

Des modèles mathématiques et statistiques pour étudier les cancers

M. Filoche

Laboratoire de Physique de la Matière Condensée

CNRS, Ecole Polytechnique

*En collaboration avec **L. Schwartz**,
Hôpital de la Pitié-Salpêtrière*

Mortalité et statistiques

Extrait du rapport du Pr Dautzenberg sur le tabagisme passif

« ...Vivre régulièrement dans une atmosphère enfumée au travail ou à la maison augmente de **26 %** le risque de décès pour les cardiaques. Et de **25 %** celui de développer un cancer du poumon chez les non-fumeurs. »

- **Statistique et probabilité** équivalentes ?
- Quels **modèles mathématiques** de la mortalité ?

Problématique

Est-il anodin de résumer les statistiques de mortalité par des probabilités ?

Quelles sont les conséquences implicites de l'emploi des probabilités ?

Peut-on détecter dans les statistiques de mortalité par cancer des « signatures » de lois de probabilité ?

Les statistiques « normales »

- Loi de probabilité possédant **moyenne** et **écart-type**
- Exemple d'une telle loi, à deux états : **la loi binomiale**

$$P(X=1)=p$$

$$P(X=0)=1-p$$

- Moyenne d'une somme de N tirages :

$$S_N = pN$$

- Ecart-type correspondant :

$$\sigma_N = \sqrt{p(1-p)N}$$

L'agrégation Gaussienne

- Si une variable X suit une loi qui possède une **moyenne** et un **écart-type**, la somme de N variables (X_1, \dots, X_N) **indépendantes** converge vers une loi **Gaussienne**

$$P(X_1 + X_2 + \dots + X_N = Y) \approx \frac{1}{\sigma \sqrt{2\pi N}} e^{-\frac{(Y - Nm)^2}{2N\sigma^2}}$$

- Valeur moyenne de la somme :

$$\langle Y \rangle = Nm$$

- Ecart-type de la somme :

$$\sigma_Y = \sqrt{N}\sigma = \frac{\sigma}{\sqrt{m}} Y^{1/2}$$

Les statistiques « anormales »

- Loi de probabilité **sans moyenne** ou **sans écart-type**
- Exemple d'une telle loi : **la loi de puissance**

$$P(x=X) \propto X^{-\beta} \quad , \quad 1 < \beta \leq 2$$

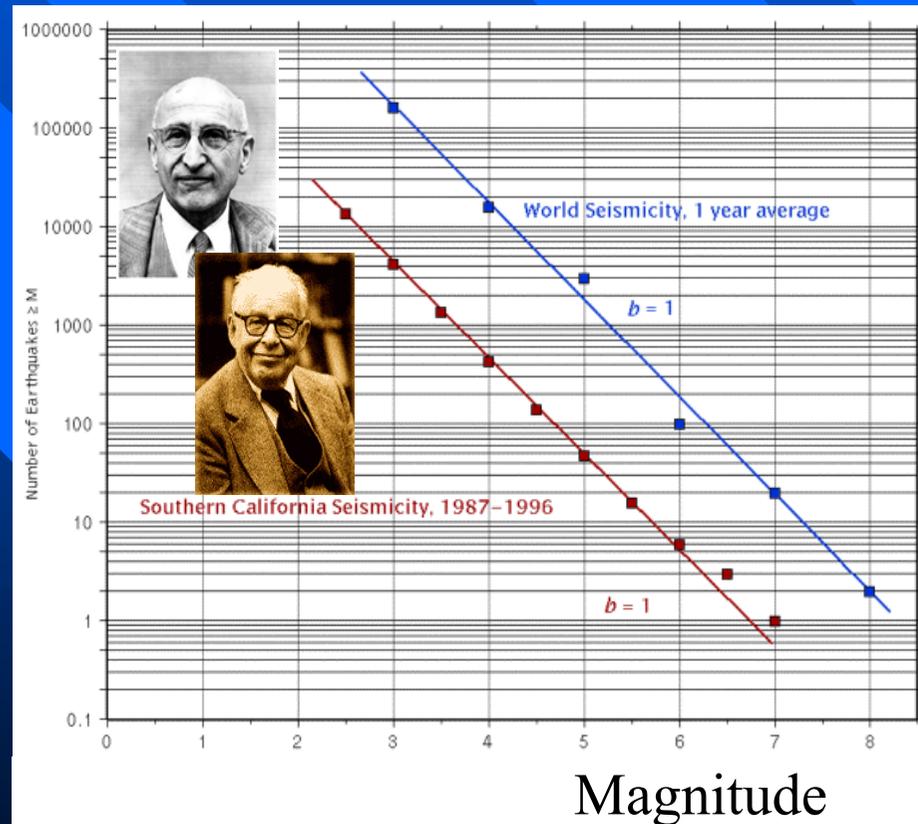
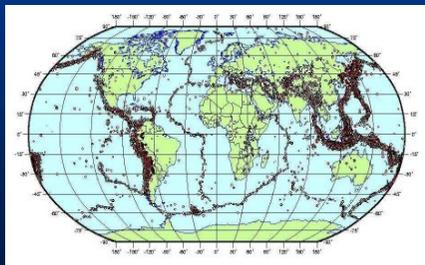
- La moyenne est dominée par l'événement le plus important.

$$\int_1^x x P(X=x) dx \propto x^{1-\beta} \xrightarrow{x \rightarrow \infty} +\infty$$

- L'écart-type varie plus vite que la racine de la moyenne :
pas d'agrégation gaussienne !

Exemple : l'activité sismique

- La loi de distribution des *tremblements de terre* en fonction de leur **énergie** est une loi de puissance (loi de *Gutenberg-Richter*, 1956)



⇒ ***Il n'existe pas de tremblement de terre moyen***

Autre exemple : la finance



- Le Krach de Wall Street en 1987 représente une fluctuation d'environ **20 fois** l'écart-type.

$$P(X = 20\sigma) \approx 10^{-87} !$$

Origines de ces fluctuations

- Interdépendance des événements : ***corrélations fortes***.
- ***Criticité*** : extrême sensibilité aux conditions extérieures, évolution par avalanches internes de toutes tailles.
- Origines de cette ***criticité*** : comportements ***coopératifs*** (finance), ***auto-organisation*** (évolution, trafic), ***vieillesse*** (fracture, corrosion).

Données brutes

% Année de décès: 79 / Département de domicile: France / Sexe: M / Groupe d'âge: 0 /

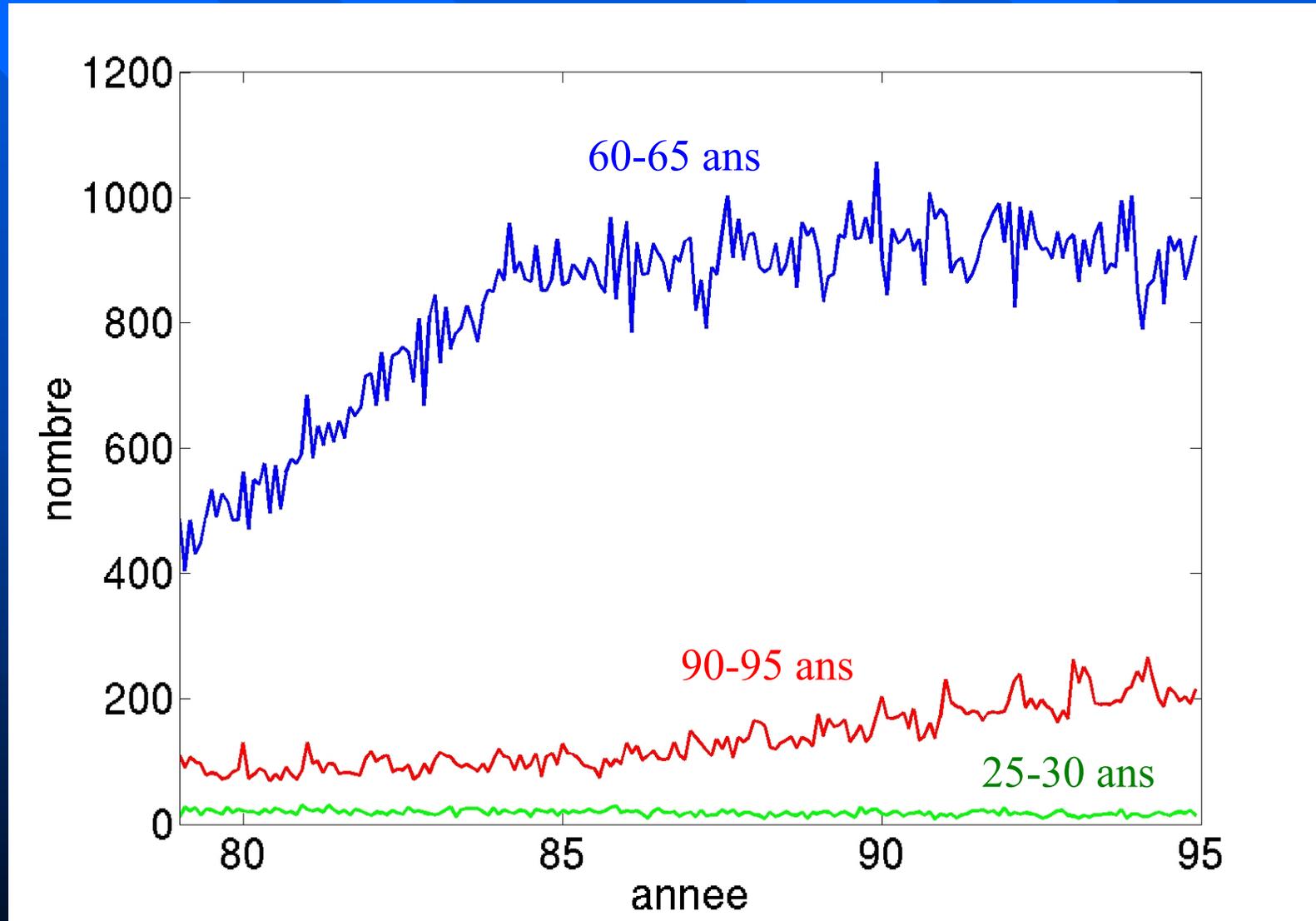
% Mois de décès / Cause principale											
%	TUMEUR	ENDOCR	SANG	MENTAL	NERVEU	CIRCUL	DIGEST	GENITO	PEAU	OSTEO	SENIL
1	3	7	2	0	7	2	1	2	0	206	48
2	0	1	3	0	7	3	3	0	0	171	41
3	2	5	3	0	12	4	2	0	0	175	45
4	3	4	3	0	5	3	3	0	0	152	39
5	1	4	1	0	15	2	5	0	0	170	30
6	0	8	3	0	5	8	3	0	0	163	27
7	5	6	2	0	13	11	5	2	0	201	32
8	1	5	0	0	5	8	3	0	0	163	26
9	2	5	1	0	5	6	3	0	0	140	37
10	1	9	0	0	12	6	10	1	0	169	45
11	1	3	2	0	3	5	4	0	0	169	53
12	0	11	0	0	3	8	7	1	0	164	69

% Année de décès: 79 / Département de domicile: France / Sexe: M / Groupe d'âge: 1 /

% Mois de décès / Cause principale											
%	TUMEUR	ENDOCR	SANG	MENTAL	NERVEU	CIRCUL	DIGEST	GENITO	PEAU	OSTEO	SENIL
1	6	2	4	0	9	4	5	0	1	11	2
2	8	2	1	0	5	5	2	0	0	6	0
3	10	4	8	0	4	2	0	0	0	18	1
4	10	2	0	0	8	2	2	1	0	13	3
5	8	4	1	0	7	3	1	0	0	11	2
6	3	3	1	0	3	2	1	1	0	15	5
7	10	3	2	0	6	7	0	0	0	8	2
8	11	0	0	0	5	2	0	0	0	10	1
9	6	2	2	1	9	1	3	0	1	9	2
10	8	3	5	0	8	4	3	1	0	11	0
11	13	4	2	0	5	0	1	1	0	5	1
12	11	2	5	0	7	2	3	0	0	8	3

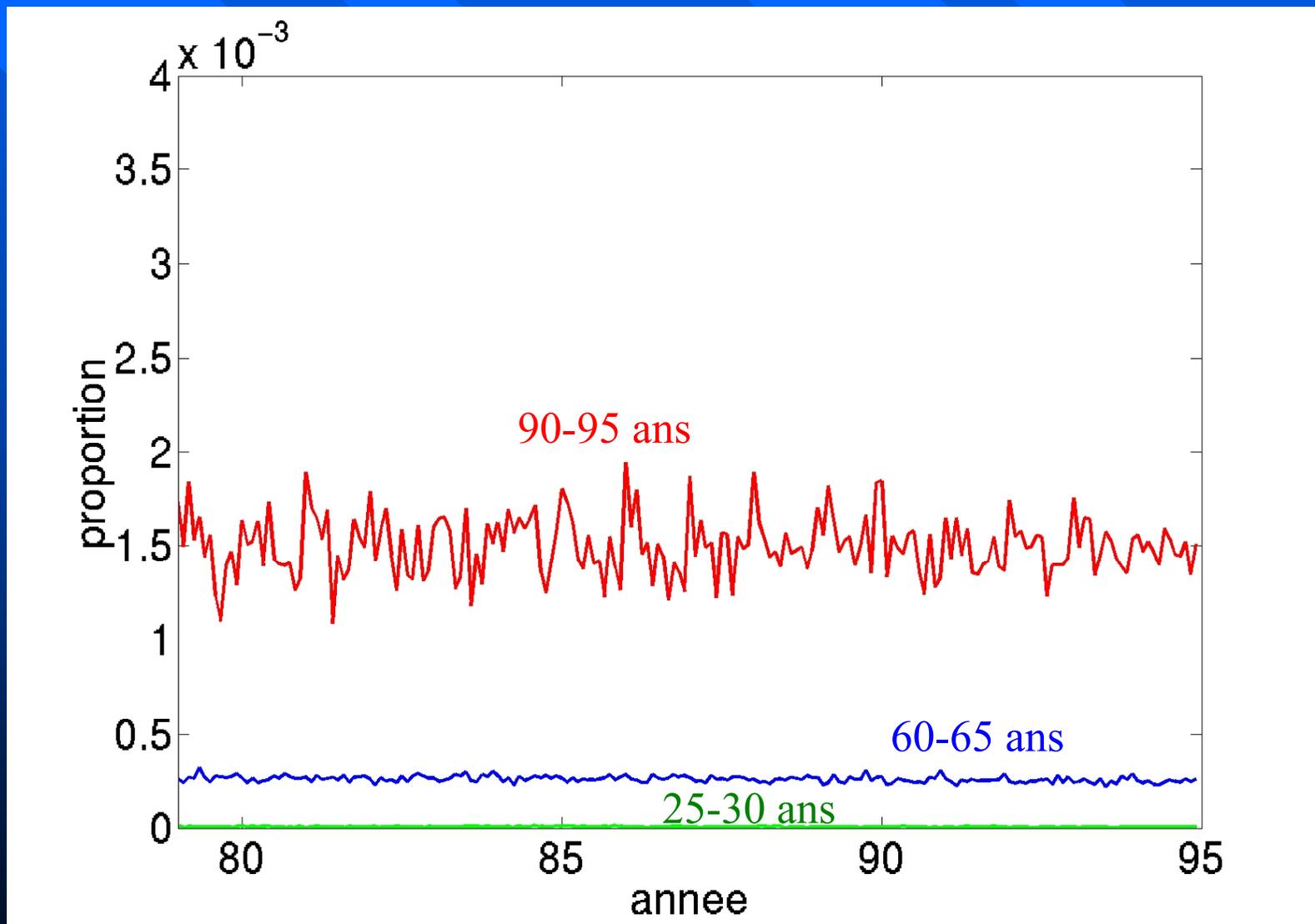
Statistiques de décès

Nombre mensuel de décès par cancer (hommes)



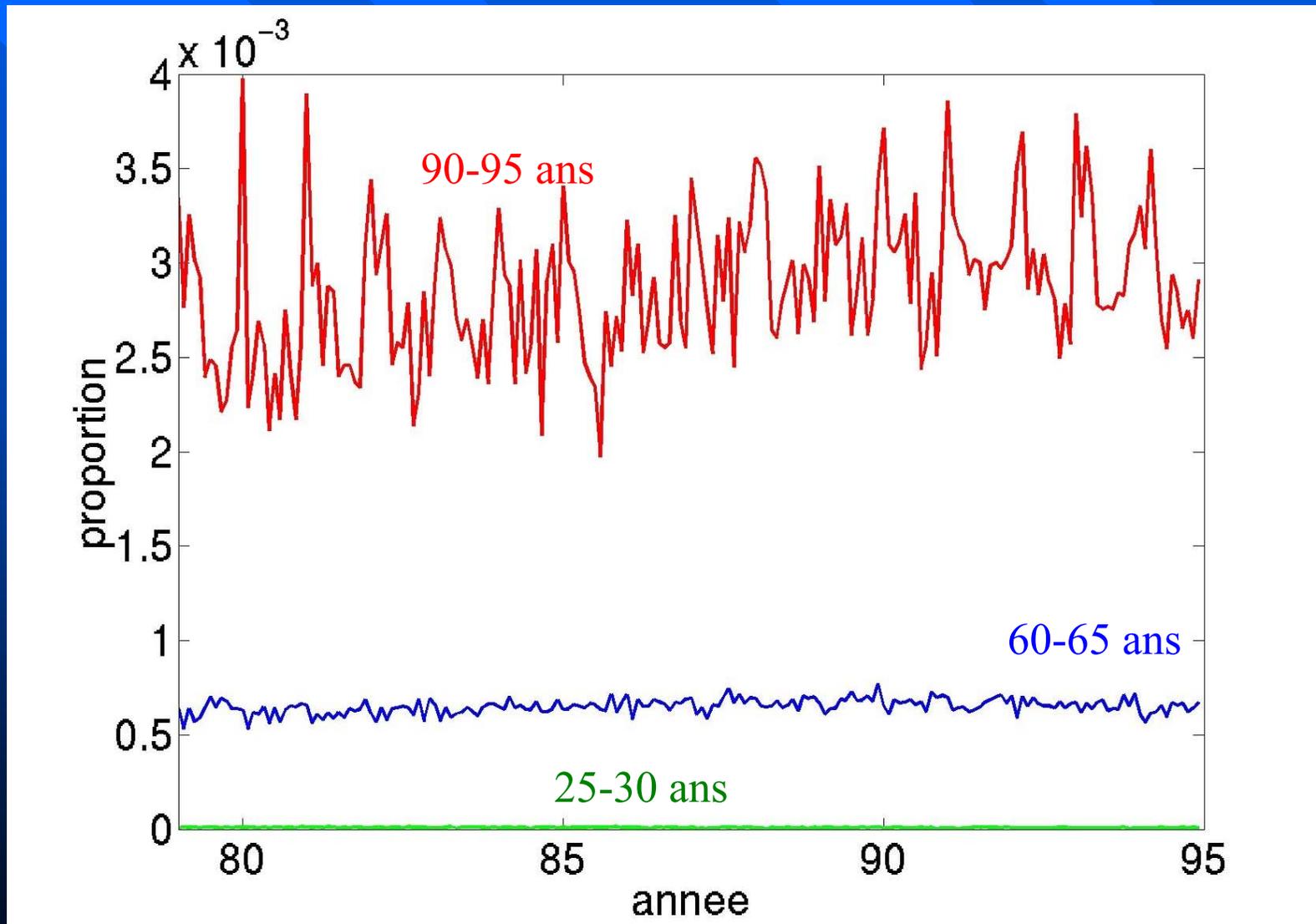
Proportions mensuelles de décès

Femmes

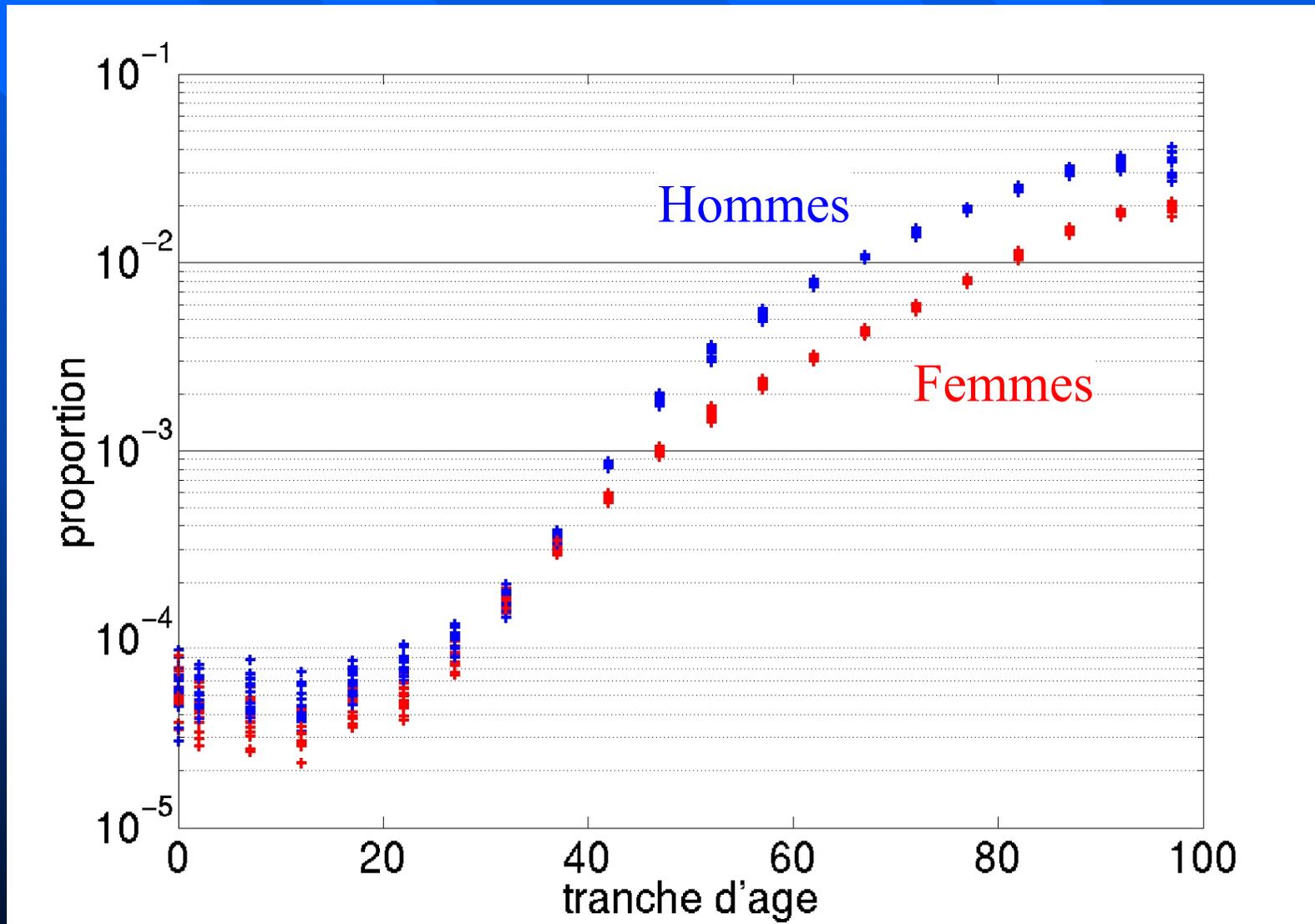


Proportions mensuelles de décès

Hommes

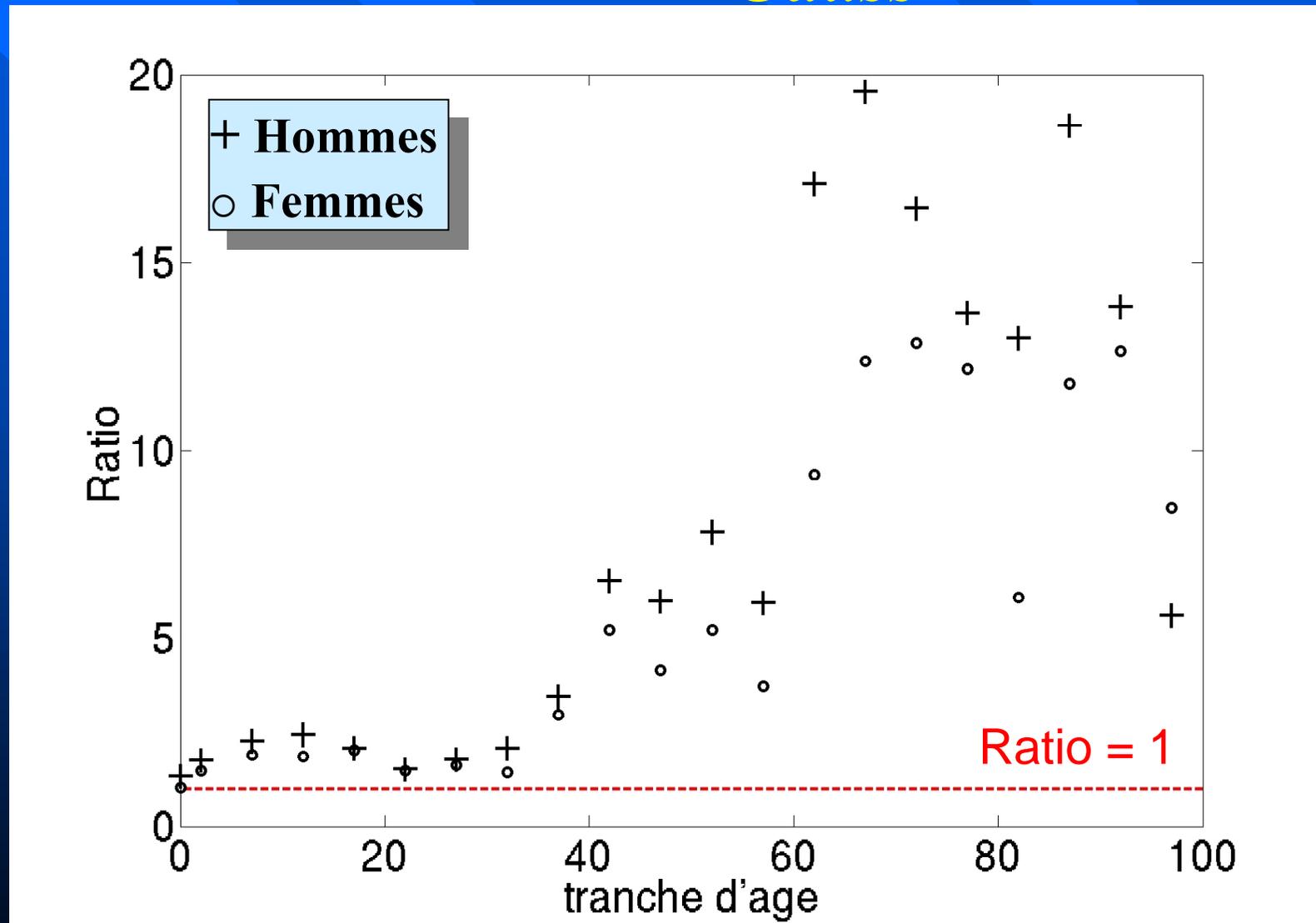


Taux annuels de décès



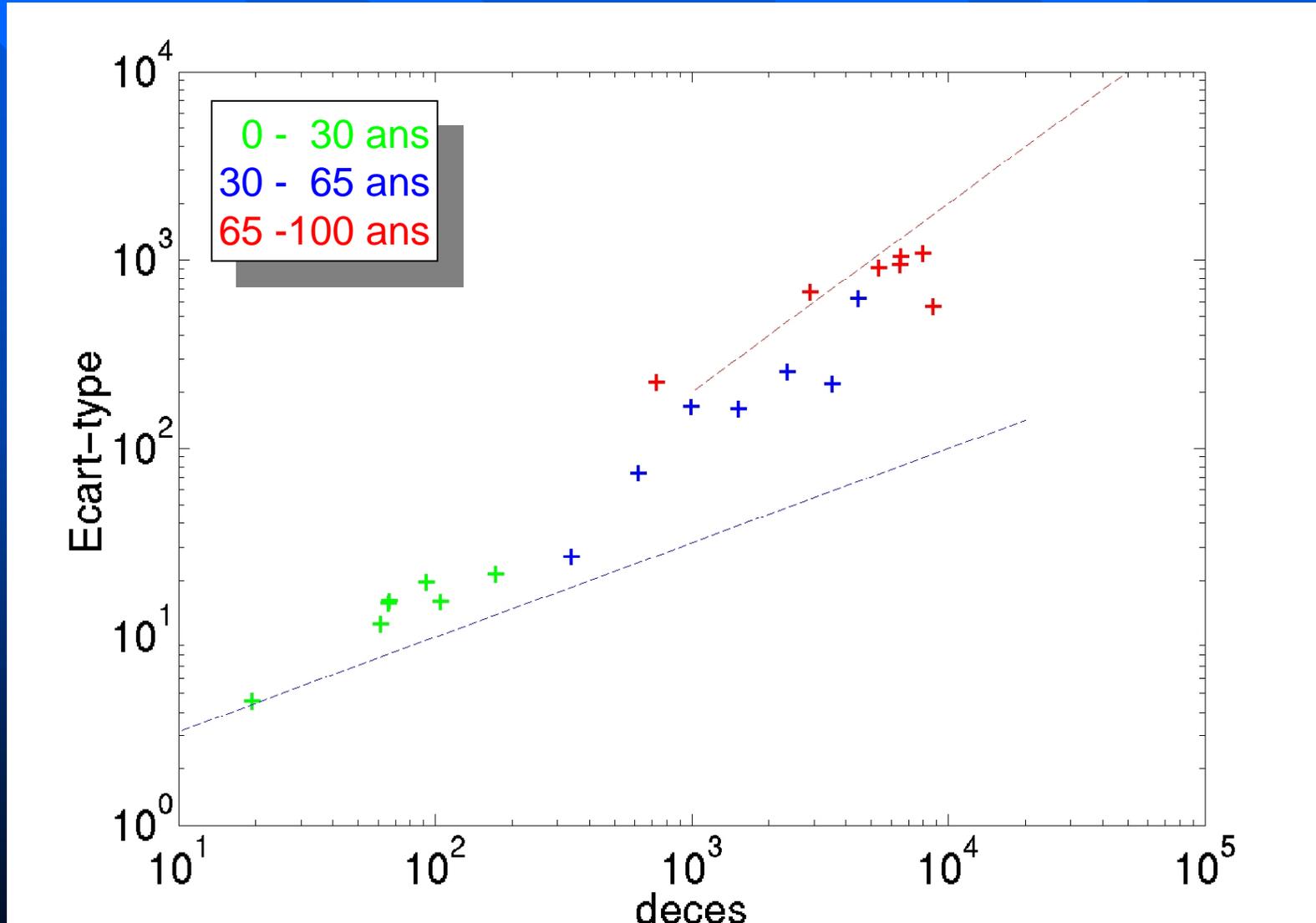
Fluctuations Gaussiennes ?

σ / σ_{Gauss}



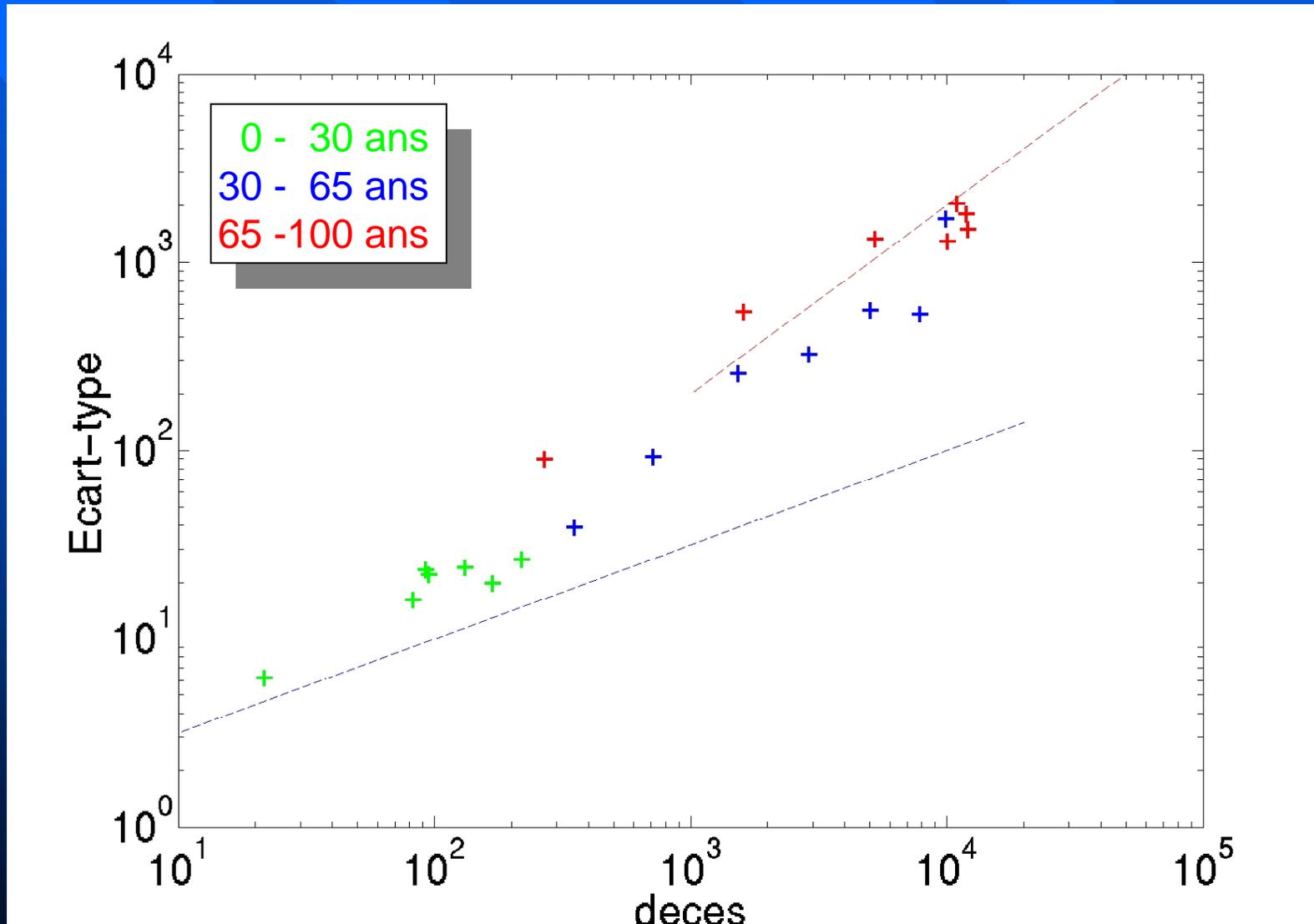
Loi anormale ?

Femmes



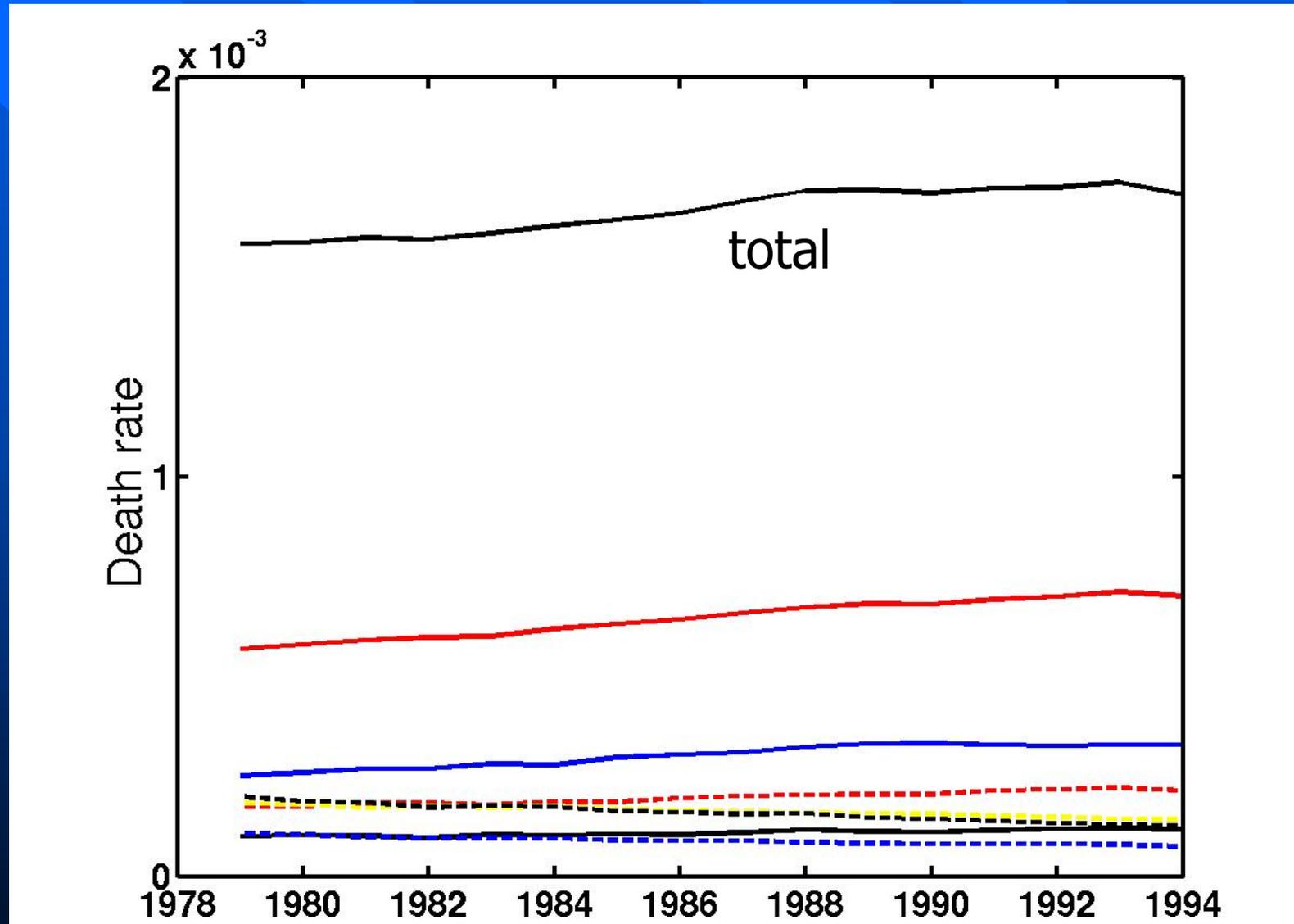
Loi anormale ?

Hommes

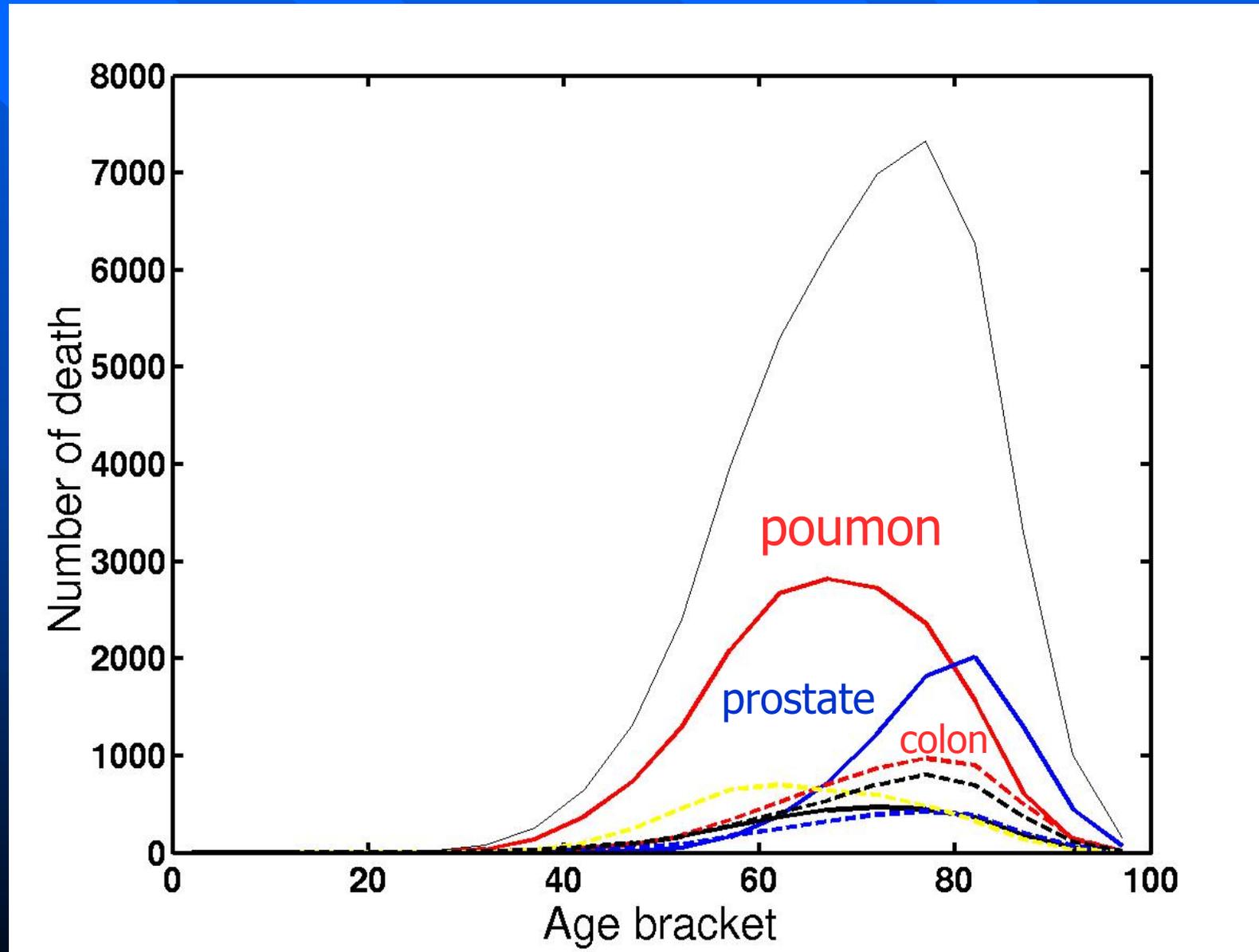


Mortalité par type de cancer

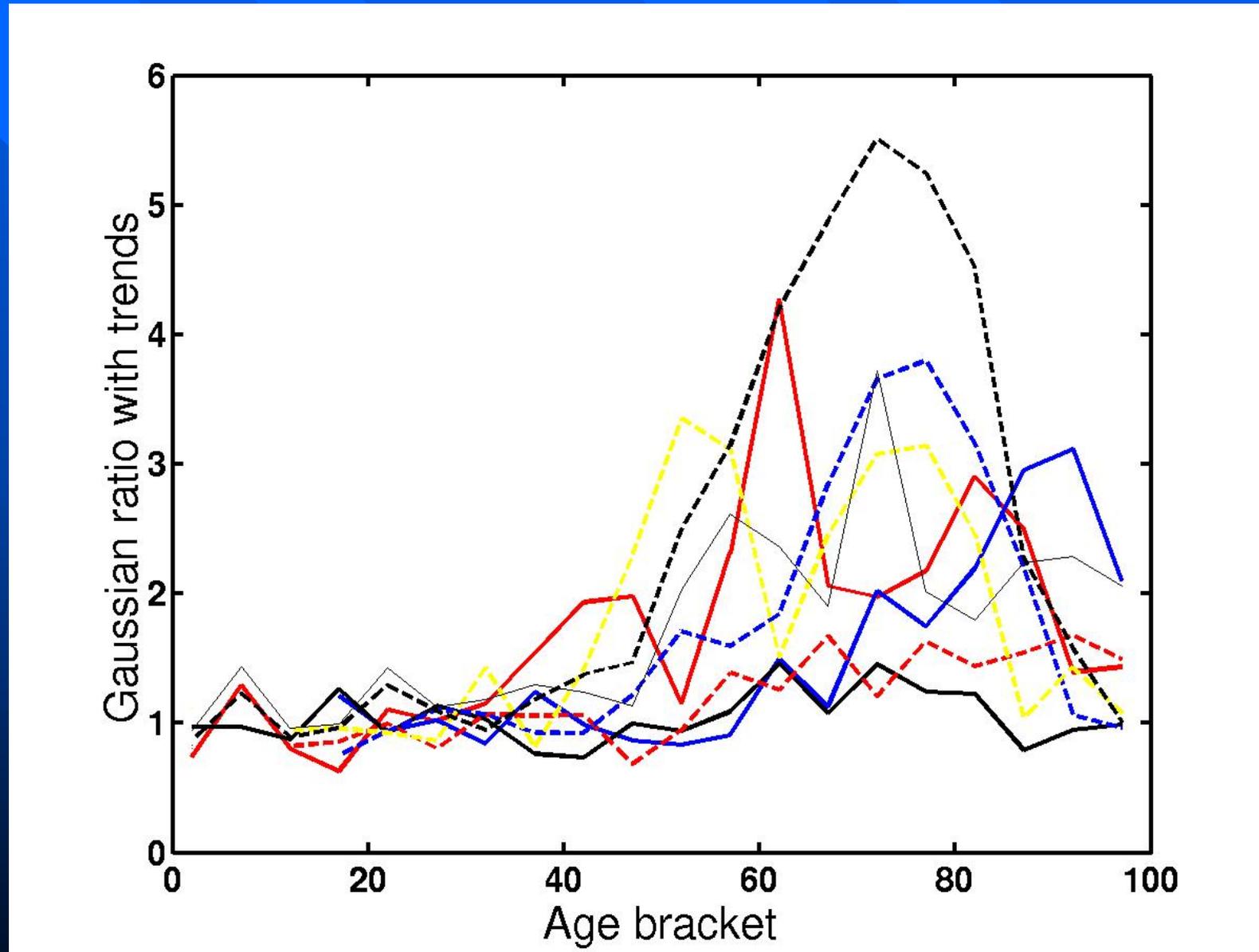
Hommes



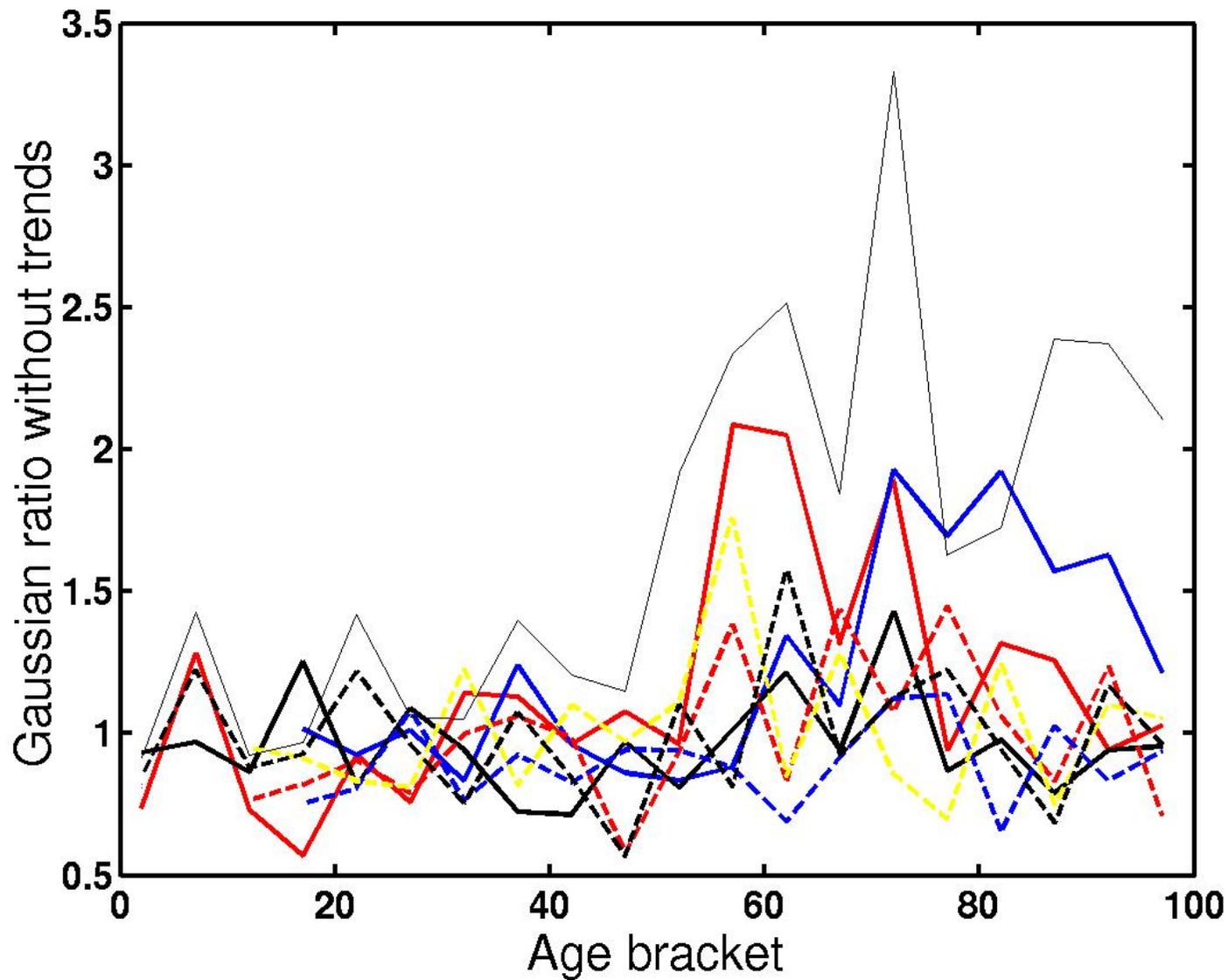
Mortalité moyenne par type de cancer par tranche d'âge



Rapport d'écart-types non corrigés



Rapport d'écart-types corrigés



Conclusion

- Les statistiques de mortalité par cancer présentent des fluctuations **importantes** à partir de 35-40 ans.
- Ces fluctuations sont **incompatibles** avec un modèle **gaussien** issu d'une probabilité individuelle $(p, 1-p)$ d'un système à deux états.
- L'étude par **types de cancer** laisse toujours apparaître des fluctuations importantes, bien que plus faibles. Quid de la **somme constante** ?

Hypothèses et perspectives

- **Vieillesse** de l'organisme, engendrant un comportement **critique** à partir d'un certain âge.
- Cancer « **déterministe** » / « **non-déterministe** »
- **Faibles doses** \Leftrightarrow « **Fatigue** » de l'organisme ?
- **Interprétation** des données très délicates.
Besoin de grandes études épidémiologiques.

Remerciements

- P.Y. Boelle (INSERM)
- J.M. Steyaert (LIX, Ecole Polytechnique)
- M. Gettler-Summa, F. Goupil (CEREMADE, Université Paris-Dauphine)