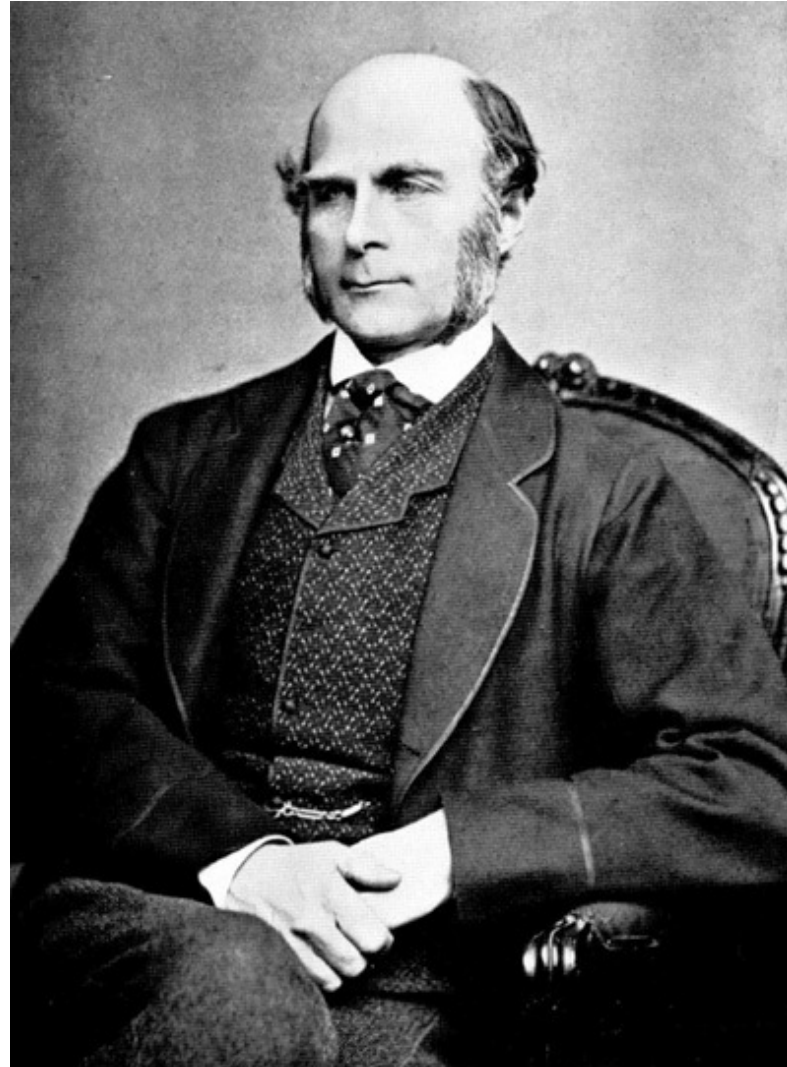


La machine de Galton de 1877
Une première dans la simulation physique de l'a posteriori
par acceptation-rejet



Sources

- Stigler S M (2011) Galton visualizing Bayesian Inference, *Chance*, 24, 8-10
- Stigler S M (2010) Darwin, Galton and the statistical enlightenment, *JRSS A*, 173, 469-482
- Stigler S M (1986) *The history of statistics: the measurement of uncertainty before 1900*. The Belknap Press of Harvard University Press, 410p

Galton

- Sir Francis Galton:1822-1911
- (1/2) Cousin de Charles Darwin par Erasmus Darwin
- Travaux en Statistique, Géographie, Météorologie, Psychométrie, Sociologie, Biométrie, Hérité
- Inventeur de la « régression » linéaire et application à la relation parent-descendant
- Inventeur de diverses machines
- Image contrastée chez les statisticiens et les généticiens (fondateur de l'eugénisme)

Galton et la régression

Table 8.1. Galton's 1885 cross-tabulation of 928 adult children born of 205 midparents, by their height and their midparent's height.

Height of the midparent in inches	Height of the adult child														Total no. of adult children	Total no. of midparents	Medians
	<61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7			
>73.0	—	—	—	—	—	—	—	—	—	—	—	1	3	—	4	5	—
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4	19	6	72.2
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—	219	49	68.2
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—	211	33	67.6
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—	78	20	67.2
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—	66	12	66.7
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—	23	5	65.8
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—	14	1	—
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	—
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—	—	—	—

Source: Galton (1886a).

Note: All female heights were multiplied by 1.08 before tabulation. Galton added an explanatory footnote to the table: "In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62.2, 63.2, &c., instead of 62.5, 63.5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents." Galton republished these data in 1889, where they are referred to as the R.F.F. Data (Record of Family Faculties); he then noted that the first row must be in error (four children cannot have five sets of parents), but he claimed that "the bottom line, which looks suspicious, is correct" (p. 208).

Galton et la régression: concept d'héritabilité

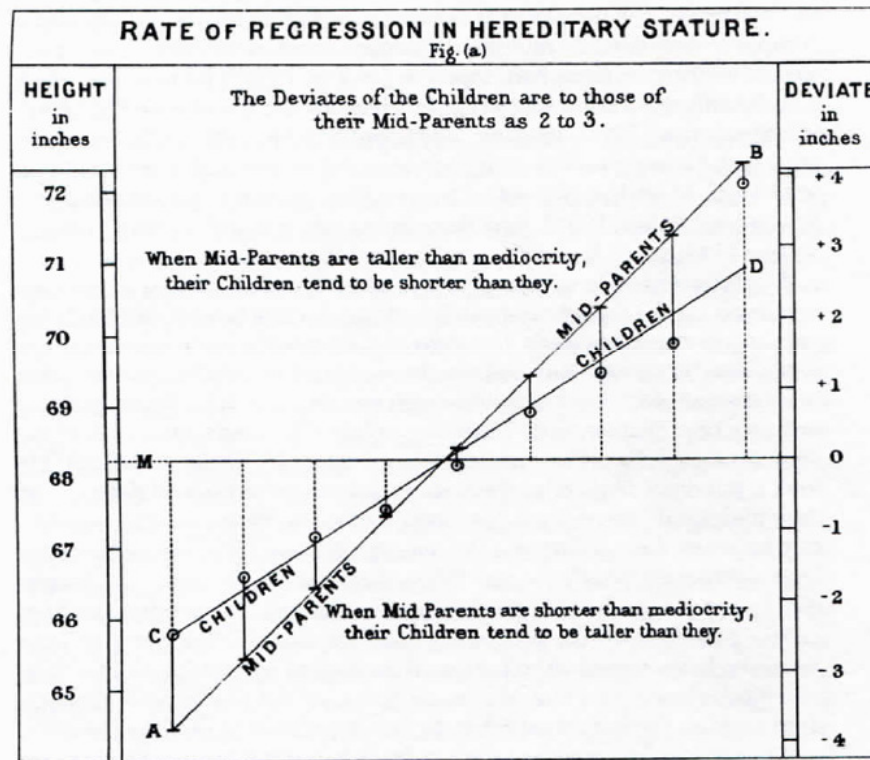


Figure 8.8. Galton's graphical illustration of regression; the circles give the average heights for groups of children whose midparental heights can be read from the line AB. The difference between the line CD (drawn by eye to approximate the circles) and AB represents regression toward mediocrity. (From Galton, 1886a.)

Galton et le théorème CL

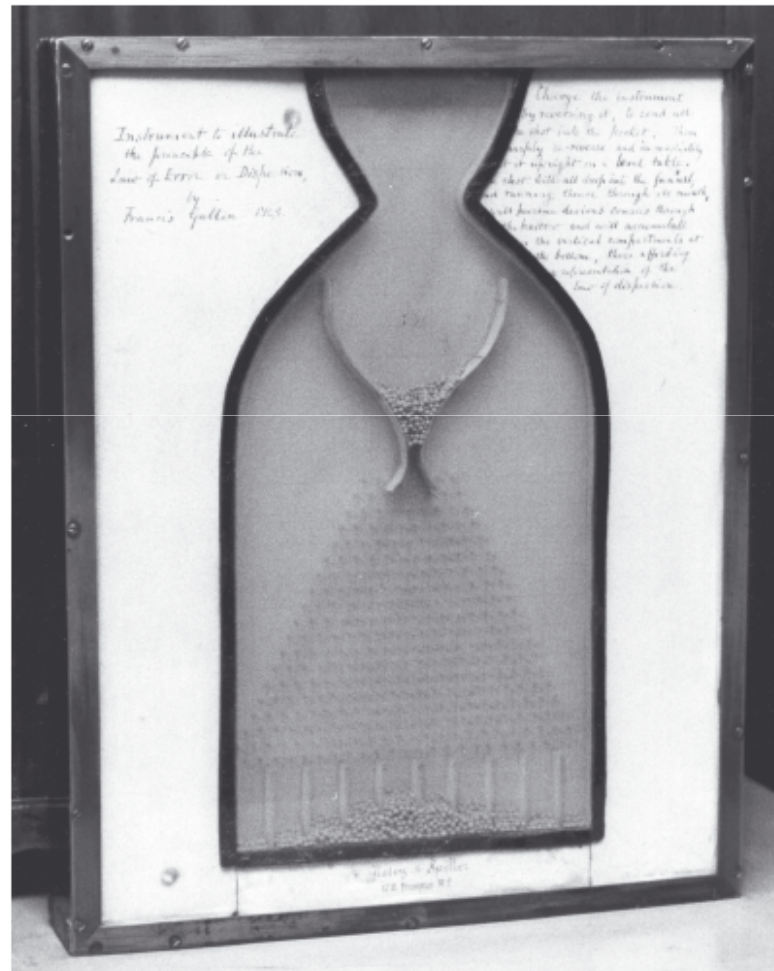
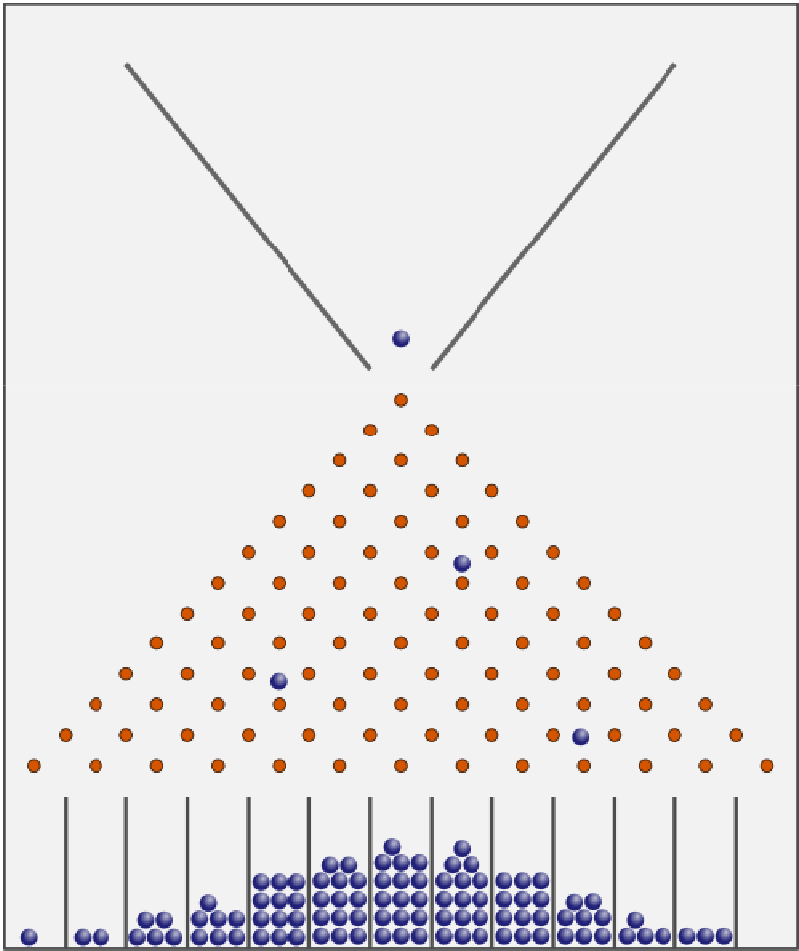
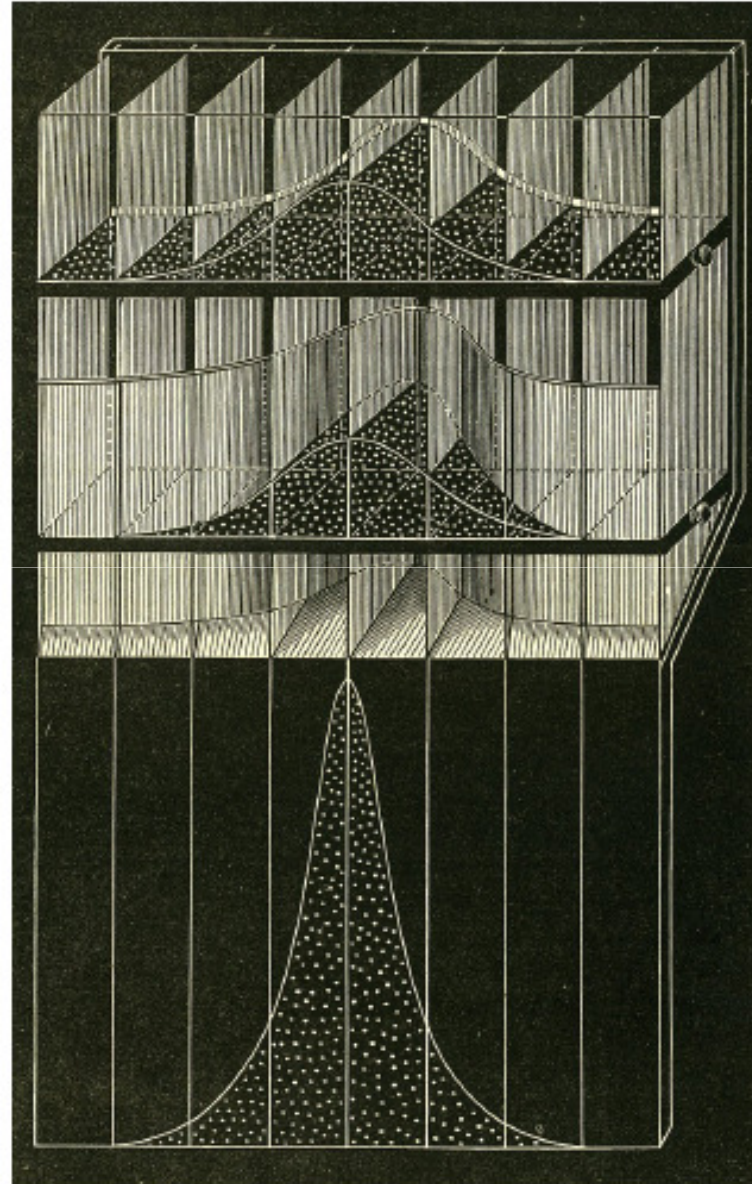


Planche de Galton



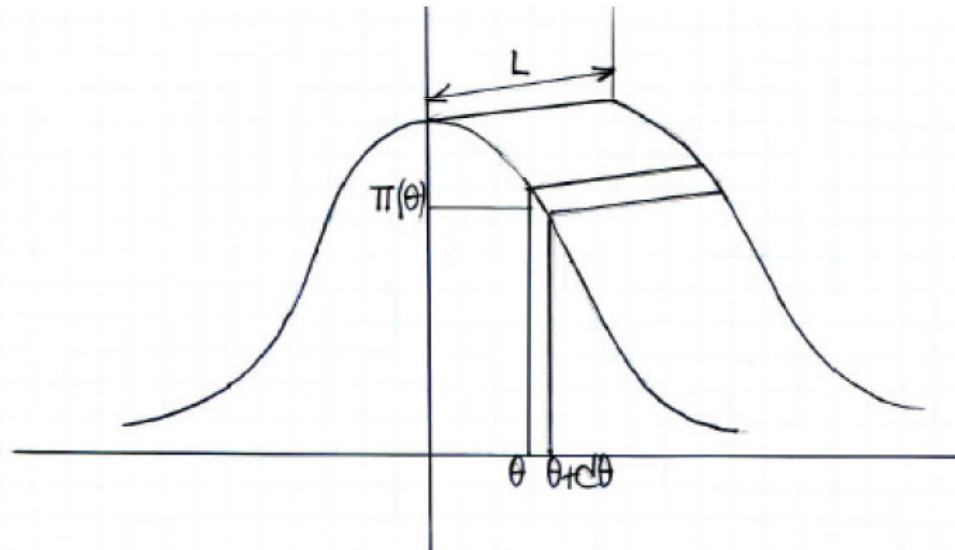
Machine de 1877



Justification

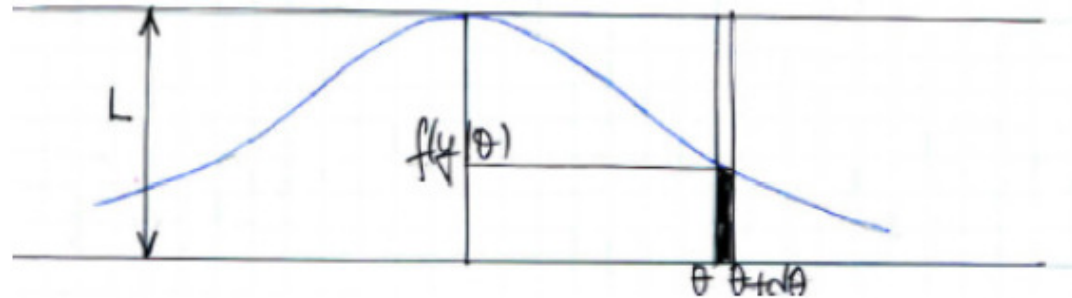
3ème étage : A-priori

Considérons l'ensemble des particules θ de la tranche de volume : $\underbrace{\pi(\theta)}_{\text{hauteur}} \underbrace{L}_{\text{profondeur}} \underbrace{d\theta}_{\text{largeur}}$, $\pi(\theta)$ désignant la densité a priori de θ



Justification

2^{ème} étage vu du dessus à la verticale



Une fois la trappe du 3^{ème} étage ouverte, les particules du volume du 3^{ème} étage se déversent au deuxième étage.

Une fraction $\frac{f(y|\theta)}{L}$ est conservée, le reste étant rejeté derrière le mur de la « vraisemblance » $f(y|\theta)$. Le

volume qui tombe dans l'intervalle $d\theta$ est donc $\pi(\theta) \cdot L \cdot d\theta \cdot \frac{f(y|\theta)}{L} = \pi(\theta) f(y|\theta) d\theta$

Cette phase est remarquable puisqu'elle mime exactement un algorithme d'acceptation rejet pour générer la distribution cible a-posteriori à partir d'un candidat tiré dans l'a priori avec une probabilité d'acceptation de

$\frac{f(y|\theta)}{\text{Max}_\theta f(y|\theta)}$ où $\text{Max}_\theta f(y|\theta) = L$: cf Gelfand & Smith, 1992, §4.1, The American Statistician, 46, 84-88

Justification

1^{er} étage. Il résulte de l'ouverture de la trappe du second étage. C'est la dernière opération qui vise à transférer hauteur x profondeur sur une seule dimension dans l'intervalle $d\theta$ et qui correspond donc, à une constante près, à l'a-posteriori

$$\pi(\theta|y)d\theta \propto \pi(\theta)f(y|\theta)d\theta$$

Rappel algo AR

$f(x)$: densité cible (à une cste de proportionnalité près)

$g(x)$: densité de la distribution instrumentale

$t(x) = Mg(x)$: enveloppe supérieure de $f(x)$

. $f(x)$ & $g(x)$ même support

. trouver un majorant à $f(x)/g(x)$

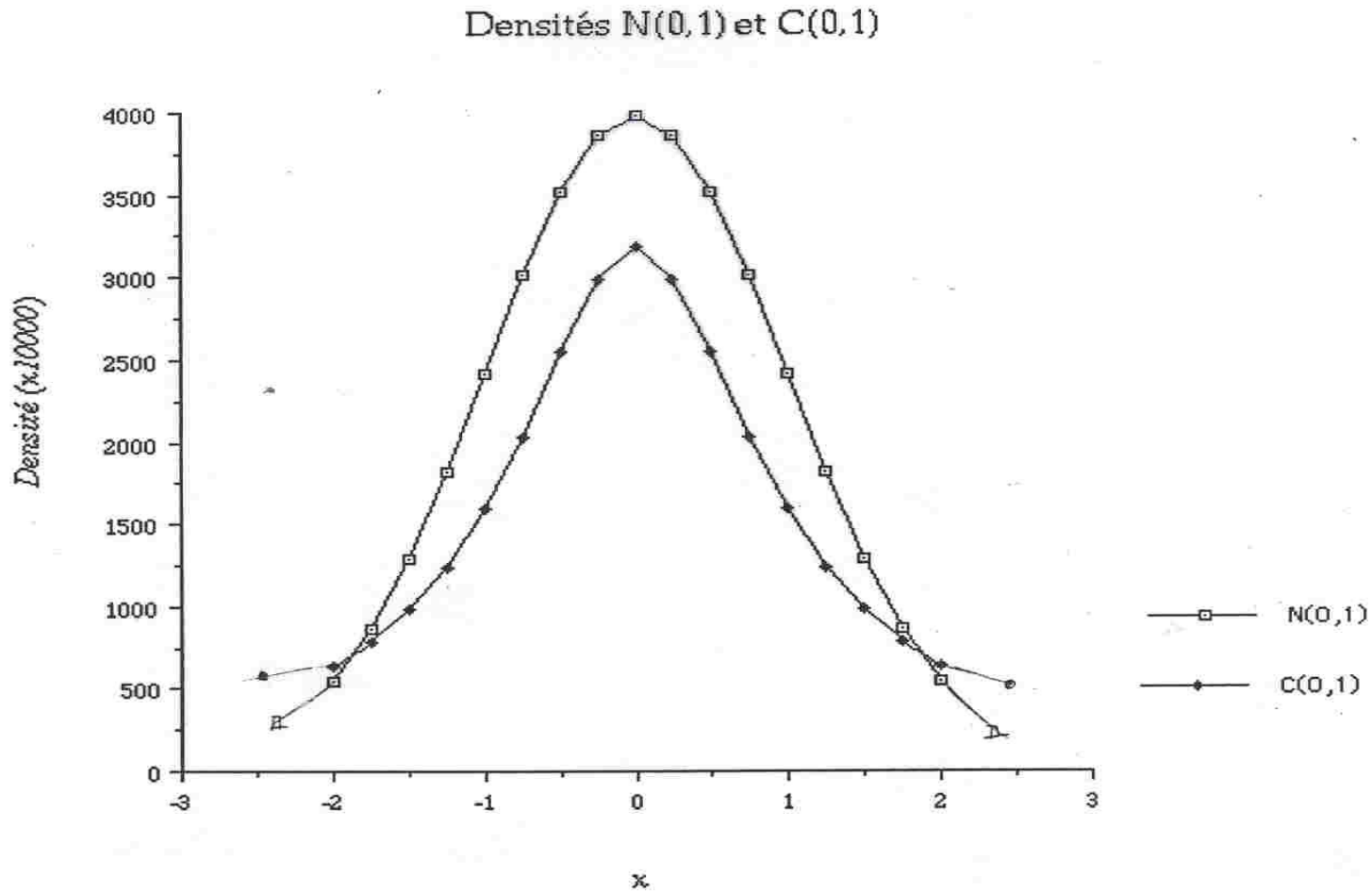
$$\text{ie } f(x)/g(x) \leq M \Rightarrow \underbrace{f(x)/Mg(x)}_{t(x)} \leq 1$$

1) Tirer U_{0-1} et x de g

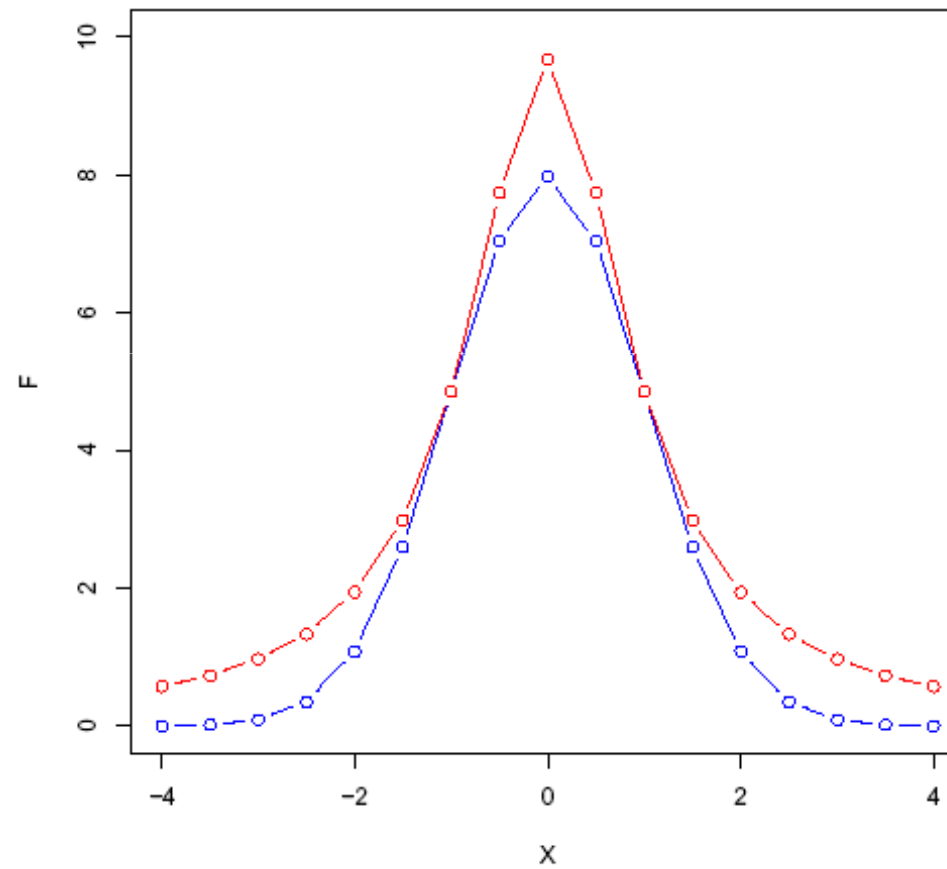
2) Calculer $\lambda(x) = \frac{f(x)}{Mg(x)}$

3) Si $u \leq \lambda(x)$, garder x , sinon revenir en 1)

Ex: simulation de la normale par la Cauchy



Ex: enveloppe de la normale via $M=1.52$ fois la Cauchy



Rappel algo AR

Prendre pour distribution cible : $\pi(\theta | y) \propto \pi(\theta) f(y | \theta)$

et distribution instrumentale : $\pi(\theta)$

Le rapport des deux a pour majorant : $M = \text{Max}_{\theta} f(y | \theta)$

(noté L sur la figure de la machine)

On accepte θ avec la probabilité

$$\frac{\pi(\theta | y)}{M \pi(\theta)} = \frac{f(y | \theta)}{\text{Max}_{\theta} f(y | \theta)}$$

Cf par ex Gelfand & Smith 1992, Bayesian statistics without tears,
The American Statistician, 46, 94-98

Extrait Gelfand & Smith, 1992

4.1 Prior to Posterior

How does Bayes's Theorem generate a posterior sample from a prior sample? For fixed x , define $f_x(\theta) = l(\theta; x)p(\theta)$. If $\hat{\theta}$ maximizes $l(\theta; x)$, let $M = l(\hat{\theta}; x)$. Then with $g(\theta) = p(\theta)$, we may immediately apply the rejection method of Section 3.1 to obtain samples from the density corresponding to the standardized f_x , which, from (1.1), is precisely the posterior density $p(\theta|x)$. Thus, we see that Bayes's Theorem, as a mechanism for generating a posterior sample from a prior sample, takes the following simple form: for each θ in the prior sample accept θ into the posterior sample with probability

$$\frac{f_x(\theta)}{Mp(\theta)} = \frac{l(\theta; x)}{l(\hat{\theta}; x)},$$

otherwise reject it.

Conclusion

- Contribution fondamentale à la statistique (lois conditionnelles)
- Approche de la génération de l'a posteriori très originale
- Jusqu'alors (ex Bayes, Laplace), utilisation d'un a priori uniforme
- 1^{er} exemple d'utilisation de lois conjuguées normales
- Importance du concret en science
 - Expérience de pensée
 - Relation science-technique
 - Main à la pâte