Jan Kašpar

La généralisation des courbes de fréquence de Pearson par Romanovsky

Aktuárské vědy, Vol. 4 (1933), No. 1, 10–18

Persistent URL: http://dml.cz/dmlcz/144591

Mitgliedern (wenn dieses Ausscheiden nicht durch Heilung geschieht) muss das Sterbegeld „1" bezahlt werden. Wir definieren daher $\beta^1_{x+t} \rightarrow \rightarrow \mu^v_{x+t}$ als Intensität der Sterblichkeit eines Unfallbeschädigten. Endlich ist noch $\beta^2_{x+t} = 0$, denn einem Unfallbeschädigten wird nie das Deckungskapital zurückgezahlt. — Die pseudo-unabhängige Ordnung $l^{(u)}_{x+t}$ ist dann dargestellt durch:

$$l^{(u)}_{x+t} = l_x \cdot e^{-\int\limits_0 w^a_{x+s} \cdot \mu^u_{x+s} ds.}$$

Sind die Gesamtheiten der Nicht-Unfallbeschädigten und der Unfallbeschädigten aus dem System (1) berechnet, so sind auch die Wahrscheinlichkeiten w^a_{x+t} nicht unfallbeschädigt und w^b_{x+t} unfallbeschädigt zu sein, gegeben. Damit ist aber auch die Möglichkeit geboten, aus der letzten Gleichung die Ordnung $l^{(u)}_{x+t}$ zu errechnen. Für das Deckungskapital folgt dann:

$${}_tV_x = \frac{1}{v^t \cdot l^{(u)}_{x+t}} \int\limits_0^t v^\lambda \cdot l^{(u)}_{x+\lambda} (w^a_{x+\lambda} \cdot P_\Delta - w^b_{x+\lambda} \cdot T - w^b_{x+\lambda} \cdot \mu^v_{x+\lambda}) \, d\lambda.$$

Wir weisen noch einmal darauf hin, dass die speziell zur Lösung der Integralgleichung eingeführten pseudo-unabhängigen Ordnungen dem Charakter nach nicht mit den Ordnungen der „Nicht-Unfallbeschädigten" und der „Unfallbeschädigten" übereinstimmen.

Schluss.

Die beiden Beispiele mögen genügen, die Anwendbarkeit unserer allgemeinen Formel zu zeigen. Durch die ganz allgemeinen Ansätze konnten die verwickelten Vorgänge in der sozialen Versicherung ein wenig geklärt und in eine engere Bindung mit den klaren Ergebnissen der Lebensversicherung gebracht werden. Eine Behandlung des Problems der Deckungskapitalberechnung unter allgemeinsten Voraussetzungen dürfte sicher dazu beitragen, der sozialen Versicherung das feste Gefüge zu verleihen, wie es die Lebensversicherung bereits aufweist.

Juni 1932.

La généralisation des courbes de fréquence de Pearson par Romanovsky.

Par le Dr. *J. Kašpar.*

Romanovsky a montré, dans un de ses multiples travaux, publiés dans la revue „Biométrie", année XVI, comment on peut généraliser les courbes de fréquence du premier, deuxième et troisième type, à savoir de la même manière, comme a été généralisée la courbe normale (septième type des courbes de Pearson) par Thiele et Charlier. Il obtient une équation de la courbe de fréquence en forme d'une série infinie

$$y = A_0 u_0 + A_1 u_1 + \ldots + A_k u_k + \ldots, \tag{1}$$

où A_k sont des constantes, u_k des fonctions indépendamment changeables de x, ayant seulement une signification et continues dans un certain intervalle. Si nous nous bornons dans (1) seulement au premier membre, alors la série a la qualité que l'on obtient une équation d'une courbe d'un des premiers trois types des courbes de Pearson, donc selon la forme des fonctions u_k. Si nous prenons dans (1) s membres, nous arrivons à l'équation

$$y = A_0 u_0 + A_1 u_1 + \ldots + A_s u_s,$$

où Romanovsky suppose, qu'elle exprime le collectif donné autant mieux que le nombre s est plus grand.

Traitons d'abord le premier type des courbes de Pearson qui a la forme de l'équation suivante

$$y = y_0 \left(1 + \frac{x}{a}\right)^{\alpha} \left(1 - \frac{x}{b}\right)^{\beta}, \tag{2}$$

où y_0, a, b, α, β sont des constantes qui se laissent exprimer à l'aide des premiers quatre moments.

Dans ce cas Romanovsky a mis

$$\left.\begin{aligned} u_0 &= (a + x)^{\alpha} (b - x)^{\beta} \\ u_k &= D^k (a + x)^{\alpha + k} (b - x)^{\beta + k} \end{aligned}\right\} \tag{3}$$

où a, b, α, β sont des constantes données; nous supposons chez les constantes α et β que $\alpha > -1$, $\beta > -1$. On peut se facilement persuader que la fonction u_k se laisse également écrire dans la forme $u_k = u_0 \Phi_k$, où Φ_k est un polynome du degré k de l'inconnue x, indépendamment changeable, ayant la forme

$$\Phi_k = (\alpha + 1)^{(-k)} (b - x)^k - \binom{k}{1} (\alpha + 2)^{(-\overline{k-1})} (\beta + k) (b - x)^{k-1} \times$$
$$\times (a + x) + \binom{k}{2} (\alpha + 3)^{(-\overline{k-2})} (\beta + k - 1)^{(-2)} (b - x)^{k-2} (a + x)^2 -$$
$$\ldots + (-1)^k (\beta + 1)^{(-k)} (a + x)^k,$$

où nous avons employé la désignation symbolique

$$x^{(-n)} = x (x + 1) (x + 2) \ldots (x + n - 1).$$

Les polynomes que nous obtenons ainsi, sont des polynomes de Jacobi, légèrement changés. Ils sont orthogonaux, i. e. ils ont les deux qualités suivantes

$$\left.\begin{aligned}\int_{-a}^{b} u_0 \Phi_h \Phi_k \, dx = 0, \qquad h \neq k \\ \int_{-a}^{b} u_0 \Phi_k^2 \, dx = I_k = k!\,(a+b)^{\alpha+\beta+2k+1}\,(\alpha+\beta+k+1)^{(-k)} \times \\ \times \frac{\Gamma(\alpha+k+1)\ \Gamma(\beta+k+1)}{\Gamma(\alpha+\beta+2k+2)}\end{aligned}\right\} \tag{4}$$

Supposons que la fonction donnée qui n'a qu'une signification et qui est continue dans l'intervalle $(-a, b)$ est développable dans une série des polynomes Φ_k que nous venons de définir

$$y = u_0\,(A_0\Phi_0 + A_1\Phi_1 + A_2\Phi_2 + \ldots). \tag{5}$$

Nous trouvons les coefficients de cette série, par ex. A_k, en multipliant les deux côtés de cette équation par le polynome Φ_k et ensuite nous intégrons dans les limites $(-a, b)$. Si nous supposons que la série est régulièrement convergente — cette supposition comme indique Romanovsky, sans une recherche détaillée, est remplie pour les fonctions d'une signification et continues dans l'intervalle $(-a, b)$ — on peut alors procéder l'intégration mentionnée à la côté droite succesivement chez chaque membre. Par raisons des qualités orthogonaux des polynoms, nous obtenons Φ_k

$$A_k = \frac{S_k}{I_k}, \tag{6}$$

où nous avons introduit la désignation

$$S_k = \int_{-a}^{b} y\Phi_k \, dx. \tag{7}$$

Jusqu'alors les constantes a, b, α, β ont été des chiffres librement choisis, seulement α, β ont subi à la condition $\alpha > -1$, $\beta > -1$. Nous prenons maintenant pour les constantes a, b, α, β les valeurs, qu'elles reçoivent comme constantes se trouvant dans l'équation du premier type des courbes de Pearson. Ce choix signifie que la courbe de Pearson (2) et la courbe exprimée par la série (5) ont les premiers quatre moments les mêmes, égale aux moments empiriques. On peut prouver que dans ces cas $S_1 = S_2 = S_3 = S_4 = 0$ ou bien $A_1 = A_2 = A_3 = A_4 = 0$. La série (5) accepte alors la forme

$$y = u_0\,(A_0 + A_5\Phi_5 + A_6\Phi_6 + \ldots),$$

où le calcul des coefficients éxige la fixation du cinquième moment et des moments supérieurs. Si l'on ne considère que le calcul, alors déjà le calcul du premier membre correctif $A_5 u_5$ est très pénible et bien souvent l'éxactitude plus grande, que nous pourrions obténir en ajoutant un nombre augmenté des membres correctifs, pourrait être perdue par la circonstance que les moments supérieurs sont chargés de fautes. Par conséquent Romanovsky a proposé une manière par laquelle on peut employer la série (5). Pour les constantes a, b, α, β nous recevons de coutume des chiffres décimaux; il faut les compter exactement à quelques fractions décimales, si l'on veut arriver à des résultats satisfaisants. Romanovsky croit que la série (5) peut être employée de sorte que nous ne prenons pas pour les constantes a, b, α, β des valeurs exactes, mais arrondies au prochain chiffre entier. En cas, où l'arrondissement au prochain chiffre entier, causerait un important changement relatif de la constante, nous arrondissons à une fraction décimale. En faisant ce choix des valeurs de constantes a, b, α, β les coefficients A_1, A_2, A_3, A_4 n'égalent plus zéro. On pourrait atteindre une simplification pour le calcul en mettant les polynoms $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ dans un tableau pour les valeurs entières des constantes a, b, α, β.

Parce que le deuxième type des courbes de Pearson

$$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m \tag{8}$$

n'est qu'un cas spécial du type premier, alors est en vigueur pour le deuxième type le même développement, que pour le type premier, nous mettons toujours $a = b$, $\alpha = \beta$.

Chez le troisième type des courbes de Pearson

$$y = y_0 \left(1 + \frac{x}{a}\right)^m e^{-\gamma x} \tag{9}$$

Romanovsky a choisi

$$\left.\begin{aligned} \mathrm{u}_0 &= (a + x)^m e^{-\gamma x} \\ \mathrm{u}_k &= D^k \left[(a + x)^{m+k} e^{-\gamma x}\right] \end{aligned}\right\} \tag{10}$$

On peut écrire à nouveau $\mathrm{u}_k = \mathrm{u}_0 \Phi_k$, où Φ_k est un polynom du degré k de x. En employant la formule de Leibnitz, nous recevons pour Φ_k dans (10) l'expression suivante

$$\Phi_k = (m + 1)^{(k)} - \binom{k}{1} (m + 2)^{(-\overline{k-1})} \gamma (a + x) +$$

$$\binom{k}{2} (m + 3)^{(-\overline{k-2})} \gamma (a + x)^2 - \ldots + (-1)^k \gamma^k (a + x)^k.$$

Les polynoms que nous obtenons ainsi sont une généralisation des polynoms de Laguerre. Ils sont de nouveau orthogonaux

$$\int_{-a}^{\infty} \mathrm{u}_0 \Phi_h \Phi_k \, \mathrm{d}x = 0, \qquad h \neq k,$$

$$\int_{-a}^{\infty} u_0 \Phi_k^2 \, dx = \frac{k!\, m^{(-\overline{k+1})}}{\gamma^{m+1}} e^{\gamma a} \Gamma(m),$$

en conséquence on peut facilement trouver les coefficients de la série

$$y = u_0 (A_0 \Phi_0 + A_1 \Phi_1 + A_2 \Phi_2 + \ldots), \tag{11}$$

dans laquelle on peut subsumer la fonction d'une signification et continue dans l'intervalle $(-a, \infty)$. Pour trouver le coefficient A_k, il suffit de multiplier les deux côtés par le polynom Φ_k et d'intégrer ensuite dans l'intervalle $(-a, \infty)$. En supposant que la série est régulièrement convergente, nous arrivons à

$$A_k = \frac{\gamma^{a+1} S_k}{k!\, \epsilon^{\gamma a}\, m^{(-\overline{k+1})}\, \Gamma(m)} \cdot$$

Si nous choisissons pour les constantes a, m, γ ces valeurs que nous recevons pour eux comme constantes du troisième type, nous trouvons alors que $A_1 = A_2 = A_3 = 0$. Il y a ici seulement trois coefficients qui sont égale à zéro, parce que le calcul des constantes a, m, γ exige la fixation des premiers trois moments et que le choix mentionné des constantes a, m, γ signifie que nous identifions les premiers trois moments de la courbe de Pearson avec les premiers trois moments de la série (11).

Premier exemple.

Pour appliquer le premier type des courbes de Pearson, on a pris le collectif formé par les hommes divorcés en Slovaquie, en tenant compte de leur âge. Ce collectif est pris du manuel „Recensement de la population en Tchécoslovaquie à la date du 15 février 1921", vol. I, page 202, où il se trouve dans le tableau XVIII „La population actuelle d'après l'âge, le sexe et la nuptualité". Les fréquences empiriques sont indiquées dans le tableau I. Des données empiriques nous trouvons que la moyenne arithmétique appartient à l'âge 43,9469 ans,

$$\mu_2 = 6{,}32491, \quad \mu_3 = 8{,}46197, \quad \mu_4 = 107{,}9899,$$
$$\beta_1 = 0{,}28299, \quad \beta_2 = 2{,}69944, \quad k = -0{,}15930,$$

il faut alors appliquer le premier type des courbes de Pearson. Pour ses constantes nous trouvons les valeurs suivantes

$$y_0 = 189{,}2479, \quad a = 3{,}101502, \quad b = 11{,}08389,$$
$$\alpha = 0{,}84412, \quad \beta = 3{,}01665,$$

le maximum de la courbe de Pearson ou bien la valeur modale du signe appartient à l'âge 37,13694 ans. La courbe de Pearson a donc l'équation suivante

$$y = 189{,}2479 \left(1 + \frac{x}{3{,}10502}\right)^{0{,}84412} \left(1 - \frac{x}{11{,}08389}\right)^{3{,}01665},$$

où le commencement du système est dans le modus. Il revient de cette équation que la courbe passe de l'âge 21,62944 ans à l'âge de 92,55639 ans. Les fréquences trouvées d'après la courbe de Pearson sont indiquées dans le tableau I, troisième colonne. Les fréquences de la première et dernière classe ont été trouvées en employant la formule de Simpson, les fréquences des autres classes à l'aide de la formule

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x \, dx = y_0 - \frac{1}{24} (\Delta_{y-1} - \Delta_{y_0}),$$

que nous obtenons, si nous remplaçons la courbe par une parabole quadratique.

Tableau I.

Age évolu	Fréquences réelles	Fréquences trouvées d'après la courbe de Pearson	Fréquences trouvées d'après Romanovsky en employant les premiers trois membres	quatre membres	cinq membres
1	2	3	4	5	6
15—19	1	—	—	—	—
20—24	42	36	28	30	32
25—29	110	133	139	144	145
30—34	195	177	193	190	184
35—39	189	188	199	191	187
40—44	177	179	181	175	177
45—49	165	156	152	150	155
50—54	120	127	119	123	128
55—59	91	97	89	95	96
60—64	65	68	63	67	65
65—69	48	43	41	43	40
70—74	26	24	24	23	21
75—79	12	11	11	10	10
80—84	3	4	4	3	3
85—89	—	1	1	—	1

Nous prenons maintenant, pour exprimer le collectif donné, le développement (5) de Romanovsky, où nous choississions pour les constantes a, b, α, β les plus proches chiffres entiers des constantes respectives de la courbe de Pearson. Nous mettons alors $a = 3$, $b = 11$, $\alpha = 1$, $\beta = 3$. Si nous mettons ces valeurs dans les premiers quatre polynomes, que nous classons d'abord d'après les potences de x, nous arrivons à

$$\Phi_1 = 10 - 6x,$$
$$\Phi_2 = -84 - 252x + 56x^2,$$
$$\Phi_3 = -14064 - 3312x + 5616x^2 - 720x^3,$$
$$\Phi_4 = -645520 + 572400x + 263520x^2 - 134640x^3 + 11880x^4.$$

Le calcul des coefficients de la série (5) exige la fixation des moments, en tenant compte du modus. Pour les premiers quatre moments, nous trouvons

$$m_1 = 1{,}36200, \quad m_2 = 8{,}17999, \quad m_3 = 36{,}83213, \quad m_4 = 227{,}9298,$$

et ensuite pour les premiers cinq coefficients

$$A_0 = 4{,}6259 \,.\, 10^{-2}, \quad A_1 = 4{,}7526 \,.\, 10^{-4}, \quad A_2 = 8{,}3595 \,.\, 10^{-6},$$
$$A_3 = 1{,}3884 \,.\, 10^{-7}, \quad A_4 = 2{,}1243 \,.\, 10^{-9}.$$

Les valeurs des premiers cinq totaux partiels de la série (5) pour les moyennes de toutes les classes sont enregistrées dans le tableau II, où il y a également dans la colonne (3) indiqué les coefficients de la courbe de Pearson. Les fréquences relatives aux classes spéciales ont été calculées de la même manière que chez la courbe de Pearson; elles sont indiquées dans le tableau I, dans les colonnes (4), (5) et (6) pour le cas, si nous nous bornons dans le développement de Romanovsky aux premiers trois, quatre et cinq membres. Le point initial de toutes les courbes correspond à l'âge de 22,13694 ans et le point final à l'âge de 92,13694 ans.

Tableau II.

Age	x	Coefficients de la courbe de Pearson	A_0u_0	$A_0u_0 + A_1u_1$	$A_0u_0 + A_1u_1 + A_2u_2$	$A_0u_0 + A_1u_1 + A_2u_2 + A_3u_3$	$A_0u_0 + A_1u_1 + A_2u_2 + A_3u_3 + A_4u_4$
1	2	3	4	5	6	7	8
22,5	— 2,92738	33,75	9,08	11,65	13,51	15,58	17,33
27,5	— 1,92738	135,21	107,20	130,94	142,75	147,65	149,41
32,5	— 0,92738	178,66	162,68	188,69	194,51	191,78	184,96
37,13691	0,00000	189,24	184,75	203,73	200,93	193,13	188,31
37,5	0,07262	189,19	185,46	203,68	200,28	192,33	187,20
42,5	1,07262	178,86	184,32	191,07	181,42	174,76	176,98
47,5	2,07262	156,12	166,95	162,77	151,74	150,13	155,45
52,5	3,07262	127,06	139,95	127,82	119,49	122,82	127,67
57,5	4,07262	96,48	108,76	92,63	89,16	94,70	95,85
62,5	5,07262	67,72	77,77	61,44	62,55	67,14	64,75
67,5	6,07262	43,12	50,21	36,56	40,65	42,41	39,07
72,5	7,07262	24,04	28,22	18,82	23,59	22,55	20,83
77,5	8,07262	10,95	12,85	7,78	11,33	9,26	9,71
82,5	9,07262	3,49	4,00	2,17	3,79	2,36	2,50
87,5	10,07262	0,47	0,48	0,23	0,46	0,15	0,51

En examinant les chiffres du tableau I nous trouvons que l'expression du collectif donné par la courbe de Pearson ou bien, en employant les premiers membres de la formule de Romanovsky est très satisfaisante, à l'exception des premières deux classes. C'est également approuvé par le critère χ. Si nous mettons les premières deux classes dans une classe et si nous remplaçons également les dernières trois classes par une classe, alors nous trouvons pour la courbe de Pearson $\chi^2 = 5{,}58$ et pour l'expression du collectif donné par les premiers membres de la formule de Romanovsky, on reçoit successivement $\chi^2 = 4{,}36$, $\chi^2 = 5{,}77$, $\chi^2 = 8{,}17$. Nous recevons alors pour ces valeurs les probabilités suivantes

$$P = 0{,}85, \quad P = 0{,}93, \quad P = 0{,}83, \quad P = 0{,}64.$$

Il est évident que lá formule de Romanovsky en employant les premiers trois membres donne les meilleurs résultats, parce que la probabilité correspondante y est la plus grande. Si, en applicant le critère χ, nous remplaçons seulement les dernières trois classes par une, nous recevons

$$\chi^2 = 9{,}42, \quad \chi^2 = 17{,}28, \quad \chi^2 = 16{,}28, \quad \chi^2 = 17{,}14$$

et en employant les chiffres des tableaux, nous recevons

$$P = 0{,}58, \quad P = 0{,}10, \quad P = 0{,}13, \quad P = 0{,}12.$$

La courbe de Pearson y donne alors une meilleure expression que les courbes de Romanovsky, ce qui est causé par les déviations remarquables dans les premières deux classes, qui sont à nouveau causées par le fait, que le point initial des courbes de Romanovsky appartient à un âge plus élevé que chez la courbe de Pearson.

Deuxième exemple.

Comme exemple pour le deuxième type des courbes de Pearson, il a été calculé le collectif des enfants, nés vifs, en Bohême dans la période de 1914—1920. Ce collectif a été pris du manuel „Recensement de la population en Tchécoslovaquie du 15 février 1921“, Ier volume, page 95. Les fréquences effectives sont indiquées dans le tableau II. Des donnés empiriques il revient:

la moyenne arithmétique corresponde à l'année 1917,396

$$\mu_2 = 5{,}02357, \quad \mu_3 = 0{,}87403, \quad \mu_4 = 37{,}64424,$$
$$\beta_1 = 0{,}006024, \quad \beta_2 = 1{,}49167, \quad k = -0{,}001629.$$

Parce que les valeurs des constantes k, β_1 sont très petites et $\beta_2 \neq 3$, on voit que le collectif donné se laisse exprimer par la courbe de Pearson du deuxième type. Nous trouvons pour les constantes de la courbe (8)

$$a = 3{,}152173, \quad m = -0{,}51103, \quad y_0 = 82{,}2814.$$

Puisque la constante m a une valeur négative, nous recevons une courbe symmétrique de la forme de la lettre U. Le commencement du système symmétrique est dans la moyenne arithmétique, qui coïncide avec l'antimodus. La courbe s'évolue de l'année 1914,244 jusqu'à l'année 1920,548. Les fréquences des classes respectives ont été trouvé à l'aide de la formule de Simpson, elles sont enregistrées dans la colonne (3) du tableau III.

En applicant la formule de Romanovsky nous mettons $m = -0{,}5$, $a = 3{,}1$. Les premiers trois polynomes ont alors la forme

$$\Phi_1 = -x, \quad \Phi_2 = 6x^2 - 28{,}83, \quad \Phi_3 = -60x^3 + 414{,}45x.$$

Le calcul des coefficients, qui y est bien simplifié, parce que $\mu_1 = 0$, nous donne les valeurs suivantes

$$A_0 = 257{,}413, \quad A_1 = 0, \quad A_2 = 0{,}81232, \quad A_3 = -0{,}13520.$$

Le point initial des courbes que nous donne la formule de Romanovsky en applicant les premiers trois ou bien quatre membres, appartient à l'année 1914,296 et le point final à l'année 1920,496. Les fréquences relatives aux diverses classes ont été trouvées de la même manière que chez la courbe de Pearson; elles sont enregistrées dans le tableau III. En applicant le critère χ nous recevons successivement pour la courbe de Pearson et pour celles de Romanovsky

$$\chi^2 = 7{,}79, \quad \chi^2 = 7{,}81, \quad \chi^2 = 7{,}73,$$

et dans les tableaux nous trouvons

$$P = 0{,}255, \quad P = 0{,}254, \quad P = 0{,}261.$$

L'expression du collectif donné est donc dans tous ces cas de la même manière satisfaisante. La deuxième courbe de Romanovsky, pour laquelle nous recevons la probabilité P un peu plus élevée, n'est plus une courbe symmétrique, mais elle caractérise beaucoup mieux le cours du collectif.

Tableau III.

Année	Nombre d'enfants nés vifs, en 1000	Fréquences trouvées d'après la courbe		
		de Pearson	de Romanovsky en applicant les premiers trois membres	de Romanovsky en applicant les premiers quatre membres
1	2	3	4	5
1914	172	183	191,6	176,9
1915	126	105,7	105,3	126,4
1916	87,7	85,8	80,8	96,1
1917	77,4	82,7	75,7	74,1
1918	73,6	88,6	83,6	65,7
1919	118	114,6	116,7	98,1
1920	154	148,3	155	171,4