Le critère (9) donne

$$\frac{\mu'_3 m'_1}{\mu'_2 (2\mu'_2 - m'_1)} = \frac{448,75}{760,48} = 0,59.$$

A la fin il faut faire remarquer qu'avec la question des équations aux différences finies dans l'analyse des fonctions de fréquence discontinues s'est occupé systématiquement M. Ragnar Frisch dans son article cité plus haut et pour les fonctions de fréquence discontinues de deux variables de nouveau M. A. Guldberg dans son travail „On discontinuous frequency functions of two variables" dans Skandinavisk Aktuarietidskrift 1933.

Les Conférences du prof. René Risser à Prague.

Monsieur René Risser, professeur au Conservatoire des arts et métiers et l'ancien chef du Actuariat du ministère du travail à Paris a été récemment à Prague, où il a eu plusieurs Conférences.

Dans la conférence, tenue à l'Institut social tchécoslovaque, il nous a donné une vue sur le développement de la loi sur les assurances sociales en France, passant par les rapports des experts jusqu'à la loi définitive de 1930. Les raisons principales qui ont exigé l'introduction de cette loi en France sont les suivantes: les conséquences de la guerre mondiale sur la morbidité de la population en France, la mortalité augmentée de la masse ouvrière, l'insuffisance des lois anciennes sur l'assurance maladie et l'assurance-pension et finalement l'aplication des lois sur les assurances sociales en Alsace et Lorraine.

La nouvelle loi sur les assurances sociales est appliquée depuis le 1er juillet 1930. Le nombre d'assuré a atteint le chiffre de 11,320.000 à la fin de l'année 1932, donc une augmentation de plus de 1,250.000 en comparaison avec l'effectif de l'année 1931. Le total des cotisations fait à la fin de l'année 1931 environ 5,2 milliards Frs et à la fin de l'année 1932 environ 8,7 milliards Frs. Le nombre des caisses d' assurances diminue; au commencement de la mise en application de la loi, le nombre des caisses d'assurance a été de 846 et à la fin de l'année 1932 leur nombre n'a fait que 775, pour les assurés non-agricoles. En 1931, on a payé pour les prestations plus de 820 millions Frs, dont 81,8% pour les prestations d'assurance-maladie.

Actuellement, on voit déjà une légère diminution de la morbidité en France, grâce à l'introduction des assurances sociales.

Les Conférences tenues au cercle des Actuaires tchécoslovaques seront publiées verbalement dans le prochain numéro de notre revue.

1ère Conférence à la Faculté des Sciences de Prague:

De la dispersion afférente à la somme de *n* erreurs dans le cas où chacune des erreurs composantes est régie par une loi simple.

Essai d'une représentation analytique. — Généralisation. Le problème en question se présente journellement aux expérimentateurs lorsqu'ils effectuent des sommations de nombres définis à une unité d'un certain ordre près; il revient à la recherche de la loi de dispersion de la somme $(x_1 + x_2 + \ldots + x_n)$, lorsque la dispersion caractérisant x_i est de la forme $\frac{1}{2a} dx_i$, le nombre auquel se rattache la dite erreur étant compris entre $(N - a)$ et $(N + a)$.

On peut se proposer tout d'abord d'évaluer la probabilité pour que la somme des n erreurs x_i soit comprise entre $(n - 2p)\,a$ et $(n - 2p + 2)\,a$, puis de rechercher ce que devient la fonction représentative lorsque le nombre des erreurs croît, de rattacher cette étude à celle d'une des lois de dispersion de Pearson, et enfin de généraliser les résultats trouvés.

1ère partie. Courbe de dispersion afférente à l'intervalle $(n - 2p)\,a$, $(n - 2p + 2)\,a$. Après avoir évalué la probabilité pour que l'erreur relative à $(x_1 + x_2)$ soit comprise entre α et $(\alpha + d\alpha)$, puis examiné le cas de trois, puis de quatre erreurs, nous avons constaté que les courbes de dispersion avec n pair ou impair pouvaient être représentees par

$$\frac{1}{(n-1)!\,(2a)^n}\,(na - \beta)^{n-1}$$

dans l'intervalle $[na, (n-2)\,a]$, et dans l'intervalle $[(n - 2p + 2)\,a, (n - 2p)\,a]$, par

$$S_1(\beta) = \frac{1}{(n-1)!\,(2a)^n}\left[(na - \beta)^{n-1} + \sum_{\gamma=1}^{p-1} (-1)^{\gamma}\, C_n^{\gamma} \left\{(n - 2\gamma)\,a - \beta\right\}^{n-1}\right]; \tag{1}$$

nous avons ensuite montré que cette loi est générale en posant

$$x_1 + x_2 + \ldots + x_n = \beta, \quad x_{n+1} + \beta = \gamma,$$

en faisant $a = 1$, et en remarquant que l'on passe de l'équation $S_1(\beta)$ de la courbe de dispersion relative à l'intervalle $[(n - 2p + 2), (n - 2p)]$, à la courbe de dispersion $S_0(\beta)$, concernant l'intervalle de gauche qui précède, soit $[(n - 2p), (n - 2p - 2)]$, en prenant

$$S_0(\beta) = S_1(\beta) + (-1)^p\, C_n^p\, (n - 2p - \beta)^{n-1},$$

puis enfin en calculant

$$\frac{d\gamma}{2}\left\{\int_{-1}^{\delta} S_1\,[n - 2p + \delta - x_{n+1}]\, dx_{n+1} + \int_{\delta}^{1} S_0\,[n - 2p + \delta - x_{n+1}]\, dx_{n+1}\right\}$$

avec $\gamma = n - 2p + \delta$, et δ variant de -1 à $+1$.

2e partie. Il nous faut déterminer — pour n croissant de plus en plus les valeurs des suites (1), (2) et (3), les suites (2) et (3) correspondant respectivement à l'intervalle (0, 2) pour n pair, et à l'intervalle $(-1, +1)$ pour n impair:

$$\frac{1}{(n-1)!\,2^n}\,[(n - \beta)^{n-1} - C_n^1\,(n - 2 - \beta)^{n-1} + \ldots$$
$$\ldots + (-1)^{\frac{1}{2}n-1}\, C_n^{\frac{1}{2}n-1}\,(2 - \beta)^{n-1}], \tag{2}$$

$$\frac{1}{(n-1)!\,2^n}\,[(n - \beta)^{n-1} - C_n^1\,(n - 2 - \beta)^{n-1} + \ldots$$
$$\ldots + (-1)^{\frac{1}{2}(n-1)}\, C_n^{\frac{1}{2}(n-1)}\,(1 - \beta)^{n-1}]. \tag{3}$$

On est conduit à la solution, sans faire appel à la fonction caractéristique, en constatant l'analogie de la fonction représentative pour le tronçon $(n - 2p)\,a$, $(n - 2p + 2)\,a$, avec celle qui apparaît dans l'étude d'un problème traité par Laplace, problème relatif à l'approximation des différences infiniment petites et finies très élevées des fonctions que met en lumière la relation (A)

$$\frac{(n+z)^{n-1} - n(n+z-2)^{n-1} + \frac{1}{2}n(n-1)(n+z-4)^{n-1} - \dots}{(n-1)!\,2^n} =$$
$$= \int_0^\infty \frac{\cos zx \left(\frac{\sin x}{x}\right)^n}{\pi} \, dx. \tag{A}$$

Faisant état de ce que les courbes de dispersion sont symétriques par rapport à l'axe des y, et tenant compte de ce que la valeur de l'intégrale I figurant dans le second membre de la relation (A) peut s'écrire:

$$I = \sqrt{\frac{3}{2n\pi}}\, e^{-{}^3/_2 z^2} \left[1 - \frac{3}{20n}(1 - 6r^2 + 3r)^4 + \dots\right] =$$
$$= \frac{(n + r\sqrt{n})^{n-1} - C_n^1 (n + r\sqrt{n} - 2)^{n-1} + C_n^2 (n + r\sqrt{n} - 4)^{n-1} - \dots}{(n-1)!\,2^n} \tag{4}$$

avec $z = r\sqrt{n}$, on trouve que les moments μ_2 et μ_4 peuvent s'exprimer ainsi qu'il suit:

$$\mu_2 = \frac{na^2}{3}, \text{ et } \mu_4 = \frac{n^2a^4}{3}\left(1 - \frac{2}{\varrho n}\right), \text{ d'ou } \frac{\mu_4}{\mu_2^2} = 3\left(1 - \frac{2}{\varrho n}\right).$$

Or si le phénomène suivait la loi de dispersion de Gauss, ou une loi de la forme $K\left(1 - \frac{x^2}{a^2}\right)^m$, on trouverait $\frac{\mu_4}{\mu_2^2} = 3$; on est ainsi conduit immédiatement à rechercher le lieu qui unit la courbe globale à la courbe de Gauss. Il suffit de revenir à l'équation (4) et de poser $r = \frac{\beta}{\sqrt{n}}$ et l'on voit alors que

$$\frac{1}{(n-1)!\,2^n}\left[(n-\beta)^{n-1} - C_n^1(n-2-\beta)^{n-1} + C_n^2(n-4-\beta)^{n-1} - \dots\right] =$$
$$= \sqrt{\frac{3}{2n\pi}}\, e^{-\frac{3\beta^2}{2n}} \left[1 - \frac{3}{20n}\left(1 - \frac{6\beta^2}{n} + \frac{3\beta^4}{n^2}\right) + \dots\right],$$

et l'on remarque en revenant à l'ancienne unité de longueur, que l'ensemble des tronçons partiels afférents à la figuration de la dispersion se rapproche de la courbe de Gauss $\frac{\sqrt{3}}{a\sqrt{2n\pi}}\, e^{-\frac{3x^2}{2na^2}}$.

Deuxième mode d'étude de la dispersion. L'emploi de la fonction caractéristique conduit en première approximation à la loi de probabilité

$$F_0(x) + \frac{3^2}{5!\,n} F^{(IV)}(x) - \frac{3^3}{7!\,n^2} F^{(VI)}(x) + \dots,$$

$F^{(IV)}$ et $F^{(VI)}$ n'étant autres que les dérivées d'ordre 4, 6, ... de la loi réduite normale $\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$, et en seconde approximation à la loi de probabilité

$$F_0(x) - \frac{1}{20n} F^{(IV)}(x) - \frac{1}{105\,n^2} F^{(VI)}(x) + \dots;$$

en se reportant à l'expression de $F^{(IV)}(x)$, on trouve qu'à la loi de probabilité simplifiée $F_0 - \frac{1}{20n} F^{(IV)}$, l'on fait correspondre le développement

$$F_0(x)\left[1 - \frac{3}{20n}\left(1 - 2x^2 + \frac{x^4}{3}\right)\right],$$

qui du fait du changement de variable $x = \beta\sqrt{\frac{3}{n}}$ devient

$$\sqrt{\frac{3}{2n\pi}}\, e^{-\frac{3\beta^2}{2n}}\left[1 - \frac{3}{20n}\left(1 - \frac{6\beta^2}{n} + \frac{3\beta^4}{n^2}\right)\right]$$

que l'on a fait apparaître ci-dessus.

3e mode d'étude. On peut traiter le problème en toute rigueur en recourant à l'emploi du coefficient restricteur de M. Galbrun.

Soit $y = f(x_1, x_2, \ldots, x_n)$ une fonction des erreurs indépendantes x_i $(i = 1, 2, \ldots, n)$, les x_i ne pouvant prendre que les valeurs de l'intervalle (a_i, b_i), et $\varphi(x_i)\,dx_i$ étant la probabilité pour que x_i soit comprise entre x_i et $(x_i + dx_i)$, avec $\int_{a_i}^{b_i} \varphi(x_i)\,dx_i = 1$.

Ceci étant, pour déterminer la probabilité P que y soit comprise entre y_1 et y_2, on introduit le multiplicateur

$$\psi(x) = \psi[H(y - u)] = \frac{2}{\sqrt{\pi}}\left[\int_0^\beta e^{-\alpha^2}\,d\alpha + \sum_{p=1}^{\infty} \frac{P_{2p}[H(y-u)]\,P_{2p-1}(\beta)\,e^{-\beta^2}}{2^{2p}\,(2p)!}\right],$$

où y et H dans le cas présent sont définies par les expressions

$$y = x_1 + x_2 + \ldots + x_n,\quad H = \frac{1}{\sqrt{2(M_2 - M_1^2)}},$$

où u est une constante quelconque, et M_1 et M_2 sont les valeurs probables de y et y^2, P_{2p-1} et P_{2p} les polynomes d'Hermite figurant dans les dérivées d'ordre $(2p - 1)$ et $(2p)$ de la fonction e^{-x^2}, on a ainsi

$$P = \int \ldots \int \psi[H(y - u)]\,\varphi_1(x_1)\,\varphi_2(x_2)\ldots\varphi_n(x_n)\,dx_n\,dx_1\,dx_2\ldots dx_n.$$

Posant $\beta = H\lambda$, $\alpha = H(\Theta - u)$, et choisissant les constantes u et β de façon que

$$u - \beta\sqrt{2M_2} = (n - 2s)\,a,\quad u + \beta\sqrt{2M_2} = (n - 2s + 2)\,a,$$

il s'ensuit que

$$P = \frac{\sqrt{3}}{a\sqrt{2n\pi}} \int_{(n-2s)a}^{(n-2s+2)a} e^{-H^2[\Theta-(n-2s+1)a]^2}\,d\Theta +$$

$$+ \frac{2}{\sqrt{\pi}}\sum_{p=1}^{\infty}\frac{P_{2p-1}(\beta)\,e^{-\beta^2}}{2^{2p}(2p)!}\,\overline{P_{2p}[H(y-u)]},$$

avec

$$\overline{P_{2p}[H(y - u)]} = \int\ldots\int P_{2p}[H(y - u)]\,\varphi_1\varphi_2\ldots\varphi_n\,dx_1\,dx_2\ldots dx_n.$$

4e mode d'étude. On peut rapprocher facilement la courbe globale de dispersion de la forme de dispersion de Pearson $Y = y'_0\left(1 - \frac{x^2}{A^2}\right)^m$

et avec telle approximation que l'on voudra du développement de Romanowsky

$$Y = y'_0 \left(1 - \frac{x^2}{A^2}\right)^m \left[1 + \sum_{k=s}^{\infty} \frac{\lambda_0}{\lambda_k} \frac{S_k}{S_0}\right].$$

Généralisation. Considérons une série d'erreurs $(X_1, X_2, \ldots, X_n)$, $(Y_1, Y_2, \ldots, Y_{n'})$, $(Z_1, Z_2, \ldots, Z_{n''})$, ... suivant respectivement les lois simples $\frac{1}{2a}\,\mathrm{d}X_i$, $\frac{1}{2b}\,\mathrm{d}Y_j$, $\frac{1}{2c}\,\mathrm{d}Z_l$, ... et supposons que les X_i, Y_j, Z_l, ... soient indépendantes.

On peut se proposer de rechercher la loi représentative de la probabilité pour que Σx_i soit comprise entre $(n - 2p)\,a$ et $(n - 2p + 2)\,a$, que Σy_j soit comprise entre $(n' - 2p')\,b$ et $(n' - 2p' + 2)\,b$, ...; du fait de l'hypothèse faite sur les x_i, y_j, z_l, ... on remarque que la probabilité cherchée s'exprime au moyen de produits de polynomes de degrés respectifs $(n - 1)$, $(n' - 1)$, $(n'' - 1)$, ..., et qu'en première approximation. l'on peut assigner à la probabilité cherchée la valeur

$$\frac{h}{\sqrt{\pi}} \int_{(n-2p)a}^{(n-2p+2)a} e^{-h^2 x^2}\,\mathrm{d}x \,.\, \int_{(n'-2p')b}^{(n'-2p'+2)b} \frac{h'}{\sqrt{\pi}}\, e^{-h'^2 y^2}\,\mathrm{d}Y \ldots$$

Une telle étude pourrait être complétée en considérant la probabilité répondant aux inégalités

$$(n - 2p)\,a^9 \leqq x_1^9 + x_2^9 + \ldots + x_n^9 \leqq (n - 2p + 2)\,a^9.$$

2ème conférence:

Les principes de la statistique mathématique.

Demeurée longtemps un moyen d'informations pour les Gouvernements, la statistique avec son domaine s'étendre à mesure que les sciences progressaient; c'est ainsi qu'elle est devenue d'un concours précieux pour tous les chercheurs adonnés aussi bien à l'étude des questions de sociologie, de démographie et d'économie politique qu'à celle des sciences naturelles et de la médecine. A une date plus rapprochée, la statistique a pris place dans la physique et dans le domaine de la physiologie.

L'étude des problèmes posés par la statistique n'a appelé l'attention des mathématiciens qu'à la suite des travaux de Galton; elle correspond à deux conceptions bien distinctes, l'une qui fait appel à l'introduction du continu et l'autre — beaucoup plus récente — à l'emploi des équations aux différences finies.

C'est cette évolution que nous nous sommes proposé d'exposer au cours de notre conférence après avoir rappelé ce que l'on entend par moyenne, médiane, quartiles, médiale et quartals, et signalé que l'examen de certaines questions se rattachant au prolongement des ensembles statistiques a pu être conduit à bien grâce à la théorie mathématique des ensembles et a eu pour heureuse conséquence une révision des concepts du calcul des probabilités.

Avant d'aborder les travaux des maîtres de l'école de statistique moderne, nous avons jugé utile de revenir sur la fonction caractéristique dont l'idée est due à Cauchy.

Polygone binomial. — Courbes de Pearson. — Séries de Bruns-Charlier. L'étude d'une série statistique se rapportant à un événement qui dépend de deux probabilités complémentaires peut être effectuée grâce au théorème de Bernoulli associée à la théorie des épreuves répétées.

Dans le cas où l'on remet dans l'urne la boule tirée après chaque tirage, on peut dire que l'équation différentielle de la courbe tangente à chaque côté du polygone binomial en son milieu est du type

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\gamma x y}{x + a}; \tag{1}$$

si au contraire, l'on ne remet pas la boule tirée dans l'urne, Pearson en employant un processus analytique à celui qui l'avait conduit à l'équation (1), est amené à résoudre l'équation

$$\frac{y'}{y} = \frac{a + x}{b_0 + b_1 x + b_2 x^2}. \tag{2}$$

Après avoir signalé les différents types de dispersion déduits de (2), et fourni une explication des suites à répartition normale et à répartition asymétrique, en mettant en lumière les idées de Kapetyn, Van Uren, Wicksell, nous avons abordé les belles recherches de Bruns-Charlier qui se rattachent à la représentation d'une série statistique par un développement de la forme

$$\mathrm{F}(z) = \alpha_0 \varphi(z) + \alpha_1 \varphi'(z) + \ldots + \alpha_n \varphi^{(n)}(z) + \ldots, \tag{3}$$

où φ représente la fonction $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, et $\varphi^{(n)}$ sa dérivée d'ordre n.

Au développement (3) de Bruns, Charlier substitue les suivants:

$$\mathrm{f}(x) = \psi(x) + c_1 \psi_1 + \ldots + c_n \psi_n + \ldots, \tag{4}$$

avec $\quad \psi(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad$ et $\quad \psi_r(x) = -\psi_{r-1}(x) + \psi_{r-1}(x-1),$

et

$$\mathrm{F}(x) = \mathrm{g}(x) + c_1 \Delta \mathrm{g} + c_2 \Delta^2 \mathrm{g} + \ldots + c_k \Delta^k \mathrm{g}, \tag{5}$$

avec

$$\mathrm{g}(x) = \binom{k}{x} p^x (1-p)^{k-x} \quad \text{et} \quad \Delta \mathrm{g} = \mathrm{g}(x-1) - \mathrm{g}(x)$$
$$\Delta^2 \mathrm{g} = \Delta_1 \mathrm{g}.$$

Alors que Pearson et ses élèves interprêtent une certaine équation différentielle, M. Guldberg revient à l'étude des propriétés des fonctions de fréquence, en procédant à la recherche d'invariants caractéristiques de ces fonctions, en utilisant les moments par rapport à l'origine et par rapport à la moyenne arithmétique

$$\sigma_r = \frac{\Sigma x_i^r}{n}, \quad m_r = \frac{\sum_{1}^{n} (x_i - m_1)^r}{n}$$

les moments factoriels $\sigma_{(r)}$ introduits par MM. Steffensen et Sheppard

$$\sigma_{(r)} = \Sigma X_i (X_i - 1) \ldots (X_i - r + 1),$$

et enfin les moments incomplets de M. Frisch

$$_t\mu_r = \sum_{x=t}^{\infty} (x - \lambda)^r \mathrm{f}(x).$$

Si l'on désigne par f(x) la fonction de fréquence, M. Guldberg cherche la relation de liaison entre f(x) et f($x + 1$), dans le cas où f est la fonction binomiale, la fonction de fréquence de Poisson et enfin la fonction de fréquence dite de Pascal; il aboutit à la relation

$$\frac{f(x+1)}{f(x)}(x+1)+\frac{\mu_1-\mu_2}{\mu_2}x=\frac{\mu_1^2}{\mu_2}, \tag{6}$$

et remarque que l'équation (6) caractérise la fonction binomiale lorsque $\mu_1 > \mu_2$, la fonction de Poisson avec $\mu_1 = \mu_2$, et enfin la fonction de Pascal, lorsque μ_1 était inférieure à μ_2.

M. Guldberg a étendu cette méthode au cas d'une distribution $f(x, y)$, et a ainsi fait apparaître la relation de récurrence

$$f(x+1, y+1)=\frac{pq}{(1-p-q)^2}\,\frac{(k-x-y)(k-x-y-1)}{(x+1)(y+1)}f(x, y). \tag{7}$$

Covariation. Corrélation. Après avoir mis en lumière l'indice de dépendance ou de covariation de Fechner, et montré comment l'on arrive à l'indice caractéristique de Galton, nous avons donné le carré moyen de contingence de Pearson, puis l'indice de Steffensen, le coefficient de corrélation, et le rapport de corrélation de Pearson, et fait apparaître le rôle des plus importants joué par le mathématicien Zochuprow dans l'étude de la corrélation.

Nous avons jugé utile de rappeler les travaux récents de l'institut international de statistique et tout spécialement la judicieuse remarque de M. Frechet afférente à l'expression du coefficient de corrélation ou plutôt du coefficient de linéarité.

Si r et η désignent respectivement le coefficient et le rapport de corrélation, l'on a

$$r=\frac{\sum_{ij} n_{ij}(x_i-a)(y_j-b)}{\sqrt{\sum_i N_i(x_i-a)^2}\sqrt{\sum_j M_j(y_j-b)^2}},$$

où (x_i, y_j) sont les valeurs des variables x et y, a et b les valeurs moyennes de x et y, et

$$N_i=\sum_j n_{ij}, \quad M_j=\sum_i n_{ij}.$$

Si l'on représente par b_i la moyenne des y_j correspondant à une valeur x_i des x, on remarque que

$$r=\frac{\sum_i (x_i-a)N_i(b_i-b)}{\sqrt{\sum_i N_i(x_i-a)^2}\sqrt{\sum_i N_i(b_i-b)^2}}\times\frac{\sqrt{\sum_i N_i(b_i-b)^2}}{\sqrt{\sum_j M_j(y_j-b)^2}},$$

et par suite $r = \varrho\eta$; ce coefficient ϱ — d'aprés la nature même de sa composition — reste le même quelle que soit la dispersion de y autour de sa valeur moyenne pour une valeur donnée de x, et est insensible au relâchement ou à la rigueur de la dépendance entre x et y.

Quant au coefficient η, on voit qu'il n'est pas affecté par la déformation de la courbe des moyennes du fait de la modification des x_i. En définitive, l'on peut dire que l'expression du facteur ϱ, dans l'expression de r, ne peut que fausser la mesure de la dépendance de (x, y).