Note on the Frequency Curves of Basic Pure Premiums.

BY

ARNE FISHER.

Introductory Remarks.

The question of a proper method of computing basic pure premiums in workmen's compensation insurance is of prime importance to all casualty statisticians and actuaries and can in no way be said to be finally solved. In the last number of this publication I gave a brief outline of a method for collecting data to be used in the calculation of pure premiums. This method, which was based on the theory of dispersion or stability of statistical series, was an attempt to determine the occupational hazard of the individual employee rather than the hazard according to industries. Mr. Mowbray had previously attacked the problem from the standpoint of frequency curves. He made the following statement: "I think we may properly consider the pure premium as made up of several elements, each having an independent probability of its own and each of which may therefore be properly considered, for purposes of discussion, alone and apart from the others and the results appropriately combined." This amounts practically to the system, originally developed by the great Laplace, for the deduction of the equation of a frequency curve. Laplace considers namely the frequency curve, $F(x)$, to have originated as the sum of a number of subsidiary frequency curves of the form $f_k(x)$ $(k=1, 2, 3, \cdots)$. It is, however, only in the statement of the origin of the final curve that Mr. Mowbray follows Laplace. Mr. Mowbray throughout the remaining part of his paper falls back upon the Gaussian Normal Curve, which is a particular case of the general Laplacean frequency curve. In a later discussion of Mr. Mowbray's article I pointed out the fact that in most cases we were not justified in regarding the frequency distribution as a normal one as we actually were dealing with a decidedly skew curve of the Charlier B Type.

In this paper I shall make an attempt to give a fuller discussion of such skew curves as are derived from the data given by the recently published experience by the Norwegian Government, viz.: "Ulykkesforsikringen for Industriarbejdere," Christiania, 1915. (Accident Assurance for Industrial Workers.) This experience covers the period from 1895 to 1912, and includes 172 groups of industries with a total payroll of 1,835,632,504 Kroner and a total loss of 31,464,034 Kroner in the above mentioned period, or a pure premium of 17.1$^0/_{00}$* of the total payroll for all industries.

The accident and invalidity insurance as practised by the Norwegian Government Institution is founded upon actuarial principles somewhat similar to those adopted by the various American companies. The losses are based upon the commuted (capitalized) values of the future benefit contingencies at the time of the accident. The actuarial tables, select as well as truncated mortality and invalidity tables for both sexes, are derived from Norwegian census data and represent without doubt the most scientifically constructed tables, which we possess at the present time. In choosing a rating system the Norwegians have wisely decided to use the level rate system instead of the assessment system. Personally, I consider this method as far superior to that of levying an assessment each year for an amount sufficient to cover the capitalized (commuted) losses incurred during the year. A certain year may be very favorable and exhibit only a few fatalities and accidents to be followed sooner or later by a very unfavorable year with great losses, and perhaps the very unfavorable experience may occur at a time of economic depression in which the financial conditions of the industries are such that they are little adapted to carry the additional burden. As an example of the great deviations I choose the following figures from the cooperage trade in the period 1909–1912.

COOPERAGE INDUSTRIES (1909–1912).

Year.	Salaries in Kroner.	Losses in Kroner.	Rate $^0/_{00}$.
1909	295,402	3,562	12.1
1910	377,483	6,123	16.2
1911	367,044	17,839	48.6
1912	189,738	223	1.2

THE QUESTION OF STABILITY.

One of the first steps in a statistical analysis is to test the stability of the series as exhibited by the actually observed data. Does the

* Expressed in terms of mills, not of per cent.—a notation used throughout this paper.

ratio of the losses to the payrolls show violent fluctuations from year to year, and is it possible to trace such fluctuations to their proper sources? I have repeatedly maintained that statistical frequency ratios are not identical with mathematical probabilities, and that it is necessary to test the stability of the observed data before using such data for future predictions. It does not suffice to rely upon the idea that a rate is safe if the number of observations is large enough. In order to determine whether the industrial conditions in Norway are such that they may be considered stable from year to year, I give below a detailed computation of the Charlier coefficient of disturbancy, which is one of the best criterions in the test for stability.

LOSSES AND CORRESPONDING PAYROLLS BY CALENDAR YEARS FROM 1895–1912. ALL INDUSTRIES. (RIKSFORSIKRINGSANSTALTEN, 1915 REPORT.)

Year.	Payroll in 1,000 Kroner, s_K.	Losses in 1,000 Kroner, m_K.	$s_K p_0$.	$\mid m_K - s_K p_0 \mid$.
1895/96	96,042	1,817	1,646	171
97	70,656	1,357	1,211	146
98	81,595	1,582	1,398	184
99	92,393	1,635	1,584	51
1900	93,518	1,710	1,603	107
01	94,037	1,626	1,612	14
02	92,894	1,453	1,592	139
03	91,529	1,568	1,568	0
04	91,760	1,464	1,573	109
1905	94,103	1,480	1,613	133
06	103,154	1,679	1,768	89
07	114,517	1,982	1,973	11
08	122,147	1,904	2,093	189
09	133,208	2,095	2,283	188
1910	140,938	2,416	2,416	0
11	153,224	2,872	2,626	246
1912	169,918	2,821	2,912	91
Totals	1,835,633	31,461		1,868

$\delta = 1.0176$, $\sigma = 1.2533$, $\delta = 1.2754$, $\sigma_B^2 = 1.5601$,

$$\rho = \frac{\sqrt{1.6266 - 1.5601}}{17.1} = .0015.*$$

The above calculation shows that the Charlier coefficient of disturbancy, 100ρ, has the low value of 0.15, which clearly indicates that for all practical purposes we may safely consider the annual total losses as a normal and stable statistical series, wherein the

* See Fisher: "Mathematical Theory of Probabilities," p. 160.

perturbations are due to sampling only. This goes to show that in Norway, at least, the various industries have reached a state of stability so far as accidents are concerned. This probably is due to factory inspection and a rigid enforcement of factory laws requiring the installation of various safety devices. Whether the same stable conditions exist in America can only be determined by actually computing the coefficient of disturbancy for a loss series corresponding to the one given above for Norway.

The Classification of Risks.

The Norwegian system of classifying risks bases the pure premium according to industries. At the time of the establishment of the Government Assurance Institution (1895) no data as derived from purely Norwegian experience were at hand, and the founders of the institution fell back upon the German system of grouping the various trades in 6 danger classes with following premium rates.

Danger Class.	Rate per 1000 of Payroll
1	5
2	7
3	11
4	15
5	20
6	25

This grouping was already in 1899 increased to 16 danger classes with following rates:

Danger Class.	Rate per 1000 of Payroll.
4	4
5	6
6	8
7	10
8	12
9	14
10	16
11	18
12	20
13	24
14	28
15	32
16	36

As the business and experience expanded additional danger classes were incorporated so that in the 1915 report we have no less than 24 classes distributed as follows:

Danger Class.	Rate per 1000 of Payroll.
3	2
4	4
5	6
6	8
7	10
8	11
9	12
10	13
11	14
12	16
13	18
14	20
15	22
16	25
17	28
18	30
19	32
20	36
21	40
22	45
23	50
24	60

The advantage of a limited system of danger classes as described above is twofold. It gives first of all a comparatively small number of pure premium rates upon which the final gross office rates of the tariff may be based. Secondly, a limited classification enables us to collect sufficient statistical data from which empirical pure premiums may be constructed.

The question which is of great importance is to what extent the various danger classes are subject to fluctuations. Each danger class may be looked upon as a sum total of several sub-classes, each subsidiary danger class possessing its own particular frequency curve. The individual frequency curves inside a certain danger class will together form a Lexian Series, that is a set of sample sets with varying probability from set to set. If we for practical purposes are justified in regarding each sub-group as a Bernoullian Series, the danger class may be represented as Lexian Series whose frequency curve will be either an A or a B curve.

For the purpose of computing the parameters of the various frequency curves I have chosen a slightly different classification than the one used in the Norwegian Manual. The less dangerous traces I have grouped in 6 danger classes and fitted to B curves of the form: $F(x) = \psi_\gamma(x) + \gamma_2 \Delta^2 \psi_\lambda(x)$ where $\psi_\lambda(x)$ is the Poisson exponential, λ and γ_2 certain parameters. In a later paper I intend to deal with the A curves of the remaining danger classes.

The following tables give the various danger classes with their subgroups and B curves.

Danger Class No. 2 (from 2 ‰–4 ‰).

Industry.	Losses.	Payroll.	Rate ‰.
Faience factories	7,276	3,825,315	2.2
Manufacture of Paris points	3,424	1,542,083	2.2
Porcelain works	6,783	2,963,836	2.3
Tobacco works	35,877	15,687,525	2.3
Book printing and lithography	79,115	35,056,063	2.3
Caoutchouc works	4,207	1,535,502	2.7
Gold and silver smiths	22,005	7,545,443	2.9
Cotton spinneries (small works)	1,841	567,377	3.1
Chocolate and candy works	15,891	4,824,228	3.3
Cotton and woolen weavers	13,698	3,483,822	3.6
Ribbon weavers	7,962	2,038,748	3.9

Fitting the above data to a Poisson-Charlier B Curve, we have: $\lambda = 2.8$, $\gamma_2 = 0.127$ and $F(x) = \psi_{2.8}(x) + 0.127 \Delta^2 \psi_{2.8}(x)$, resulting in following values:

x	$F(x)$
0	.0685
1	.1764
2	.2331
3	.2119
4	.1493
5	.0870
6	.0435
7	.0191
8	.0074
9	.0027
10	.0008
11	.0002
12	.0001

The number x denotes the loss per 1000 of salary and $F(x)$ the probability of the occurrence of such a loss. The accompanying figure 1, of the curve shows it is decidedly skew.

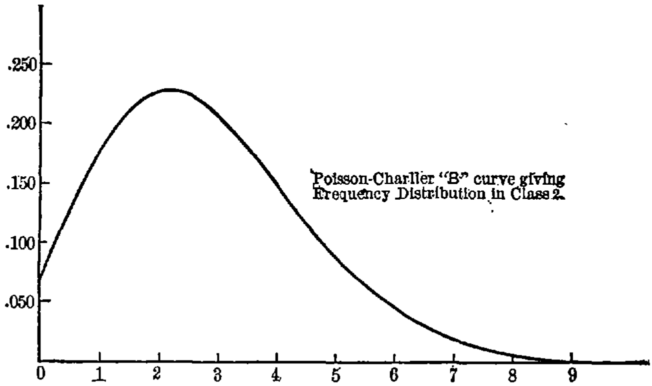

Figure 1.

DANGER CLASS No. 3 (FROM 4 ‰–6 ‰).

Industry.	Losses.	Payroll.	Rate ‰.
Knitting works	10,932	2,758,529	4.0
Brush factories	5,788	1,361,198	4.2
Shoe factories	71,241	16,636,010	4.3
Glass works	64,013	14,679,270	4.4
Textile works	123,235	28,131,256	4.4
Book binding	16,785	3,633,445	4.6
Soap works (with motor)	5,734	1,217,803	4.7
Manufacture of gas and sewer mains	8,631	1,799,768	4.8
Manufacture of mats, hemp and jute	36,753	7,603,414	4.8
Tanneries	37,972	7,689,349	4.9
Manufacture of nails, screws, etc.	70,319	13,874,259	5.1
Manufacture of spices, coffee roasting, etc.	7,689	1,494,860	5.1
Metal works (brass foundry)	15,308	2,924,874	5.2
Match factories	51,436	9,781,926	5.2
Frame and panel works	7,107	1,310,020	5.4
Work under the navy	87,925	16,281,365	5.4
Bakeries and confectioners	120,794	21,566,643	5.6

The parameters as fitted to a B curve are: $\lambda = 4.8$, $\gamma_2 = 0.085$ and $F(x) = \psi_{4.8}(x) + 0.085\Delta^2\psi_{4.8}(x)$. See Figure 2.

x	$F(x)$
0	.008930
1	.041462
2	.096849
3	.151825
4	.179773
5	.171550
6	.137446
7	.095098
8	.057992
9	.031654
10	.015650
11	.007050
12	.002948
13	.001139
14	.000138
over 15	.000496

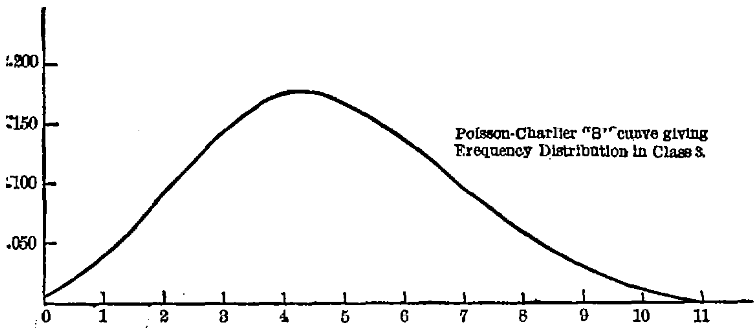

Figure 2.

Danger Class No. 4 (from 6 ‰–8 ‰).

Industry.	Losses.	Payroll.	Rate ‰.
Woolen weavers	12,769	2,075,216	6.1
Condensed milk works	34,998	5,603,452	6.2
Piano works (with motor)	13,436	2,143,135	6.3
Oeleomargarine works	38,982	5,927,310	6.6
Installation of small electric works	25,812	3,748,121	6.9
Manufacture of fishing net	10,617	1,542,953	6.9
Railway wagon works	42,692	6,041,482	7.1
Soap and perfume works	5,717	800,242	7.1
Works, under Army	114,942	15,907,629	7.2
Potteries	8,494	1,159,214	7.3
General woolen works	183,061	25,201,093	7.3
Dairies	77,472	10,412,116	7.4
Small mechanical shops	22,519	2,938,184	7.7
Hemp, jute and linen spinneries	53,001	6,873,647	7.7

Parameters as fitted to a B curve are: $\lambda = 7.0$, $\gamma_2 = 0.234$. Hence we have:

x	$F(x)$
0	.0012
1	.0075
2	.0248
3	.0554
4	.1006
5	.1271
6	.1455
7	.1440
8	.1258
9	.0930
10	.0695
11	.0462
12	.0280
13	.0158
14	.0083
15	.0041
16	.0019
17	.0008
18 and over	.0005

The curve is shown in Figure 3.

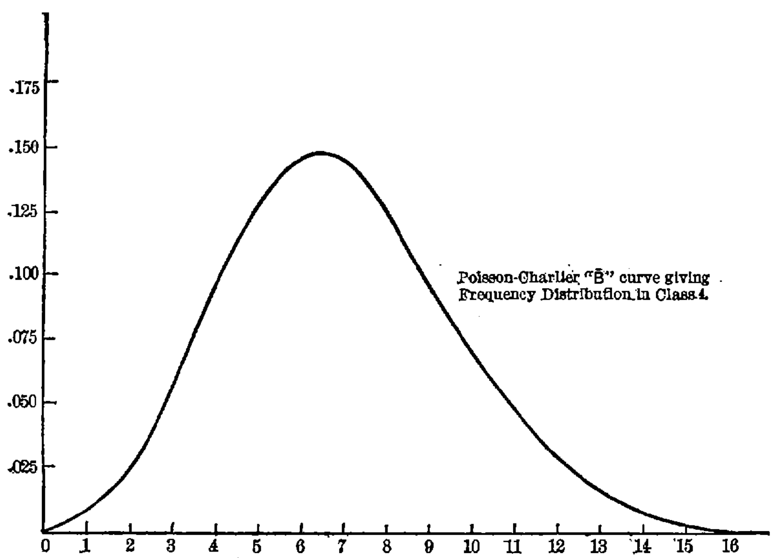

Figure 3.

Danger Class No. 5 (from 8 ‰–10 ‰).

Industry.	Losses.	Payroll.	Rate ‰.
Street car service outside power house	99,679	10,474,345	9.5
Work in store house, loading and unloading of ships	408,307	41,309,317	9.9
Brickyards	61,891	6,781,763	8.4
Iron and steel foundry without model shops	144,584	16,431,262	8.8
Manufacture of tools and cutlery	15,151	1,599,854	9.5
Woolen spinneries	15,087	1,782,739	9.6
Dye works (with motor and stamping)	21,151	3,323,405	9.1
Manufacture of paper and paste board	16,882	2,046,040	8.2
Conserves manufacture (without box making)	19,495	2,288,443	8.5
Conserves manufacture (with box making)	155,984	19,376,285	8.1
Butcheries, sausage works with motor	53,270	6,361,006	8.4

The parameters are here: $\lambda = 8.9$, $\gamma_2 = 0.27$. See Figure 4.

x	$F(x)$
0	.0002
1	.0015
2	.0063
3	.0178
4	.0381
5	.0657
6	.0949
7	.1183
8	.1299
9	.1276
10	.1136
11	.0926
12	.0698
13	.0488
14	.0319
15	.0195
16	.0113
17	.0062
18	.0032
19	.0016
20 and over	.0012

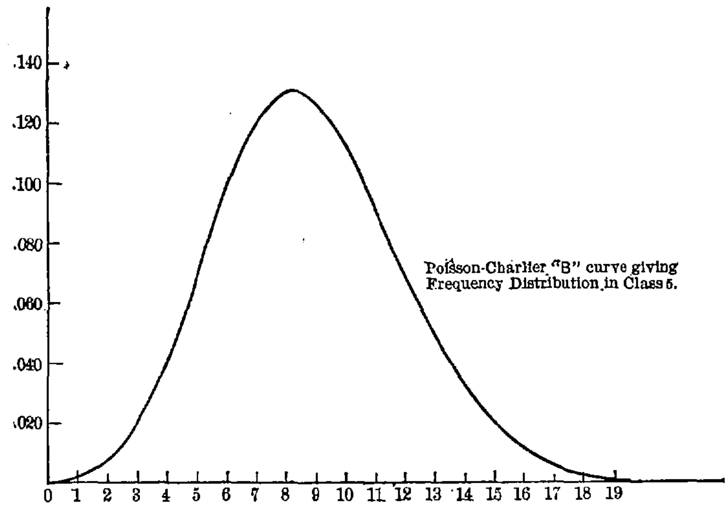

Figure 4.

DANGER CLASS NO. 6 (FROM 11 ‰–14 ‰).

Industry.	Losses.	Payroll.	Rate ‰.
Railroading	143,453	12,513,539	11.5
Storage work (exclusive of ship transport)	250,345	19,306,977	13.0
Storage work (connected with ship transport)	30,060	1,880,778	13.5
Iron works (furnaces)	15,840	1,542,062	10.3
Steel works (rolling)	9,980	890,013	11.2
Mechanical shops	787,088	67,377,472	11.7
Wagon factories (with motor)	24,127	2,049,263	11.8
Manufacture of electric light and power supplies	88,598	8,016,661	11.1
Electro-chemical works	43,984	3,497,863	12.6
Paper and carton works	454,293	39,619,658	11.5
Saw mills (Group I)	1,429,922	106,474,329	13.4
Planing mills	190,703	14,507,952	13.1
Mills (flour, groats, etc.)	262,540	19,052,627	13.8
Distilleries	35,240	2,473,218	13.5
Breweries	342,795	28,704,711	11.9
Painter (building trade)	196,578	16,748,653	11.7
Gas, water and sewer works	87,135	6,383,275	13.7
Installation of telegraph and telephone lines	70,457	5,476,201	12.9
Chimney sweeps	15,068	1,410,373	10.7
Government works	33,919	2,687,056	12.6

A computation of the parameters gives: $\lambda = 12.3$, $\gamma_2 = 1.04$ and the following values for $F(x)$. See Figure 5.

x	$F(x)$
0	.0000
1	.0001
2	.0007
3	.0022
4	.0062
5	.0141
6	.0268
7	.0438
8	.0632
9	.0818
10	.0963
11	.1043
12	.1050
13	.0989
14	.0878
15	.0737
16	.0589
17	.0446
18	.0323
19	.0224
20	.0148
21	.0094
22	.0057
23	.0033
24	.0018
25	.0010
26	.0005
27	.0003
28	.0005

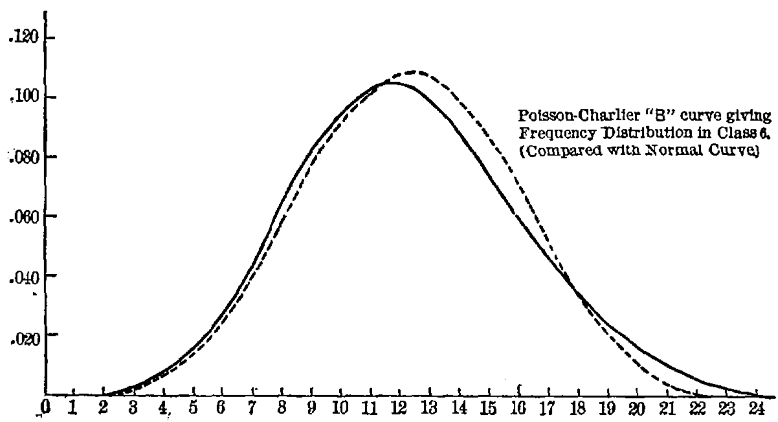

Figure 5.

COMPARISON OF EXPECTED AND ACTUAL LOSSES.

The equations of the above frequency curves are derived by giving equal weight to the various sub-classifications. This is, strictly speaking, not exact when the payrolls differ greatly, since we in such cases give equal weight to small and large payrolls. Introduction of weighted systems would indeed offer no serious obstacles. It may, however, be of interest to compare the expected losses as derived from the equations of the frequency curves as based upon the unweighted data with the actual losses incurred in the 17-year interval.* The losses in each danger class may be looked upon as a mathematical expectation. "A mathematical expectation is the product of an contingent gain (loss) in actual value and the mathematical probability of obtaining such a gain (loss)."†

The computed values of $F(x)$ for each danger class as given above makes the calculation of the expected losses quite simple. Take for instance danger class No. 2. The probability of the occurrence of a loss of $1.00 per 1000 of payroll is 0.1764, that of a loss of $2.00 is 0.2331, that of $3.00 is 0.2119, and so forth. The total expected loss per 1000 of payroll is therefore:

$$E = \Sigma x F(x) = 0 \times .0685 + 1 \times .1764 + 2 \times .2331 + \cdots + 12 \times .0001.$$

Multiplying this with the sum total of the payroll, we obtain the total expected losses or $P \times E$, where P is the total payroll.

An actual calculation for the various danger classes gives the following results:

Danger Class.	Sum Total of Payrolls.	Expected Losses Computed from B Curves.	Actual Losses.	Excess (+) or Deficient (−).	Per Cent. of Actual Loss.
2	78,070,000	218,603	198,079	+ 20,524	10.36
3	131,176,000	629,290	620,850	+ 8,440	1.36
4	90,373,000	629,077	644,422	− 15,345	2.37
5	111,773,000	994,489	1,011,481	− 16,992	1.67
6	360,613,000	4,387,494	4,492,125	−104,631	2.33
All classes	772,005,000	6,858,953	6,966,957	−118,004	1.69

The deficiency for all classes is 1.69 per cent. of the total losses,

* In this connection see the remarks by Mr. Joseph H. Woodward on page 478 of Volume II of the *Proceedings*.

† Fisher, "Mathematical Theory of Probabilities," page 49.

which is a rather close fit despite the fact that an unweighted series was used in the determination of the parameters. The various curves seem therefore safe approximations for the pure premiums, which with proper loadings ought to serve quite satisfactory as a basis for office rates. The curves show how important an element is the fluctuations due to random sampling. Take for instance danger class No. 5. The probability that the loss will be less than $5.00 is .0637, but the probability it will be $11.00 or more per 1000 is .2861, or we might expect that in 286 out of 1000 cases the loss will be higher than 11 dollars per 1000 of payroll.