna

Universidad Carlos III de Madrid

I: STRANGE CASE OF COMPANIES X AND Y

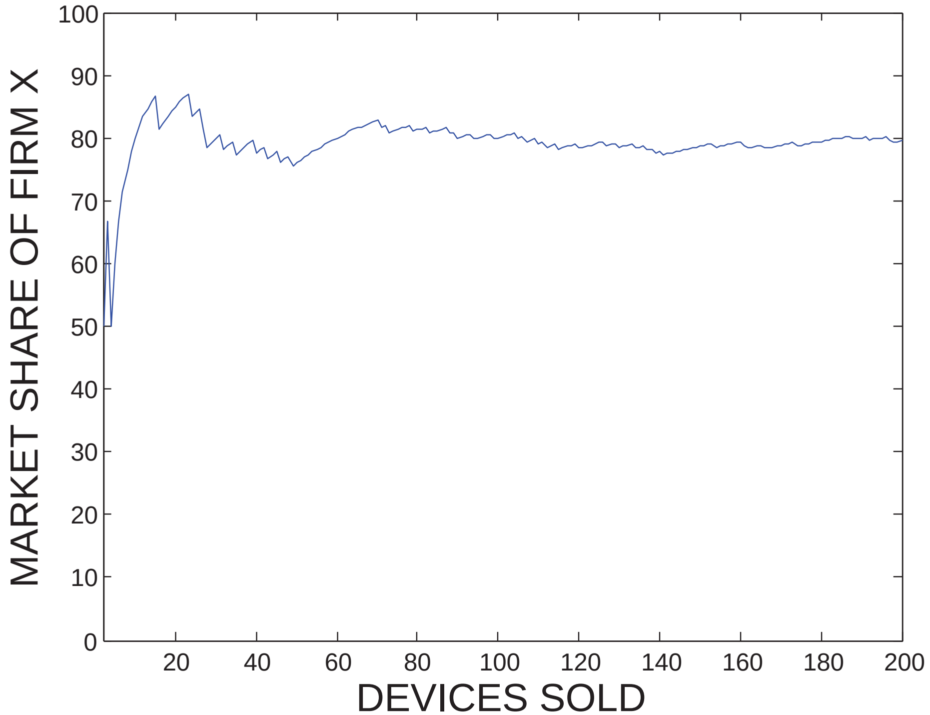
• In Imaginetown two rival companies, X, Y, launched simultaneously their fourth generation E-LOTs.

• No publicity: costumer satisfaction thought to be best ad.

• Ann bought an X E-LOT and Ben got a Y. After that, each interested person talked to one, randomly met, E-LOT owner and immediately bought X or Y according to the information received.

• Next buyers were Carla X, Daniel Y, Elaine X, Francis X, Gordon X, . . .

- This was the evolution of the market share of X:



- After an initial period with fluctuations (comprising the 25 or 50 first sales), the share of X stabilized at approx. 80%.

- What is the reason for the success of X?

Some possible explanations:

- E-LOT X is superior to Y?

- But perhaps E-LOT X was bought by more persuasive people?

- or . . .

- The truth: E-LOTS X e Y are indistinguishable.

- All buyers were very pleased with their purchases and always recommended the brand they own.

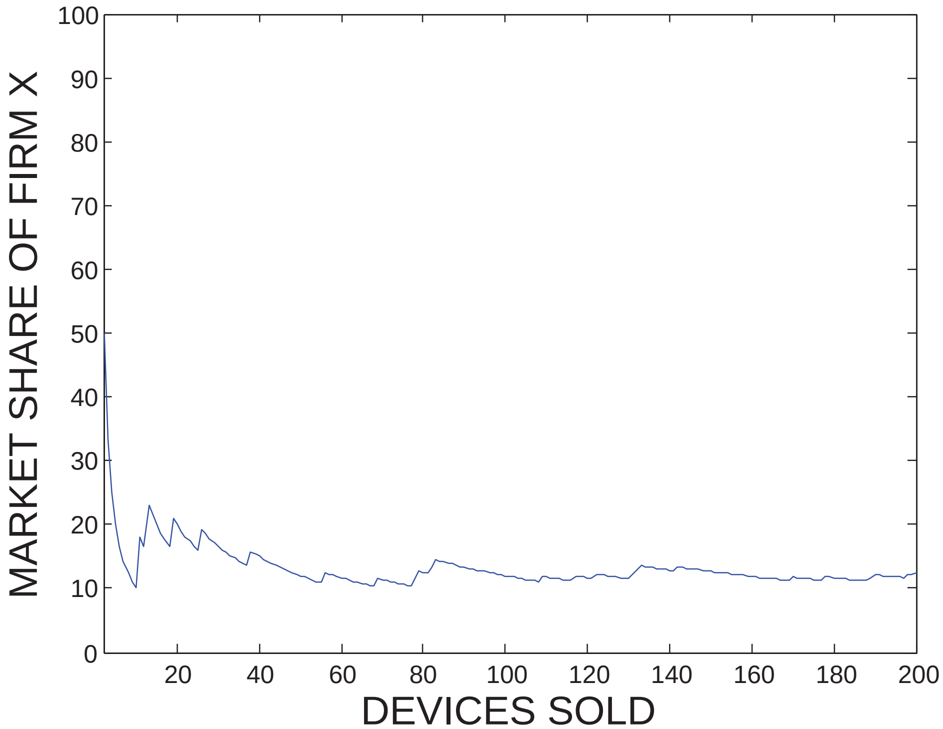
- X is bought by people who randomly ran into X owners, and likewise for Y.

- e.g. Carla bought X because she asked Ann que who had X, had she run into Ben she would have bought Y.

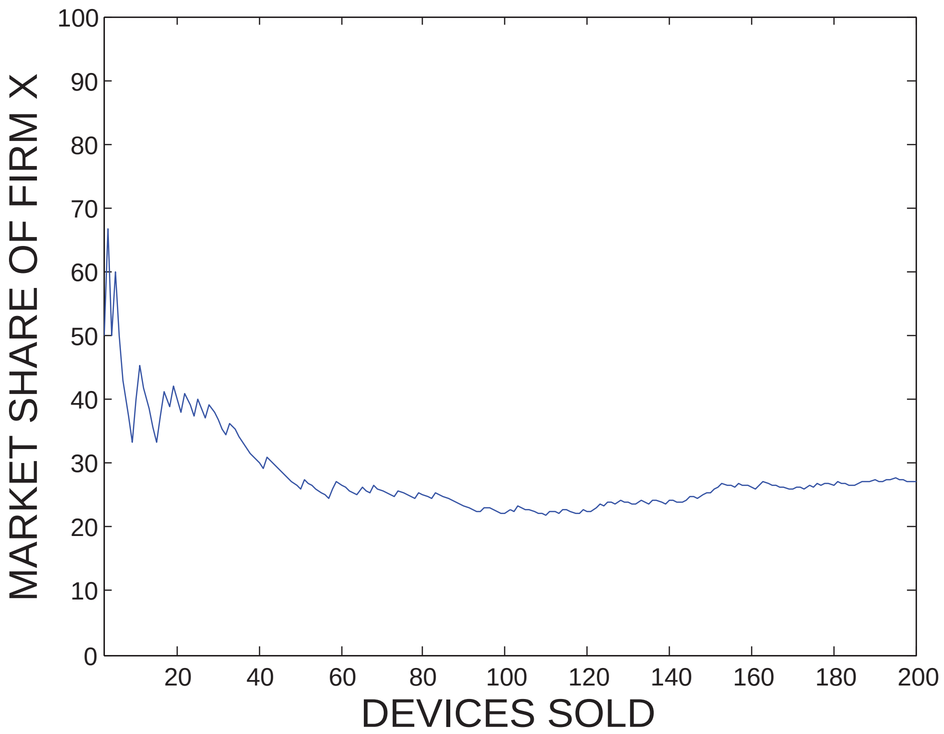
- The graph we saw is entirely the *result of randomness!!!*

- If the experiment is rerun and the initial encounters are different . . .

. . . things are thus:

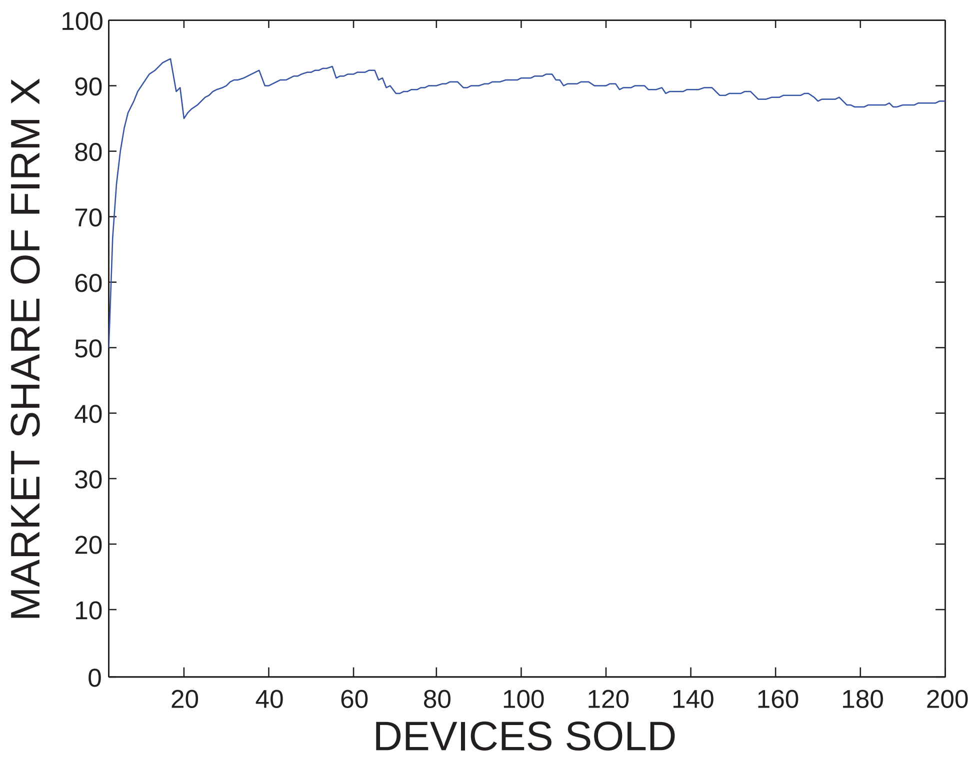


. . . or thus:



MARKET SHARE OF FIRM X
100
90
80
70
60
50
40
30
20
10
0
20
40
60
80
100
120
140
160
180
200
DEVICES SOLD

. . . or:


MARKET SHARE OF FIRM X
DEVICES SOLD
0
10
20
30
40
50
60
70
80
90
100
20
40
60
80
100
120
140
160
180
200

• MATHS: May prove rigorously (Polya urn martingale) that each time the experiment is performed the market share of X (and hence that of Y) will approach a limit as the number of devices sold increases.

• The value of the limit changes from one repetition of the experiment to the next.

• The distribution of the limit is uniform in the interval from 0% to 100%.

Two points:

(1) Random procedures that involve sufficiently many steps (or many atoms, genes, individuals, . . .) often result in regular patterns.

(2) We quickly come up with causal explanations of those patterns.

- We are surrounded by examples of regularities induced by randomness.

- The purposeless random motion of Na and Cl ions leads to uniformly salted water.

- Likewise we have to thank randomness for the fact that molecules of oxygen in this room do not all go to one of the corners. [Molecules do not decide where to go.]

- Individuals of an animal species distribute themselves evenly in a given habitat when they move randomly.

Finding causes:

“. . . the desire to find those causes is implanted in man’s soul.” (L. Tolstoi, War and Peace, Chapter 1, Book XIII).

“. . . communally led into the belief that world events have identifiable causes . . . ” (L. Valiant).

- The popular book *Thinking fast and slow* (Kahneman) reports many experiments showing our bias toward finding causes.

• This bias may well have resulted from evolutional advantages:

- More generally, the learning algorithms in our heads are biased towards generalization/construction of theories, even when such generalizations are not warranted.

- The algorithms are very successful most of the time. Hence our surprise when they do not work well.

- Listening to this talk made one of my colleagues quickly came up with a generalization: always try and be the first in the market. This may be more important than the quality of product you sell.

II: COMPUTING INTEGRALS

- Wish to compute (approximately)

$$\int_0^\infty f(x)\rho(x)\,dx,$$

where the weight function $\rho(x) = \mathrm{exp}(-x)$ is seen as fixed once and for all. We use a quadrature rule with m (f-independent) abscissas x_i/ weights w_i:

$$\sum_{i=1}^{m} w_i f(x_i).$$

• Optimal choice of the x_i: zeros of Laguerre polynomials:

$$L_2(x) = \frac{1}{2}(x^2 - 4x + 2)$$
$$L_3(x) = \frac{1}{6}(-x^3 + 9x^2 - 18x + 6)$$
$$L_4(x) = \frac{1}{24}(x^4 - 16x^3 + 72x^2 - 96x + 24)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

• E.g. for $m = 5$ the abscissas are approximately:

$$\begin{array}{r} 0.263560319718 \\ 1.413403059107 \\ 3.596425771041 \\ 7.085810005859 \\ 12.640800844276 \end{array}$$

• I test the rule in the simple example $f(x) = \cos x$, where I know the true value of the integral. The (small) relative errors are:

	$m = 2$	$m = 3$	$m = 4$	$m = 5$
	$1.4(-1)$	$-4.7(-2)$	$5.0(-3)$	$1.1(-3)$

• For $m = 5$, if the abscissa at $0.26\ldots$ moves to 0.30 the error increases from $1.1(-3)$ to $9.5(-3)$.

NOTE:

- Construction of rule needed knowledge of weight function.
- Much ‘theory’ needed to find abscissas and weights (integrand approximated by polynomial, theory of orthogonal polynomials, . . .)
- Little work needed to apply the rule to a given integrand f (formula useful in precomputer days).

- *Let's make things harder . . .*

• For $f(x) = x^{1/5}$ (non differentiable at $x = 0$):

$m = 2$	$m = 3$	$m = 4$	$m = 5$
$3.9(-2)$	$2.4(-2)$	$1.7(-2)$	$1.3(-2)$

[before we had

$m = 2$	$m = 3$	$m = 4$	$m = 5$
$1.4(-1)$	$-4.7(-2)$	$5.0(-3)$	$1.1(-3)$

].

• Consider d-fold integral

$$\int_{[0,\infty)^d} F(x_1,\ldots,x_d)\,\rho(x_1,\ldots,x_d)\,dx_1\cdots dx_d,$$
$$\rho(x_1,\ldots,x_d) = \exp(-x_1-\cdots-x_d),$$

and extend the rule:

$$\sum_{i_1=1}^{m}\cdots\sum_{i_d=1}^{m} w_{i_1}\cdots w_{i_d}F(x_{i_1},\ldots,x_{i_d}).$$

- For the smooth integrand

$$F(x_1, \ldots, x_d) = \cos x_1 \cdots \cos x_d,$$

the relative errors would be:

d	$m = 2$	$m = 3$	$m = 4$	$m = 5$
1	1.4(−1)	−4.7(−2)	5.0(−3)	1.1(−3)
4	$***$	$***$	2.0(−2)	4.3(−3)
16	$***$	$***$	8.0(−2)	1.7(−2)
64	$***$	$***$	$***$	7.1(−2)

- The number of times F would have to be evaluated is . . .

d	$m=2$	$m=3$	$m=4$	$m=5$
1	$1.4(-1)$ 2	$-4.7(-2)$ 3	$5.0(-3)$ 4	$1.1(-3)$ 5
4	$***$ 16	$***$ 81	$2.0(-2)$ 256	$4.3(-3)$ 3125
16	$***$ 65536	$***$ 4.3(7)	$8.0(-2)$ 4.2(9)	$1.7(-2)$ 1.5(11)
64	$***$ 1.8(19)	$***$ 3.4(30)	$***$ 3.4(38)	$7.1(-2)$ 5.4(44)

- For the nonsmooth

$$F(x_1, \ldots, x_d) = (x_1 \cdots x_d)^{1/5}$$

the relative errors would be even worse:

d	$m = 2$	$m = 3$	$m = 4$	$m = 5$
1	3.9(−2)	2.4(−2)	1.7(−2)	1.3(−2)
4	$***$	$***$	7.1(−2)	5.4(−2)
16	$***$	$***$	$***$	$***$
64	$***$	$***$	$***$	$***$

- We conclude that the ‘clever’ Gauss-Laguerre rule can only be applied if the dimensionality is low.

- Let us use an alternative, not-so-clever quadrature rule.

• Monte Carlo (Metropolis et al. 1953):

$$\int_{[0,\infty)^d} F(\mathbf{x})\, \rho(\mathbf{x})\, d\mathbf{x} \approx \frac{1}{N} \sum_{n=1}^{N} F(\mathbf{x_n}),$$

with the quadrature nodes $\mathbf{x}_n$ chosen randomly. (The approximate value will then be itself random.)

• Here are the relative errors for two runs (nonsmooth integrand) with $N = 1,000,000$.

d	First run	Second run
1	$8.1(-4)$	$8.0(-4)$
4	$-7.7(-3)$	$-1.2(-2)$
16	$5.9(-3)$	$-1.3(-2)$
64	$7.0(-2)$	$-4.0(-3)$

• Algorithm for finding the nodes ($\mathbf{z}_n$ standard normal, u_n uniform, mutually independent) ($h = 0.20$):

```
x_0 = 0 % or any other point
for n = 1 to N
    x_n^* = |x_{n-1} + h z_n|
    if ρ(x_n^*)/ρ(x_{n-1}) > u_n,  x_n = x_n^*
    else x_n = x_{n-1}
next n
```

$\mathbf{x}_0 = \mathbf{0}$ % or any other point

for $n = 1$ to N

$\mathbf{x}_n^* = |\mathbf{x}_{n-1} + h\mathbf{z}_n|$

if $\rho(\mathbf{x}_n^*)/\rho(\mathbf{x}_{n-1}) > u_n$, $\mathbf{x}_n = \mathbf{x}_n^*$

else $\mathbf{x}_n = \mathbf{x}_{n-1}$

next n

NOTE:

• No a priori knowledge of weight function: will work for all or them. Randomness used systematically to 'discover' locations of large weight and place quadrature nodes accordingly.

• No great theory needed to design algorithm.

• Application of algorithm requires repetition of simple steps (formula useful in computer days).

“Learning is achieved in many steps that are plausible but innocuous when viewed one by one in isolation. These steps work because there is an overall algorithmic plan. In combination the steps achieve something, in particular some kind of convergence” (L. Valiant).

Also implications for evolution in biology.

Are our clever theories (Gaussian quadrature) an emergent result of not-so-clever, randomized algorithms operating in the human mind?

Is our blindness to randomness the result of the application of not-so-clever, randomized algorithms operating in the human mind?