

STATISTICAL MODELS OF CLAIM DISTRIBUTIONS IN FIRE INSURANCE

LARS-GUNNAR BENCKERT and JAN JUNG

Stockholm

SUMMARY

The authors have studied the combined data on claims in fire insurance of dwelling houses reported 1958-1969 by Swedish fire insurance companies. The claims were cleared of deductibles and adjusted according to a suitable index. Only losses above the largest deductible (in real value) applied during the observation period were included.

The material contains four different classes according to the fire resistibility of the building construction. For international comparisons, the pure classes B_1 ("stone" dwellings) and B_4 (wooden houses) are of interest. The distribution of the claims could be well approximated by the log-normal distribution in B_1 and by the Pareto distribution in B_4 . An equally good or better fit was obtained by assuming the original loss, reported or not, being distributed according to these distributions and applying the distributions, conditioned by the loss being larger than the deductible. In both cases the distribution parameters are functions of the insurance amount in such a way, that the mean value of the loss is described as a power of this amount.

The authors refrain from any theoretical arguments for the general applicability of the distributions used. They observe, however, the good approximation by wellknown parametric distributions which facilitates many actuarial tasks, such as the determination of first loss premiums, deductible premium factors, excess-of-loss premiums etc. The agreement between model and reality make these functions fit for use in the models underlying the general risk theory and in the more comprehensive models of the non-life insurance business.

I. NOTATIONS

A	insurance amount	
D	Deductible (300 Skr 1965)	
Y	loss, reported or not	$(0 < Y < A)$
L	reported loss	$(D < L \leq A)$
$C = L - D$,	claim	$(0 < C \leq A - D)$
n	number of claims	
\ln	natural logarithm	
$d.f.$	cumulative distribution function	
$\phi(x)$	normal $d.f.$	
$G(y)$	$d.f.$ of Y	

$H(y)$	<i>d.f.</i> of L	$H(y) = \frac{G(y) - G(D)}{1 - G(D)}$
$P(y)$	<i>d.f.</i> of C	$P(y) = H(y + D)$
i	interval of claim amount	
l_i	upper limit of reported losses in i	
$c_i = l_i - D$	upper limit of claims in i	
P_i	cumulated frequency of claims $< c_i$ or losses $< l_i$	
π	frequency of losses $< D$	
Skr	Swedish "kronor" (approx. 0.1 £)	
Skr 1965	, index adjusted to real value 1965	
hkr	100 Skr	
tkr	1000 Skr	

2. INTRODUCTION

The actuary is expected to know as much as possible about the future claims in a portfolio. This knowledge is condensed in a "mathematical model", which in most non-life branches should include the random nature of the outcome. The model also serves as a guide for assembling and arranging the risk statistics, which should give us information when the real development deviates from the expected.

Risk statistics involves a race against time and is not complete until all losses are reported and the claims settled. In some branches the actuary may even be forced to make prognoses of past losses, e.g. the I.B.N.R. claims (Incurred But Not Reported). This applies i.a. to liability insurance. In fire insurance this problem is negligible, as fires are easily observed, but the settlement of large claims may be considerably delayed.

The sum of claims S for a future period, may be expressed as

$$S = n \cdot \frac{S}{n},$$

where n is the number of claims.

The extension can be further refined by introduction of insurance amounts in the portfolio and in the policies hit by damage etc., but if the portfolio is subdivided in reasonably

homogeneous classes, especially with respect to size, the description by number of claims and mean claim will suffice for our purpose. For a thorough survey of these questions we refer to the lecture by G. Benktander at the 1972 congress of actuaries [Ref. 3].

The actuary in a medium sized company normally gets sufficient information on the incidence of fires to make a forecast of n , but as the distribution of claims is very skew the mean S/n depends heavily on the scarce large losses.

In order to obtain the best information on S/n we should try to estimate the distribution function $P(y)$ of the individual claim, given all information of the policy. The function $P(y)$ is fundamental for the application of the collective risk theory and also for the everyday decisions regarding deductibles, first loss amounts, loadings, retentions and other questions of reinsurance. As these decisions are based on the tails of the distribution, it is essential that the estimation is based on as large statistics as possible. Thus the task of estimating $P(y)$ is suitable for the cooperation of competing companies. In Sweden the companies keep their own records of the portfolio and of the claims, but also pool all their claims experience to "Centralstället för Svensk Brandskadestatistik".

This common data pool comprises the statistical data on the losses and on the policies hit. This material has been used in this study.

3. STATISTICAL DATA

The statistics comprises all claims in fire insurance for dwelling houses paid by the nation-wide companies during 1958-1969, numbering 78,940 in total. Thus the contents are not included.

In order to make the figures from different years and companies comparable, the influence of inflation and varying deductibles should be eliminated. The highest deductible (in real value) occurring during the period was *Sk*r 300, applied since 1965. Consequently all losses less than *Sk*r 300 after conversion to the money value of 1965, should be disregarded. The choice of a suitable index, however, is not evident since the claims depend of costs of building and repair material as well as of earnings of workers for reparation or construction.

We found that the rise of claim costs corresponded reasonably well to the index number I_t based on "average hourly earnings of workers in mining and manufacturing", which index is published yearly in the Statistical abstract of Sweden by the National Central Bureau of Statistics.

Some data illustrating the application of this index are given in Table 1 below

TABLE 1

Index I_t (earnings of workers in mining and manufacturing) and resulting corrections in number of claims and average losses.

Year	Index I_t ($I_{1965} = 300$)	Registered no. of claims	Registered no. of losses	Average of losses $> I_t$ <i>Skv</i>	Indexcorrected average of losses $> I_t$ <i>Skv 1965</i>
1958	178	6176	5795	2849	4804
59	186	5961	5580	3635	5863
60	197	6762	6195	3140	4779
61	213	7421	6590	3363	4773
62	232	8041	7046	3984	5161
63	249	9844	8486	3657	4406
64	271	8503	7181	4141	4581
65	300	8060	6859	4888	4888
66	328	8090	6955	5984	5475
67	357	8046	6546	5729	4814
68	382	7192	5714	8326	6585
69	415	7777	5980	8052	5821

After this preliminary adjustment all claims, where the total loss (= claim + deductible) was less or equal to I_t , were eliminated, and for the remaining claims the following information was registered:

1. Building class, B :

1. Stone and brick houses with fire resisting flooring.
2. Stone and brick houses with wooden flooring.
3. Wooden houses with plastered walls.
4. Wooden houses.

2. Insurance amount, A (not index-adjusted).

3. Index-adjusted total loss, L (= claim C + deductible D), expressed in *Skv* 1965.

The losses were also classified according to a semi-logarithmic two-figure code. Thus the code $c_1 c_2$ denotes the interval

$$c_2 \cdot 10^{c_1-1} Skr \leq L < (c_2 + 1) \cdot 10^{c_1-1} Skr.$$

An introductory study showed that the distribution functions of the losses differed between the building classes, and that the general shape could be described as

log normal	in $B = 1,$
partial log normal	in $B = 2,$
between log normal and Pareto	in $B = 3,$
Pareto	in $B = 4.$

As building constructions vary between geographic areas, we have thought that the pure classes $B = 1$ and $B = 4$ should be most fit for international comparisons and hence we have in this connection restricted the discussion to these classes.

In most tables and diagrams the intervals are put together in the following way (losses below 300 *Skr* 1965 are omitted!).

Interval code <i>i</i>	Lower limit <i>Skr</i> 1965 (included)	Upper limit <i>Skr</i> 1965 (excluded)
33—34	300	500
35	500	600
36	600	700
37—39	700	1,000
41	1,000	2,000
42	2,000	3,000
43—44	3,000	5,000
45—49	5,000	10,000
51	10,000	20,000
52	20,000	30,000
53—54	30,000	50,000
55—59	50,000	100,000
61	100,000	200,000
62	200,000	300,000
63—64	300,000	500,000
65—69	500,000	1,000,000
71	1,000,000	

4. MODELS OF THE LOSS DISTRIBUTIONS FOR DWELLING HOUSES OF STONE OR BRICK

The distribution of the loss amounts L in building Class 1 ("Stone houses") is given in the "Total" columns of Table 2 on page 7. In Diagram 1 on page 8, the cumulated frequencies (per cent) are plotted on a normal-probability paper against functions of the loss.

In the continuous curve (1) the abscissa x represents the natural logarithm of the loss L (in *hkr* 1965), thus starting at $x = 1.10$ (\ln of the deductible 3 *hkr*). The deviation from the log normal distribution for small x is obvious and natural, as this distribution should be positive over the whole positive x -axis.

In the lashed curve (2), x represents the natural logarithm of the claim (= loss minus 3 *hkr*), which covers the real axis. Although the curve does not deviate ostensibly from a straight, there is a significant concavity, which should not discourage the model builder. If there are reasons to expect a certain structure of the loss distribution (e.g. the specific model of log-normality as proposed and justified by i.a. Giovanna Ferrara [5]), this structure should be independent of the deductible and refer to the real loss, reported or not.

If the *d.f.* of the loss Y is $G(y)$ we have only observed the conditioned *d.f.* $H(y) = G(y | Y > D)$

$$= \frac{G(y) - G(D)}{1 - G(D)}$$

where D is the deductible which in our material is 3 *hkr*.

If $\pi = G(D)$, the probability of no claim or the loss being less than the deductible, were known, we could calculate $G(y)$ from the equation

$$G(y) = \pi + (1 - \pi) \cdot H(y) \quad (y > D) \quad (1)$$

The curves (3), (4) and (5) in Diagram 1 represents this transformation with $x = \ln y$ and π being chosen as 0.3, 0.4 and 0.5 respectively. Although we should expect a decent linear approximation, as we have chosen the parameter π for that purpose, the curve (4), where $\pi = 0.4$, shows an astonishing good fit to a straight

TABLE 2

Building class 1 (Stone dwellings). Frequency distributions of claim and loss amounts.

Magnitude group				1		2		3		4		Total	
Insurance amount (1,000 Skr):				0 — 249		250 — 1,249		1,250 — 2,449		2,500 — ∞			
Loss interval, i	Max loss $l_i = c_i + 3$ (100 Skr)	$\ln l_i$	$\ln c_i$	Number	Cum. fr.								
				of claims n_i	% F_i								
33—34	5	1.61	0.69	517	21.1	625	14.7	451	14.2	507	14.2	2,100	15.6
35	6	1.79	1.10	175	28.2	268	21.0	193	20.4	219	20.3	855	22.0
36	7	1.94	1.39	147	34.2	256	27.0	187	26.3	189	25.6	779	27.8
37—39	10	2.30	1.94	343	48.2	521	39.3	391	38.6	380	36.3	1,635	40.0
41	20	3.00	2.83	528	69.7	834	58.9	589	57.3	679	55.4	2,630	59.6
42	30	3.40	3.30	264	80.5	400	68.4	296	66.6	345	65.1	1,305	69.3
43—44	50	3.91	3.85	225	89.6	488	79.9	299	76.1	357	75.1	1,369	79.5
45—49	100	4.61	4.57	146	95.6	443	90.3	345	87.0	338	84.6	1,272	89.0
51	200	5.30	5.28	61	98.1	234	95.8	210	93.6	279	92.4	784	94.8
52	300	5.70	5.69	16	98.7	74	97.5	81	96.2	93	95.0	264	96.8
53—54	500	6.21	6.21	20	99.5	51	98.7	62	98.1	72	97.0	205	98.3
55—59	1,000	6.91	6.90	10	99.9	40	99.6	42	99.5	59	98.7	151	99.4
61	2,000	7.60	7.60	2	100.00	12	99.93	12	99.84	24	99.4	50	99.79
62	3,000	8.00	8.00	—	—	1	99.95	3	99.94	8	99.6	12	99.88
63—64	5,000	8.51	8.51	—	—	2	100.00	1	99.97	6	99.83	9	99.95
65—69	10,000	9.21	9.21	—	—	—	—	1	100.00	4	99.94	5	99.99
71	20,000	9.90	9.90	—	—	—	—	—	—	2	100.00	2	100.00
Total				2,454		4,249		3,163		3,561		13,427	
Mean loss (100 Skr 1965)				27.47		47.72		61.96		91.07		58.87	
Mean insurance amount (1,000 Skr)				120		716		1,730		6,131		2,254	

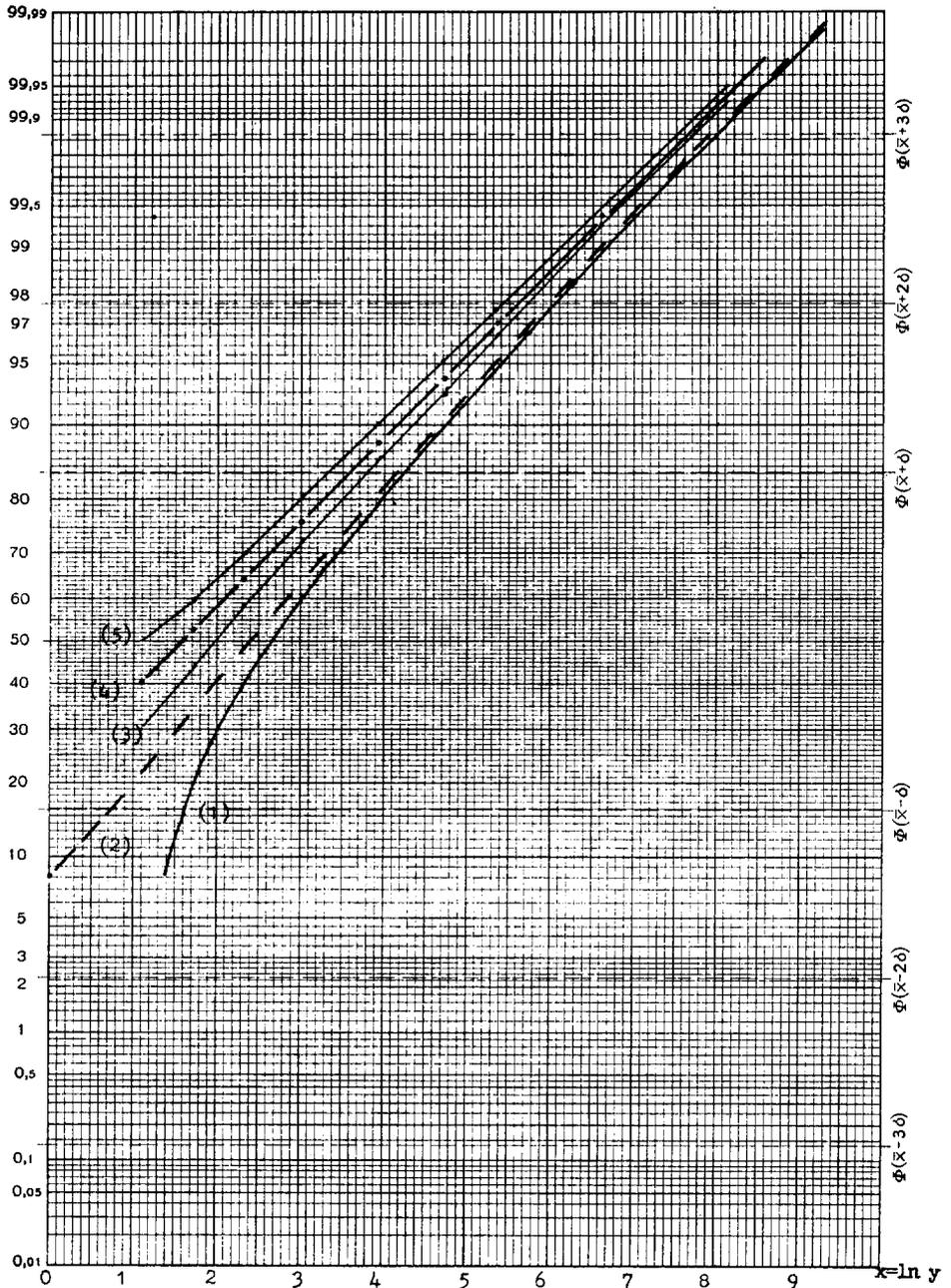


Diagram 1

Building class 1
(Stone dwellings)

Distribution of loss and claim amounts

- 1) $H(y)$, $y = \text{loss (in } hkr)$
- 2) $P(y)$, $y = \text{claim (in } hkr)$
- 3) $G(y) = 0.3 + 0.7 H(y)$, $y = \text{loss (in } hkr)$
- 4) $G(y) = 0.4 + 0.6 H(y)$, $y = \text{loss (in } hkr)$
- 5) $G(y) = 0.5 + 0.5 H(y)$, $y = \text{loss (in } hkr)$

Number of claims 13,427

from which we could estimate the parameters of the normal distribution:

$$\begin{aligned}\mu &= 1.60 \\ \mu + 2\sigma &= 5.58 \text{ and thus} \\ \sigma &= 1.99.\end{aligned}\tag{2}$$

Assuming the $\ln Y$ being normal with parameters μ and σ , the *d.f.* of Y is

$$G(y) = \phi\left(\frac{\ln y - \mu}{\sigma}\right), \text{ where } \phi(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau.$$

The r :th moment of Y is wellknown:

$$E(Y^r) = \int_{-\infty}^{\infty} y dG(y^r) = \int_{-\infty}^{\infty} e^{rx} d\phi\left(\frac{x - \mu}{\sigma}\right) = e^{r\mu + \frac{r^2\sigma^2}{2}}\tag{3}$$

For the variable L , the corresponding moment is

$$E(L^r) = E(Y^r | Y > D) =$$

$$= \frac{1}{1 - \phi\left(\frac{\ln D - \mu}{\sigma}\right)} \int_D^{\infty} y^r dG(y)$$

$$= \frac{1}{1 - \phi\left(\frac{\ln D - \mu}{\sigma}\right)} \int_{\log D}^{\infty} e^{rx} d\phi\left(\frac{x - \mu}{\sigma}\right).$$

$$\text{As } e^{rx} d\phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{rx - \frac{(x - \mu)^2}{2\sigma^2}} dx$$

$$\equiv e^{r\mu + \frac{r^2\sigma^2}{2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu - r\sigma^2)^2}{2\sigma^2}} dx, \text{ we get}$$

$$E(L^r) = e^{r\mu + \frac{r^2\sigma^2}{2}} \cdot \frac{1 - \phi\left(\frac{\ln D - \mu - r\sigma^2}{\sigma}\right)}{1 - \phi\left(\frac{\ln D - \mu}{\sigma}\right)}$$

$$= \frac{\phi\left(\frac{\mu - \ln D}{\sigma} + r\sigma\right)}{\phi\left(\frac{\mu - \ln D}{\sigma}\right)} \cdot e^{r\mu + \frac{r^2\sigma^2}{2}} \quad (4)$$

Introducing the estimates (2), which give $\phi\left(\frac{\mu - \ln D}{\sigma}\right) = 1 - \pi 0.6$ we obtain from (3) and (4)

$$\begin{aligned} E(Y) &= 35.9 \\ E(L) &= 59.1 \text{ and thus} \\ E(C) &= 56.1 \end{aligned}$$

The identity

$$E(Y) = Pr(Y \leq D) \cdot E(Y | Y \leq D) + Pr(Y > D) [D + E(Y - D | Y > D)] \text{ gives}$$

$$\begin{aligned} 35.9 &= 0.44 + 0.60 (3 + 56.1) \\ &= 0.44 + 1.80 + 33.66 \end{aligned}$$

which is a decomposition of a random loss in three parts

- a) the mean value of losses $\leq D$
- b) mean value of deductible (when claims occur)
- c) mean value of positive claims (occurring with probability 0.6).

Both the losses $< D$ and the deductible are small compared to the claims, when they occur. The role of the deductible is mainly to avoid the administration of all small claims, estimated to 40 per cent of all losses.

The log-normal model described has been subject to a χ^2 -test. Thus the frequencies in Table 1 (Total column) were compared with the frequencies deduced from the log-normal model ($\mu = 0.4$, $\mu = 1.60$ $\sigma = 1.99$). All claims above 50.000 *Skkr* (intervals 55 —) were added into one single group. We got $\chi^2 = 25.8$ with $12 - 3 = 9$ degrees of freedom, a value falling between the 99.5 and the 99.9 per-cent value of the one-sided test. This does certainly not give reason to accept the log-normal distribution as an hypothesis for the loss distribution, but it shows that for the total loss material used the model might give a fairly good description.

We have not hitherto used our knowledge of the insurance amounts which certainly affect the distributions. In many studies the statistics are based on the extent of damage, i.e. the loss as a fraction of the insurance amount. As the index-adjusted insurance sums are not included in the material at our disposal, we have based further analysis on a subdivision according to groups of magnitude, defined as follows.

Magnitude group	Insurance amount A (tkr)		Approx. interval
	1958-1963	1964-1969	Skv 1965
1	$A < 200$	$A < 300$	0- 249
2	$200 \leq A < 1000$	$300 \leq A < 1500$	250-1249
3	$1000 \leq A < 2000$	$1500 \leq A < 3000$	1250-2449
4	$2000 \leq A$	$3000 \leq A$	2500-

In the diagrams 2:1-2:4 on following pages the cumulated frequencies in the four magnitude groups have been plotted on a normal probability paper against 1) ln claim (hkr) and 2) ln loss (after estimating the probability π of the loss being less than the deductible).

The original estimates of π gave the following results:

Group:	1	2	3	4	Total
$\pi^* =$	0.5	0.3	0.4	0.4	0.4

As the estimates are very rough [judged from the linear tendency among several trial transforms as the curves 3), 4) and 5) in Diagram 1] the value $\pi = 0.4$ was accepted in all groups. A common value implies, that independent of the value of the dwelling house insured, and of the frequency of fire outbreaks, such an outbreak has a certain probability (≈ 0.6) of causing a loss larger than the deductible (3 hkr 1965).

The diagrams show, that also the distributions of the subgroups may be fairly well described by a log-normal distribution as well for the claims C as for the losses $L > D$.

The parameters, as estimated from the normal-probability paper, are given in Table 3 on page 16.

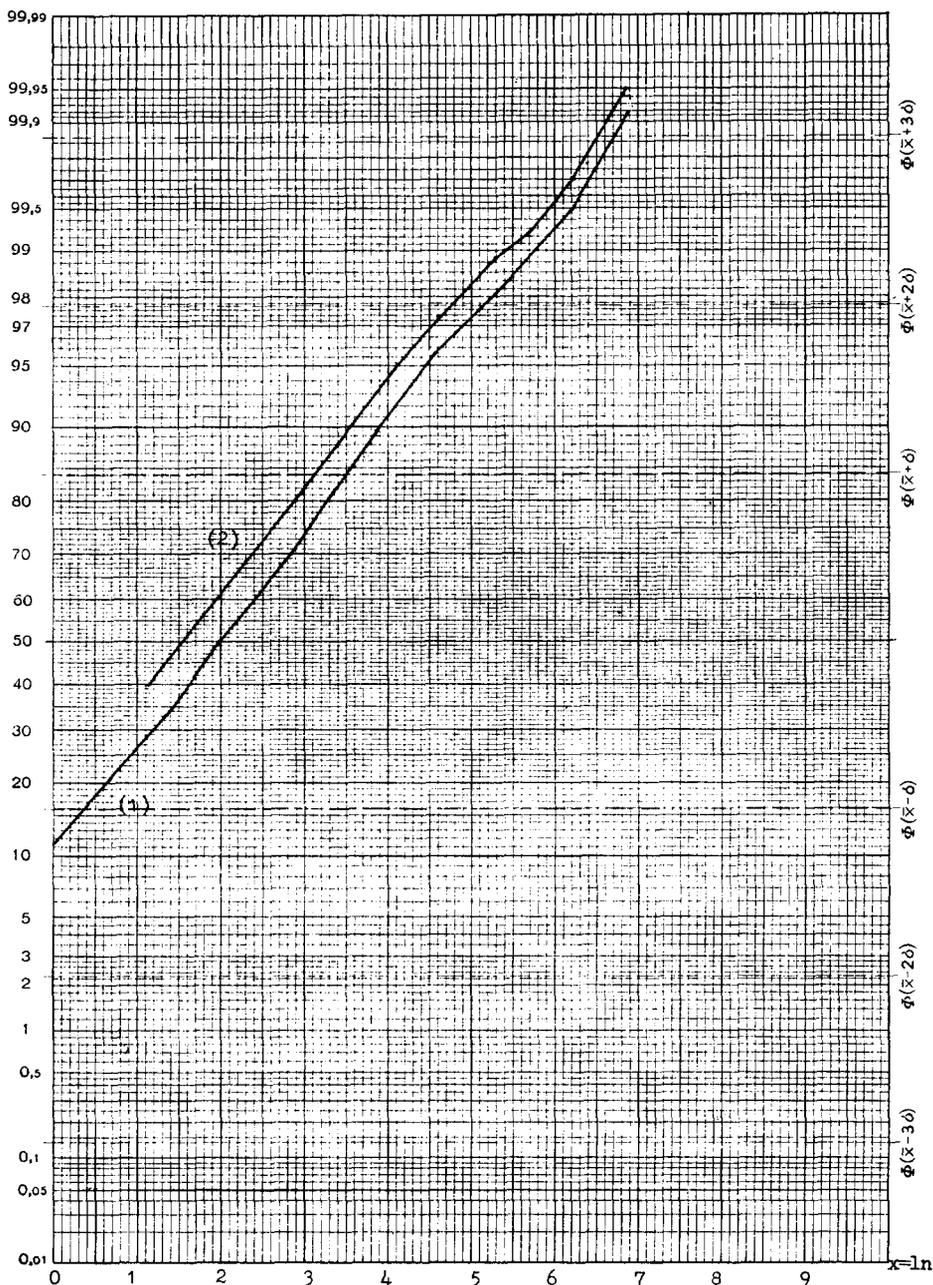


DIAGRAM 2:1

Building class I
(Stone dwellings)

Group I
Insurance amount
c-249 *tkr*

Distribution of loss and claim amounts

1) $P(y)$, $y = \text{claim (in } hkr)$

2) $G(y) = 0.4 + 0.6 H(y)$, $y = \text{loss (in } hkr)$

Number of claims 2,454

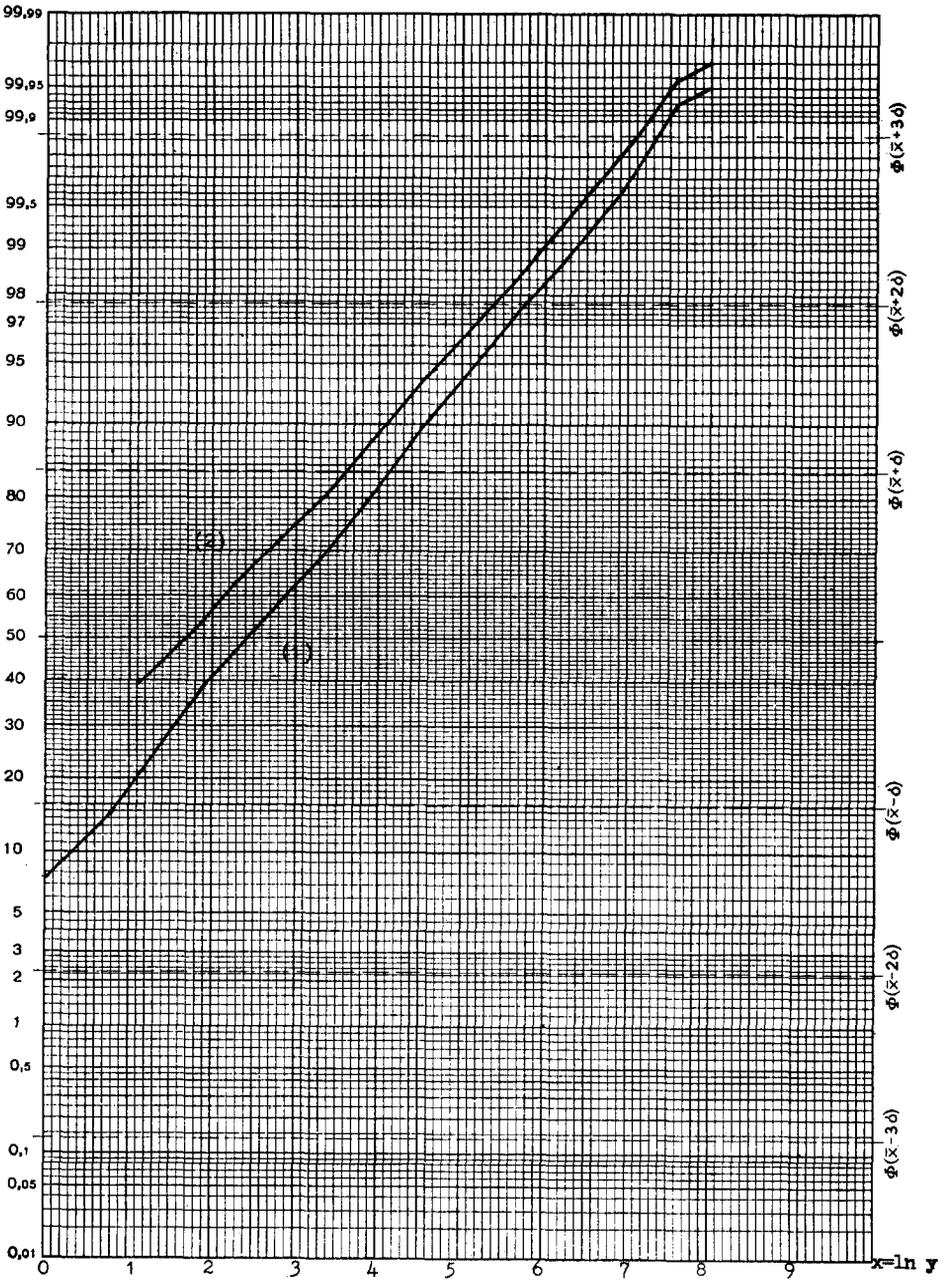


DIAGRAM 2:2

Building class 1
(Stone dwellings)

Group 2

Insurance amount
250-1,249 *tkr*

Distribution of loss and claim amounts

1) $P(y)$, $y = \text{claim (in } hkr)$

2) $G(y) = 0.4 + 0.6 H(y)$, $y = \text{loss (in } hkr)$

Number of claims 4,249

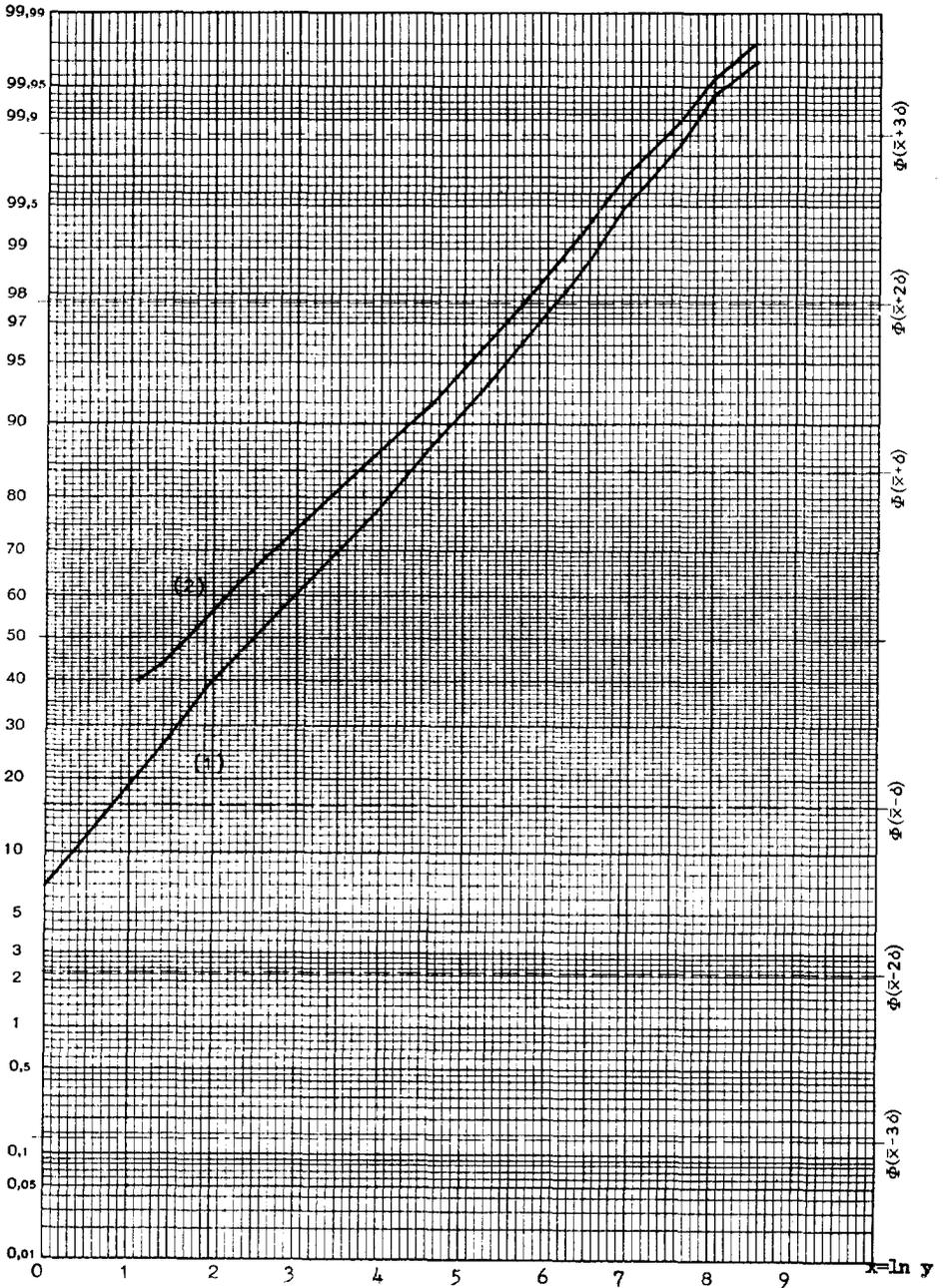


DIAGRAM 2:3

Building class 1
(Stone dwellings)

Group 3
Insurance amount
1,250-2,499 tkv

Distribution of loss and claim amounts

1) $P(y)$, $y = \text{claim (in hkr)}$

2) $G(y) = 0.4 + 0.6 H(y)$, $y = \text{loss (in hkr)}$

Number of claims 3,163

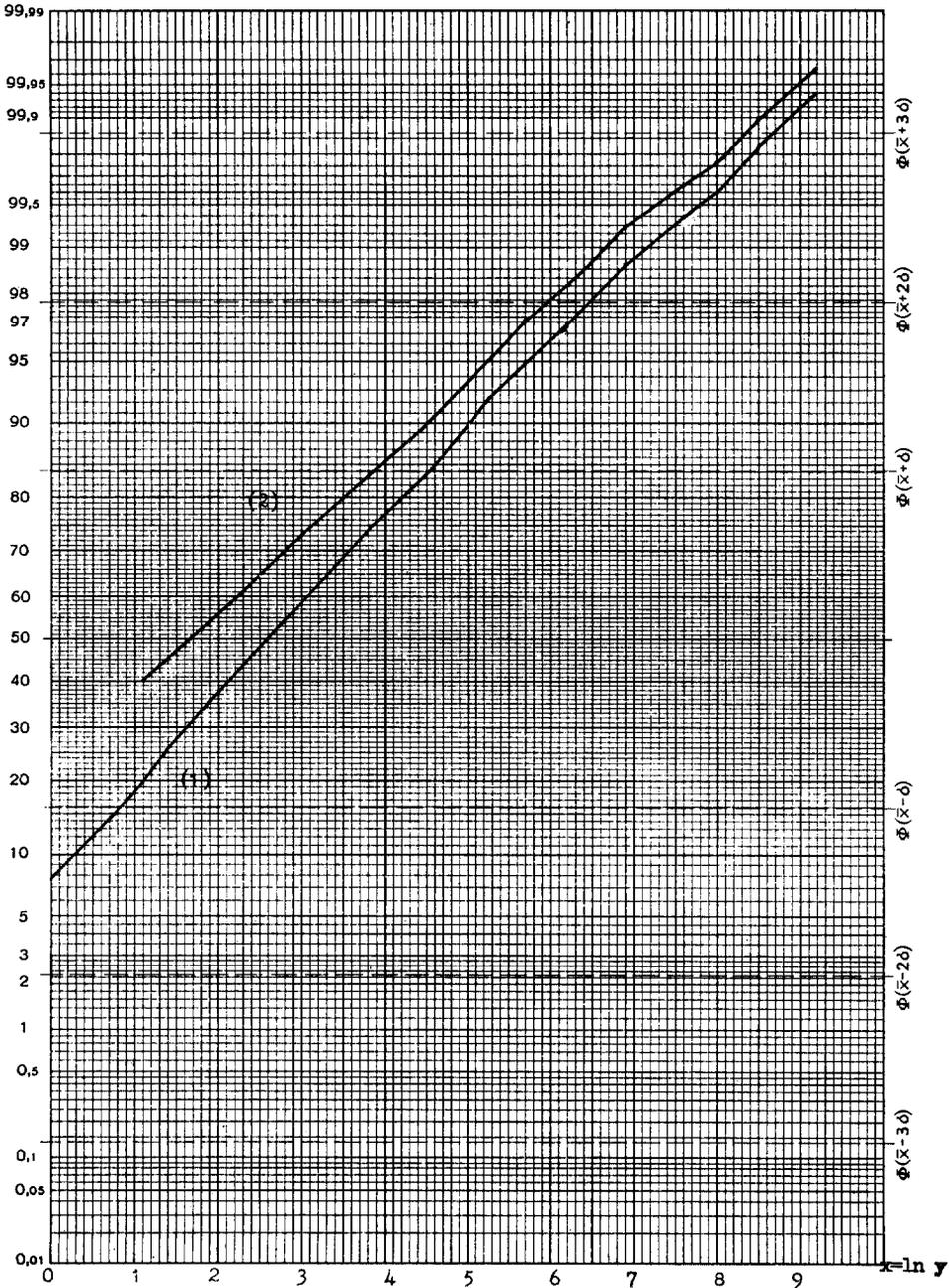


DIAGRAM 2:4

Building class 1
(Stone dwellings)

Group 4
Insurance amount
2,500 tkr —

Distribution of loss and claim amounts

1) $P(y)$, $y = \text{claim (in hkr)}$

2) $G(y) = 0.4 + 0.6 H(y)$, $y = \text{loss (in hkr)}$

Number of claims 3,561

TABLE 3. *Building class 1 (Stone dwellings)*
Comparison between the log-normal models and the observed distributions.

Group	1	2	3	4	Total B1
Insurance amount					
A <i>tkr</i>	0-249	250-1,249	1,250-2,449	2,500	
Mean ins. amount					
\bar{A} <i>tkr</i>	120	716	1,730	6,131	2,254
<i>Model A:</i>					
$\ln C$ „normal“ (μ, σ)					
μ^*	2.05	2.40	2.50	2.55	2.40
σ^*	1.50	1.65	1.75	1.95	1.80
$E(C) = \exp(\mu + \sigma^2/2)$	23.9	43.1	56.3	85.7	55.7
Mean \bar{C} observed (<i>hkr</i>)	24.4	44.7	59.0	88.1	55.9
<i>Model B:</i>					
$\ln Y$ „normal“ (μ, σ)					
π^* (= Prob $Y < 3hkr$)	.4	.4	.4	.4	.4
μ	1.50	1.60	1.60	1.70	1.60
σ^*	1.65	1.90	2.00	2.15	1.99
$E(L)$ [cf.(4)]	28.2	49.4	60.3	91.3	59.0
Mean \bar{L} observed (<i>hkr</i>)	27.4	47.7	62.0	91.1	58.9
\bar{L} (not index-adjusted)					
<i>hkr</i>	24.9	43.7	55.8	90.2	56.2

There is an obvious tendency in both models, that μ and σ , and thus \bar{C} and \bar{L} , increase with the mean insurance amount \bar{A} . As these amounts are not index-adjusted, we have studied the relation between the means of not index-adjusted losses L , given in the last line of the table. This relation is well described by the formula

$$\bar{L} = 5.35 \cdot (\bar{A})^{0.32} \quad (5)$$

This is in agreement with the wellknown experience that the average extent of damage L/A is proportional to a negative power of \bar{A} (c.f. Depoid, ref. [4] p. 463 f.f.).

This formula for \bar{L} also gives a good approximation of \bar{C} , as can be seen from the following comparison.

Group	\bar{A} <i>tkr</i>	$5.35 \cdot \bar{A}^{0.32}$	L	C
1	120	24.8	24.9	24.4
2	716	43.8	43.7	44.7
3	1,730	58.0	55.8	59.0
4	6,131	87.2	90.2	88.1
Total	2,254	63.1	56.2	55.9

Studying how μ^* and σ^* of Model A depends of \bar{A} , we get the following approximations:

$$\begin{aligned} \mu^{**} &= 1.48 + 0.13 \ln \bar{A} \\ \sigma^{**2} &= 0.40 + 0.38 \ln \bar{A} \end{aligned}$$

Group	μ^{**}	μ^*	σ^{**}	σ^*
1	2.10	2.05	1.49	1.50
2	2.33	2.40	1.70	1.65
3	2.45	2.50	1.80	1.75
4	2.61	2.55	1.93	1.95

5. MODELS FOR WOODEN BUILDINGS

The distribution of claims and losses are given in Table 4 on page 19.

The "Total" column shows the cumulated frequencies F_i of all claims $< c_i$ (or reported losses $< l_i$).

In Diagram 3 on page 20 the values of $1 - F_i$ are plotted against $\log l_i$, curve 1), on a logarithmic chart. The curve does not show the linear character of a normed Pareto distribution. Now this is hardly to be expected as heed has been paid neither to the effect of the deductible nor to the truncation at $y = A$. We have therefore used a slightly altered *d.f.* starting at $y = 0$.

Now suppose that the *d.f.* of the original loss Y , is

$$G(y) = 1 - \left(1 + \frac{y}{\sigma}\right)^{-\alpha} \quad \begin{matrix} 0 < \alpha \\ 0 < y < \infty \end{matrix} \quad (6)$$

For the reported and registered loss L , we get the *d.f.*

$$H(y) = G(y | Y > D) = \frac{G(y) - G(D)}{1 - G(D)} = \begin{cases} 0 & y < D \\ 1 - \left(\frac{\sigma + y}{\sigma + D}\right)^{-\alpha} & D \leq y < \infty \end{cases} \quad (7)$$

and thus for the claim $C = L - D$ the *d.f.*

$$P(y) = H(y + D) = 1 - \left(1 + \frac{y}{\sigma + D}\right)^{-\alpha} \quad (8)$$

We also have to truncate the distribution at a truncation point T , determined by insurance amounts.

Thus the distribution applied is

$$H(y; T) = \begin{cases} 0 & y < D \\ 1 - \left(\frac{\sigma + y}{\sigma + D}\right)^{-\alpha} & D \leq y < T \\ 1 & T \leq y. \end{cases} \quad (7')$$

The corresponding mean value is

$$E(L) = D + \frac{\sigma + D}{1 - \alpha} \left[\left(\frac{\sigma + T}{\sigma + D}\right)^{1-\alpha} - 1 \right] \quad \alpha \neq 1 \quad (9a)$$

$$= D + (\sigma + D) \ln \frac{\sigma + T}{\sigma + D} \quad \alpha = 1 \quad (9b)$$

A rough estimation shows, that $\sigma^* = 2 \text{ hkr}$ gives a good approximation of the distribution. Thus on Diagram 3 we have plotted $1 - F_i$ against $x = \ln(l_i + 2)$ in the curve 2), which for not too large values of x can be approximated by the straight

$$\ln [1 - H(y)] = 0.8624 - 0.785 \ln(y + 2),$$

corresponding to

$$H(y) = 1 - \left(\frac{y + 2}{5}\right)^{-0.785}, \quad 3 \leq y < T$$

Thus $\alpha^* = 0.785$.

For an individual insurance, T could be chosen as A and for a group of insurances with limited variation of the A values, T could be chosen so as to obtain a correct mean value. This should preferably be applied to separate magnitude groups, but to illustrate the method the observed mean $\bar{L} = 51.7 \text{ hkr}$ (cf. Table 4) substituted for $E(L)$ in (9a) gives for $\alpha = 0.785$ $T = 970 \text{ hkr}$, belonging to the loss interval 60 (900 — 1000 hkr). In our material only 233 claims out of 40,859 or 0.6 per cent of the claims fall above this interval.

In building class 4, where the insurance amounts are smaller than in class 1, we have used three magnitude groups, defined as

Magnitude group	Insurance amount A (tkr)		Approx. interval Skr 1965
	1958-1963	1964-1969	
1	$A < 100$	$A < 150$	0-124
2	$100 \leq A < 200$	$150 \leq A < 300$	125-249
3	$200 \leq A$	$300 \leq A$	250-

TABLE 4

Building class 4 (Wooden dwellings). Frequency distributions of claim and loss amounts.

Magnitude group: Insurance amount (1,000 Skr):				1 0 — 124		2 125 — 249		3 250 —		Total	
Loss inter- val, i	Max loss $l_i = c_i + 3$ (hkr)	$\ln l_i$	$\ln c_i$	Number of claims n_i	Cum. fr. % F_i						
33—34	5	1.61	0.69	6,930	20.0	731	17.0	233	12.9	7,894	19.3
35	6	1.79	1.10	2,469	27.0	293	23.8	110	19.0	2,872	26.4
36	7	1.94	1.39	2,073	33.0	247	29.5	108	25.0	2,428	32.3
37—39	10	2.30	1.94	4,403	45.7	559	42.5	241	38.4	5,203	45.0
41	20	3.00	2.83	7,710	67.9	848	63.4	362	58.5	8,970	67.0
42	30	3.40	3.30	3,244	77.2	428	73.3	185	68.8	3,857	76.4
43—44	50	3.91	3.85	2,841	85.5	387	82.3	177	78.6	3,405	84.8
45—49	100	4.61	4.57	2,116	91.5	326	89.9	141	86.4	2,583	91.1
51	200	5.30	5.28	1,321	95.0	135	93.0	90	91.4	1,456	94.6
52	300	5.70	5.69	583	96.7	59	94.4	34	93.3	676	96.3
53—54	500	6.21	6.21	560	98.30	59	95.7	31	95.1	650	97.9
55—59	1,000	6.91	6.90	530	99.82	71	97.4	31	96.8	632	99.42
61	2,000	7.60	7.60	62	100.00	96	99.61	21	97.9	179	99.86
62	3,000	8.00	8.00	—	—	17	100.00	19	99.00	36	99.95
63—64	5,000	8.51	8.51	—	—	—	—	11	99.61	11	99.98
65—69	10,000	9.21	9.21	—	—	—	—	7	100.00	7	100.00
71	20,000	9.90	9.90	—	—	—	—	—	—	—	—
Total				34,752		4,306		1,801		40,859	
Mean loss (hkr 1965)				43.2		83.0		149.7		51.7	
Mean insurance amount (tkr)				54		164		597		88	

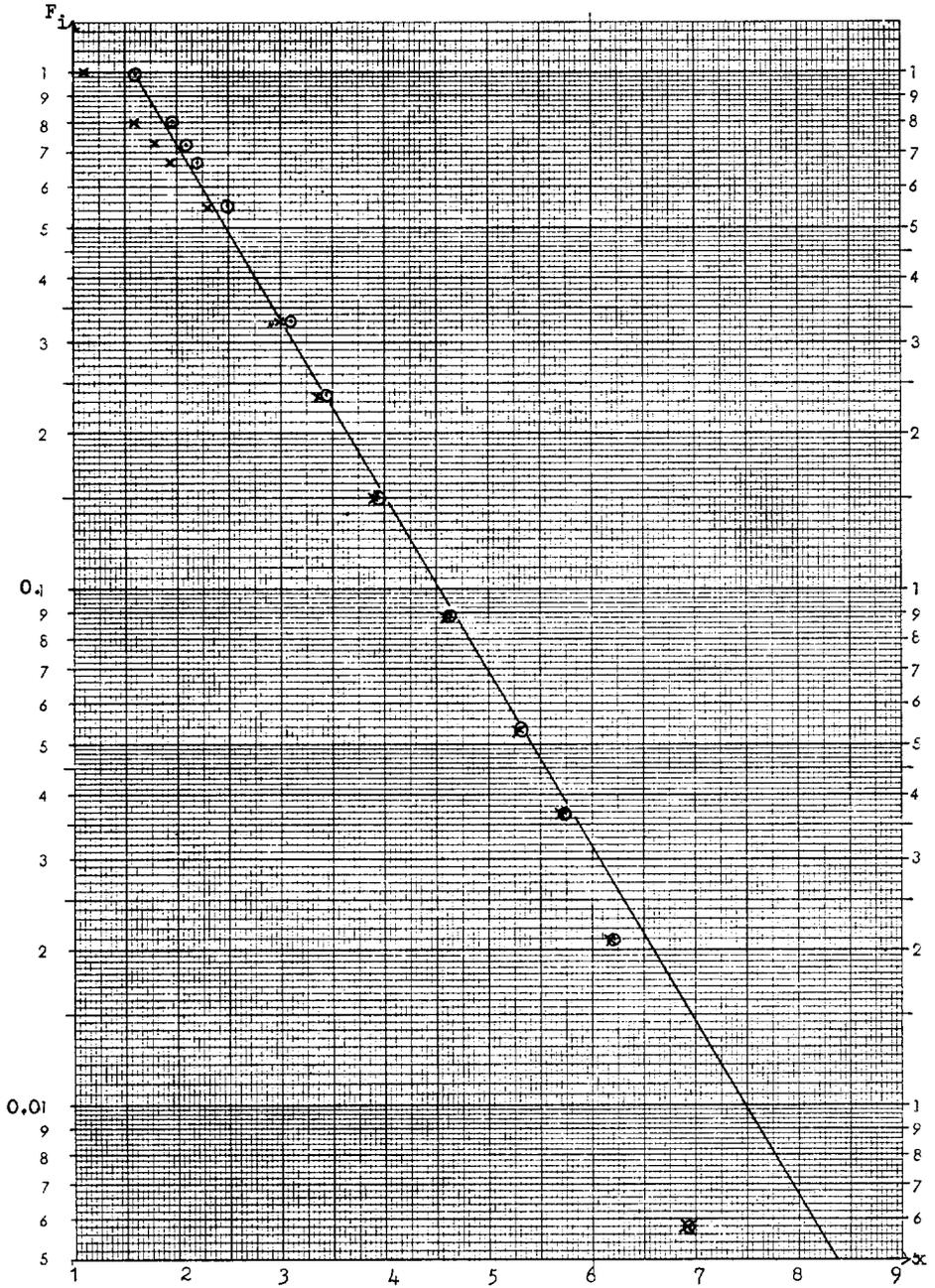


DIAGRAM 3

Building class 4
(Wooden dwellings)

- 1) $I - F_i, x = \ln l_i$
- 2) $I - F_i, x = \ln (l_i + 2)$
- 3) $I - H(y), x = \ln (y + 2)$

$\times \times \times \times$
 $\circ \circ \circ \circ$

Number of claims 40,859

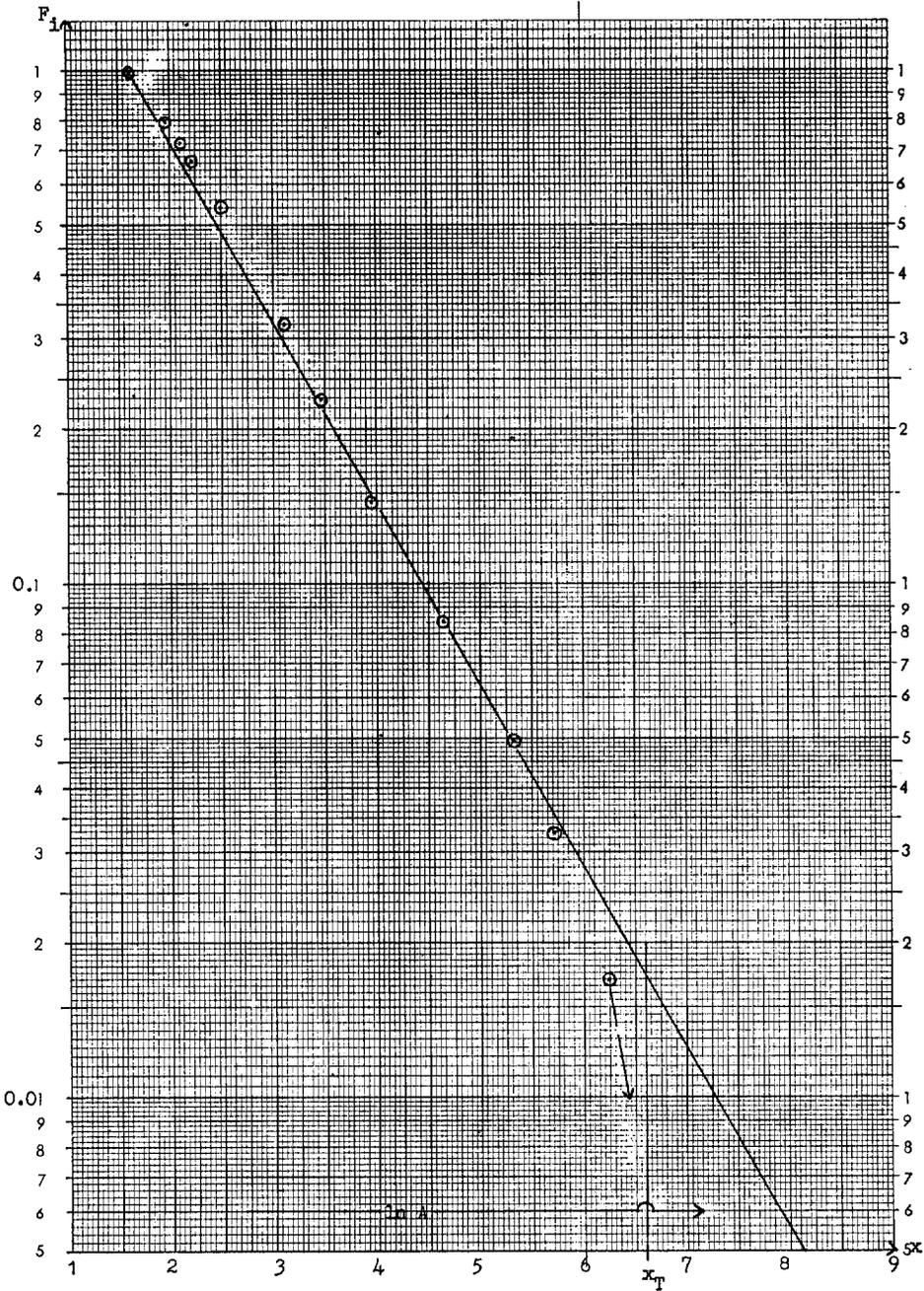
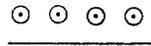


DIAGRAM 4:1

Building class 4
(Wooden dwellings)

- 1) $F_i - F_i, \quad x = \ln(l_i + 2)$
- 2) $F_i - H(y), \quad x = \ln(y + 2)$



Group 1
 $A < 125 \text{ thr}$

Number of claims 34,752

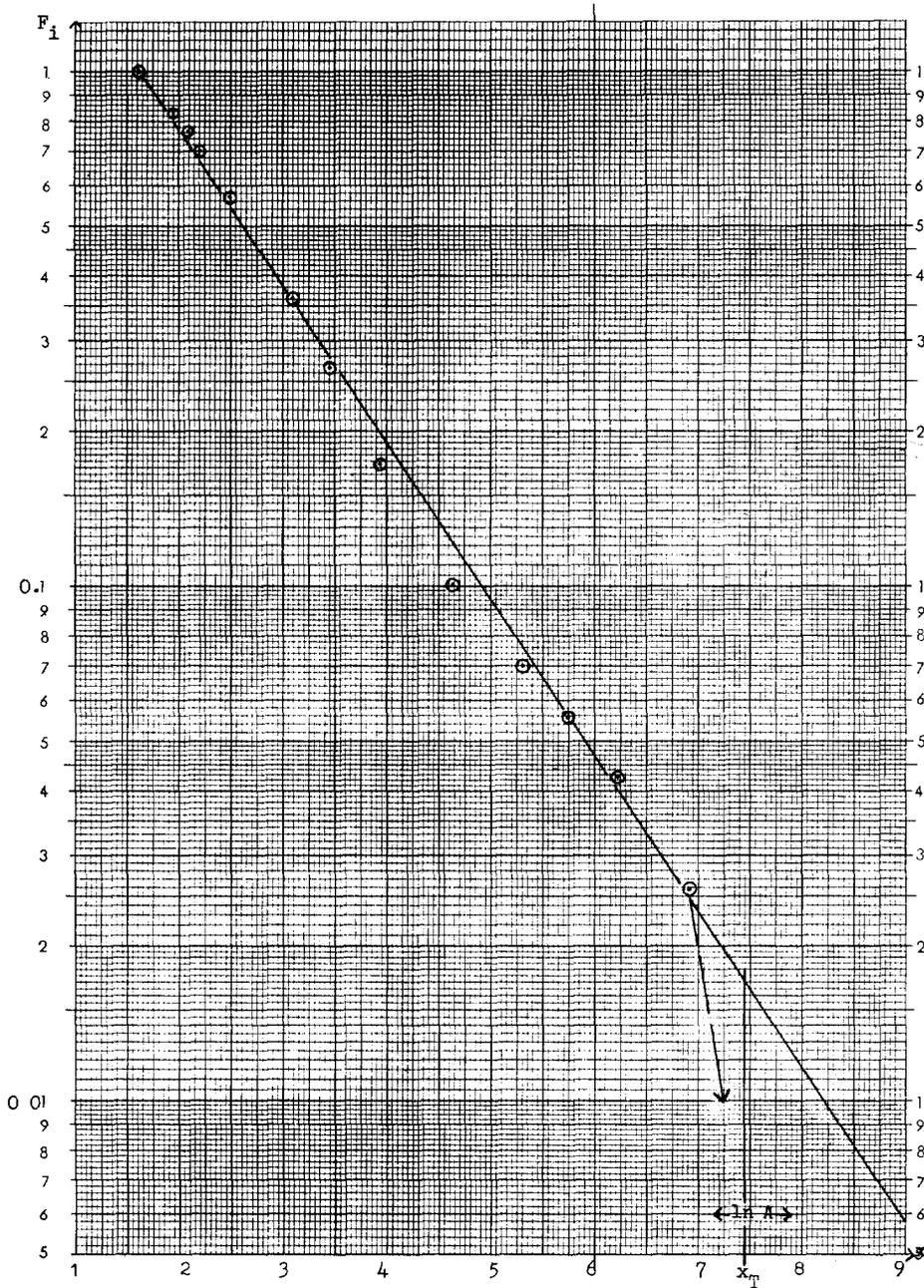


DIAGRAM 4:2

Building class 4
(Wooden dwellings)

- 1) $I - F_i, \quad x = \ln(l_i + 2)$
- 2) $I - H(y), \quad x = \ln(y + 2)$



Group 2
 $125 \text{ thr} \leq A < 250 \text{ thr}$

Number of claims 4,306

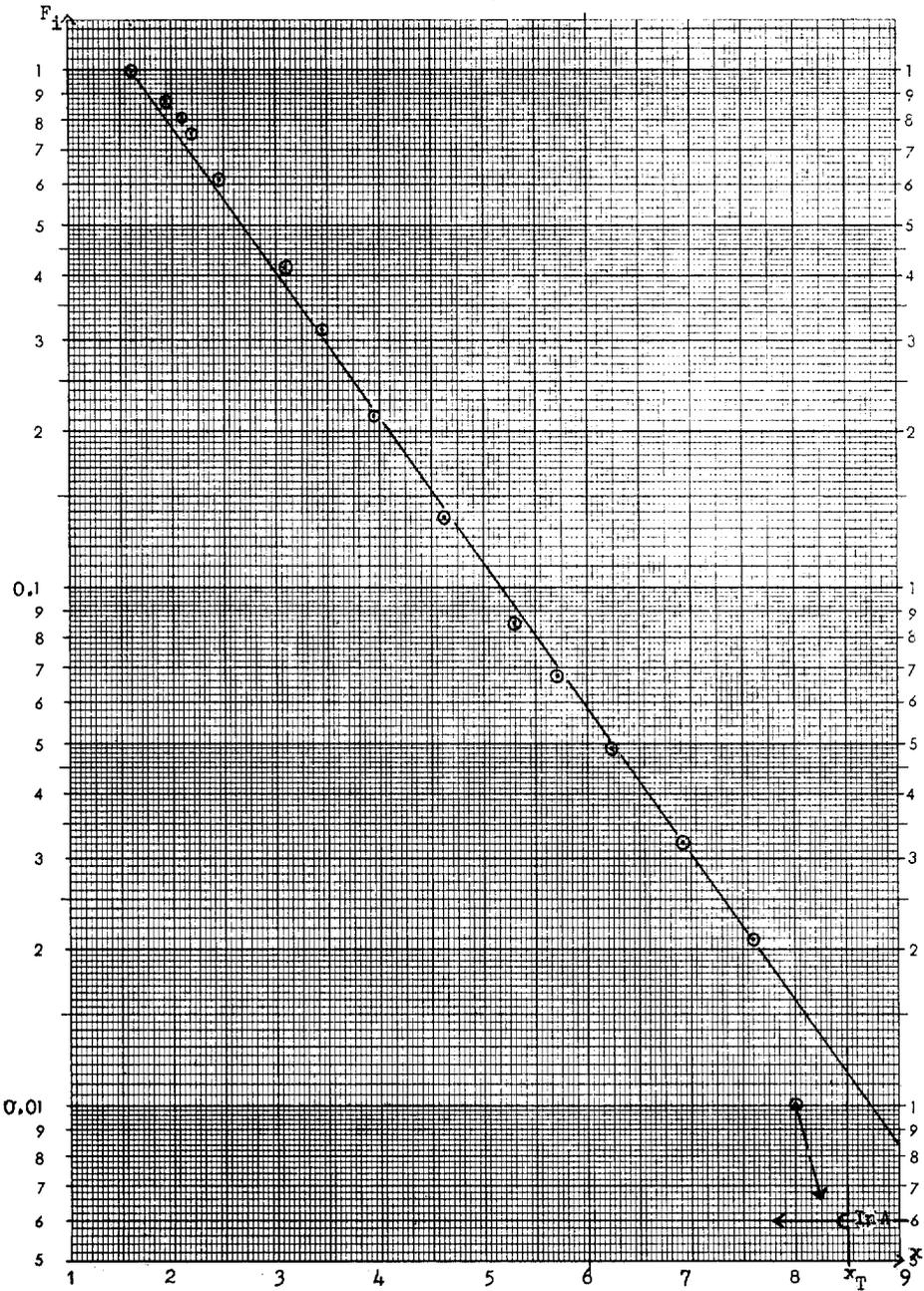
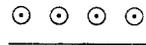


DIAGRAM 4:3

Building class 4
(Wooden dwellings)

- 1) $1 - F_i, x = \ln(l_i + 2)$
- 2) $1 - H(y), x = \ln(y + 2)$



Group 2
 $250 \text{ tkr} \leq A$

Number of claims 1,801

The observed distribution of claims and losses are given in Table 4, p. 20. The relation between mean loss L and mean insurance amount A can be roughly described by

$$\bar{L} \cong 6.26 (\bar{A})^{0.5} \quad (10)$$

Group:	1	2	3
\bar{A} (thr)	54	164	597
$6.26(\bar{A})^{0.5}$	46.0	80.1	153.0
\bar{L}	43.2	83.0	149.7

We found that the parameter $\sigma = 2$ served as well in the different groups as in total class. Thus in the diagrams 4:1 — 4:3 on the following pages, for each magnitude group the "tail" values $1 - F_i$ have been plotted against $\ln(l_i + 2)$ in the curve 1).

The observed distributions represented by the curve 1) in diagrams 4:1 — 4:3 can for all three groups be approximated by a straight line 2) in the logarithmic chart, and thus corresponding to the original loss distribution $G(y)$ according to (6) and the distribution of reported and registered losses $H(y)$ according to (7).

For the parameter P we obtained the estimates

Group:	1	2	3	Total
$\alpha^* =$	0.815	0.699	0.647	0.785

In Table 5 on the next side the observed distribution $1 - F_i$ is compared to the untruncated Pareto approximation $1 - H(y)$.

The dependence between the parameter α and the mean insurance amount \bar{A} can be approximated by the formula

$$\alpha \sim 1.14 (\bar{A})^{-0.09}$$

with the following result

Group:	1	2	3
\bar{A} thr:	54	164	597
$1.14(\bar{A})^{-0.09}$:	0.796	0.720	0.642
α^* :	0.815	0.699	0.647

As described on p. 22 for the total loss, we can use the observed values of the mean loss \bar{L} together with the estimated parameters α^* to obtain estimates for the truncation points T .

We get the following values:

Group:	1	2	3
$T^*(hkr)$	690	1700	4900
$x_T = \ln(T^* + 2)$	6.54	7.44	8.50

In the diagrams 4:1 — 4:3 these values of x_T are marked together with the upper and lower limits of $\ln A$, (A in hkr) in the different groups.

TABLE 5

Comparison between Pareto Model and observed loss distribution for building class 4.

Group:	1		2		3		Total	
Loss y_i hkr	$1-H(y_i)$	$1-F_i$	$1-H(y_i)$	$1-F_i$	$1-H(y_i)$	$1-F_i$	$1-H(y_i)$	$1-F_i$
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	.756	.800	.787	.830	.802	.871	.763	.807
6	.684	.730	.719	.762	.741	.810	.691	.736
7	.619	.670	.664	.705	.684	.750	.631	.677
10	.492	.540	.543	.575	.571	.616	.507	.550
20	.298	.321	.357	.366	.383	.415	.313	.330
30	.221	.228	.275	.267	.301	.312	.235	.236
50	.148	.145	.194	.173	.221	.214	.159	.152
100	.085	.085	.121	.101	.142	.136	.093	.089
200	.050	.050	.076	.070	.092	.086	.055	.054
300	.035	.033	.057	0.56	.071	.067	.040	.037
500	.023	.017	.040	.043	.051	.049	.027	.021
1000	.013	.002	.025	.026	.032	.032	.016	.006
2000	.008	—	.015	.004	.021	.021	.009	.0014
3000	.005	—	.011	—	.016	.010	.007	.0005
5000	.004	—	.008	—	.011	.004	.004	.0002
10000	.002	—	.005	—	.007	—	.002	—

REFERENCES

[1] ANDERSSON, H. An analysis of the development of the fire losses in the Northern countries after the Second World War. A.B. 6:1, pp. 25-30.
 [2] BENCKERT, L.- G. The lognormal model for the distribution of one claim. A.B. 2:1, pp. 2-23.
 [3] BENKTANDER, G. Schadenhäufigkeit und Risikoprämiensatz als Funktion der Grösse. Trans. XIX I.C.A., Oslo 1972, part 5, pp. 179-192.
 [4] DEPOID, P. Applications de la statistique aux assurances accidents et dommages. Berger-Levrault 1967.
 [5] FERRARA, G. Distributions des sinistres selon leur coût. A.B. 6:1, pp. 31-41.
 [6] FLACH, D.—STRAUSS, J. Analyse der Deutschen Feuer-Industriestatistik. BDGFV, IX:4, 1970, pp. 40-46.