

Statistiques de comptage

par Jörg W. Müller

Bureau International des Poids et Mesures, F-92310 Sèvres

Parmi les travaux accomplis dans ce domaine, nous en choisissons deux de caractères assez différents, à titre d'illustration. Dans le premier cas il s'agit d'un problème qui a déjà surgi dans divers contextes, mais n'a guère reçu de traitement adéquat. L'autre exemple concerne les débuts d'un nouveau développement qui, par une généralisation appropriée, permet d'établir un lien entre deux modèles de temps mort qui semblaient s'exclure mutuellement. Ainsi, le vieux domaine des statistiques de comptage se renouvelle tout en gardant sa jeunesse.

Sur le choix d'une impulsion au hasard

A première vue, il paraît difficile d'admettre que choisir tout à fait "au hasard" une impulsion parmi toutes celles formant un processus aléatoire pourrait poser un problème sérieux. Puisqu'aucun événement ne paraît être dissemblable des autres, n'importe lequel devrait nous convenir. Il faut, cependant, être prudent et regarder de plus près comment un tel choix se fait en réalité. Le procédé le plus couramment utilisé consiste à prendre, à partir d'un moment déterminé arbitrairement dans le temps, la première impulsion qui se présente. Or, il se trouve qu'un tel choix est aléatoire dans le temps, mais pas dans les intervalles.

Considérons un simple processus de Poisson, à taux de comptage ρ , pour lequel on sait que la densité pour un intervalle d'ordre k est donnée par

$$f_k(t) = \frac{\rho(\rho t)^{k-1}}{(k-1)!} e^{-\rho t}, \quad \text{avec } k = 1, 2, \dots$$

Pour l'intervalle t entre impulsions consécutives on a donc la simple forme exponentielle

$$f_1(t) = \rho e^{-\rho t}, \quad \text{pour } t > 0.$$

Puisque la fréquence avec laquelle on "tombe" sur un certain intervalle est proportionnelle à sa durée t , la densité effective de cet intervalle

(qui précède l'impulsion prise pour origine du temps) est donnée par

$$g(t) = f \frac{t f_1(t)}{\int_0^{\infty} t f_1(t) dt} = \rho^2 t e^{-\rho t} = f_2(t) .$$

Il s'ensuit que l'impulsion que l'on croyait choisie au hasard est en réalité précédée d'un intervalle dont la longueur moyenne correspond au double de la distance qui sépare deux événements dans la séquence originelle.

Puisque dans un processus de renouvellement les intervalles sont indépendants les uns des autres, tous ceux qui suivent ou qui précèdent ce premier intervalle, décrit par la densité $g(t)$, sont "normaux" et donc donnés par $f_1(t)$. Pour l'intervalle numéro k qui précède l'origine du temps $t = 0$ on a ainsi la densité (avec $x \equiv -t$)

$$g_k(x) = g(x) * f_{k-1}(x) = f_{k+1}(x) .$$

La densité totale est alors

- pour le domaine positif du temps:

$$F(t) = \sum_{k=1}^{\infty} f_k(t) = \rho e^{-\rho t} \sum_{k=1}^{\infty} \frac{(\rho t)^{k-1}}{(k-1)!} = \rho ,$$

- pour le domaine négatif ($t < 0$, d'où $x > 0$):

$$G(x) = \sum_{k=1}^{\infty} g_k(x) = \sum_{k=2}^{\infty} f_k(x) = \rho - f_1(x) .$$

Ce raisonnement se généralise facilement et on peut montrer que si l'on prend comme point de départ (et origine du temps) l'impulsion numéro $K \geq 1$, on trouve pour la densité totale

$$G(x) = \rho - f_K(x) = \rho \left[1 - \frac{(\rho x)^{K-1}}{(K-1)!} e^{-\rho x} \right] .$$

Pour le domaine $t > 0$ la densité reste toujours à sa valeur initiale ρ , indépendamment du choix de K .

Il est facile de vérifier ce comportement expérimentalement en accumulant dans un analyseur multicanal, utilisé en mode multiéchelle, les temps d'arrivée des impulsions. Comme le montre la figure 1, la répartition ainsi obtenue correspond bien à ce que l'on doit attendre.

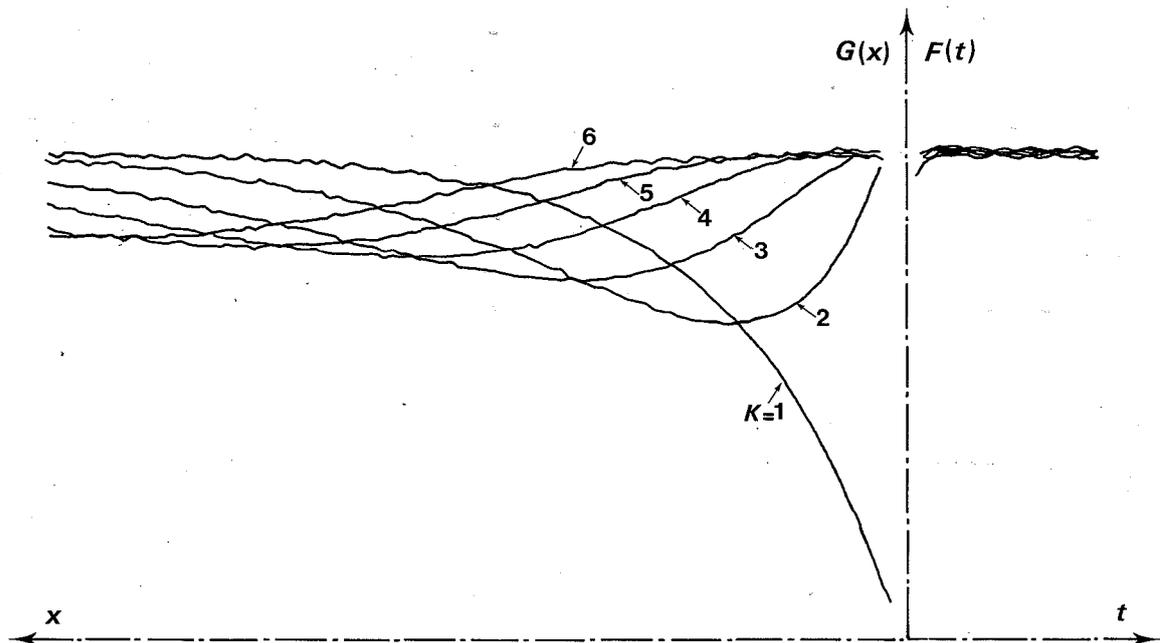


Fig. 1.- Répartition observée des temps d'arrivée des impulsions au voisinage d'une impulsion choisie au hasard. K est le rang de l'impulsion qui définit l'origine du temps.

Cependant, il reste un petit problème ennuyeux, car on ne sait toujours pas comment faire pour choisir une impulsion qui soit précédée d'un intervalle sans biais. En pratique, l'utilisation de $K \gg 1$ peut être acceptable, mais ce subterfuge manque d'élégance. La situation peut être améliorée à l'aide d'une méthode qui, au lieu d'un intervalle, vise directement une impulsion, comme c'est le cas pour les coïncidences. D'autre part, pour un processus perturbé par un temps mort τ il existe un moyen simple et rigoureux pour s'assurer d'un comportement "normal" dans le domaine de x à 0: il suffit de choisir $K > x/\tau$. Pour plus de détails, nous renvoyons au Rapport BIPM-85/4.

Intervalles après un temps mort généralisé

Pour la généralisation d'un procédé on peut normalement envisager un grand nombre de possibilités, mais ce choix est souvent fortement restreint par des conditions supplémentaires que des considérations pratiques nous imposent. Dans le cas présent, où l'on veut se libérer des modèles trop spécifiques d'un temps mort de type étendu ou non étendu, on aimerait, par exemple, disposer d'une généralisation qui inclut ces deux types comme cas limites. De plus, il serait fort souhaitable d'avoir affaire à un modèle qui nous permette de trouver des solutions exactes pour les grandeurs qui sont les plus utiles dans les applications pratiques. Si l'on veut respecter ces contraintes, un seul modèle est actuellement connu; il a été proposé il y a longtemps, mais est resté peu exploité.

La généralisation indiquée par Albert et Nelson admet que la prolongation d'un temps mort, par l'arrivée d'une nouvelle impulsion, se fait avec une probabilité θ , supposée indépendante des décisions précédentes. Pour ce modèle, et en admettant à l'entrée du dispositif un processus de Poisson, Takács a indiqué une expression pour la transformée de Laplace de la densité des intervalles à la sortie. Elle a la forme

$$\tilde{\theta}f(s) = \frac{\theta\rho+s}{\rho+s} \frac{\rho e^{-(\theta\rho+s)\tau}}{s + \theta\rho e^{-(\theta\rho+s)\tau}},$$

où ρ est le taux de comptage originel tandis que les paramètres τ et θ caractérisent le temps mort généralisé.

Sous cette forme, cependant, la formule est d'une utilité très limitée. Nous avons réussi à trouver l'original correspondant, mais puisque la dérivation de ce résultat est assez longue, nous nous contentons d'en indiquer ici le résultat (pour les détails techniques, voir le Rapport BIPM-85/5) qui peut être écrit sous la forme

$$\theta f(t) = \sum_{j=1}^J \left\{ \frac{1}{\theta} A_j(t) + \left(\frac{1-\theta}{\theta} \right) B_j(t) \right\},$$

$$\text{où } A_j(t) = \frac{\theta\rho}{(j-1)!} (-\theta T_j)^{j-1} e^{-j\theta\rho\tau}$$

$$\text{et } B_j(t) = \rho(\theta e^{-\theta\rho\tau})^j \left[e^{-T_j} + (-1)^j \sum_{k=1}^{j-1} \frac{(-1)^k}{(j-1-k)!} T_j^{j-1-k} \right],$$

avec $T_j = \rho(t-j\tau)$ et où J est le plus grand nombre entier inférieur à t/τ .

L'évaluation pratique des coefficients A_j et B_j s'effectue aisément à l'aide d'un petit programme d'ordinateur.

Une étude plus approfondie montre que la fonction de densité ainsi obtenue peut presque toujours être suffisamment bien approchée par une superposition linéaire des densités limites, c'est-à-dire par

$$\theta f(t) \approx (1-\theta) {}_0f(t) + \theta {}_1f(t).$$

On peut prouver que cette relation est exacte pour $\theta = 0,5$ et $t = 1,5 \tau$, situation dans laquelle on a

$${}_{1/2}f(3\tau/2) = \frac{\rho}{2} e^{-\rho\tau} (1 + e^{\rho\tau/2}).$$

Grâce à la construction d'un dispositif électronique qui permet de simuler le comportement d'un temps mort généralisé, il a été possible d'enregistrer des distributions expérimentales d'intervalles (fig. 2), et l'accord avec les prévisions est excellent. Dans une étape ultérieure, il s'agira d'étudier la distorsion introduite par un tel temps mort généralisé dans la loi de Poisson, et en particulier d'évaluer les trois premiers moments que l'on peut également mesurer expérimentalement.

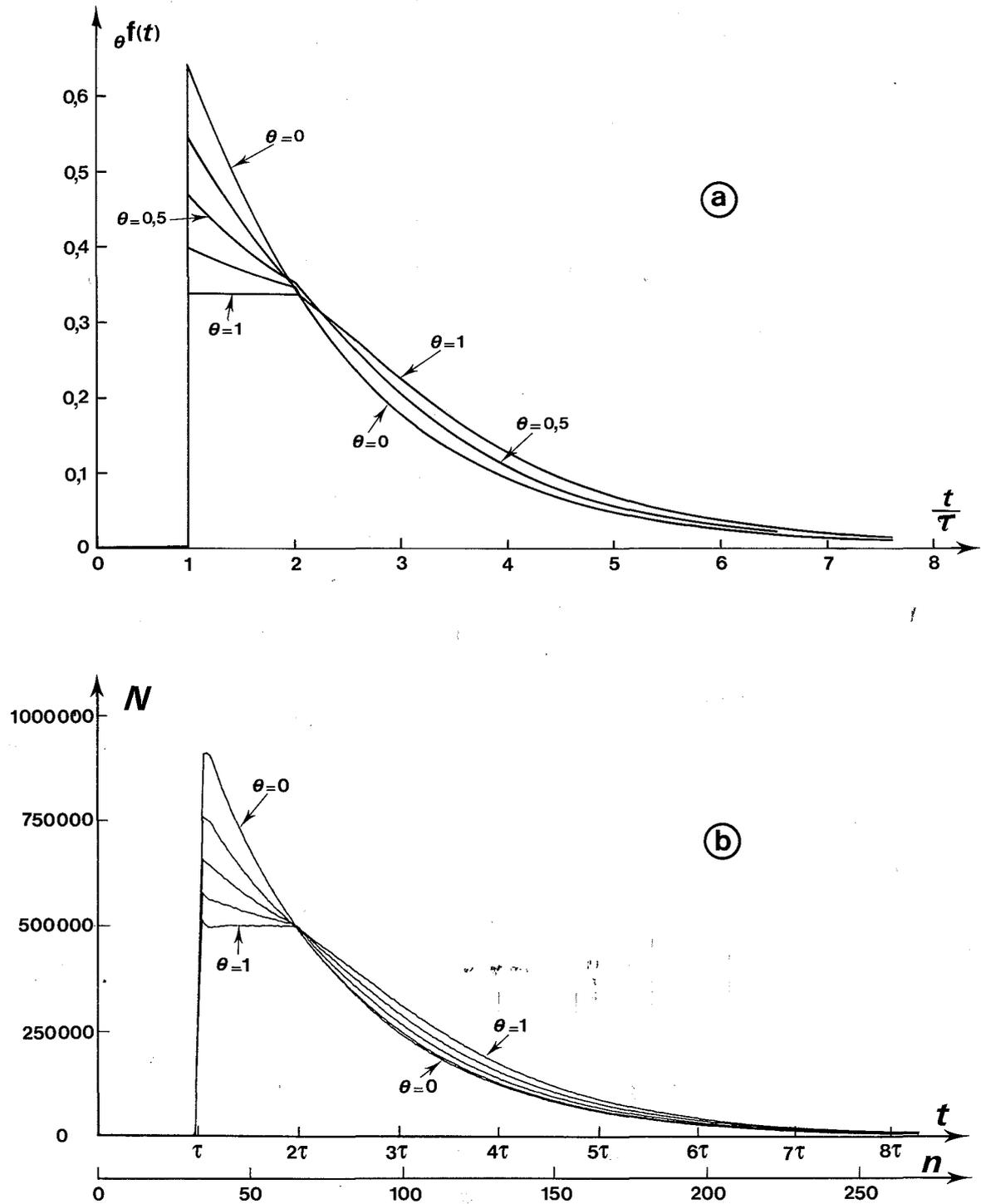


Fig. 2.- Distribution des intervalles entre impulsions après un temps mort généralisé pour diverses valeurs du paramètre θ , pour $\rho = 40\,000\text{ s}^{-1}$ et $\tau = 16\ \mu\text{s}$.
 a, distribution théorique;
 b, distribution observée: n , numéro du canal; N , nombre d'intervalles par canal (un canal correspond à $0,5\ \mu\text{s}$).

Autres travaux

L'utilité pratique du concept d'un temps mort généralisé a été mise en évidence dans la note BIPM WPN-227, où l'on décrit comment les deux paramètres τ et θ ont pu être mesurés pour notre chaîne électronique qui sert à toutes les mesures absolues d'activité. Un simple algorithme numérique permettant d'obtenir le taux de comptage originel à partir du taux mesuré (et pour des paramètres connus) est indiqué dans la note BIPM WPN-228. Quelques compléments mathématiques concernant le problème de trouver une fonction réciproque (étudié et utilisé précédemment) sont rassemblés dans le Rapport BIPM-85/1.

Les questions qui se rapportent à l'évaluation, l'expression et l'utilisation d'incertitudes de mesures expérimentales continuent à préoccuper le BIPM. Une revue de l'état actuel de ces problèmes, intitulée "The treatment of measurement uncertainties", a été préparée pour un cours de métrologie qui s'est tenu en Autriche. De plus, le problème d'une éventuelle mauvaise pondération de résultats de mesure a été discuté et illustré dans le Rapport BIPM-84/7.

(Octobre 1985)