Détermination d'une dérivée pour une répartition empirique (J.W.M.)

Nous voudrions faire une comparaison détaillée de la distribution expérimentale dite de Baerg, c'est-à-dire de la fonction de densité $F(t)$ des intervalles de temps qui résulte de la superposition d'un processus de Poisson et d'une fréquence fixe, avec la prévision théorique (voir Rapport du Directeur 1968/9).

Il est assez évident que cela peut se faire de préférence à l'aide des dérivées des logarithmes des deux répartitions F et f. Leur rapport correspondrait alors au changement de l'exposant, et pour celui-ci on s'attend à une dépendance de la forme

$$\frac{\frac{d}{dt} \ln F(t)}{\frac{d}{dt} \ln f(t)} = 1 + \frac{1}{2 + \rho\,(T-t)} \qquad \text{pour } 0 < t < T \,.$$

Mais bien que l'enregistrement automatique de la répartition des intervalles sur un sélecteur multicanaux ait été prolongé jusqu'à 60 heures environ, la fluctuation relative du nombre d'impulsions par canal - en particulier dans la "queue" de la répartition, qui est pour nous la partie la plus intéressante -, est loin d'être négligeable. Elle ne permet donc pas une détermination directe de la dérivée, par exemple en formant les différences de points consécutifs.

Après quelques essais peu convaincants sur l'ordinateur, un procédé beaucoup plus primitif que celui que nous avions envisagé au début (ajustement de polynômes par moindres carrés) s'est avéré utile et efficace. Il consiste à lisser d'abord les n données expérimentales (ou leurs logarithmes), dont les arguments t sont supposés équidistants, en les remplaçant tout simplement par leur moyenne locale, c'est-à-dire par exemple par

$$f(t_i) \longrightarrow \bar{f}(t_i) = \frac{1}{5} \sum_{k=-2}^{2} f(t_{i+k}) \,, \qquad i = 3,\ 4,\ \ldots,\ n-2,$$

et analogue pour $F(t)$.

Ce lissage peut être répété un certain nombre de fois, si nécessaire.

Quant à la dérivée, elle peut alors se former en calculant par exemple

$$\frac{d}{dt} f(t_j) \cong \frac{1}{10} \sum_{k=-2}^{2} k \, \overline{f} (t_{j+k}) .$$

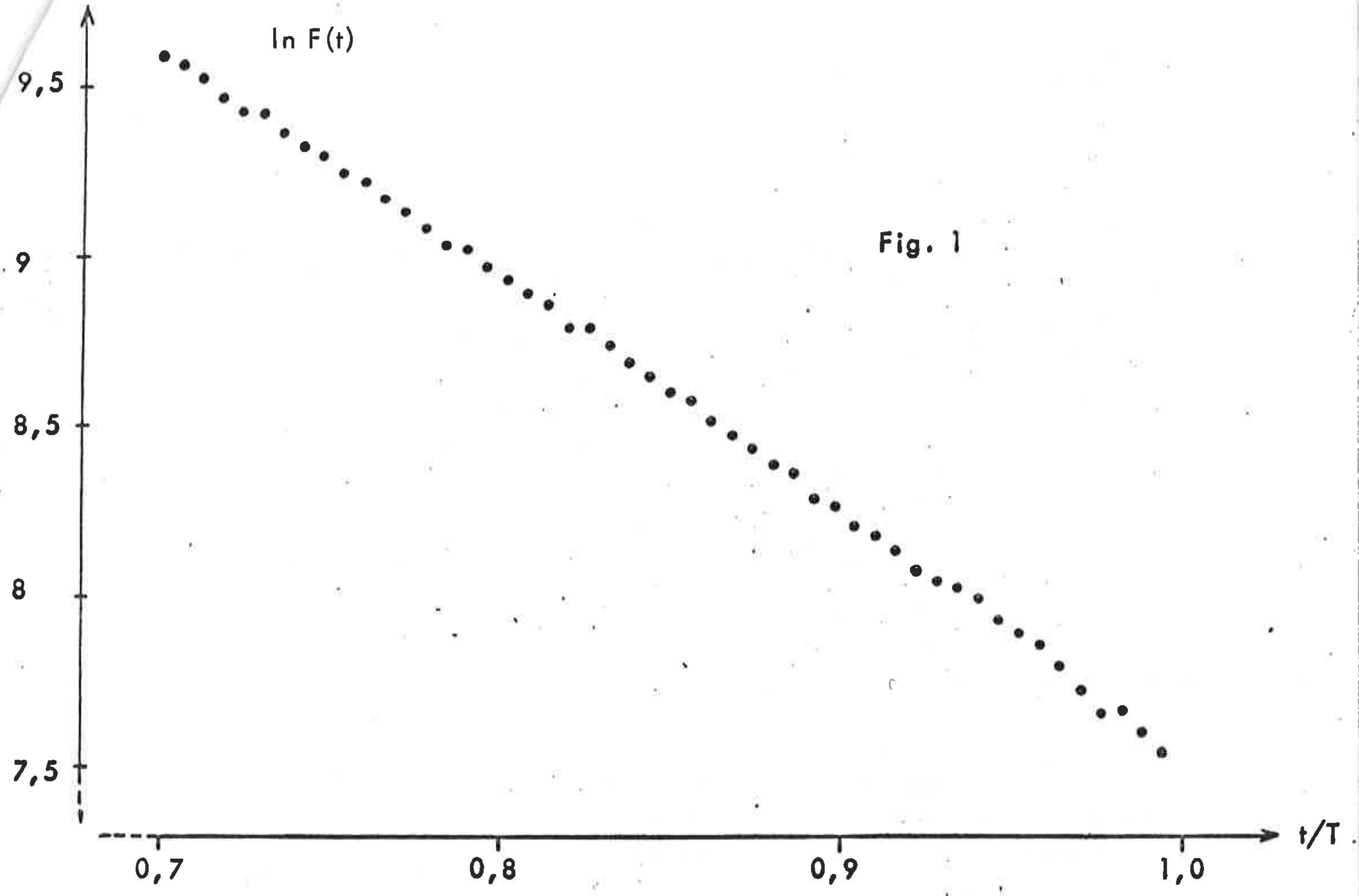
Cette méthode de calcul, qui correspond à la différentiation d'un polynôme obtenu par ajustement plutôt que par interpolation, est supposée moins sensible aux erreurs aléatoires des données expérimentales.

Finalement, les résultats de cette différentiation numérique peuvent être soumis de nouveau à un lissage analogue à celui décrit plus haut.

Grâce à l'ordinateur il est facile d'ailleurs d'exécuter des calculs de ce type pour des lissages répétés, de même qu'avec des formules qui se basent sur des polynômes d'ordres différents. Cependant, les distributions limites ne sont pas toujours nécessairement les mêmes. Pour des ordres trop élevés nous avons même observé qu'on arrive parfois à des structures apparemment "inventées" qui n'ont guère de ressemblance avec les données de départ.

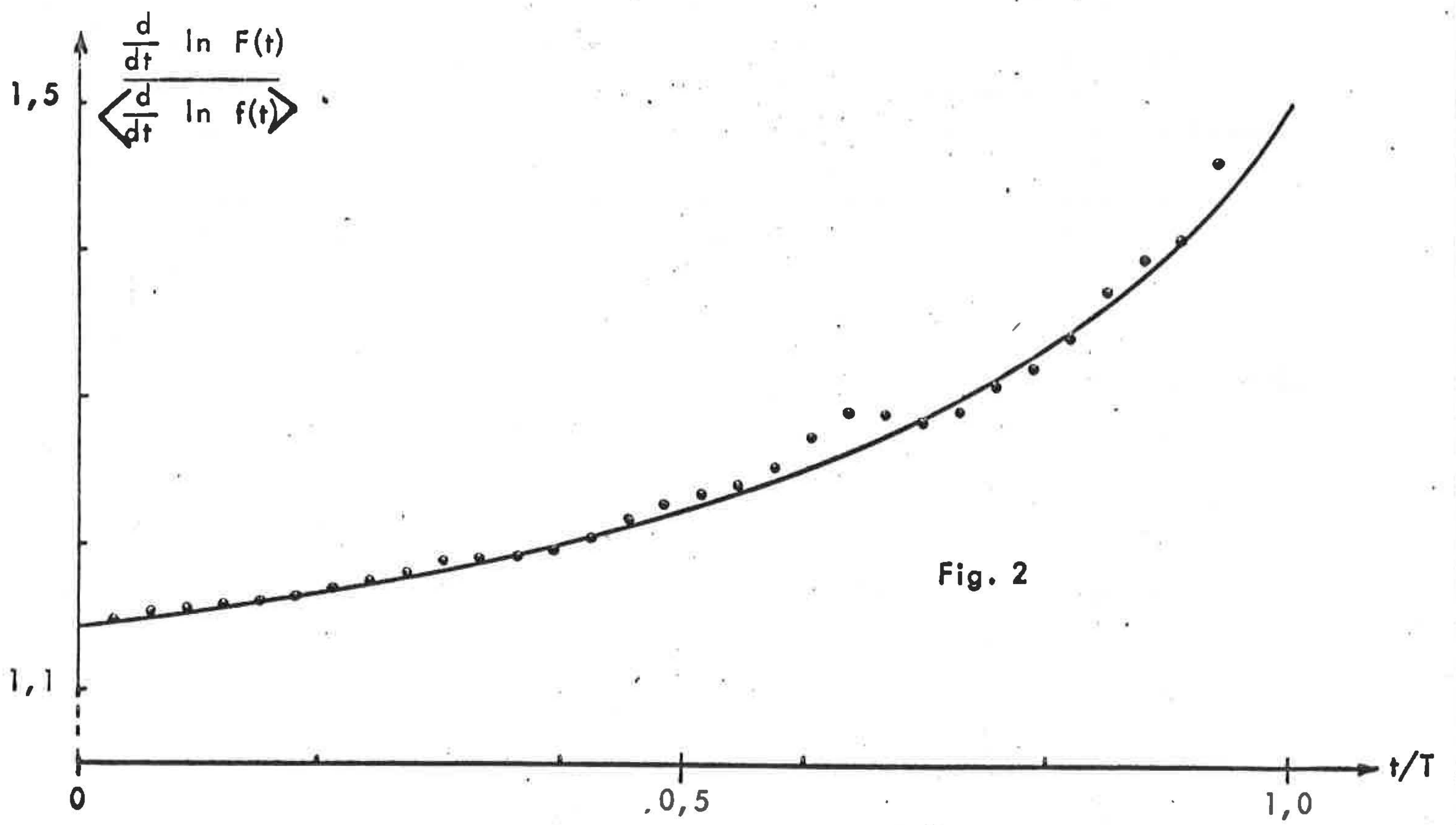
Pour illustrer le fonctionnement de ce procédé de calcul, la fig. 1 représente une partie des données originales F(t), enregistrées par le multicanal, pour une distribution de Baerg (avec $\rho T = 5$), tandis que la fig. 2 nous en donne (pour 10 lissages avant et après la différentiation) la dérivée numérique, de même que l'espérance théorique. L'accord est très satisfaisant.

Octobre 1969



ln F(t)
9,5
9
8,5
8
7,5
0,7
0,8
0,9
1,0
t/T

Fig. 1



$\dfrac{\frac{d}{dt}\ \ln\ F(t)}{\left\langle \frac{d}{dt}\ \ln\ f(t) \right\rangle}$
1,5
1,1
0
0,5
1,0
t/T

Fig. 2