IX. LOIS DE PROBABILITÉ ASYMPTOTIQUES.

Nous avons rencontré aux chapitres **VII** et **VIII** des situations où apparaissent des lois de probabilité complexes, qu'il est non seulement très difficile, voire impossible d'exprimer analytiquement, mais qu'il est en outre *inutile* ou même nuisible de connaître exactement.

Au chapitre **VII** il s'agissait de la loi d'une somme d'un grand nombre de variables aléatoires indépendantes. Une telle loi peut bien sûr s'exprimer de manière mathématiquement exacte dans certains cas exceptionnels où les régularités sont très fortes (par exemple la loi binômiale), mais en général elle est dans son détail prodigieusement complexe, chaotique, et irréductible à une formule analytique simple (voir figure 17 par exemple). Cependant en distinguant dans la loi exacte un signal (la densité) et du bruit (le "bruit discret"), on peut mettre en évidence une régularité à grande échelle (la forme gaussienne de la densité).

Au chapitre **VIII** nous avons étudié la loi d'un *processus en cascade* (du type réaction de fission nucléaire en chaîne) après un grand nombre de générations. Là aussi, on obtient une densité limite, caractérisée explicitement par sa transformée de Fourier ou fonction caractéristique. Nous avons vu que, si la loi exacte et détaillée est en général prodigieusement complexe et irréductible à une formule analytique, on peut cependant calculer à l'aide d'algorithmes simples toute l'information *utile*, qui consiste en

a) la probabilité de l'extinction ultime ;

b) la densité de probabilité du nombre de particules-filles de la n^{e} génération.

Ces deux situations ont en commun que les lois exactes sont complexes, que les valeurs possibles de la variable considérée sont très nombreuses et qu'il importe peu de connaître la probabilité individuelle de chaque valeur possible ; ce qui est utile est de pouvoir calculer, pour n'importe quel intervalle grand par rapport aux distances séparant les valeurs individuelles, la probabilité pour que ces valeurs se situent dans cet intervalle. Cette information est fournie par la densité limite.

On appelle *lois de probabilités asymptotiques* ces lois limites, qui sont souvent, mais pas toujours, des densités continues. Parfois elles comportent à la fois une partie discrète et une partie continue : par exemple dans le cas du processus en cascade étudié au chapitre **VIII**, la loi asymptotique était composée d'une partie discrète (la probabilité que $Z_n = 0$) et d'une partie continue (pour les valeurs autres que zéro).

Dans ce chapitre nous nous proposons d'étudier quelques lois de probabilité asymptotiques, qu'on rencontre souvent et qu'il faut connaître.

La loi normale est un cas typique. Un autre type classique de loi limite, entièrement discrète, est la loi de Poisson. Les lois limite qu'on rencontre dans la pratique se ramènent souvent à ces deux cas. Par exemple les lois asymptotiques du χ^2 ou de Student (voir chapitre **XI**) ne sont que des *déguisements* (par changement de variable) de la loi asymptotique normale. Dans la section **1** nous verrons un cas particulier de la loi de Student, la *loi de Cauchy*. Dans la section **2** nous étudierons la *loi de Poisson*, qui est une loi limite très fréquente, entièrement discrète : elle régit généralement les événements caractérisés par comptage, tels que le nombre de véhicules qui se présentent chaque minute à un péage d'autoroute.

Il y a une différence de nature importante entre d'une part les lois asymptotiques telles que la loi normale et la loi de Poisson, et d'autre part les lois limite de processus en cascade. En effet, les premières ont un caractère universel : ainsi la somme d'un grand nombre de variables aléatoires indépendantes aura toujours une densité gaussienne, quelle que soit la loi particulière des variables individuelles. De même le nombre de véhicules qui se présentent par minute à un péage sera toujours une loi de Poisson, quelle que soit la loi de probabilité particulière qui régit l'heure ou le lieu de départ de chaque véhicule. Par contre les secondes — comme le montre la diversité des figures à la fin du chapitre **VIII** — sont très variées et dépendent de la loi génératrice Z_1. Mais les premières sont des sommes de variables indépendantes (ce qui a pour effet de brouiller l'information contenue dans les lois particulières), tandis que les secondes ne sont pas des sommes de termes indépendants, et amplifient au contraire certaines caractéristiques de la loi originelle.

IX.1. Loi de Cauchy.

Soient X et Y deux variables aléatoires de lois approximativement gaussiennes. Par exemple X et Y peuvent être binômiales ou être la somme d'un grand nombre de termes indépendants, mais on supposera que la probabilité pour que $Y = 0$ est nulle.

Si X et Y sont stochastiquement indépendantes, quelle est la loi approximative (la densité) de $Z = X/|Y|$?

Le problème est de trouver la densité de la loi de Z sans passer par les lois exactes, à partir de la seule connaissance des densités de X et de Y.

Ces densités sont caractérisées par

$$\begin{aligned} \mathcal{P}\,(a < X < b) \;&\simeq\; \int_a^b \sqrt{\frac{\alpha}{\pi}}\; e^{-\alpha x^2}\; dx \\ \mathcal{P}\,(a < Y < b) \;&\simeq\; \int_a^b \sqrt{\frac{\beta}{\pi}}\; e^{-\beta y^2}\; dy \end{aligned} \qquad (IX.1.)$$

Supposons un instant qu'on connaisse les lois discrètes exactes de X et de Y : $\mathcal{P}\,(X = x_j) = p_j$ et $\mathcal{P}\,(Y = y_k) = q_k$. Alors

$$\mathcal{P}\,(a < Z < b) \;=\; \sum_{j,k \in E} p_j q_k \qquad (IX.2.)$$

où E est l'ensemble des couples d'indices j, k tels que $a|y_k| < x_j < b|y_k|$. Bien entendu, que les termes de la somme ci-dessus soient les produits $p_j q_k$ est dû à l'indépendance stochastique des variables X et Y : sinon on aurait une loi conjointe $r_{j,k}$ non factorisable.

Quoique étant à deux dimensions, cette loi conjointe peut être lissée par convolution : après lissage, la loi discrète $\{x_j, p_j\}$ donnera une densité $\sqrt{\alpha/\pi}\, e^{-\alpha x^2}$ et la loi $\{y_k, q_k\}$ une densité $\sqrt{\beta/\pi}\, e^{-\beta y^2}$, de sorte que la loi conjointe $\{x_j, y_k, p_j q_k\}$ donnera une densité $(\sqrt{\alpha\beta}/\pi)\, e^{-\alpha x^2 - \beta y^2}$. Par ailleurs, l'événement $a < X/|Y| < b$ correspond, dans le plan des coordonnées x, y, au domaine $a|y| < x < b|y|$, donc l'approximation des lois discrètes par des densités donnera

$$\mathcal{P}\,(a < Z < b) \quad \simeq \iint\limits_{a|y| < x < b|y|} \frac{\sqrt{\alpha\beta}}{\pi}\; e^{-\alpha x^2 - \beta y^2}\; dx\, dy \qquad (IX.3.)$$

Il ne reste plus qu'à calculer cette intégrale double.

Pour cela il suffit d'effectuer de bons changements de variable ; la figure 34 montre quel est le domaine d'intégration $a|y| < x < b|y|$ (secteur hachuré, délimité par les droites $x = a|y|$ et $x = b|y|$). La figure montre le cas où a et b sont positifs. Introduisons les angles u et v tels que $\tan u = a$ et $\tan v = b$. En coordonnées polaires r, θ telles que $x = r \sin\theta$ et $y = r\cos\theta$ le secteur hachuré correspond à $0 \leq r < \infty$, $u < \theta < v$. Par conséquent, dans ces coordonnées polaires l'intégrale précédente devient :

$$2\, \frac{\sqrt{\alpha\beta}}{\pi} \int_0^\infty\!\!\int_u^v e^{-[\alpha \sin^2\theta + \beta\cos^2\theta]\, r^2}\; r\, dr\, d\theta$$

Le facteur 2 devant provient de ce que dans $IX.3$ on intègre aussi sur le secteur $-ay < x < -by$ correspondant à $y < 0$ alors que ci-dessus (voir

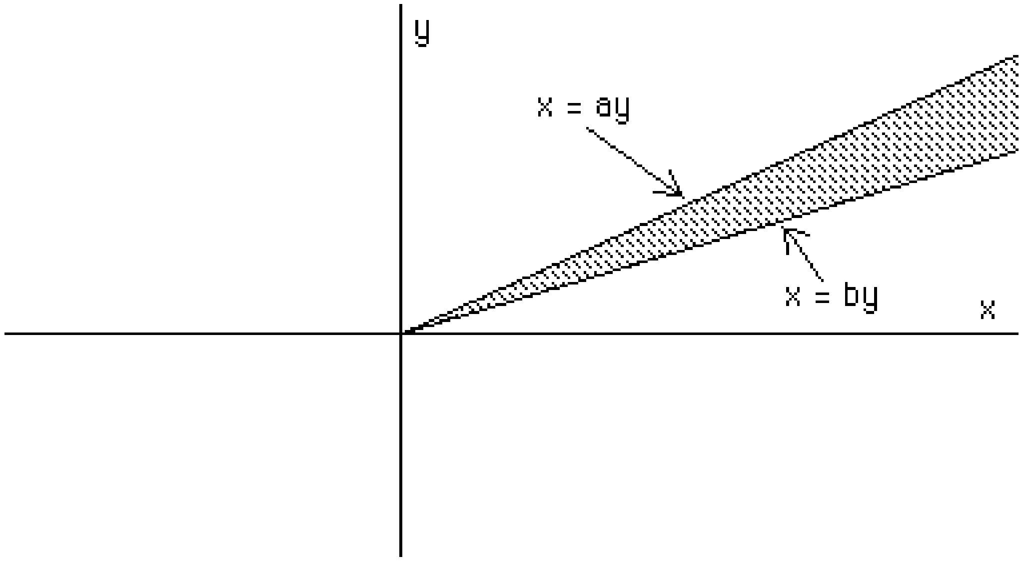

figure 34

figure 34) on n'a retenu que le cas $y > 0$. L'intégration par rapport à la variable r est immédiate, et on obtient

$$\frac{\sqrt{\alpha\beta}}{\pi}\int_u^v \frac{1}{\alpha\sin^2\theta+\beta\cos^2\theta}\,d\theta$$

On effectue alors un second changement de variable: $t = \tan\theta$; l'intégrale devient maintenant

$$\frac{\sqrt{\alpha\beta}}{\pi}\int_a^b \frac{1}{\alpha\,t^2+\beta}\,dt$$

On peut donc écrire en conclusion:

$$\mathcal{P}\,(a < Z < b) \;\simeq\; \frac{\sqrt{\alpha\beta}}{\pi}\int_a^b \frac{1}{\alpha\,t^2+\beta}\,dt \qquad (IX.4.)$$

ce qui signifie simplement que la variable Z a pour densité la fonction

$$\frac{\sqrt{\alpha\beta}}{\pi}\,\frac{1}{\alpha\,t^2+\beta}$$

On pourra remarquer que

$$\frac{\sqrt{\alpha\beta}}{\pi}\int_{-\infty}^{+\infty} \frac{1}{\alpha\,t^2+\beta}\,dt \;=\; 1$$

comme il se doit.

Cette densité (ou loi asymptotique) est appelée *densité de Cauchy*.

La densité de Cauchy décroît lentement à l'infini (en t^{-2}). Cela entraîne que la moyenne ou espérance mathématique est donnée par une intégrale divergente, $\int [t/(\alpha\, t^2 + \beta^2)]\, dt$; à plus forte raison encore la variance. Mais bien sûr la loi discrète exacte de Z, n'ayant qu'un nombre fini de valeurs possibles, a bien une moyenne, qui est d'ailleurs pratiquement nulle, et une variance (très grande).

Les variables X et Y avaient chacune une loi de probabilité gaussienne (de variances $1/2\alpha$ et $1/2\beta$ respectivement). Pour donner une illustration concrète à cela, imaginons que deux personnes $\mathcal{A}$ et $\mathcal{B}$ jouent chacune à pile ou face avec des partenaires respectifs $\mathcal{A}'$ et $\mathcal{B}'$, en payant 1 euro au partenaire chaque fois que pile sort, et en recevant 1 euro du partenaire chaque fois que face sort. Sur un grand nombre N de lancers, il est très peu probable que les gains ou les pertes de $\mathcal{A}$ soient supérieures à trois ou quatre fois $\sqrt{N}$ (voir chapitre **VII**) : c'est la loi gaussienne. Mais supposons que le joueur $\mathcal{A}$, au lieu de compter son gain en euros, le rapporte au gain en valeur absolue de $\mathcal{B}$, qu'il dise par exemple : "j'ai gagné le double de $\mathcal{B}$" ($Z = 2$), ou bien "j'ai gagné 3.7 fois ce que $\mathcal{B}$ a perdu" ($Z = 3.7$), ou encore "j'ai perdu 1.75 fois plus que $\mathcal{B}$" ($Z = -1.75$), etc. Afin qu'il n'y ait pas de division par zéro, on supposera N impair, car on sait (voir chapitre **III**) que dans ce cas les gains sont toujours un nombre impair d'euros et ne peuvent donc pas être nuls. Alors ce gain relatif est distribué selon la densité de Cauchy. Or la densité de Cauchy, au contraire de la densité gaussienne, ne décroît pas très rapidement lorsque t tend vers l'infini. Les rapports $X/|Y|$ élevés ont certes des probabilités plus faibles que les rapports proches de 1, mais pas infinitésimales. Cela n'est pas paradoxal : $X/|Y|$ est élevé si X est grand (ce qui est très peu probable), mais aussi si Y est petit (ce qui par contre est probable).

La variable aléatoire Z, c'est-à-dire le rapport des gains de $\mathcal{A}$ à ceux (en valeur absolue) de $\mathcal{B}$ est bien sûr une variable aléatoire discrète, qui ne prend jamais qu'un nombre fini de valeurs ; ces valeurs sont des fractions et sont d'autant plus nombreuses et serrées que N (le nombre de lancers) est plus grand. Pour N grand ces valeurs sont alors bien trop nombreuses et proches les unes des autres pour que la connaissance mathématiquement exacte de la probabilité de chacune soit utile ; en outre son expression exacte est compliquée. C'est pourquoi la densité de Z est à nouveau la seule information intéressante.

IX.2. Pourquoi la densité gaussienne est-elle si répandue ?

Le calcul par lequel nous avons obtenu la densité de Cauchy n'était qu'une affaire de changements de variables. Il laisse donc le sentiment que n'importe quelle fonction positive et dont l'intégrale vaut 1 peut être la densité d'une variable aléatoire gaussienne, après avoir effectué sur celle-ci une transformation non linéaire adéquate. Nous retrouvons ici encore le phénomène qui était à l'origine du paradoxe des cordes prises au hasard sur un cercle: si une certaine variable aléatoire a une densité uniforme, son carré ou sa racine carrée n'auront pas une densité uniforme. Ici, au lieu de densités uniformes, nous considérons ce qu'il advient des densités gaussiennes après une transformation non linéaire. N'importe quoi (pourvu que ce soit positif et d'intégrale 1) pourrait après une transformation non linéaire adéquate se ramener à une densité gaussienne. D'un point de vue purement mathématique, cela est vrai; donnons-nous en effet une densité arbitraire $f(x)$. Tout ce que nous lui imposons est que

$$f(x) \;\geq\; 0 \qquad \text{et} \qquad \int f(x)\,dx \;=\; 1$$

Prenons alors une variable aléatoire X de densité gaussienne $\sqrt{\alpha/\pi}\,e^{-\alpha\,x^2}$ et posons $Y = h(X)$, h étant une transformation non linéaire que nous supposerons inversible (disons croissante pour fixer les idées), avec la fonction inverse $h^{-1} = g : X = g(Y)$. On peut alors écrire:

$$\mathcal{P}\,(a < Y < b) \quad = \quad \mathcal{P}\,(a < h(X) < b) \quad \simeq \int\limits_{a<h(x)<b} \sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha\,x^2}\;dx$$

En effectuant le changement de variable $x = g(y)$ dans cette intégrale on obtient

$$\mathcal{P}\,(a < Y < b) \;\simeq\; \int_a^b \sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha\,g(y)^2}\;g'(y)\;dy \qquad (IX.5.)$$

Cette égalité met bien en évidence que la densité de $Y = h(X)$ est la fonction $\sqrt{\alpha/\pi}\,\exp\{-\alpha\,g(y)^2\}\;\,g'(y)$, donc pour que cette densité soit la densité prescrite $f(y)$ il suffit d'avoir

$$\sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha\,g(y)^2}\;g'(y) \;=\; f(y) \qquad (IX.6.)$$

ce qu'on peut interpréter comme une équation différentielle dont la fonction inconnue est g. Elle est facile à résoudre en introduisant la fonction

$$\mathcal{N}\,(x) \;=\; \int_{-\infty}^{x} \sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha\,t^2}\;dt$$

car alors l'équation différentielle peut s'écrire sous la forme équivalente

$$\frac{d}{dy}\,\mathcal{N}\,(g(y)) \;=\; f(y)$$

En introduisant la primitive $F(y)$ de $f(y)$ (celle qui s'annule pour $y = -\infty$) l'équation se résoud par quadrature :

$$g(y) = \mathcal{N}^{-1}(F(y)) \qquad (IX.7.)$$

La fonction $\mathcal{N}$ est en effet croissante, donc a un inverse $\mathcal{N}^{-1}$.

En conclusion, si Y est une variable aléatoire densité arbitraire $f(y)$, il suffit de choisir $g(y)$ selon $IX.7$ pour que la variable $X = g(Y)$ soit de densité gaussienne. Cela est un résultat purement mathématique, qui ne signifie évidemment pas que n'importe quelle densité se rencontre dans des problèmes concrets. L'expérience montre que les densités gaussiennes sont fréquentes dans la nature, alors que la plupart des autres qu'on pourrait imaginer mathématiquement ne se rencontrent pas. Or, s'il suffit d'observer le carré, la racine carrée, ou le logarithme, pour que la densité cesse d'être gaussienne, on peut se demander pourquoi on rencontre si souvent la densité gaussienne. Cette question mérite une discussion quelque peu approfondie. Lorsqu'on rencontre des distributions de probabilités "dans la nature", c'est sous la forme d'une distribution statistique ; en effet, on peut faire beaucoup d'expériences déterministes, par exemple si on mesure dans un conducteur donné le rapport de l'intensité à la tension entre les bornes, on trouvera toujours pratiquement la même valeur. Dans un tel cas on ne dit pas que le résultat est aléatoire. Par contre si on ne retient pas la valeur mesurée de l'intensité, mais seulement le petit écart, de l'ordre des erreurs de mesure, par rapport à cette valeur, on constate une grande variabilité et on dit que cet écart est aléatoire. De même, si on enregistre le nombre de naissances qui ont lieu chaque jour dans une maternité, on observera une certaine variabilité : certains jours il y aura eu cinq naissances, d'autres jours dix ; toutefois les jours sans aucune naissance, ou les jours avec vingt naissances, seront rarissimes. Dans ce cas, on dira que le nombre de naissances est aléatoire, mais bien sûr tous les nombres ne sont pas équiprobables et certains sont plus fréquents que d'autres. On peut donc représenter le nombre de naissances par jour par une *variable aléatoire*. Si au lieu de prendre dix valeurs possibles, la variable aléatoire en prend un nombre très grand, par exemple le nombre de naissances par semaine et non par jour, on en représentera la répartition par une densité. On observera alors que le nombre de naissances par semaine, qui varie d'une semaine à l'autre, suit une densité gaussienne, avec un écart-type qu'on pourra mesurer en faisant une étude statistique sur dix ans (bien sûr il faudra la corriger par rapport aux évolutions saisonnières). Il semblerait farfelu de vouloir plutôt étudier la répartition statistique des carrés ou des logarithmes de ces nombres, qui auraient alors une densité de répartition autre, mais on constate que c'est justement avec la variable qui semble la plus naturelle qu'on observe la densité gaussienne.

D'autre part, si le nombre qu'on mesure est petit (par exemple si dans la maternité on compte le nombre de naissances *par jour* et non par an), on ne peut obtenir une densité gaussienne pour la simple raison que les valeurs possibles ne sont pas assez nombreuses : une densité n'apparaît que si les valeurs de la variable aléatoire sont nombreuses. Mais on observera presque toujours dans de tels cas une loi de Poisson (voir section suivante). Que ces deux lois, la loi normale pour les variables qui prennent beaucoup de valeurs, et la loi de Poisson pour les variables qui prennent des valeurs entières et petites, soient aussi répandues dans la nature, doit pouvoir se comprendre.

Les conclusions du chapitre **VII** conduisent à penser que la distribution gaussienne autour d'une valeur moyenne est due à l'addition d'un grand nombre d'effets indépendants les uns des autres (et nous verrons plus loin que c'est le cas aussi pour la loi de Poisson). Par ailleurs, le constat purement mathématique que nous avons fait ci-dessus nous conduit à penser que de telles grandeurs à fluctuations gaussiennes devraient être la somme d'un grand nombre d'effets indépendants, et non le carré ou le logarithme d'une telle somme.

Lorsqu'on définit une grandeur physique, on ne la choisit pas arbitrairement. L'intensité d'un courant électrique pourrait, d'un point de vue purement métrologique, être tout aussi bien mesurée par sa racine carrée ou son carré : au lieu de la définir comme proportionnelle à la tension ($I = U/R$), on la définirait comme proportionnelle à la puissance ($I^2 = P/R$). Mais en agissant de la sorte à tort et à travers, on changerait complètement la nature des lois physiques et on perdrait les invariances spatio-temporelles (voir au chapitre **I** les remarques — empruntées à H. Poincaré — à propos du principe de relativité de Galilée et la mesure du temps). Les grandeurs physiques ne sont pas définies à la légère et sont choisies de manière à refléter les invariances fondamentales, sans lesquelles la physique perdrait son sens ; plus exactement : elles sont choisies de manière à avoir les lois les plus simples possibles, et les invariances fondamentales sont l'expression de ce choix. Lorsque des variations aléatoires *indépendantes* s'ajoutent, elles expriment implicitement ces invariances : soit il s'agit d'une accumulation au cours du temps, soit d'une addition de déplacements dans l'espace. En toute dernière instance, une grandeur se ramène toujours à cela puisque, quel que soit l'instrument de mesure, on lira toujours la valeur mesurée sur une graduation, un écran, etc. L'indépendance stochastique de ces variables qui s'ajoutent spatialement les unes aux autres est une expression de la causalité et de la séparabilité dans le temps et l'espace : que dans n'importe quel phénomène complexe on observe des fluctuations gaussiennes (erreurs de mesures, dispersion aléatoire, etc.) reflète d'une part que de nombreux effets s'ajoutent spatialement, et d'autre part que ces nombreux effets sont séparables dans le temps ou dans l'espace, de sorte qu'ils ne s'influencent pas mutuellement (comme les boules qui tombent dans les boîtes sans s'influencer).

Nous avons cependant étudié des situations quantiques où cette séparabilité n'existe pas (par exemple **II. 6** *la loi de Planck*) et pourtant on y rencontre aussi des fluctuations gaussiennes ; l'analyse mathématique détaillée (voir **II.6.**) faisait apparaître ces fluctuations gaussiennes comme une propriété des factorielles. Or les factorielles interviennent par l'intermédiaire des formules de dénombrement (coefficients binômiaux, etc.), qui s'appliquent dès lors qu'une *équiprobabilité* a été postulée. Mais l'équiprobabilité présuppose toujours une invariance, comme cela a été largement discuté au chapitre **I**. Dans les statistiques quantiques, il n'y a pas de séparabilité, mais des invariances de nature différente (et aujourd'hui encore mystérieuses). Dans tous ces cas, aussi bien lorsque des causes très nombreuses et causalement séparées (au sens de la Mécanique classique) s'ajoutent parce que leur effet est spatial, que lorsqu'on analyse les statistiques quantiques où une invariance d'origine autre que spatiale conduit à des factorielles, on rencontrera la forme gaussienne. On pourrait dire que la forme gaussienne des fluctuations statistiques est le symptôme d'une invariance sous-jacente.

Il n'est donc pas surprenant que les fluctuations aléatoires qu'on observe dans la nature soient si souvent gaussiennes.

La discussion serait bien sûr plus concrète sur un exemple ; en voici un. Pour mesurer la quantité d'information qu'on peut stocker sur une mémoire on utilise une unité qui est l'octet (ou le bit) ; dans un octet on peut ranger 256 signes différents, alors pourquoi ne dit-on pas que la quantité d'information est 256 ? Sur $1\,Mo$ on peut ranger environ $4.26 \cdot 10^{2525222}$ textes différents, mais on préfère mesurer la quantité d'information par le logarithme en base 2 de ce nombre ; c'est certes plus commode, mais est-ce seulement une question de commodité ?

En analysant les choses de plus près, on peut remarquer que la mémoire mesurée en octets est proportionnelle au volume que la mémoire physique occupe dans l'espace (à sa longueur si la mémoire est stockée sur une ligne, à la superficie si elle est stockée sur un disque, à un volume si la mémoire est stockée en trois dimensions). Ainsi la quantité d'information mesurée en octets se ramène à de l'espace, alors que la quantité mesurée par le nombre de textes possibles ne s'y ramène pas linéairement. Si on remplit une mémoire progressivement en y vidant une centaine de fois un buffer, l'espace disponible ne sera pas rempli complètement, il restera toujours des trous dûs au fait que l'adressage des buffers successifs ne coïncide pas avec un pavage géométriquement parfait de l'espace disponible. Ces trous auront donc un volume assez petit, mais qu'on peut avec raison considérer comme aléatoire ; il n'est peut-être pas absolument évident que les trous successifs soient stochastiquement indépendants, mais si le remplissage de l'espace-disque est suffisamment chaotique, c'est une hypothèse qui se tient, parce qu'on peut admettre — en vertu justement du principe d'invariance — qu'il n'y a pas de région privilégiée de l'espace (voir à ce sujet **IV. 6** *l'effacement de la causalité*).

Après avoir vidé quelques centaines de fois le buffer, le volume total de ces trous sera une variable aléatoire de densité gaussienne (somme des volumes des trous accumulés), dont l'écart-type représente alors une incertitude sur la mémoire disponible. Si on avait compté en nombre de textes possibles, on n'observerait pas une fluctuation gaussienne, puisque la relation entre les deux grandeurs est logarithmique et non linéaire. Mais le niveau où se produit une *addition* de variables indépendantes est celui du remplissage de l'espace, et non celui du nombre de textes possibles, qui serait multiplicatif. Ainsi, la grandeur qui a été choisie pour mesurer l'information est celle qui reflète le remplissage spatial et par conséquent aussi celle qui présentera des fluctuations gaussiennes.

On pourra faire le même constat pour l'intensité électrique (nombre d'électrons passant *par seconde* à travers une *section* de conducteur), etc.

C'est la véritable raison de l'universalité de la densité gaussienne. Il s'y ajoute encore une raison secondaire, mais très importante : même lorsqu'une grandeur ne se ramène pas linéairement à des déplacements dans l'espace, ou plus généralement à des formules de dénombrement contenant des factorielles, les fluctuations peuvent rester gaussiennes tout simplement parce qu'elles sont petites. En effet, soit x une grandeur subissant des fluctuations gaussiennes ε : on mesure donc x, mais avec un bruit ε qui s'ajoute à x. Si on considère maintenant la grandeur $y = f(x)$ qui dépend non linéairement de x, celle-ci fluctuera selon $f(x+\varepsilon)$. Si ε est grand, la non linéarité de f déformera les fluctuations, de telle sorte que celles-ci ne seront plus gaussiennes, comme l'étude précédente l'a montré. Mais si ε est petit et f différentiable, on aura $f(x+\varepsilon) \simeq f(x) + f'(x)\,\varepsilon$, donc la fluctuation sera néanmoins transformée linéairement. En fin de compte, il est même plutôt difficile de trouver des grandeurs qui échappent à toutes ces bonnes raisons.

Malgré tout, nous avons pu constater que les processus en cascade engendraient une bien plus grande variété de densités asymptotiques ; mais la variable Z_n du chapitre **VIII** n'est pas une somme de variables indépendantes, le principe de composition est différent et n'efface pas entièrement l'information sur la loi initiale de Z_1.

IX.3. Loi de Poisson.

La loi de Poisson est la plus simple de toutes les lois de probabilité asymptotiques ; elle est entièrement discrète.

Considérons le problème très simple (et purement mathématique) suivant : on a n variables aléatoires $X_1,\ X_2,\ \ldots\ X_n$, stochastiquement indépendantes, et ne prenant que les deux valeurs 0 et 1 avec les probabilités respectives p et q $(p+q=1)$. La loi de leur somme $S = X_1 + X_2 + \cdots + X_n$ est

$$\mathcal{P}\,(S=k) = \binom{n}{k} p^{n-k} q^k$$

(loi de Bernoulli). Mais supposons que n soit très grand et q très petit, de l'ordre de $1/n$. On pourra poser $q = \lambda/n$ avec λ ni grand ni petit. Alors, pour $k \ll n$, $p^{n-k} = (1-\lambda/n)^{n-k} \simeq e^{-\lambda}$ et $\binom{n}{k} \simeq n^k/k!$ De sorte que

$$\mathcal{P}\,(S=k) \simeq \frac{\lambda^k}{k!}\,e^{-\lambda} \qquad (IX.8.)$$

On constate d'abord que pour $k \gg \lambda$ ces probabilités sont pratiquement nulles (aussi bien sous leur forme approchée que sous leur forme exacte). Seules sont significatives les probabilités correspondant aux petites valeurs de k.

Plus généralement, supposons que les X_j n'aient pas toutes exactement la même loi, mais que toutes, comme ci-dessus, prennent la valeur 1 avec une probabilité très petite ε_j, et la valeur zéro avec une probabilité $1-\varepsilon_j$ presque égale à 1.

La fonction génératrice de chaque X_j étant $G_j(x) = 1 + \varepsilon_j\,(x-1)$, la somme $S = X_1 + X_2 + \cdots + X_n$ aura pour fonction génératrice (si les X_j sont indépendantes) :

$$G(x) = \prod_{j=1}^{j=n} [1 + \varepsilon_j\,(x-1)]$$

Pour approcher ce type d'expression on prend bien sûr le logarithme ; or

$$\ln\left(1 + \varepsilon_j\,(x-1)\right) \simeq \varepsilon_j\,(x-1) - \frac{1}{2}\varepsilon_j^2\,(x-1)^2$$

où on a conservé le terme d'ordre deux pour une évaluation de l'erreur. On obtient donc en revenant aux exponentielles :

$$G(x) \simeq \exp\Big\{ \sum_{j=1}^{j=n} \varepsilon_j\,(x-1) - \frac{1}{2} \sum_{j=1}^{j=n} \varepsilon_j^2\,(x-1)^2 \Big\}$$

Introduisons les paramètres

$$\lambda = \sum_{j=1}^{j=n} \varepsilon_j \qquad \text{et} \qquad \mu = \sum_{j=1}^{j=n} \varepsilon_j^2 \tag{IX.9.}$$

Le premier, λ, est la somme des moyennes des X_j, donc la moyenne de S. On a supposé les ε_j très petits, supposons maintenant que l'ordre de grandeur de n est tel que la somme λ ait une valeur appréciable. Alors, les carrés ε_j^2 étant bien plus petits que les ε_j, leur somme sera très petite par rapport à λ. Ainsi la fonction génératrice $G(x)$ sera approximativement donnée par

$$G(x) \simeq \exp\{\lambda\,(x-1)\} \tag{IX.10.}$$

avec une erreur relative de l'ordre de $-\frac{1}{2}\mu\,(x-1)^2$. Les probabilités pour que $S = 0, 1, 2, 3, \ldots$ sont donc à nouveau données par la loi de Poisson, avec une erreur relative de l'ordre de μ. Toutes ces approximations ne sont justifiées que pour les petites valeurs de S et non pour $S = 1\,000$ ou $S = 10\,000$, mais pour les grandes valeurs, bien que l'erreur *relative* ne puisse plus être considérée comme petite, les probabilités aussi bien exactes qu'approchées sont de toute façon si prodigieusement petites (ce que montrent les factorielles au dénominateur de la loi de Poisson) que cela n'a pas d'importance.

Nous voyons donc apparaître le résultat suivant : lorsqu'on considère une somme d'un grand nombre n de variables indépendantes qui ne prennent que les deux valeurs 0 et 1, avec une très faible probabilité (de l'ordre de $1/n$) pour 1, alors leur somme obéit à la loi de Poisson.

A première vue, on pourrait penser que cela contredit les conclusions du chapitre **VII** : on devrait en effet avoir une loi gaussienne. Mais en étudiant le passage à la limite, nous avons remarqué que, surtout si la loi des X_j est fortement dissymétrique — ce qui est le cas ici — il fallait que le nombre N de variables à sommer soit vraiment très grand pour arriver à la loi gaussienne ; nous allons voir qu'il doit être beaucoup plus grand que le nombre n qui intervient ici. Nous en avons donné une estimation explicite $(VII.5)$: il fallait que N soit grand par rapport à $400\,M_3^2/M_2^3$, M_2 et M_3 étant les moments d'ordre deux et trois de la loi de X. Ces moments ne sont pas les mêmes pour toutes les X_j ; mais pour chaque X_j, ils sont tous trois égaux à ε_j ($M_1 = M_2 = M_3 = \varepsilon_j$), et sous nos hypothèses ils ont donc tous le même ordre de grandeur $1/n$. De sorte que $M_3^2/M_2^3 \sim n^3/n^2 = n$. D'après $VII.5$ il faudrait donc, pour arriver à la densité gaussienne, que $N \gg 400\,n$; autrement dit, n n'est pas assez grand et c'est pour cela qu'on obtient la loi de Poisson au lieu de la densité gaussienne.

Un tel phénomène est très courant pour les lois asymptotiques : il y en a qui sont vraiment des limites lorsque n tend vers l'infini : c'est le cas pour la loi normale du chapitre **VII**, ainsi que pour la loi limite de $Z_n/E(Z_n)$ du chapitre **VIII**. Mais il y en a aussi, comme ici la loi de Poisson, qui sont une étape intermédiaire sur la route vers l'infini. Tant que n est de l'ordre de grandeur des $1/\varepsilon_j$ on a la loi de Poisson, puis, lorsque n poursuit son voyage vers l'infini et atteint un ordre de grandeur nettement supérieur aux nombres $400/\varepsilon_j$, on arrive à la loi gaussienne.

On remarquera cependant que la loi de Poisson est aussi la loi d'une somme de variables aléatoires indépendantes, mais de variables qui ne prennent que les deux valeurs 0 et 1, la seconde avec une faible probabilité : la somme de ces variables sera donc un nombre entier petit (comme le nombre de naissances par jour ou par heure dans une maternité). Les remarques présentées à la section **2** à propos de la loi normale expliquent donc aussi pourquoi la loi de Poisson s'observe si fréquemment dans la nature, dès lors que les nombres qu'on mesure sont entiers et petits.

Jusqu'ici nous sommes restés dans l'abstraction : nous avons considéré des sommes de variables aléatoires sans nous préoccuper de ce qu'elles peuvent représenter.

Voici maintenant un problème concret : les queues qui peuvent se former à un péage d'autoroute. Mettons qu'il faut une minute pour effectuer les paiements au guichet. On supposera qu'il n'y a qu'un guichet ; s'il y en a plusieurs, le problème n'est pas essentiellement différent ; la durée de paiement de une minute n'a ici aucune signification pratique, n'importe quelle unité de temps eût convenu. On voudrait connaître la loi de probabilité du nombre de voitures qui se présentent par minute, sachant qu'en moyenne il s'en présente 0.76. L'intérêt de ce problème est le suivant : cette moyenne est inférieure à 1, donc "en moyenne" il ne se forme pas de queue : au bout de cent minutes, environ 76 véhicules sont passés, alors que le guichet aurait pu en absorber cent. Il est bien clair que si la moyenne avait été 1.22, il y aurait un résidu d'environ 22 véhicules au bout de cent minutes, et on serait donc sûr qu'une queue se formerait et s'allongerait en moyenne de 22 véhicules par cent minutes. Toutefois, lorsque la moyenne est inférieure à 1, la probabilité qu'il se forme des queues même longues n'est pas nulle ; si cette probabilité est supérieure à 5% pour une durée de trois heures (180 minutes) par exemple, la compagnie peut avoir intérêt à améliorer le péage. En effet une probabilité de 5% pour trois heures signifie que toutes les soixante heures environ se forme une queue longue, soit tous les trois jours. Cela peut dissuader bon nombre d'usagers d'emprunter l'autoroute et faire perdre de l'argent à la compagnie. Celle-ci peut améliorer le péage de deux façons : soit construire un second guichet, ce qui aura pour effet de faire tomber la moyenne de 0.76 à 0.38, soit améliorer son fonctionnement pour que la durée de paiement soit raccourcie : si celle-ci est abaissée à une demi-minute, la moyenne tombera aussi à 0.38. Il est assez évident a priori que la probabilité de formation de queues de longueur fixée doit être une fonction décroissante du nombre moyen de véhicules par minute. Mais cette intuition a priori ne permet pas de connaître la loi de cette décroissance. C'est cette loi que nous nous proposons de découvrir par le Calcul des probabilités.

Qu'il puisse se former des queues même lorsque le nombre moyen de véhicules par minute est plus petit que 1 n'est pas paradoxal, il suffit d'imaginer que si la moyenne est 0.76, il peut parfois arriver deux, trois, ou même quatre voitures en une seule minute. Cela va alors faire durer les formalités plus d'une minute, puisqu'il y aura plus d'une voiture au guichet ; or pendant ces deux ou trois minutes vont arriver en moyenne encore deux ou trois fois 0.76 voitures, qui à leur tour bloqueront les guichets deux minutes au lieu d'une, pendant lesquelles arriveront encore des voitures, etc.

Pour aborder ce problème il faut commencer par trouver la loi de probabilité pour qu'il arrive zéro, un, deux, trois, ... véhicules. C'est seulement avec la connaissance quantitative de cette loi qu'on pourra ensuite étudier quantitativement la loi de formation des queues.

Or il est facile de montrer que cette loi doit être une loi de Poisson, du moins si on est assuré que le nombre de voitures par minute est petit (qu'il n'est pas de l'ordre de mille ou un million, auquel cas il faut procéder autrement).

Supposons d'abord que le flux est constant au cours de la journée ; cela est toujours faux en pratique, où il y a des heures de pointe (aux environs de huit heures ou de dix-huit heures) et des heures creuses, mais cette hypothèse permettra déjà de trouver la loi de probabilité pour une moyenne constante ; on agit de même lorsqu'on veut définir la vitesse en cinématique : on commence par la définir dans le cas où le mouvement est uniforme, puis on étend la définition en partant du principe que tout mouvement peut être considéré comme uniforme pendant un instant assez court.

Sous cette hypothèse, les automobilistes quittent leur domicile et prennent l'autoroute à des instants aléatoires, mais uniformément distribués au long de la journée (car si ces instants n'étaient pas uniformément distribués on ne pourrait avoir un flux constant). Chaque automobiliste agit indépendamment des autres, donc les départs du domicile (ou les arrivées au péages) de tous ces usagers sont stochastiquement indépendants.

Lorsqu'on dit que le flux moyen au péage est de λ voitures par minute, cela signifie que si dans la journée N voitures en tout sont passées, $\lambda =$

$N/1\,440$, puisqu'il y a 1 440 minutes dans une journée. Si on prend *au hasard* une voiture parmi les N, la probabilité pour que cette voiture arrive au péage au cours d'une minute donnée parmi les 1 440 que compte la journée est $\varepsilon = 1/1\,440$. L'hypothèse d'un flux constant se traduit en effet par l'équiprobabilité a priori de chacune des minutes de la journée. Inversement, la probabilité pour que cette voiture arrive en dehors de la minute donnée, est $1 - \varepsilon = 1\,439/1\,440$. On peut donc associer à chaque automobiliste j ($j = 1, 2, 3, \ldots N$) une variable aléatoire X_j qui vaut 1 si l'automobiliste j arrive au péage au cours de la minute donnée, et 0 si l'automobiliste arrive au péage en dehors de cette minute donnée. Le fait que chaque automobiliste agit indépendamment se traduit par l'indépendance stochastique des X_j, et les X_j ont toutes la même loi :

$$X_j = \begin{cases} 1 & \text{avec probabilité } \varepsilon\,; \\ 0 & \text{avec probabilité } 1-\varepsilon\,; \end{cases} \tag{IX.11.}$$

La somme des X_j, $S = \sum_{j=1}^{N} X_j$, est donc le nombre de voitures qui arrivent pendant la minute donnée. Nous avons vu précédemment que, si N est du même ordre de grandeur que $1/\varepsilon$, la loi de la somme S est approximativement une loi de Poisson :

$$\mathcal{P}\,(S = k) \quad = \quad \frac{\lambda^k}{k!}\, e^{-\lambda} \tag{IX.12.}$$

Puisque $\lambda = N/1\,440 = N\varepsilon$, la condition que N soit du même ordre de grandeur que $1/\varepsilon$ signifie simplement que λ ne doit être ni trop grand, ni trop petit. Ceci est la loi de probabilité pour une minute donnée, mais bien entendu toutes les minutes de la journée sont équivalentes et la loi sera la même pour n'importe quelle minute de la journée.

Revenons alors à l'exemple qui a introduit la discussion, où λ valait 0.76. Les valeurs numériques de la loi de Poisson sont données par le tableau suivant :

k	$\mathcal{P}\,(S = k)$
0	0.4677
1	0.3554
2	0.1351
3	0.0342
4	0.0065
5	0.0010
6	0.0001
7	0.0000

$(IX.13.)$

Au-delà de 6 les probabilités sont inférieures au dix-millième. En faisant la somme des six dernières, on voit que la probabilité pour qu'il se présente deux véhicules ou plus au péage est 0.177 ; la probabilité pour qu'il se présente trois véhicules ou plus est 0.0418. Avec les capacités d'absorption que nous avons posées, il suffit qu'une seule fois se présentent plus de deux véhicules pour que l'engorgement momentané ne puisse être résorbé par la suite que si on a la chance d'avoir moins que deux véhicules dans les minutes qui suivent ; or la probabilité pour qu'il se présente plus de deux véhicules étant 0.177, cela va se produire en moyenne toutes les six minutes. Par cette estimation purement qualitative, on comprend que des bouchons puissent se former bien que le flux moyen soit inférieur à un véhicule par minute.

Nous nous proposons maintenant d'analyser cela quantitativement.

IX.4. La loi des queues.

À chaque instant n de la journée (les instants étant les minutes successives, $n = 1, 2, 3, \ \ldots \ 1\,440$), la longueur de la queue est égale au nombre de véhicules déjà arrivés moins le nombre de véhicules déjà autorisés. Mais le nombre de véhicules déjà autorisés n'est en général pas n, car si par exemple aucun véhicule ne s'est présenté pendant les trois premières minutes, puis qu'il s'en présente trois d'un coup pendant la quatrième minute, la longueur de la queue sera de deux véhicules, et non de $0 + 0 + 0 + 3 - 4$; à la fin de la quatrième minute, le nombre de véhicules autorisés aura été 1 et non $n = 4$. En effet, si pendant une période il se présente moins de véhicules que prévu, il y aura des temps morts au guichet, mais ces temps morts sont perdus et ne peuvent plus être recomptés plus tard pour absorber un trafic plus dense. Il faut donc tenir compte de ces temps morts.

Appelons Q_n la variable aléatoire “longueur de la queue à l'instant n”, et X_n le nombre de véhicules arrivés *pendant* la n^{e} minute (les X_n ont toutes la loi de Poisson mentionnée ci-dessus et sont indépendantes entre elles). Si par exemple $Q_{n-1} = 1$ et qu'il arrive zéro véhicules pendant la minute suivante, Q_n sera égale à 0 ; dans ce cas on obtient Q_n en ajoutant $X_n - 1$ à Q_{n-1}. De même si $Q_{n-1} = 0$ et $X_n = 1$, Q_n sera nulle : là aussi $Q_n = Q_{n-1} + X_n - 1$. Par contre si $Q_{n-1} = 0$ et $X_n = 0$, on n'aura pas $Q_n = Q_{n-1} + X_n - 1 = -1$, mais $Q_n = 0$ (la queue ne peut jamais avoir une longueur négative). Pour trouver la récurrence qui fait passer de Q_{n-1} à Q_n, le mieux est de décomposer les événements en réunions disjointes adéquates. Dans de tels problèmes *il ne faut jamais essayer de deviner*, mais toujours suivre un procédé systématique.

L'événement $\{Q_n = 0\}$ est la réunion de $\{Q_{n-1} = 0\} \cap \{X_n = 0\}$, de $\{Q_{n-1} = 0\} \cap \{X_n = 1\}$, et de $\{Q_{n-1} = 1\} \cap \{X_n = 0\}$. Ces trois événements sont disjoints : on ne peut évidemment pas avoir à la fois $Q_{n-1} = 0$ et $Q_{n-1} = 1$, ni $X_n = 0$ et $X_n = 1$. D'autre part, les variables aléatoires Q_{n-1}

et X_n sont stochastiquement indépendantes (les véhicules qui arrivent au cours de la n^{e} minute “ignoraient” ce qui s'était produit jusque là). Donc

$$\begin{aligned}
\mathcal{P}\left(Q_n=0\right) &= \mathcal{P}\left(Q_{n-1}=0 \text{ et } X_n=0\right)+\\
&\qquad +\mathcal{P}\left(Q_{n-1}=0 \text{ et } X_n=1\right)+\\
&\qquad +\mathcal{P}\left(Q_{n-1}=1 \text{ et } X_n=0\right)\\
&= \mathcal{P}\left(Q_{n-1}=0\right)\times\mathcal{P}\left(X_n=0\right)+\\
&\qquad +\mathcal{P}\left(Q_{n-1}=0\right)\times\mathcal{P}\left(X_n=1\right)+\\
&\qquad +\mathcal{P}\left(Q_{n-1}=1\right)\times\mathcal{P}\left(X_n=0\right)
\end{aligned} \tag{IX.14.}$$

Cette expression de $\mathcal{P}\left(Q_n = 0\right)$ diffère de ce qu'elle serait si Q_n était simplement la somme $Q_{n-1}+X_n-1$: le premier des trois termes serait alors absent.

En revanche, pour les probabilités $\mathcal{P}\left(Q_n=k\right)$ avec $k\geq 1$, tout se passe comme si Q_n était bien la somme $Q_{n-1}+X_n-1$; Q_n ne peut en effet être égal à $k\geq 1$ que si $Q_{n-1}+X_n=k+1$, car il n'y a pas dans ce cas de temps mort au guichet : il y a Q_{n-1} véhicules qui attendent (Q_{n-1} pouvant être zéro), il en arrive X_n de plus, et un passera. C'est seulement lorsque Q_{n-1} et X_n sont tous deux nuls qu'il aura un temps mort et que Q_n sera égal à 0 au lieu de -1. Ainsi l'événement $\{Q_n=k\}$ pour $k\geq 1$ sera la réunion pour j variant de 0 à $k+1$ des événements disjoints $\{Q_{n-1}=j\}\cap\{X_n=k+1-j\}$, de sorte que

$$\mathcal{P}\left(Q_n=k\right)=\sum_{j=0}^{k+1}\mathcal{P}\left(Q_{n-1}=j\right)\times\mathcal{P}\left(X_n=k+1-j\right) \tag{IX.15.}$$

tout comme si Q_n était simplement la somme de Q_{n-1} et de X_n-1. En regroupant (IX.14.) et (IX.15.) on obtient la récurrence qui fait passer de la loi de Q_{n-1} à celle de Q_n. Cette récurrence permet d'écrire un programme qui calculera récursivement les lois de Q_n à partir de celle de Q_1, qui est évidemment la suivante :

$$\begin{aligned}
&\mathcal{P}\left(Q_1=0\right)=\mathcal{P}\left(X_1=0\right)+\mathcal{P}\left(X_1=1\right) \quad \text{pour } k=0\\
&\mathcal{P}\left(Q_1=k\right)=\mathcal{P}\left(X_1=k+1\right) \quad \text{pour } k\geq 1
\end{aligned}$$

Si on exécute ce programme, on s'aperçoit que la loi de Q_n se stabilise au bout d'une centaine d'itérations (voir figures 36, 37, 38), ce qui correspond au fait que la loi de Q_n tend vers une limite lorsque n tend vers l'infini. Le nombre d'itérations avant stabilisation dépend évidemment de la précision : plus on prend en compte de décimales, plus grand est ce nombre. La centaine correspond à trois ou quatre décimales, et la limite n'est, au sens mathématique, jamais atteinte. Ainsi, il s'établit au bout d'un certain temps un régime stable. Mais ceci n'est vrai que si le flux moyen est inférieur à un véhicule par minute, sinon évidemment les queues vont s'allonger indéfiniment (c'est-à-dire que la probabilité d'avoir une queue inférieure à n'importe quelle longueur fixée tendra vers zéro). L'établissement d'un régime stable peut se comprendre a priori : dans le phénomène de formation des queues, il n'y a aucune raison de penser que la loi de probabilité de Q_n puisse dépendre de l'instant n, puisque les conditions (le flux moyen, la durée du paiement, etc.) sont supposées constantes ; seul le début du processus présente des conditions spéciales, car on a imposé artificiellement que la queue soit de longueur nulle à l'instant zéro. La longueur de la queue (c'est-à-dire la variable aléatoire Q_n) redevient nulle indéfiniment à des instants aléatoires ; mais les distributions possibles de ces instants de retour à zéro sont toutes équiprobables, il n'y a pas de distribution privilégiée. En imposant que l'instant $n = 0$ soit obligatoirement l'un de ces instants on détruit l'équivalence des instants. C'est pourquoi pendant une période de l'ordre de la centaine de minutes la loi de Q_n évolue selon un régime transitoire avant de retrouver la loi constante comme limite.

On peut exprimer commodément la récurrence des lois des Q_n et leur limite en considérant les fonctions génératrices. En effet, les variables Q_n prennent des valeurs entières non négatives, donc l'usage des fonctions génératrices est tout à fait recommandé.

Appelons $G_n(z)$ la fonction génératrice inconnue de Q_n et $F(z)$ celle, connue, des X_n. La loi commune des X_n étant la loi de Poisson, $F(z)$ est la fonction $\exp(\lambda[z-1])$. Les relations $(IX.14.)$ et $(IX.15.)$ se traduisent en termes de fonctions génératrices comme suit :

$$z\,G_n(z) - F(z)\,G_{n-1}(z) \;=\; (z-1)\,F(0)\,G_{n-1}(0) \qquad (IX.16.)$$

On peut montrer mathématiquement que, pour $0 < \lambda < 1$, $G_n(z)$ doit tendre vers une limite, mais on se contentera de l'admettre. La démonstration est assez longue et peu instructive, j'estime donc qu'elle ne vaut pas la peine de rallonger excessivement cette section. De toutes façons le programme qui calcule la récurrence $IX.14$ – $IX.15$ montre une stabilisation de la loi de Q_n (voir figures 36 à 39).

Sachant que $G_n(z)$ tend bien vers une limite finie $G(z)$, on peut faire tendre n vers l'infini dans la relation ci-dessus, ce qui donne

$$[\,z - F(z)\,]\,G(z) \;=\; (z-1)\,F(0)\,G(0) \qquad (IX.17\,a.)$$

d'où on déduit ce que doit être la fonction génératrice de la loi limite si cette limite existe :

$$G(z) \;=\; F(0)\,G(0)\,\frac{z-1}{z-F(z)} \qquad (IX.17\,b.)$$

La fonction $(z-1)/(z-F(z))$ se prolonge analytiquement en $z=1$, où elle vaut $1/(1-F'(0))$; en prenant $z=1$ dans l'expression ci-dessus, on obtient $G(1)=F(0)\,G(0)/(1-F'(0))$, et bien sûr $G(1)=1$ puisque $G(1)$ est la limite des $G_n(1)$ qui sont tous égaux à 1. Donc $F(0)\,G(0)=1-F'(0)$. On remarquera en passant que la relation $IX.16$ est valable quelle que soit la loi de la variable X_n, c'est-à-dire pour des fonctions génératrices $F(z)$ arbitraires ; de sorte que la loi limite aura pour fonction génératrice

$$G(z) \;=\; [1-F'(0)]\,\frac{z-1}{z-F(z)} \qquad (IX.17\,c.)$$

résultat obtenu presque sans calculs. Dans le cas qui nous concerne la loi des X_n est la loi de Poisson, pour laquelle $F(z)=\exp\left(\lambda[z-1]\right)$; dans ce cas $1-F'(0)=1-\lambda$. Mais les expressions $IX.17\,b$ ou $IX.17\,c$ sont valables pour n'importe quelle loi initiale.

Nous voyons ici, avec la formation de queues, un nouveau phénomène conduisant à une loi de probabilité asymptotique discrète : pour une loi donnée des arrivées (qui est généralement, dans les cas ayant un sens pratique, une loi de Poisson), la loi de la longueur des queues après stabilisation est déterminée par $IX.17\,a, b, c$. Notez bien que les Q_n ne sont pas des sommes de variables indépendantes.

Si on veut calculer la loi limite, il suffit en principe de calculer les coefficients de Taylor de cette fonction analytique $G(z)$, ce qui est aisé pour les premiers, mais de plus en plus difficile pour les suivants (il n'y a pas de formule simple). Toutefois la relation $(IX.17\,c.)$ contient toute l'information.

Le problème des queues à un guichet a été abordé par Émile Borel(1). Son approche est différente : il ne prend pas en considération la variable aléatoire "longueur de la queue", dont la loi est certes exprimable analytiquement (formule $IX.17\,c.$ ci-dessus), mais ne permet pas une expression simple pour chaque probabilité. Il faut dire qu'à l'époque on ne disposait pas de *personal computers* sur lesquels on pouvait écrire en vitesse un petit programme qui calcule tout.

Borel a donc trouvé le biais suivant : au lieu de considérer les queues, il considère ce qu'il appelle une série : c'est une suite d'usagers qui se

(1) Émile Borel *Sur l'emploi du théorème de Bernoulli, pour le calcul d'une infinité de coefficients. Application au problème d'attente à un guichet.* Comptes-rendus de l'Académie des Sciences, mars 1942.

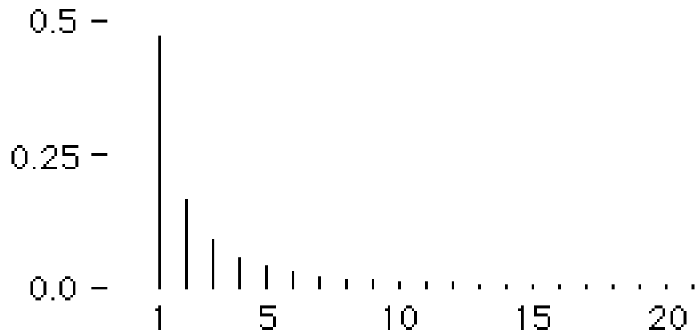

figure 35

Le graphique ci-dessus illustre la loi de Borel pour la longueur des séries d'usagers; ici pour la valeur $\lambda = 0.76$ (celle de notre exemple). Le premier trait vertical correspond à la probabilité $e^{-\lambda} \simeq 0.46$ d'avoir une série de longueur 1, c'est-à-dire un usager qui se présente au guichet et qui n'est pas immédiatement suivi d'un autre; le second trait correspond à la probabilité $\simeq 0.17$ d'avoir une série de longueur 2: un usager se présente, puis un second arrive avant que le premier n'ait terminé, après quoi le guichet reste inoccupé au moins pendant la minute suivante. Si le flux λ reste constant toute la journée, on peut conclure d'après la loi des grands nombres qu'au cours de la journée environ 46% des séries seront réduites à un seul usager, 17% à deux usagers consécutifs, etc. Bien entendu les queues longues ne peuvent se former que dans des séries au moins aussi longues.

succèdent au guichet consécutivement sans que le guichet se libère; dès que le guichet redevient libre (même une seule minute) la série se termine, mais une nouvelle série se produira avec le premier usager qui se présente après le temps mort. Une série peut se réduire à un seul usager. Les queues longues ne peuvent se former qu'à l'intérieur d'une série longue, mais il n'y a aucune relation mathématique précise entre la longueur d'une série et le maximum de la longueur de la queue pendant la série: une série peut être très longue sans que jamais plus de deux usagers ne fassent la queue (si les usagers se présentent de façon assez régulièrement espacée), et inversement une queue longue peut se former dans une série relativement courte (si beaucoup d'usagers arrivent presque en même temps). Il y a toutefois une corrélation assez forte entre la longueur des queues et la longueur des séries dans lesquelles elles se forment. On peut donc utiliser la probabilité α_n d'avoir des séries de longueur supérieure à un seuil donné n comme un critère quantitatif pour décider si le nombre de guichets est suffisant, comme nous l'avons proposé pour la probabilité pour que se forme une queue de longueur supérieure à n.

Borel a obtenu la formule explicite suivante pour la probabilité p_n pour qu'une série “prise au hasard” soit formée de n usagers :

$$p_n = e^{-n\lambda} \lambda^{n-1} \frac{n^{n-2}}{(n-1)!}$$

où λ est le nombre moyen d'usagers par minute ($\lambda = 0.76$ dans notre exemple). Cette formule n'est valable que si $\lambda < 1$. Elle permet de calculer les durées d'attente moyennes pour un usager, mais ne permet pas de calculer la loi de la longueur des queues $IX.17\,c.$ que nous avons obtenue par une autre méthode.

Revenons à la loi des queues et à la fonction génératrice $G(z)$ de la formule $(IX.17\,c)$.

Pour calculer numériquement les coefficients de Taylor de $G(z)$ (c'est-à-dire les probabilités limite de la loi de la longueur des queues) d'ordre élevé l'expression $(IX.17\,c.)$ n'est guère commode. Pour le calcul numérique effectif de la loi limite il est préférable d'exécuter le programme itératif basé sur la récurrence $(IX.14, 15)$. Par contre l'expression analytique $(IX.17\,c)$ de la fonction génératrice est très pratique pour calculer la moyenne et la variance de la loi limite qui sont données par $G'(1)$ et $G''(1)$, ainsi que les premiers coefficients de Taylor qui sont données par $G'(0)$ et $G''(0)$. Notons Q sans indice pour la variable aléatoire limite. Rappelons que $\mathbf{E}(Q) = G'(1)$ et $\mathbf{Var}(Q) = G''(1) + \mathbf{E}(Q) - \mathbf{E}(Q)^2$ (cf $VI.6$ et $VI.7$).

Dérivant formellement $(IX.17\,c.)$ on obtient

$$\begin{aligned}
G'(z) &= [1-F'(1)]\, \frac{1-F(z)+(z-1)F'(z)}{[z-F(z)]^2} \\
G''(z) &= [1-F'(1)] \times \qquad\qquad\qquad (IX.18.) \\
&\times \frac{[1-F'(1)](z-1)F''(z) - 2[1-F'(z)]\,[1-F(z)+(z-1)F'(z)]}{[z-F(z)]^3}
\end{aligned}$$

Pour $z = 0$ ces expressions prennent les valeurs particulières suivantes :

$$G(0) = \mathcal{P}\,(Q=0) = \frac{1-F'(0)}{F'(0)} = (1-\lambda)\, e^{\lambda}$$

$$\begin{aligned}
G'(0) = \mathcal{P}\,(Q=1) &= [1-F'(0)] \frac{1-F(0)-F'(0)}{F'(0)^2} \\
&= (1-\lambda)\, e^{\lambda}\, (e^{\lambda} - 1 - \lambda)
\end{aligned}$$

$$G''(0) = \mathcal{P}\,(Q=2) = (1-\lambda)\,[2\,(1-\lambda)\, e^{2\lambda}\,(e^{\lambda} - 1 - \lambda)]$$

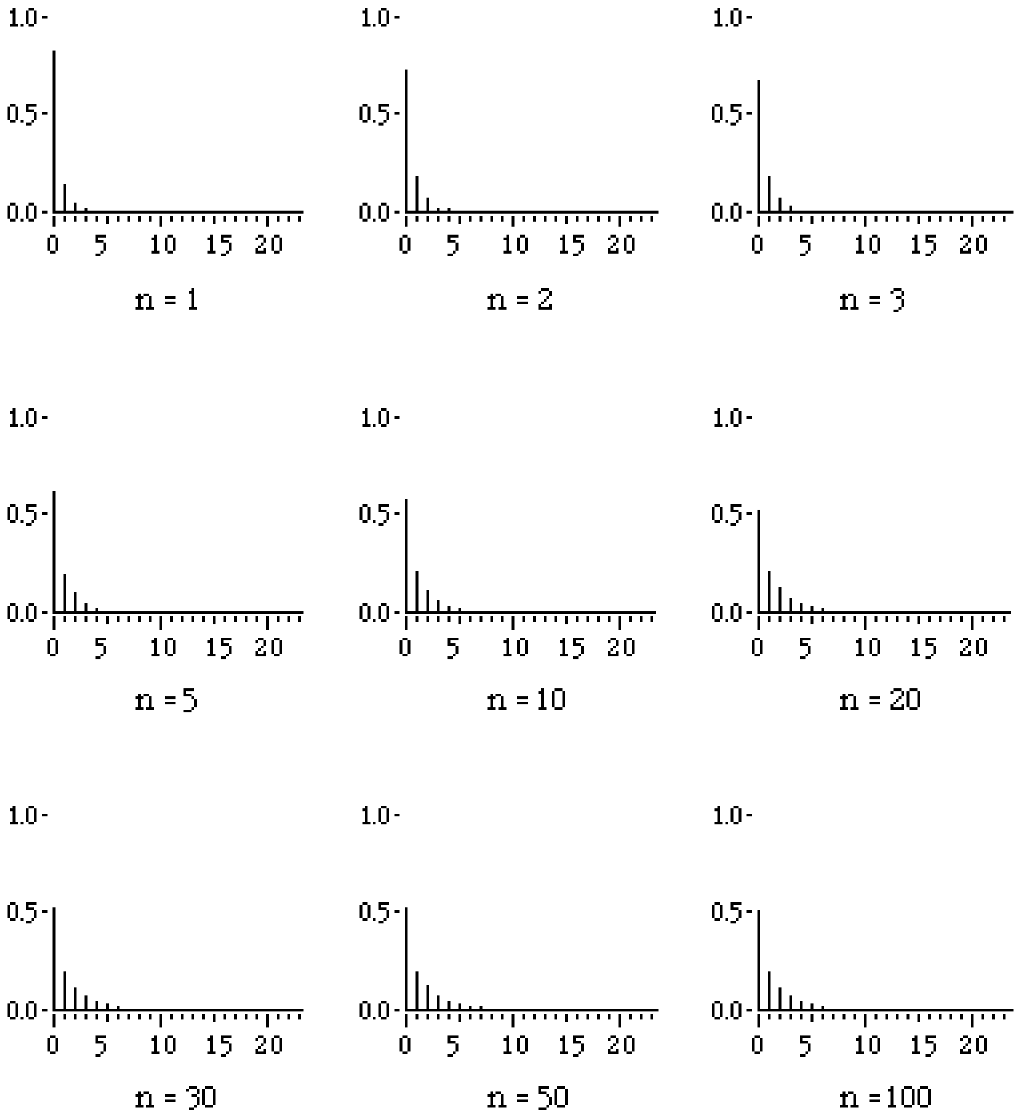

figure 36

Les neuf graphiques ci-dessus représentent les lois de probabilité de la longueur des queues pour $\lambda = 0.76$ après 1, 2, 3, 5, 10, 20, 30, 50, et 100 minutes : comme d'habitude, la hauteur des barres est proportionnelle à la probabilité de l'abcisse correspondante. L'échelle des abcisses est visualisée par une graduation (chaque trait de graduation correspond à une valeur – entière – de la variable aléatoire "longueur de la queue"). L'échelle des ordonnées est visualisée également : comme il s'agit de probabilités, celle-ci va de 0.0 à 1.0. On peut observer l'évolution vers une loi limite : pour n petit (donc peu après le début), la barre d'abscisse zéro est encore haute, puis elle diminue de plus en plus en plus au profit des autres ; le graphique n'évolue pratiquement plus au-delà de $n = 100$.

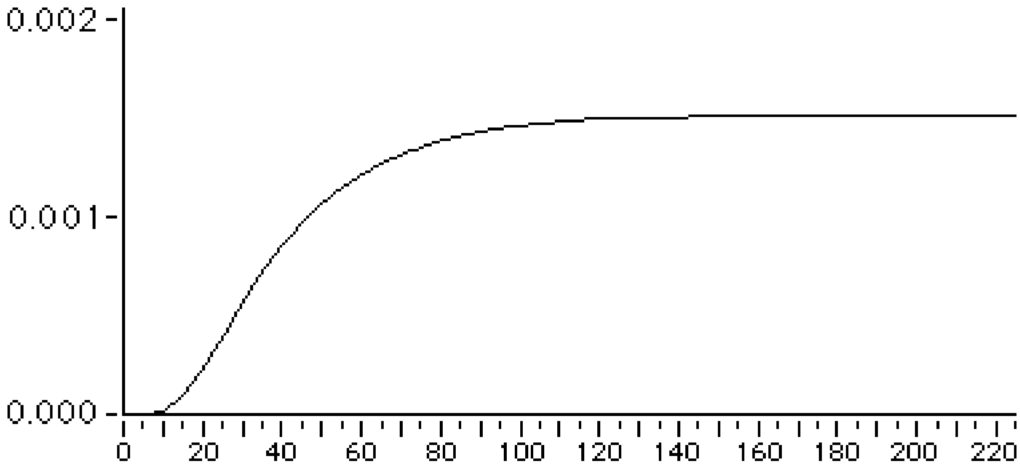

figure 37

Afin de visualiser l'évolution de la loi de Q_n vers la loi limite, on peut aussi représenter la probabilité que la queue ait une longueur supérieure ou égale à un seuil fixé, en fonction du temps écoulé. La courbe ci-dessus correspond à un seuil de 12 (avec toujours $\lambda = 0.76$) : l'abscisse représente le temps écoulé en minutes et l'ordonnée la probabilité que la queue comporte au moins douze véhicules. Bien sûr la courbe n'est qu'une interpolation des valeurs successives de ces probabilités, puisque le temps est un nombre entier de minutes.

On voit qu'au début cette probabilité est pratiquement nulle : elle est donnée par la loi de Poisson : pour que la queue soit d'au moins douze véhicules dès le premier instant, il faut que douze véhicules ou plus se présentent pendant la première minute, événement dont la probabilité est

$$\sum_{k \geq 12} \frac{\lambda^k}{k!} e^{-\lambda}$$

Mais après cette probabilité augmente et tend vers une limite, égale à 0.0015 environ. Ce qui est minuscule et ne justifie pas qu'on construise un second guichet. Toutefois pour $\lambda = 0.9$, la probabilité limite d'avoir une queue d'au moins douze véhicules est de 7% environ.

Nous n'avons pas *démontré* que les probabilités $\mathcal{P}\left(Q_n = k\right)$ tendaient vers une limite, mais le calcul numérique montre bien une stabilisation des valeurs.

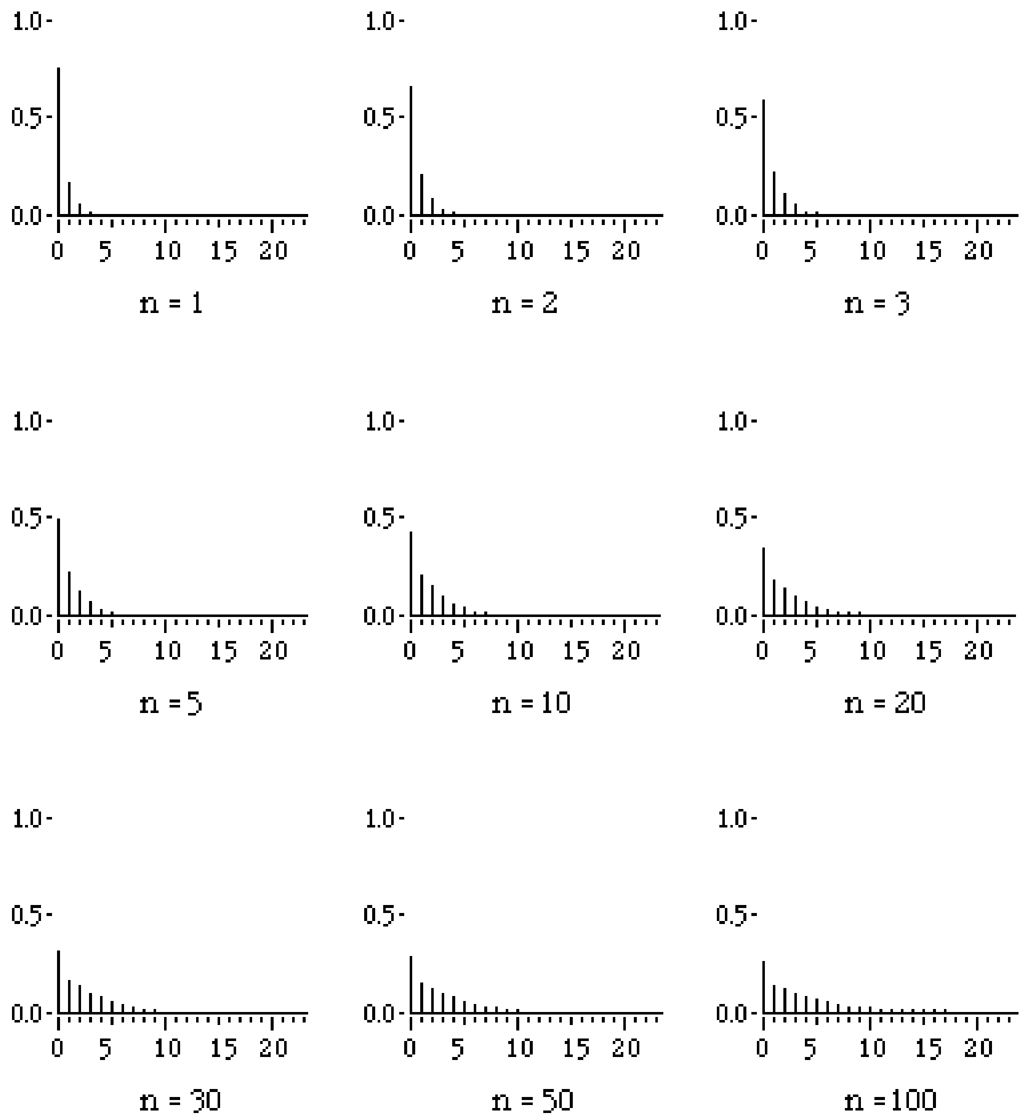

figure 38

Ici on représente la même chose que sur la figure 36, mais pour $\lambda = 0.9$ au lieu de 0.76. En comparant les deux figures, on peut observer que la barre d'abscisse zéro – qui représente la probabilité de n'avoir aucune queue – diminue ici plus vite, et les autres barres augmentent plus vite, ce qui est bien conforme à l'évidence. Mais on peut calculer quantitativement : par exemple la probabilité limite pour que la queue dépasse douze véhicules (obligeant ainsi les usagers à attendre leur tour plus de douze minutes) est ici de 7%, alors qu'elle était insignifiante pour $\lambda = 0.76$.

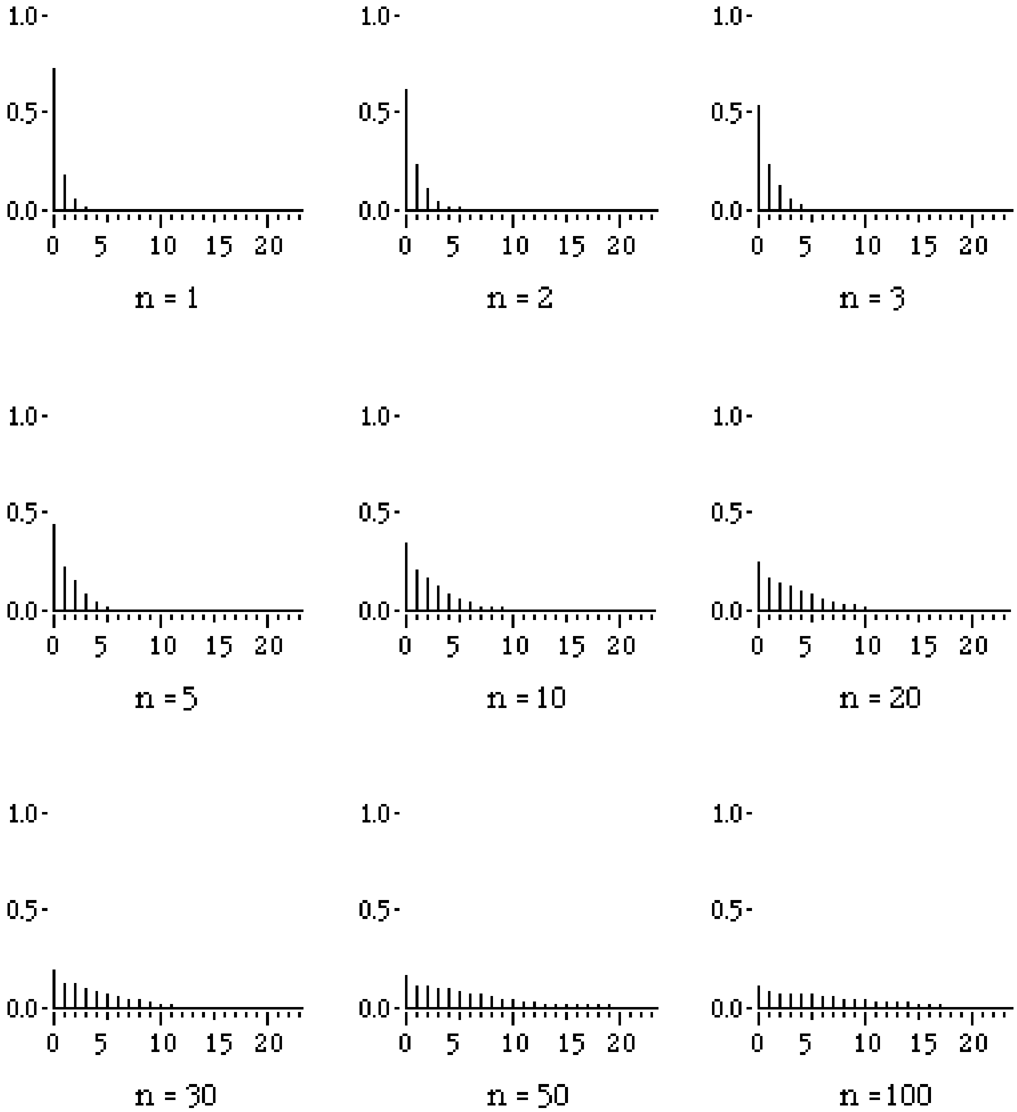

figure 39

Même chose que dans les figures 36 et 38, mais avec cette fois $\lambda = 0.99$.

On voit que $G''(0)$ est déjà relativement compliquée et cela permet d'imaginer à quoi pourrait ressembler par exemple $\mathcal{P}\,(Q = 12)$ si au lieu de le calculer numériquement par itérations, on en voulait une expression analytique formelle en fonction de λ.

Pour les valeurs en $z = 1$ on obtient des rapports indéterminés en $0/0$; on peut lever l'indétermination en recourant à la règle de l'Hospital, mais il est plus commode de revenir à $IX.16.$ qu'on dérive trois fois ; ainsi :

$$
\begin{gathered}
[z - F(z)]G(z) = (z-1)F(0)G(0) \\
[1 - F'(z)]G(z) + [z - F(z)]G'(z) = F(0)G(0) \\
-F''(z)G(z) + 2[1 - F'(z)]G'(z) + [z - F(z)]G''(z) = 0 \\
-F'''(z)G(z) - 3F''(z)G'(z) + 3[1 - F'(z)]G''(z) + [z - F(z)]G'''(z) = 0
\end{gathered}
$$

remplaçant après coup z par 1, on obtient $0 = 0$ pour la première ligne et $1 - \lambda = F(0)G(0)$ pour la seconde (ce qui n'apporte rien de nouveau), mais $-F''(1) + 2\,(1-\lambda)\,G'(1) = 0$ pour la troisième et $-F'''(1) - 3F''(1)G'(1) + 3\,(1-\lambda)\,G''(1) = 0$ pour la quatrième, ce qui donne aisément ce qu'on cherche :

$$
\begin{aligned}
G'(1) = \mathbf{E}(Q) = \frac{F''(1)}{2\,(1-\lambda)} &= \frac{\lambda^2}{2\,(1-\lambda)} \\
G''(1) = \frac{F'''(1) + 3F''(1)G'(1)}{3\,(1-\lambda)} &= \frac{\lambda^3(\lambda+2)}{6\,(1-\lambda)^2}
\end{aligned}
$$

d'où la variance

$$
\mathbf{Var}(Q) = G''(1) + \mathbf{E}(Q) - \mathbf{E}(Q)^2 = \frac{6\lambda^2 - 2\lambda^3 - \lambda^4}{12\,(1-\lambda)^2}
$$

Le moment est maintenant venu de résumer et de critiquer notre approche de ce problème des queues aux guichets. Pour le résoudre nous avons copieusement recouru à des approximations, qui nous étaient offertes sous la forme de lois asymptotiques, en l'occurrence la loi de Poisson et un passage à la limite.

La loi de Poisson est d'abord apparue comme celle (évidemment approchée) du nombre de véhicules qui arrivent pendant une minute, à condition de supposer, ce qui est essentiel, que la densité du trafic, définie par le flux moyen instantané $\lambda = 0.76$ véhicules par minute, est stable. En admettant que le flux instantané reste stable pendant une heure, on peut le mesurer en comptant le nombre de véhicules qui sont passés pendant une heure, puis en divisant ce nombre par soixante ; si le flux instantané était resté stable pendant deux heures, on aurait compté le nombre de véhicules passés pendant ces deux heures, et on aurait divisé ce nombre par cent-vingt. S'il n'était resté stable que pendant une demie-heure, on aurait divisé le nombre de véhicules passés pendant la demie-heure par trente. Plus haut, nous avons comparé cette idée de flux instantané à celui de vitesse instantanée en cinématique. On rencontre cependant ici une différence essentielle avec la cinématique.

Lorsqu'on veut définir la vitesse instantanée d'un mobile, on commence, comme nous avons fait ici, par définir la vitesse moyenne au cours d'un intervalle de temps ; disons pour fixer les idées qu'on la mesure en mètres par seconde. Sur une heure, c'est la distance parcourue en mètres, divisée par 3 600. Sur une minute, c'est la distance parcourue divisée par 60. On admet que lorsque l'intervalle de temps devient petit (par exemple une seconde), cette vitesse moyenne tend vers une limite, au point que si on mesurait la distance parcourue pendant un milliardième de seconde, et qu'on la divisait par 10^{-9}, on trouverait la même limite (ou une valeur plus exacte de cette limite, d'autant plus exacte que l'intervalle serait plus court). Avec le flux il n'en va plus de même, car le flux moyen calculé sur un intervalle de temps long (deux heures, une heure, une demie-heure) a des chances de rester à peu près le même ($\simeq 0.76$) à cause de la loi des grands nombres ; par contre, si on compte le nombre de véhicules qui passent pendant une minute, on n'obtiendra pas 0.76, mais des valeurs entières aléatoires $0, 1, 2, 3, \ldots$ distribuées justement selon la loi de Poisson. Loin de se rapprocher d'une limite, on obtiendra des valeurs de plus en plus divergentes lorsque la durée de l'intervalle sera rendue plus courte : par exemple, le nombre de véhicules arrivant pendant une seconde sera presque toujours 0, rarement 1, et pratiquement jamais 2 (il sera donné par la loi de Poisson de paramètre $\lambda/60 = 0.76/60 \simeq 0.01267$, qui donne 0.987 pour la probabilité d'avoir zéro véhicule, mais seulement 0.0125 pour la probabilité d'avoir un seul véhicule, et 0.000079 pour la probabilité d'avoir deux véhicules).

Le flux instantané ne peut pas être mesuré s'il varie sans cesse, car la loi des grands nombres est indispensable pour sa détermination ; il faut donc s'assurer de sa stabilité par des mesures préalables. Dans le cas du péage d'autoroute, il se peut que le flux instantané varie beaucoup au cours de la journée, mais se répète d'une journée à l'autre (en excluant évidemment les dimanches et jours fériés, les départs en vacances, etc.). Par exemple le flux peut être constamment variable, être différent à 8 heures, 8 heures trente, 9 heures, 9 heures trente, ..., mais être identique tous les jours à 8 heures, ou tous les jours à 8 heures trente. On pourra donc mesurer le flux moyen en additionnant sur vingt jours le nombre de véhicules qui arrivent entre 8 *h.* et 8 *h.* 05, puis entre 8 *h.* 05 et 8 *h.* 10, etc. Tous les procédés statistiques utilisables auront en commun de reposer sur une hypothèse d'invariance : soit stabilité sur une durée suffisamment longue, soit similitude des différentes journées entre elles. Comme cela avait été longuement expliqué au chapitre **I**, la notion de probabilité exige une hypothèse d'invariance ; nous voyons sur l'exemple présent qu'il en va de même lorsqu'on veut *mesurer*.

L'invariance permet de calculer des probabilités a priori, que nous avons opposées aux probabilités empiriques ; mais lorsqu'on veut mesurer de telles probabilités empiriques il faut aussi s'assurer qu'il y a une invariance quelque part. Ici, c'est l'invariance au cours du temps ou *reproductibilité*.

Au chapitre **I** nous avions signalé une analogie entre l'équiprobabilité qui résulte d'une invariance, et le temps physique qui résulte des invariances galiléennes. Poursuivant cette comparaison, nous pouvons dire que la mesure du temps (c'est-à-dire la réalisation d'appareils mesurant le temps) exige un principe physique préalable d'invariance : la mesure du temps présuppose une conviction (que seule une perception théorique peut donner) quant à l'homogénéité de l'écoulement du temps. De même on ne peut prétendre mesurer de probabilités empiriques sans postuler la loi des grands nombres, et pour que ce postulat soit valide, il faut une conviction que les conditions se reproduisent, que par exemple le trafic se reproduit à l'identique d'un jour à l'autre.

Il va de soi que la précision de mesures ainsi basées sur une hypothèse de reproductibilté très approximative et sur la loi des grands nombres ne peut en aucun cas atteindre la seconde décimale ; ainsi la valeur supposée de 0.76 pour le flux instantané supposé constant pourrait aussi bien avoir été de 0.7 ou 0.8 véhicules par minute. Il est très rare que des moyennes statistiques ou des probabilités empiriques soient connues avec une précision plus grande. En effet, la loi des grands nombres, avec les fluctuations gaussiennes étudiées au chapitre **VII**, nous a montré que l'écart-type des fluctuations était de l'ordre de la racine carrée du grand nombre. Cela signifie que si on détermine la moyenne λ en comptant le nombre de véhicules passés pendant une heure (qui est donc $0.76 \times 60 \simeq 105$), l'incertitude relative sera de l'ordre de $1/\sqrt{105}$, c'est-à-dire du dixième.

Il faut donc partir du principe que dans des problèmes comme celui du péage d'autoroute, ainsi que dans tous les problèmes de statistiques humaines ou industrielles, les probabilités sont déterminées au dixième, ou à la rigueur dans des cas exceptionnels, au centième près. Il n'y a guère que la physique statistique où on peut aller au delà, le grand nombre étant celui d'Avogadro ($\sim 10^{24}$) ; les grandeurs sont alors en principe définies (mais rarement *effectivement connues*) à 10^{-12} près.

On peut conclure que dans le calcul des probabilités une approximation même grossière est *toujours* meilleure qu'une expression mathématique compliquée. Il ne faut donc pas hésiter à utiliser l'approximation gaussienne, même pour la somme de dix ou vingt variables indépendantes, ou la formule de Stirling pour des factorielles : pour $4! = 24$ la formule de Stirling donne 23.506, pour $3! = 6$ elle donne 5.836, pour $2! = 2$ elle donne 1.919, et même pour $1! = 1$ elle donne 0.922 : à 10% près, la formule de Stirling est correcte pour *n'importe quelle factorielle*, $0!$ excepté.

On peut également utiliser l'approximation plus précise

$$n! \simeq n^n\, e^{-n} \sqrt{2\pi\,(n+1/6)}$$

qui donne 1.075 827 pour $1!$, 2.078 921 pour $2!$, et 6.160 443 pour $3!$; même $0!$ est alors approché par 1.023 327. Une précision relative de 2% dès $n = 0$, et qui s'améliore encore quand n augmente : 1% pour $n = 8$, 0.5% pour $n = 16$, 0.2% pour $n = 40$, ...

IX. 5. Autres lois asymptotiques.

En discutant l'exemple de la densité de Cauchy, nous avons remarqué que celle-ci apparaissait plutôt comme un déguisement de la loi normale, par l'effet d'une déformation non linéaire. Nous avons même pu démontrer mathématiquement que n'importe quelle densité fixée à l'avance pouvait être obtenue comme une déformation non linéaire de la densité gaussienne. Il en va de même pour la loi de Poisson : on pourrait démontrer que n'importe quelle loi discrète peut être obtenue artificiellement comme une déformation non linéaire de la loi de Poisson. Mais il ne s'agirait que de constructions artificielles qui ne répondraient pas à la véritable question : "quelles lois limites (asymptotiques) rencontre-t-on dans des problèmes ayant une pertinence réelle ?" (cette question a été discutée à la fin de la section **2**).

Bien que ce chapitre soit spécialement consacré aux lois asymptotiques, la plupart de ces lois ont été ou seront rencontrées dans d'autres chapitres à propos de problèmes particuliers. C'est pourquoi nous conclurons par un petit catalogue des lois asymptotiques restantes, avec référence au chapitre où elles sont traitées.

La loi du dernier retour

Elle a été rencontrée au chapitre **III**. Nous avions calculé alors que pour une marche aléatoire de $2n$ pas, la probabilité R_k pour que la marche soit passée *pour la dernière fois* par zéro à l'instant $2k$ est

$$R_k = 2^{-2n} \cdot \binom{2k}{k} \cdot \binom{2n-2k}{n-k} \qquad (IX.18.)$$

On en déduisait que la probabilité pour que le dernier retour avant l'instant $2n$ se soit produit *avant* l'instant 2ℓ est

$$S_\ell = \sum_{k=1}^{k=\ell} R_k$$

Lorsque n est grand on peut approcher R_k par $1/[\pi\sqrt{k(n-k)}]$ en approchant les factorielles des coefficients binômiaux par la formule de Stirling, puis remarquer que la somme des R_k est la somme de Riemann de l'intégrale

$$\int_0^{\ell/n} \frac{1}{\pi\sqrt{t\,(1-t)}}\,dt$$

d'où on déduit que

$$S_\ell \simeq \frac{2}{\pi}\arcsin\left(\sqrt{\frac{\ell}{n}}\right)$$

Ces résultats font donc apparaître une densité asymptotique : lorsque n est grand, la loi de $\frac{1}{n}\,R_k$ a pour densité la fonction

$$f(t) = \begin{cases} \dfrac{1}{\pi\sqrt{t\,(1-t)}} & \text{si } 0 < t < 1\,; \\ \\ 0 & \text{si } t \leq 0 \text{ ou } t \geq 1. \end{cases} \qquad (IX.19.)$$

La variable R_k ne pouvant prendre des valeurs en dehors de $0 < k < n$, il est logique que la densité f soit nulle (ou non définie) en dehors de $0 < t < 1$.

L'approximation qui conduisait à la densité $(IX.19)$ reposait sur la formule de Stirling $n! \simeq n^n e^{-n}\sqrt{2\pi n}$. Celle-ci n'est en principe correcte que pour n grand ; elle a été appliquée à l'expression $(IX.18)$ de la probabilité R_k, qui contient les factorielles $(2k)!$, $k!$, $(2n-2k)!$, et $(n-k)!$ donc il faut que k et $n-k$ soient tous deux assez grands. Voyons ce que cela donne en pratique : pour $n = 1000$ et $k = 1$ on obtient $R_1 = 0.0089$ (valeur exacte) et $R_1 \simeq 0.0101$ (valeur approchée), soit une différence relative de 12% ; pour $k = 2$, $R_2 = 0.0067$ (valeur exacte) et $R_2 \simeq 0.0071$ (valeur approchée), soit une différence relative de 6.3%. Pour $k = 10$, $R_{10} = 0.00316$ (valeur exacte) et $R_{10} \simeq 0.00320$ (valeur approchée), soit une différence relative de 1.2%. La seule erreur vraiment grosse concerne les extrémités, $k = 0$ ou $k = n$, car alors la valeur approchée est infinie et la valeur exacte $R_0 = 0.0178$. On peut constater sur les valeurs numériques données ci-dessus que la valeur approchée est toujours plus grande que la valeur exacte calculée par les factorielles ; pour de petites valeurs de k, la différence ne diminue pas lorsque n tend vers l'infini, mais ces valeurs représentent par contre une proportion

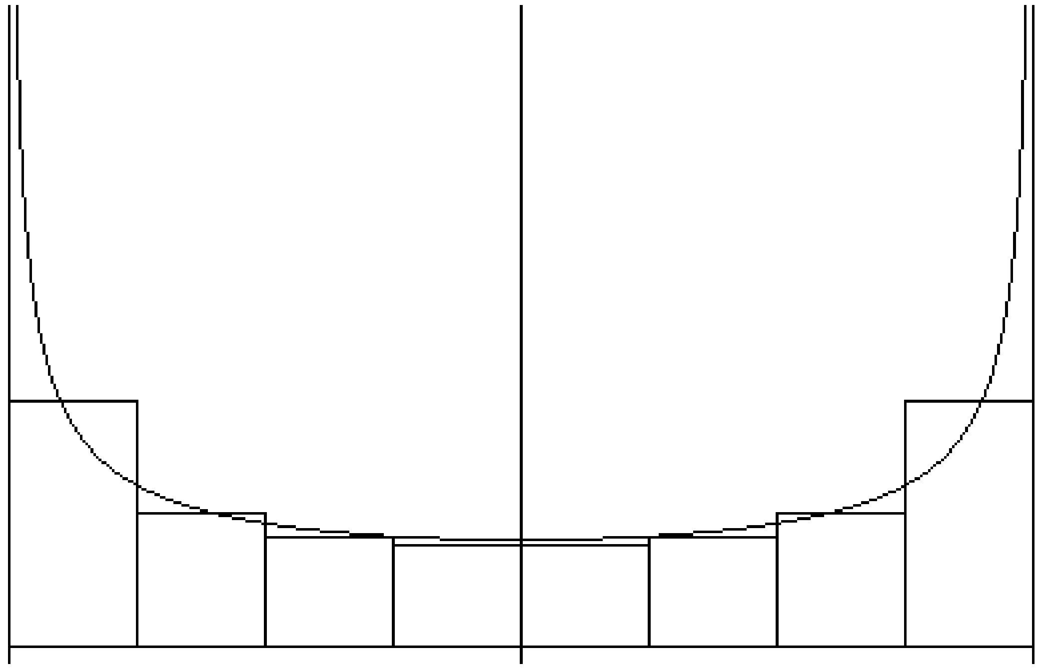

figure 40a

Ce graphique compare les probabilités de dernier retour R_k pour $k = 0 \dots 7$ ($n = 7$), représentées par les hauteurs des huit rectangles, à la densité $f(x) = 1/\pi\sqrt{x(1-x)}$.

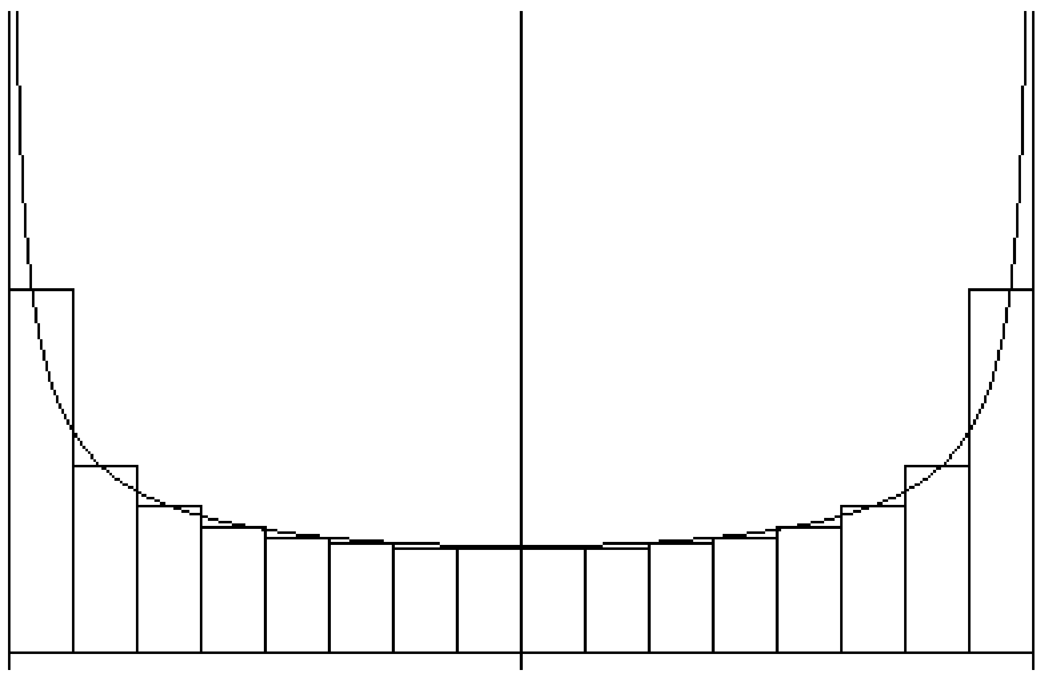

figure 40b

Ici $n = 15$, les probabilités R_k sont représentées par les hauteurs des seize rectangles, la courbe de la densité $f(x)$ est inchangée.

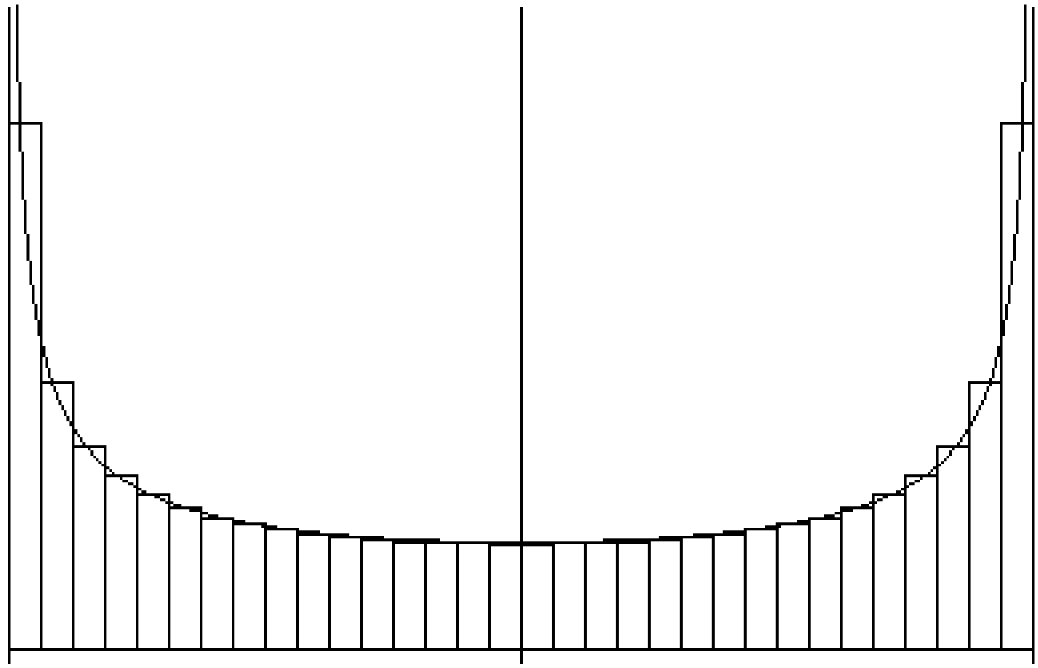

figure 40c

Même chose pour $n = 31$.

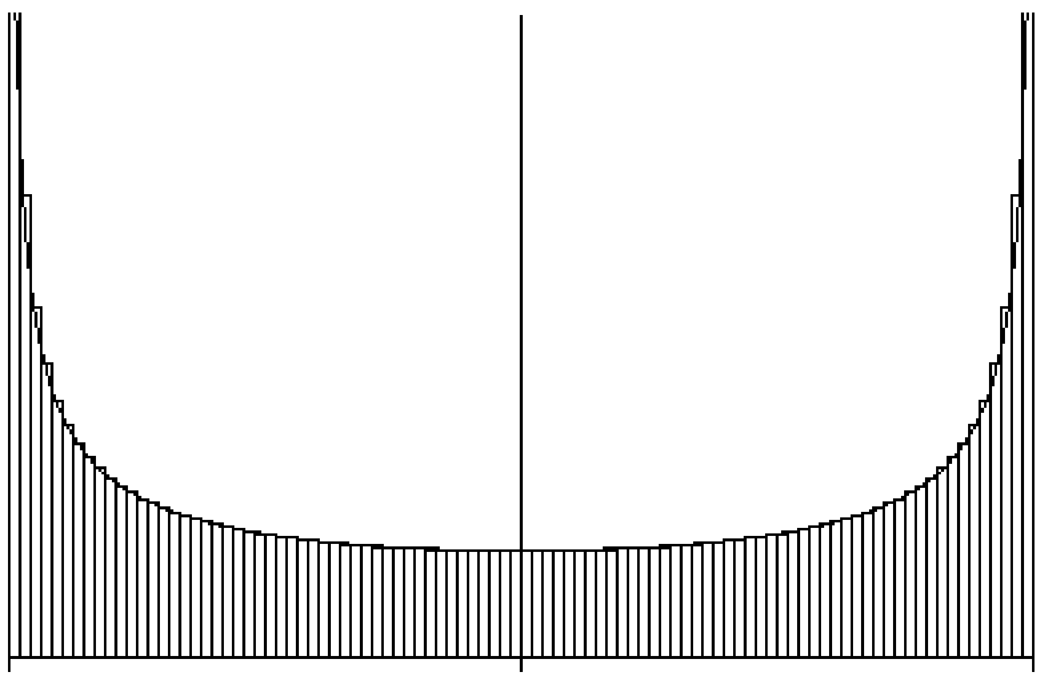

figure 40d

$n = 95$.

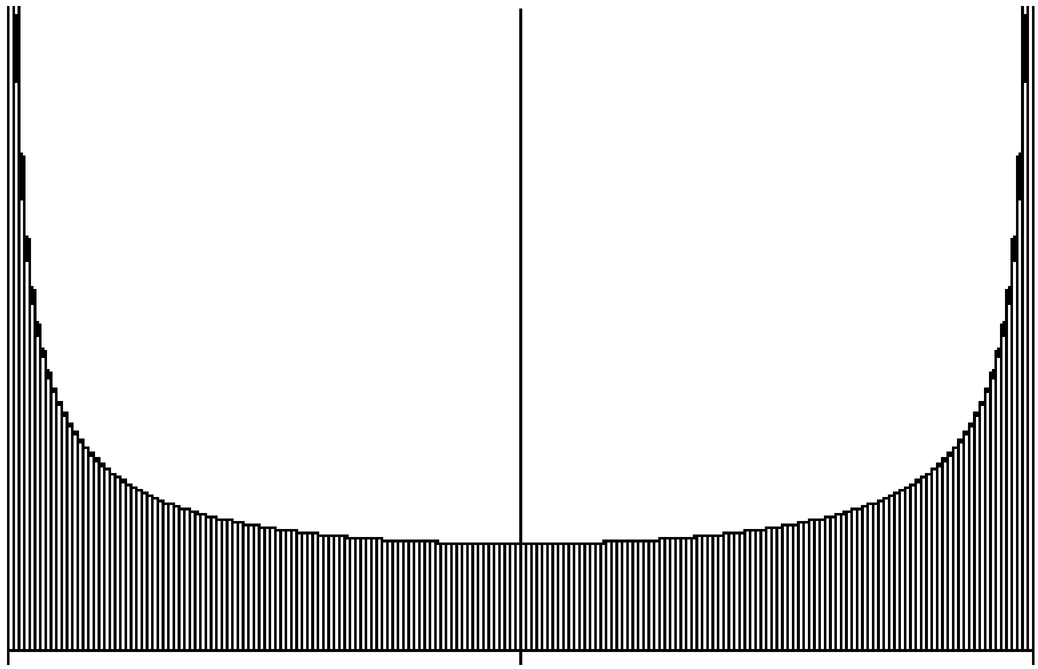

figure 40e

$n = 191$. Il est devenu impossible de distinguer les sommets des rectangles de la courbe.

de plus en plus petite de l'ensemble de toutes les valeurs (qui vont de zéro à n).

On peut voir sur les figures 40 (a, b, c, d, e) comment la loi discrète se rapproche de la densité asymptotique. Ces figures se passent de commentaire.

La loi de Bernoulli.

En statistique, un procédé classique est le *sondage par échantillon*. On veut étudier un certain caractère dans une population (par exemple combien d'électeurs vont voter pour le Parti du Progrès). Au lieu d'interroger tout le monde, on prend au hasard un échantillon de deux mille personnes. Le principe est que la probabilité pour que la proportion des partisans du Progrès dans l'échantillon diffère sensiblement de la proportion réelle dans la population totale, est faible. Mais cela n'est vrai que pour des échantillons assez gros.

On peut utiliser les méthodes qui ont été présentées dans ce livre pour calculer cette probabilité : on cherche une invariance, puis on calcule par dénombrement, ce qui conduira à une formule pleine de factorielles ; ensuite, on cherche une approximation simple qui s'applique aux grands entiers, ce qui donnera une loi asymptotique.

L'invariance résulte du choix *au hasard* de l'échantillon : "au hasard" signifie que tous les échantillons sont équiprobables. Dans une population de N personnes, il y a $\binom{N}{n}$ échantillons de taille n. Le problème sera étudié en détail au chapitre **X**, section **2** (*la théorie des échantillons de Bernoulli*), de sorte que le lecteur y est renvoyé pour les calculs. Si p est le nombre de partisans du Progrès dans la population totale, la probabilité pour que l'échantillon de n personnes tiré au hasard en contienne k est

$$p_k = \frac{\binom{n}{k}\cdot\binom{N-n}{p-k}}{\binom{N}{p}} = \frac{\dfrac{n!}{k!\,(n-k)!}\cdot\dfrac{(N-n)!}{(p-k)!\,(N-p-n+k)!}}{\dfrac{N!}{p!\,(N-p)!}}$$

Ceci est la loi "mathématiquement exacte", qui est valable quels que soient les nombres N, p, n, et k. On l'appelle *la loi hypergéométrique*.

Cette loi gouverne les problèmes de tirage du type : *Une urne contient N boules, dont p blanches et $N-p$ noires. On tire au hasard et sans remise n boules dans l'urne ; quelle est la probabilité pour que l'échantillon tiré contienne k boules blanches et $n-k$ boules noires ?*

Si n est beaucoup plus petit que N, on peut faire les approximations suivantes :

$$\begin{aligned}
(N-n)! &= \frac{N!}{N(N-1)(N-2)\cdots(N-n+1)} \simeq \frac{N!}{N^n}\\
(p-k)! &= \frac{p!}{p(p-1)(p-2)\cdots(p-k+1)} \simeq \frac{p!}{p^k}\\
(N-p-n+k)! &= \frac{(N-p)!}{(N-p)(N-p-1)\cdots(N-p-n+k+1)}\\
&\simeq \frac{(N-p)!}{(N-p)^{n-k}}
\end{aligned}$$

(ces approximations supposent évidemment que $n \ll N$, $k \ll p$, et $n-k \ll N-p$). Par conséquent, si on pose $\alpha = p/N$ et $\beta = (N-p)/N = 1-\alpha$, on aura :

$$p_k = \binom{n}{k}\cdot\frac{\binom{N-n}{p-k}}{\binom{N}{p}} \simeq \binom{n}{k}\,\alpha^k\,\beta^{n-k}$$

Cette approximation, valable pour N grand, est donc une loi asymptotique, appelée *loi de Bernoulli*. On peut aussi la rencontrer sous une forme non asymptotique dans le problème de l'urne cité plus haut, mais pour des tirages *avec* remise : si on tire successivement et avec remise n boules dans l'urne, à chaque tirage la probabilité de tirer une boule blanche est $\alpha = p/N$; la fonction génératrice pour chaque tirage est donc $G(z) = \alpha + \beta\, z$; les tirages avec remise sont indépendants, donc la fonction génératrice pour n tirages est $(\alpha + \beta\, z)^n$; en développant selon la formule du binôme, on retrouve bien la loi de Bernoulli.

Il n'est évidemment pas surprenant que la loi sans remise donne asymptotiquement pour N grand la même loi qu'avec remise : en effet, si $n \ll N$, le tirage de n boules sans remise modifie à peine les proportions de boules noires ou blanches, de sorte qu'on doit obtenir à peu près le même résultat qu'avec remise.

Nous ne mentionnons cette loi ici que pour insister sur le fait qu'elle peut aussi se rencontrer sous forme asymptotique. De toute façon elle sera étudiée en détail au chapitre **X**.

Cette loi décrit le sondage par échantillon lorsque $n \ll N$; mais signalons — ce sera étudié au chapitre **X** — que si n devient lui-même grand, ainsi que k et $n - k$, la loi de Bernoulli devient gaussienne : comme la loi de Poisson, la loi de Bernoulli est une loi asymptotique de l'échelle intermédiaire ; si n est grand, mais que k reste petit et $\beta \exp\{\alpha/\beta\} \simeq 1$ (par exemple $k \leq 10$, et $10 \ll n \ll N$) alors la loi de Bernoulli devient une loi de Poisson. Cela montre que la même loi hypergéométrique peut se transformer en plusieurs sortes de lois asymptotiques selon l'échelle, c'est-à-dire selon les ordres de grandeur relatifs des nombres entiers N, p, n, et k. Mais on constate que l'on retrouve toujours la même famille : loi gaussienne ou loi de Poisson. Cela provient de ce que la loi de Bernoulli est la loi d'une somme de variables indépendantes.

Les lois du khi-deux et de Student.

Enfin, deux lois asymptotiques à densité continue que nous signalons sans insister : la loi du khi-deux ou χ^2 est la densité d'une somme de *carrés* de variables gaussiennes indépendantes. Si X_1, X_2, … X_n sont n variables aléatoires indépendantes, ayant chacune une loi de densité gaussienne de variance 1 et de moyenne 0, alors la somme des carrés $S = X_1^2 + X_2^2 + \cdots + X_n^2$ aura la densité χ^2 à n degrés de liberté. Cette densité sera étudiée en détail au chapitre **XI**. Elle est à la base d'un test statistique simple et bien connu.

La loi de Student est la loi du quotient d'une variable gaussienne par la racine carrée d'une somme de carrés de gaussiennes : si Y, X_1, X_2, … X_n sont $n + 1$ variables aléatoires gaussiennes indépendantes, de variance 1 et de moyenne 0, alors la variable

$$T = \frac{Y}{\sqrt{X_1^2 + X_2^2 + \cdots + X_n^2}}$$

aura une loi de densité de Student. Cette densité aussi est, tout comme celle du χ^2, à la base d'un test statistique bien connu, et sera étudiée dans cet esprit au chapitre **XI**.

Comme le montrent leurs définitions même, ces lois ne sont, comme la loi de Cauchy, qu'un déguisement de la loi normale. Leur fonction est de simplifier le travail des statisticiens en fournissant des tests *clés en main*. Ces densités ne se rencontrent pas directement "dans la nature" ; elles apparaissent parce qu'on choisit délibérément, pour la commodité du test, de calculer les grandeurs de manière non linéaire (somme des carrés des écarts).

On peut noter que la loi de Cauchy est le cas particulier de la loi de Student pour $n = 1$.