STATISTICAL MEMOIRS
Institute of Mathematics
University of Oslo

No. 2
February 1969

RISIKO VED FORSIKRINGSVIRKSOMHET
NOEN VIDEREGÅENDE EMNER

av

Erling Sverdrup

(Bygger på forfatterens "Risiko ved forsikringsvirksomhet".
Opprinnelig utgave 1961. Korrigert utgave 1966).

INNHOLDSFORTEGNELSE

I. Sammensatt Poisson-prosess.

I.1. Definisjon av prosessen.

Det hender at en begivenhet som tilfeldig inntreffer fra tid til annen utløser en verdi av en variabel størrelse. Således kan en bedriftsulykke medføre et visst antall drepte. Et produksjonsuhell kan medføre tap av råstoff eller fast utstyr. En forsikret ulykke (brann, biluhell, dødsfall, osv.) medfører et krav overfor forsikringsselskapet, og erstatningens størrelse avhenger av ulykkens omfang og hvor høy forsikringssummen (risikosummen) på angjeldende polise er. Lignende situasjoner har man visstnok også i fysikken. Situasjoner av denne type kan ofte beskrives ved en "sammensatt Poisson-prosess". Vi skal i kapitel I behandle slike prosesser generelt, men vi kommer til å benytte uttrykkene krav og erstatning fra forsikringsvirksomheten.

For et vilkårlig intervall $(\tau, \tau+\Delta\tau)$ på tidsaksen la $\Lambda_1(\tau,\tau+\Delta\tau)$ være sannsynligheten for akkurat ett krav i intervallet, og la $\Lambda_2(\tau,\tau+\Delta\tau)$ være sannsynligheten for minst 2 krav i intervallet. Vi forutsetter at

$$\lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau} \Lambda_1(\tau,\tau+\Delta\tau) = \lambda_\tau$$
$$\lim_{\Delta\tau\to 0} \frac{1}{\Delta\tau} \Lambda_2(\tau,\tau+\Delta\tau) = 0 \tag{1}$$

Det vil si at vi gjør de vanlige forutsetninger ved Poisson-prosesser at sannsynligheten for ett krav i intervallet

$(\tau,\tau+\Delta\tau)$ er asymptotisk lik $\lambda_\tau \Delta\tau$, mens sannsynligheten for mer enn ett krav i intervallet er forsvinnende liten. Men vi forutsetter at kravsintensiteten λ_τ kan avhenge av τ, slik som det faktisk vil være i mange tilfeller. Vi antar at λ_τ er begrenset for $\tau \geq 0$.

Vi gjør også den vanlige forutsetning fra Poisson-punkt-prosesser om uavhengighet. La (τ_i,τ_i'); $i = 1,2,\ldots$ være vilkårlige ikke-overlappende tidsintervaller. Da skal sekvensen av begivenhetene

"krav i (τ_i,τ_i')"; $i = 1,2,\ldots$

være stokastisk uavhengige.

Når krav er oppstått på tidspunktet τ vil erstatningen $S(\tau)$ være en viss størrelse. (Den er vanligvis positiv, men kan i visse tilfeller - ved f.eks. livrenteforsikringer - være negativ.) Vi forutsetter at $S(\tau)$ har en kumulativ sannsynlighetsfunksjon F(y) som er uavhengig av τ. F kalles erstatningsfordelingen. F(y) er altså den betingede sannsynlighet for begivenheten $S(\tau) \leq y$ gitt at krav er oppstått i tidsintervallet $(\tau,\tau+\Delta\tau)$. Vi forutsetter at $S(\tau)$ er uavhengig av hva som er inntruffet tidligere med hensyn til krav.

Ovenstående definerer vår modell. Vi skal foreta en forenkling av modellen idet vi innfører effektiv tid t istedenfor kalendertid τ, ved transformasjonen

$$t = \int_0^\tau \lambda_v dv \tag{2}$$

Dette innebærer at vi lar "tiden løpe fortere" når det er stor kravintensitet. Hvis det i en bedrift som arbeider døgnet rundt er dobbelt så stor sjanse for bedriftsulykker om natten som om dagen, settes én natt-time lik to dagtimer. Det innebærer også at hvis bedriften står stille en tid, regnes denne tid ikke med. I et forsikringsselskap "går tiden fortere" når selskapet har mange risikable poliser.

Av (2) følger nå

$$\lambda_\tau d\tau = dt$$

Altså er sannsynligheten for krav i et effektivt tidsintervall $(t,t+dt)$ lik dt. Etter innføring av effektiv tid er kravintensiteten konstant lik 1. Vi setter nå $S(\tau) = Y(t)$, dvs. $Y(\int_0^\tau \lambda_v dv) = S(\tau)$. Det er åpenbart at den kumulative sannsynlighetsfunksjon for $Y(t)$ er $F(y)$.

For vilkårlig $t > 0$ og $\Delta t > 0$ har vi nå ifølge vår forutsetning om uavhengighet

$$\Pr[\text{Intet krav i } (0,t+\Delta t)] = \Pr[\text{Intet krav i } (0,t)] \cdot \Pr[\text{Intet krav i } (t,t+\Delta t)].$$

Setter vi nå $\Pr[\text{Intet krav i } (0,t)] = Q(t)$, har vi altså asymptotisk for små Δt, $Q(t+\Delta t) = Q(t)(1-\Delta t)$, dvs. $Q'(t) = -Q(t)$. Herav

$$Q(t) = e^{-t}$$

idet $Q(0) = 1$. La T være tidspunktet for første krav (etter 0-punktet). Da har vi

$$\Pr[t < T \leq t+dt] = \Pr[\text{Intet krav i } (0,t)] \cdot$$
$$\cdot \Pr[\text{Ett krav i } (t,t+dt)] = e^{-t} \cdot dt$$

e^{-t} er altså sannsynlighetstettheten (frekvensfunksjonen) for tidspunktet for første krav, og dermed er <u>e^{-t} også sannsynlighetstettheten for avstanden mellom på hverandre følgende krav</u>. Gjennomsnittlig tidsavstand mellom to på hverandre følgende krav blir $ET = \int_0^\infty te^{-t}dt = 1$. Vår tids-"enhet" er altså valgt lik forventet tidsavstand mellom kravene.

La nå $g_n(t)$ betegne sannsynlighetstettheten for tidspunktet for n-te krav. Vi har da

$$g_n(t) = \int_0^t g_{n-1}(t-v)e^{-v}dv \tag{4}$$

mens $g_1(t) = e^{-t}$. Herav finner vi lett ved induksjon

$$g_n(t) = \frac{t^{n-1}}{(n-1)!} e^{-t} \tag{5}$$

Hvis $b(n;t)$ betegner sannsynligheten for n krav i et tidsintervall av lengde t (f.eks. $(0,t)$), finner vi lett ved et enkelt sannsynlighetsteoretisk resonnement

$$b(n;t) = \int_0^t g_n(v)e^{-(t-v)}dv = \frac{t^n}{n!} e^{-t} \tag{6}$$

Her er $g_n(v)dv = \Pr[n\text{-te krav i } (v,v+dv)]$ og $e^{-(t-v)} = \Pr[\text{Intet krav i } (v,t)]$.

Ifølge (6) er altså <u>antall krav i et tidsintervall av lengde t Poisson-fordelt med parameter t</u>. Forventet antall krav i et tidsintervall av lengde t blir $\sum_{n=0}^{\infty} nb(n;t) = t$; hvilket også er en direkte følge av vårt valg av enhet.

Anta at vår "bestand" er heterogen slik at vi kan inndele den i R risikogrupper med kravintensiteter henholdsvis $\lambda_{\tau}^{(1)},\ldots,\lambda_{\tau}^{(R)}$, relativt til den opprinnelige kalendertid τ, og med erstatningsfordelinger henholdsvis $F_{(1)},\ldots,F_{(R)}$. Kravintensiteten λ_{τ} for hele bestanden finnes av

$$1-\lambda_{\tau}d\tau = \prod_{i=1}^{R}(1-\lambda_{\tau}^{(i)}d\tau)$$

og herav fåes

$$\lambda_{\tau} = \sum_{i=1}^{R}\lambda_{\tau}^{(i)}$$

Den betingede kumulative sannsynlighetsfunksjon for $Y(t)$ blir

$$F(y) = \frac{Pr(Krav \cap (Y \leq y))}{Pr(Krav)} = \frac{1}{\lambda_{\tau}}\sum_{i=1}^{R}\lambda_{\tau}^{(i)}F_{(i)}(y) \qquad (7)$$

som er uavhengig av τ såfremt <u>forholdet mellom kravintensitetene i de forskjellige risikogrupper er uavhengig av τ</u>. I så fall kan altså det heterogene tilfelle tilbakeføres til det homogene med en sammensatt Poisson-prosess som vi allerede har beskrevet.

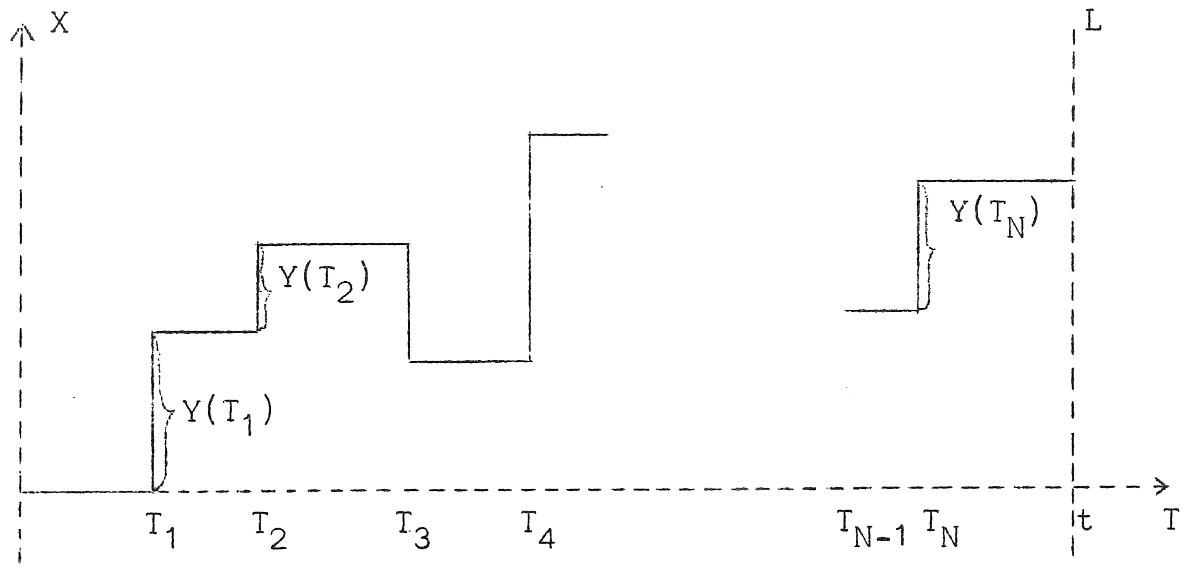
I.2. Samlet erstatning i et gitt intervall.

A. Den eksakte fordeling.

Betrakt igjen en observasjonsperiode av lengde t, f.eks. tidsintervallet $(0,t)$. La $T_1, T_2, \ldots, T_N$ være tidspunkter da krav oppstår. Samlet erstatningsutbetaling i tidsintervallet $(0,t)$ er

$$X(t) = \sum_{j=1}^{N} Y(T_j) \tag{1}$$

Vi bemerker at både antall ledd N og de enkelte ledd i (1) er stokastiske variable. Det grafiske billede av $X(t)$ vil være omtrent slik:



$X(t)$ er prosessens sampelfunksjon. Vi skal finne sannsynlighetsfordelingen for $X(t)$ for en gitt t, dvs. vi er interessert i å finne sannsynlighetsfordelingen for sampelfunksjonens skjæring med den loddrette linje L. Vi kaller den kumulative sannsynlighetsfunksjon $F(x,t)$,

$$F(x,t) = \Pr\left[X(t) \leq x\right]$$

Bemerk at sannsynlighetsfordelingen for N er gitt ved (6) og at følgelig sannsynligheten for at sampelfunksjonen skal ha et endelig antall "hopp" (diskontinuitetspunkter) i et intervall av lengde t er $\sum_{n=0}^{\infty} \frac{t^n}{n!} e^{-t} = 1$.

La oss først finne den betingede fordeling $F_n(x)$ for $X(t)$ gitt $N = n$. For gitt n er leddene i

$$X(t) = \sum_{j=1}^{n} Y(T_j)$$

uavhengige, ifølge forutsetningene. Vi har derfor ved et enkelt sannsynlighetsteoretisk resonnement*) (konvolusjonsformelen)

*) De som ikke kjenner til teorien for Lebesgue- eller Stieltjes-integral kan oppfatte $\int_a^b (\quad)dF$ på én av to måter:

1. Når F er absolutt kontinuerlig med tilhørende sannsynlighetstetthet f, sett $dF = fdx$.

2. Når F er diskret med positiv sannsynlighetsmasse for heltallige verdier, sett $\int_a^b (\quad)dF = \sum_{a<x\leq b} (\quad)f$, hvor f er den elementære sannsynlighetsfunksjon. Legg merke til at øvre grense b er <u>med</u>, mens nedre grense a <u>ikke</u> er med i summasjonen. Hvis spesielt X antar verdiene $0, \pm 1, \pm 2, \ldots$ og a og b er hele tall, er altså $\int_a^b (\quad)dF = \sum_{x=a+1}^{b} (\quad)f$.

$$\Pr\left[X(t) \leq x \mid N = n\right] = F_n(x) = \int_{-\infty}^{\infty} F_{n-1}(x-y)\,dF(y) \qquad (2)$$

Hvis vi innfører funksjonen $F_o(x)$ lik 0 når $x < 0$ og lik 1 når $x \geq 0$, samt setter $F = F_1$, gjelder (2) for $n = 1,2,\ldots$.

På den annen side har vi

$$\Pr(N = n) = \frac{t^n}{n!}e^{-t}$$

Herav

$$\Pr\left[(X(t) \leq x) \cap (N = n)\right] = \frac{t^n}{n!}e^{-t}F_n(x)$$

Altså finner vi den viktige formel <u>for den kumulative sannsynlighetsfunksjon for samlet erstatningsutbetaling i tidsrommet t</u>.

$$\Pr\left[X(t) \leq x\right] = F(x;t) = e^{-t}\sum_{n=0}^{\infty}\frac{t^n}{n!}F_n(x) \qquad (3)$$

Bemerk spesielt betydningen av det første leddet $e^{-t}F_o(x)$. Vi har også

$$1-F(x;t) = e^{-t}\sum_{n=0}^{\infty}\frac{t^n}{n!}(1-F_n(x)) \qquad (3)'$$

(3) og (3)' kan brukes til direkte beregning av $F(x,t)$ såfremt t ikke er for stor, idet i så fall leddene i (3) raskt blir neglisjable og vi bare trenger noen få ledd. Koeffisientene $e^{-t}\frac{t^n}{n!}$ kan enten beregnes rekursivt eller av Table 39 i <u>E.S. Pearson and H.O. Hartley</u>: Biometrika Tables. Med 6 desimalers nøyaktighet trenges høyst følgende antall ledd i (3):

t :	0.1	0.5	1.0	2.0	5.0	10.0	15.0
Antall ledd :	5	8	10	13	20	30	38

Det kan være at vi trenger et mindre antall ledd enn dette maksimum. Det avhenger av $F_n(x)$.

Eksempel: Erstatningen er kji-kvadrat-fordelt med 4 fr.gr., dvs. $dF = \frac{1}{2}\, x\, e^{-\frac{x}{2}}$, $t = 0.5$. Vi skal beregne $1-F(x,t)$ for $x = 7$. Vi ser at F_n er den kumulative kji-kvadrat fordeling, med $4n$ frihetsgrader. Vi finner ved bruk av Pearson and Hartley tabeller (Table 7 & 39):

n	$\frac{t^n}{n!}e^{-t}$	$1 - F_n$
0	0.606531	0.00000
1	0.303265	0.13589
2	0.075816	0.53663
3	0.012636	0.85761
4	0.001580	0.97326
5	0.000158	0.99669
6	0.000013	0.99971
7	0.000001	0.99998

Herav finner vi

$$1-F(x;t) = 0.09444.$$

<u>Forventning og varians</u> for de stokastiske variable N og $X(t)$ kan lett finnes uttrykt ved t og forventningen og variansen for enkelterstatningen Y. La momentene for enkelterstatningen Y være

$$p_\nu = \int_{-\infty}^{\infty} y^\nu \, dF \tag{4}$$

Vi har altså $EY = p_1$, $\operatorname{var} Y = p_2 - p_1^2$. Videre

$$EN = t, \quad \operatorname{var} N = t \tag{5}$$

Vi har nå for de betingede forventninger

$$E[X(t)|N] = NEY = Np_1$$

$$E[X(t)^2|N] = \operatorname{var}[X(t)|N] + \{E[X(t)|N]\}^2 =$$

$$= N \operatorname{var} Y + N^2 p_1^2 = Np_2 - Np_1^2 + N^2 p_1^2$$

Herav

$$EX(t) = EE[X(t)|N] = ENp_1 = p_1 t \tag{6}$$

$$EX(t)^2 = EE[X(t)^2|N] = (p_2 - p_1^2)EN + p_1^2 EN^2 =$$

$$= (p_2 - p_1^2)t + p_1^2(t + t^2) = p_2 t + p_1^2 t^2$$

Altså

$$\operatorname{var} X(t) = p_2 t \tag{7}$$

For mange formål vil vi også komme til å få bruk for den momentgenererende funksjon for $X(t)$.

La Z være en stokastisk variabel med kumulativ sannsynlighetsfunksjon $G(z)$. La videre s være en kompleks variabel. Da kalles

$$\mu(s) = \int_{-\infty}^{\infty} e^{zs} dG(z) = Ee^{Zs}$$

den momentgenererende funksjon for Z. Vi har åpenbart at $\mu(iv) = \varphi(v)$ er den karakteristiske funksjon for Z. Hvis μ eksisterer for reelle s rundt origo har vi også $EZ^{\nu} = \mu^{(\nu)}(0)$.

La nå den momentgenererende funksjon for enkelterstatningen Y være:

$$p(s) = Ee^{sY} = \int_{-\infty}^{\infty} e^{sy} dF(y) \tag{8}$$

Vi skal finne den momentgenererende funksjon $Ee^{sX(t)}$ for $X(t) = \sum_{j=1}^{N} Y_j(T_j)$. Siden leddene $Y_j(T_j)$ er uavhengige og identisk fordelte, har vi:

$$E\left[e^{sX(t)} \middle| N\right] = \left[Ee^{sY}\right]^N = p(s)^N$$

Altså

$$Ee^{sX(t)} = Ep(s)^N = \sum_{n=0}^{\infty} p(s)^n \frac{t^n}{n!} e^{-t}$$

dvs.

$$Ee^{sX(t)} = e^{t\left[p(s)-1\right]} \tag{9}$$

Herav kunne (6) og (7) vært utledet ved å derivere m.h.p. s og sette $s = 0$.

For kumulantene $\mathcal{K}_\nu$ for $X(t)$ (se I.4.A) finner vi:

$$\mathcal{K}_\nu(X(t)) = \frac{d^\nu}{ds^\nu} \log Ee^{sX(t)}\Big|_{s=0} = tp_\nu \tag{10}$$

mens de standardiserte kumulanter γ_ν er gitt ved

$$\gamma_\nu = \left(\frac{1}{\sqrt{tp_2}}\right)^\nu \mathcal{K}_\nu = \frac{p_\nu}{t^{\frac{\nu}{2}-1} p_2^{\frac{\nu}{2}}} \; ; \; \nu = 2,3,\ldots$$

B. Tilnærmelse ved Edgeworth rekke.

Den eksakte formel (3) viser seg i mange tilfeller tungvint i bruk og fører med store verdier av t til omfattende regninger.

Vi skal derfor utlede noen tilnærmelsesformler. For det første skal vi approksimere fordelingen til $X(t)$ med Edgeworth rekke. La g og G være den gaussiske tetthet og det gaussiske integral:

$$g(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} , \; G(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v} e^{-\frac{v^2}{2}} dy$$

og la $H(v)$; $v = 1,2,\ldots$ være de Hermite-ske polynomer (se I.4.B). Spesielt trenger vi her

$$\begin{aligned} H_2(v) &= v^2-1 \\ H_3(v) &= v^3-3v^2 = v(v^2-3) \\ H_5(v) &= v^5-10v^3+15 = v\left[v^2(v^2-10)+15\right] \end{aligned} \tag{11}$$

idet den siste form egner seg best ved numeriske regninger.

Vi finner nå lett Edgeworth rekke for $F(x;t)$ (se I.4.D (32) og (33)).

$$F(x;t) = G(v) - \frac{p_3}{6p_2^{\frac{3}{2}}}\left(\frac{1}{t}\right)^{\frac{1}{2}} G^{(3)}(v) + \frac{p_4}{24p_2^2 t} G^{(4)}(v) + \frac{p_3^2}{72p_2^3}\frac{1}{t} G^{(6)}(v) \tag{12}$$

eller

$$F(x;t) = G(v)-g(v)\left[\frac{p_3}{6p_2^{\frac{3}{2}}}\left(\frac{1}{t}\right)^{\frac{1}{2}}H_2(v) + \frac{p_4}{24p_2^2 t} H_3(v) + \frac{p_3^2}{72p_2^3}\frac{1}{t} H_5(v)\right] \tag{13}$$

hvor

$$v = \frac{x-tp_1}{\sqrt{tp_2}} \tag{14}$$

Tilnærmelsen blir bedre jo større t er. En kortfattet oversiktstabell over $G^{(n)}(x)$ for $n = 0,1,2,3,4,5,6,7$; $x = 0.0(0.1)4.0$ finnes i Harald Cramer: Mathematical methods of Statistics, Uppsala 1945, p.557. Ved nøyaktigere beregninger bør en benytte (13) idet en beregner H_ν av (11). For omfattende tabelleringer med samme p_2, p_3, p_4 og t vil det lønne seg å skrive ut hakeparentesen i (13) som et polynom og så omforme dette tilsvarende de siste former i (11), altså

$$a_0x^5+a_1x^3+a_2x^2+a_3x+a_4 = \left[\left[\left[a_0x^2+a_1\right]x+a_2\right]x+a_3\right]x+a_4$$

<u>Eksempel</u>: Hvis enkelterstatningen er kji-kvadratfordelt med ϱ frihetsgrader, har vi

$$p(s) = (1-2s)^{-\frac{\varrho}{2}} = \sum_{\nu=0}^{\infty} \varrho(\varrho+2)\ldots[\varrho+2(\nu-1)]\frac{s^\nu}{\nu!}$$

Herav

$$p_\nu = \varrho(\varrho+2)\ldots[\varrho+2(\nu-1)]$$

Spesielt for $\varrho = 4$, $t = 0.5$, $x = 7$, får vi

$$1-F = 1-G+g\cdot\left[0.384900H_2+0.277778H_3+0.074074H_5\right]$$

$$v = 1.443376, \quad v^2 = 2.083334$$

$$H_2 = 1.083334, \quad H_3 = -1.323093, \quad H_5 = -2.15505$$

Av Pearson and Hartley Table 1 (Lineær interpolasjon gir 5 riktige desimaler),

$$G(v) = 0.92554, \quad g(v) = 0.14077$$

Herav $1-F(x,t) = \underline{0.10401}$, mens det eksakte var 0.09444.

Anta nå at vi kan inndele vår bestand i R risikogrupper med konstante kravsintensiteter henholdsvis $\lambda_1, \lambda_2, \ldots, \lambda_R$ og med kumulative sannsynlighetsfunksjoner for enkelterstatningene henholdsvis $F_{(1)},\dots,F_{(R)}$. La de ν-te momenter for erstatningene være $p_{\nu 1},\dots,p_{\nu R}$. Som bemerket i I.1 representerer dette ikke noen ny situasjon sammenlignet med den vi allerede har behandlet. Det er imidlertid bekvemt å benytte følgende regel som lett avledes:

Sannsynlighetsfordelingen for den samlede erstatning i de R grupper over en tidsperiode på t er gitt ved (12), (13), (14) hvor p_ν erstattes med $\sum_{i=1}^{R} \lambda_i p_{\nu i}$; $\nu = 1,2,3,4$.

[Bevis: La $F(y)$ betegne den betingede kumulative sannsynlighetsfunksjon for $Y(t)$, gitt ved I.1.(7). Det ν-te ordens moment for $Y(t)$ er da

$$p_\nu = EY^\nu = \int y^\nu dF(y) = \int y^\nu \sum_{i=1}^{R} \frac{\lambda_i dF_{(i)}(y)}{\lambda} = \sum_{i=1}^{R} \frac{\lambda_i}{\lambda} p_{\nu i}$$

$$(\lambda = \sum_{i=1}^{R} \lambda_i) .$$

Etter (10) er ν-te ordens kumulant κ_ν lik tp_ν, forutsatt at kravsintensiteten er 1. I den aktuelle situasjon er imidlertid kravsintensiteten λ, og man må derfor innføre effektiv tid $\tau = \lambda t$. Det gir da

$$\kappa_\nu(Y(t)) = \tau \cdot p_\nu = t \cdot \sum_{i=1}^{R} \lambda_i p_{\nu i}$$

mens de standardiserte kumulanter blir

$$\gamma_\nu(Y(t)) = t^{-\frac{\nu}{2}+1} \frac{\sum \lambda_i p_{\nu i}}{\left[\sum \lambda_i p_{2i}\right]^{\frac{\nu}{2}}}$$]

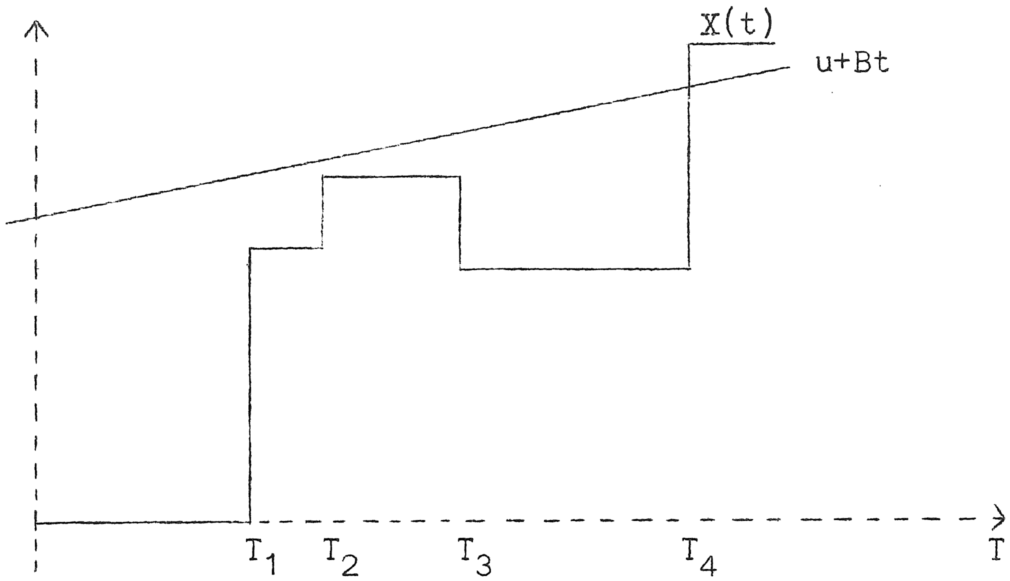
I.3. Ruinsannsynligheten.

A. Problemet.

La $X(t)$ være sampelfunksjonen for prosessen definert for alle verdier av $t > 0$ slik som beskrevet i I.2.A. Vi skal være interessert i å finne sannsynligheten for at

$$X(t) > u+Bt \tag{1}$$

for en eller annen $t > 0$, hvor u og B er gitte tall. Vi ønsker m.a.o. å finne sannsynligheten for at det grafiske billede av sampelfunksjonen skal skjære den rette linje $u+Bt$.



Vi skal betegne denne sannsynlighet med $\psi(u)$. Den avhenger naturligvis både av F og av B; men da u blir den eneste størrelse vi kommer til å variere i våre resonnementer, anfører vi bare u som argument. $\psi(u)$ er åpenbart en ikke-tiltagende funksjon av u.

$1-\psi(u)$ er altså sannsynligheten for at $X(t)$ i hele sitt forløp fra $t = 0$ til $t = \infty$ skal holde seg under den rette linje $u+Bt$.

Størrelsen av $\psi(u)$ kan være av interesse ved forskjellige slags praktiske anvendelser.

Således kan det tenkes at en produksjonsprosess normalt trenger en råstofftilgang på B_1 pr. tidsenhet. Fra tid til annen skjer imidlertid et uhell i produksjonen som ødelegger en viss del av råstoffene, slik at en ekstra råstofftilgang er nødvendig. Her er altså $F(0) = 0$. Hvis uhellene inntreffer på tidspunktene $T_1, T_2, \ldots$ og ekstratilgangen som trenges må være henholdsvis $Y(T_1), Y(T_2), \ldots,$ vil den råstofftilgang opptil tidspunktet t som trenges for å sikre kontinuerlig drift være

$$B_1 t + \sum_{j=1}^{N} Y(T_j) = B_1 t + X(t)$$

hvor T_N er den største T_j som er $\leq t$. Anta nå at produsenten på tidspunktet 0 har en råstoffreserve på u og at den <u>tilgjengelige</u> tilgang på råstoffer er B_2 pr. tidsenhet. Kontinuerlig drift er da sikret hvis

$$u+B_2 t \geq B_1 t + X(t)$$

dvs.

$$X(t) \leq u+(B_2-B_1)t = u+Bt$$

for alle $t > 0$. Denne begivenhet har altså sannsynlighet $1-\psi(u)$.

I et forsikringsselskap er selskapets erstatningsutbetalinger opptil tidspunktet t lik $X(t)$, mens på den annen side selskapet har en premieinntekt på B pr. tidsenhet og en sikkerhetsreserve på u på tidspunktet 0. Selskapet disponerer da på tidspunktet t et beløp

$$V(t) = u+Bt-X(t) \tag{2}$$

Hvis $V(t)$ blir negativ er selskapet ruinert og slåes konkurs. $\psi(u)$ er altså ruinsannsynligheten, dvs. sannsynlighet for at selskapet skal bli ruinert en eller annen gang i fremtiden (etter tidspunktet 0). Vi skal nedenfor anvende terminologien fra forsikringsvirksomheten og snakke om ruinsannsynligheten $\psi(u)$. $V(t)$ vil vi kalle risikoreserven og u sikkerhetsreserven. Situasjonen i forsikringsvirksomheten vil forøvrig bli grundigere utredet senere. Her skal vi bare beskjeftige oss med de matematiske problemer.

Vi skal også være interessert i sannsynligheten $\psi(u,\tau)$ for ruin en eller annen gang senest på tidspunktet τ $(0 < t \leq \tau)$. Det er da åpenbart at vi har

$$\psi(u) = \lim_{\tau \to \infty} \psi(u,\tau) \tag{3}$$

Det viser seg at det vil være bekvemt innledningsvis å se på følgende problem:

La $\Delta > 0$ og betrakt tidspunktene $\Delta, 2\Delta, 3\Delta, \ldots$. Disse tidspunktene vil vi kalle oppgjørs"dagene". Vi lar $\psi_\Delta(u)$ være sannsynligheten for ruin på en eller annen fremtidig oppgjørsdag, dvs. $\psi_\Delta(u)$ er sannsynligheten for at $X(\nu\Delta) > u+B\nu\Delta$ for minst én heltallig $\nu > 0$. Tilsvarende er $\psi_\Delta(u;\tau)$ sannsynligheten for ruin på en eller annen fremtidig oppgjørsdag senest på tidspunktet τ. I forsikringsvirksomheten ville ψ_Δ være den sannsynlighet som interesserer, forutsatt at selskapet eventuelt kunne holde kreditorene på avstand til neste oppgjørsdag i håp om da å ha opparbeidet et overskudd ved et gunstig forløp av forsikringene.

Vi har åpenbart

$$\psi_\Delta(u,\tau) \leq \psi(u,\tau) \qquad (4)$$

Når $\Delta \to 0$ vil oppgjørsdagene komme tettere og tettere på hverandre i tid og vi har

$$\begin{aligned} \lim_{\Delta \to 0} \psi_\Delta(u,\tau) &= \psi(u,\tau) \\ \lim_{\Delta \to 0} \psi_\Delta(u) &= \psi(u) \end{aligned} \qquad (5)$$

B. Ruin på en oppgjørsdag.

$\psi_\Delta(u;m\Delta)$ er definert som sannsynligheten for ruin senest på m'te oppgjørsdag etter tidspunktet $t = 0$. Denne begivenhet kaller vi S_o. La R være ruin på første oppgjørsdag. Da er altså S_o unionen av de to begivenheter

$S_o \cap R = R =$ Ruin allerede på første oppgjørsdag, dvs. for $t = \Delta$.

$S_o \cap \bar{R} =$ Ikke ruin for $t = \Delta$, men ruin inntreffer senest på $(m-1)$'te oppgjørsdag etter $t = \Delta$.

Vi har for det første:

$$\Pr(S_o \cap R) = \Pr(X(\Delta) > u+B\Delta) = 1-F(u+B\Delta;\Delta)$$

For å finne $\Pr(S_o \cap \bar{R})$ betrakter vi sannsynligheten for snittet av begivenhetene

$\bar{R}(x)$ = Totalerstatningen på tidspunktet $t = \Delta$ ligger i intervallet $(x,x+dx)$, hvor $x \leq u+B\Delta$.

S_Δ = Ruin senest på $(m-1)$'te oppgjørsdag etter $t = \Delta$.

Vi har nå*)

$$\Pr(\bar{R}(x)) = dF(x;\Delta), \quad \text{og}$$

$$\Pr(S_\Delta | \bar{R}(x)) = \psi_\Delta(u+B\Delta-x;(m-1)\Delta).$$

Sannsynligheten for snittet av $\bar{R}(x)$ og S_Δ blir derfor

$$\Pr(\bar{R}(x) \cap S_\Delta) = \psi_\Delta(u+B\Delta-x;(m-1)\Delta)dF(x;\Delta) \tag{7}$$

$\Pr(S_o \cap \bar{R})$ finnes ved å integrere (7) over x. Kombineres dette med

$$\psi_\Delta(u;m\Delta) = \Pr(S_o) = \Pr(S_o \cap R)+\Pr(S_o \cap \bar{R}),$$

får man

$$\psi_\Delta(u;m\Delta) = 1-F(u+B\Delta;\Delta)+\int_{-\infty}^{u+B\Delta}\psi_\Delta(u+B\Delta-x;(m-1)\Delta)dF(x;\Delta) \tag{8}$$

Lar vi $m \to \infty$, får vi følgende integralligning til bestemmelse av $\psi_\Delta(u)$:

*) Se fotnote side 7.

$$\psi_{\Delta}(u) = 1-F(u+B\Delta;\Delta) + \int_{-\infty}^{u+B\Delta} \psi_{\Delta}(u+B\Delta-x)dF(x;\Delta) \qquad (9)$$

Her er naturligvis

$$\psi_{\Delta}(u;\Delta) = 1-F(u+B\Delta;\Delta) \qquad (10)$$

Vi bemerker at den øvre grense $u+B\Delta$ er inkludert i integrasjonsintervallet.

C. En fundamental ulikhet om ruinsannsynligheten.

Forventet erstatningsutbetaling i et tidsintervall av lengde dt er gitt ved

$$dt\int ydF = p_1 dt,$$

mens den premie som faktisk innbetales i samme tidsrom er Bdt. Vi kan altså oppfatte p_1 og B som henholdsvis netto- og bruttopremien.

Vi skal bevise at hvis $B > p_1$, så er

$$\psi_{\Delta}(u;m\Delta) \leqq e^{-Ru} \qquad (11)$$

<u>såfremt R kan bestemmes som et reelt positivt tall som er slik at</u>

$$1+BR = p(R) = \int_{-\infty}^{+\infty} e^{Ry}dF \qquad (12)$$

$R > 0$ er altså en reell løsning av <u>prosessens tilordnede</u>

<u>algebraiske ligning</u>

$$1+Ba = p(a) \tag{13}$$

Røttene i (13) kaller vi prosessens <u>egenverdier</u>.

Vi skal først undersøke eksistensen av en reell positiv løsning av (13). Vi forutsetter $F(0) < 1$. Siden F er en kumulativ sannsynlighetsfordeling, er den kontinuerlig fra høyre, og da har vi at det eksisterer et tall $\varepsilon > 0$, slik at $F(\varepsilon) < 1$.

Vi har åpenbart for alle $t > 0$, såfremt $p(a)$ eksisterer for en $a > 0$:

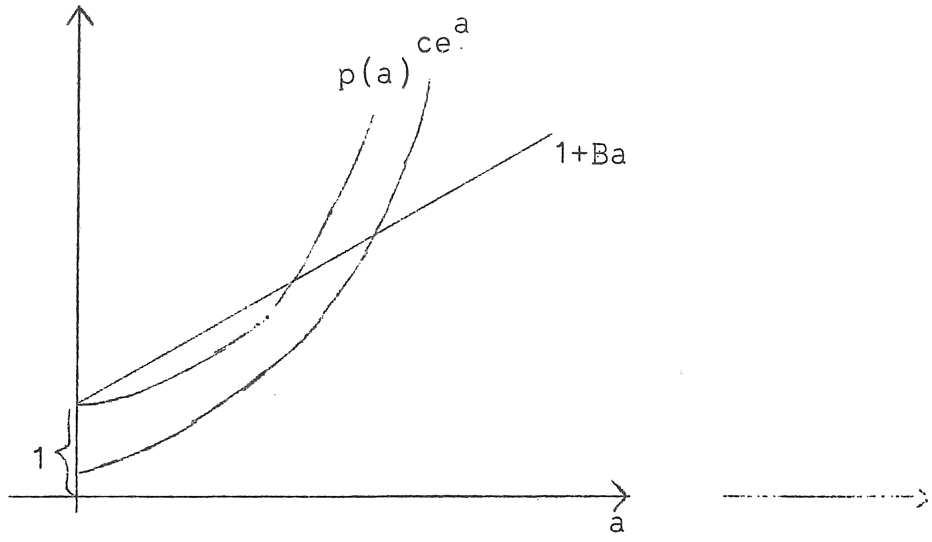
$$p(a) = \int_{-\infty}^{+\infty} e^{ay} dF \geqq \int_{t}^{\infty} e^{ay} dF \geqq e^{at} \int_{t}^{\infty} dF = \left[1-F(t)\right] e^{at} \tag{14}$$

og herav

$$p(a) \geqq ce^{a\varepsilon}$$

hvor $0 < c \leqq 1$. (Nedre grense i integralene er ikke regnet med ved integrering.) Dessuten har vi

$$p'(0) = p_1 < B$$



Vi ser av det grafiske billede at hvis $p(a)$ eksisterer for alle a, har (13) en positiv reell rot. Det samme gjelder hvis $p(a)$ eksisterer for tilstrekkelig stor a, dvs. en a slik at $p(a) > 1+Ba$.

Setter vi nå i I.2(9) $t = \Delta$ og $s = R$ og benytter (12), får vi

$$E\left[e^{X(\Delta)R}\right] = e^{\Delta(p(R)-1)} = e^{BR\Delta}$$

eller

$$E\left[e^{(X(\Delta)-B\Delta-u)R}\right] = e^{-uR} \tag{15}$$

Etter en enkel omforming av (8) finner vi

$$\begin{aligned} e^{-Ru} - \psi_{\Delta}(u;m\Delta) &= e^{-Ru} - \psi_{\Delta}(u;\Delta) - \int_{-\infty}^{u+B\Delta} e^{(x-B\Delta-u)R} dF(x;\Delta) + \\ &+ \int_{-\infty}^{u+B\Delta} \left[e^{(x-B\Delta-u)R} - \psi_{\Delta}(u+B\Delta-x;(m-1)\Delta) \right] dF(x;\Delta) \end{aligned} \tag{16}$$

Vi betrakter de tre første ledd på høyre side av (16):

$$K(u) = e^{-Ru} - \psi_{\Delta}(u;\Delta) - \int_{-\infty}^{u+B\Delta} e^{+(x-B\Delta-u)R} dF(x;\Delta)$$

Innsettes her (15) og trekkes første og siste ledd sammen, fåes

$$K(u) = \int_{u+B\Delta}^{\infty} e^{(x-B\Delta-u)R} dF(x;\Delta) - \psi_{\Delta}(u;\Delta)$$

(hvor nedre grense ikke er regnet med ved integreringen). Vi ser at i integrasjonsintervallet er $x > B\Delta+u$. Da dessuten $R > 0$, får vi

$$K(u) \geq \int_{u+B\Delta}^{\infty} dF(x;\Delta) - \psi_{\Delta}(u;\Delta) = 1-F(u+B\Delta,\Delta) - \psi_{\Delta}(u;\Delta) = 0$$

Betrakter vi nå definisjonsligningen for $K(u)$, ser vi at $K(u) \geq 0$ medfører

$$\psi_{\Delta}(u;\Delta) \leq e^{-Ru}$$

Ligning (11) er dermed bevist for $m = 1$. Av $K(u) \geq 0$ og ligning (16) følger dessuten at

$$e^{-Ru} - \psi_{\Delta}(u;m\Delta) \geq \int_{-\infty}^{u+B\Delta} \left[e^{-(u+B\Delta-x)R} - \psi_{\Delta}(u+B\Delta-x),(m-1)\Delta) \right] dF(x;\Delta)$$

Hvis (11) gjelder for $m-1$, gjelder den derfor også for m. Dermed er (11) bevist ved induksjon.

Av (11) følger ved å la $m \rightarrow \infty$

$$\psi_{\Delta}(u) \leq e^{-Ru} \tag{17}$$

Ved å la $\Delta \rightarrow 0$ i (11) og (17) følger

$$\psi(u;\tau) \leq e^{-Ru}, \quad \psi(u) \leq e^{-Ru}$$

Vi ser også at

$$\lim_{u\to\infty} \psi(u;\tau) = \lim_{u\to\infty} \psi(u) = 0$$

Eks.1. Anta at hvis krav opptrer vil det alltid være av størrelse 1. $X(t)$ angir da rett og slett antall krav i tidsintervallet av lengde t og er derfor Poissonfordelt med parameter t, dvs. $F(x;t) = \sum_{n\leq x} \frac{t^n}{n!} e^{-t}$, som det også fremgår av den generelle formel.

Vi har $F(y) = 0$ når $y < 1$ og $F(y) = 1$ når $y \geq 1$. R bestemmes av

$$1+BR = \int e^{Ry} dF = e^R$$

Vi har altså

$$\psi(u) \leq e^{-Ru} \quad \text{hvor } R \text{ er bestemt av } 1+BR = e^R.$$

Hvis spesielt $B = 2$ (dvs. bruttopremien det dobbelte av nettopremien $= p_1 = 1$) får vi

$$\psi(u) \leq e^{-1.257u}.$$

Eks.2. La $dF(y) = e^{-y}dy$ for $y \geq 0$. Vi finner $p_1 = 1$. Vi har til bestemmelse av $R < 1$,

$$1+BR = \int_{-\infty}^{+\infty} e^{yR} dF = \int_0^{\infty} e^{-(1-R)y} dy = \frac{1}{1-R}$$

Herav $R = \frac{B-1}{B}$. Da $B > 1$ er åpenbart $R < 1$. Vi finner altså

$$\psi(u) \leq e^{-\frac{B-1}{B}u}$$

Spesielt for $B = 2$

$$\psi(u) \leq e^{-0.5u}$$

dvs. en betydelig høyere skranke enn i eks.1.

D. En integralligning for ruinsannsynligheten.

Vi skal se nærmere på bestemmelsen av $\psi(u)$, dvs. sannsynligheten på tidspunktet 0, hvor sikkerhetsreserven er u, for ruin en eller annen gang i fremtiden. Vi antar $B > p_1$. Hvis $u < 0$ er $\psi(u) = 1$. Vi ser derfor på tilfellet $u \geq 0$.

Vi betrakter hva som skjer i tidsintervallet 0 til dt. Det er åpenbart at enten oppstår det i dette intervall intet krav, eller det oppstår et krav som ikke er stort nok til å ruinere selskapet, eller det oppstår et krav som umiddelbart ruinerer selskapet. Kjennetegnet S = "Ruin for en eller annen $t > 0$" er derfor en disjunksjon av tre kjennetegn

S_1 = Intet krav oppstår i intervallet 0 til dt, men ruin inntrer for en $t > dt$.

S_2 = Et krav som ikke er stort nok til å ruinere oppstår i 0 til dt, men ruin inntrer for en $t > dt$.

S_3 = Krav som er stort nok til å ruinere selskapet oppstår i intervallet 0 til dt.

Vi finner herav, idet de tre ledd på høyre side er sannsynlighetene for henholdsvis S_1, S_2 og S_3:

$$\psi(u) = (1-dt)\psi(u+Bdt)+dt \int_{-\infty}^{u+Bdt} \psi\,(u+Bdt-y)dF(y)+ \\ +dt\left[1-F(u+Bdt)\right]+o(dt) \tag{18}$$

Annet ledd = $Pr(S_2)$ er oppstått slik: Sannsynligheten for krav med en erstatning mellom y og y+dy er dtdF(y). Hvis kravet er y, er risikoreserven på tidspunktet dt lik u+Bdt-y, idet inntekten og utgiften i intervallet (0,dt) har vært henholdsvis Bdt og y. Den betingede sannsynlighet for ruin gitt at erstatningen i intervallet 0 til dt er mellom y og y+dy er da $\psi(u+Bdt-y)$. Sannsynligheten for S_2 med en erstatning mellom y og y+dy er derfor

$$dtdF(y)\psi(u+Bdt-y)$$

hvor $y \leq u+Bdt$, siden ruin ikke inntrer umiddelbart ifølge S_2. Sannsynligheten for S_2 alene, får vi nå ved å integrere over alle y fra $y = -\infty$ til u+Bdt.

Vi antar nå $B \neq 0$ og setter Bdt = du. Etter en liten omforming blir (18)

$$B\left[\psi(u+du)-\psi(u)\right] = \psi(u+du)du\cdot\left[1-F(u+du)\right]du - du\int_{-\infty}^{u+du}\psi(u+du-y)dF(y) \qquad (19)$$

Summeres her over ekvidistante verdier av u fra v til $v+K$ med ekvidistans du, og lar man så du gå mot 0, fremkommer

$$B\int_v^{v+K} d\psi(u) = B\left[\psi(v+K)-\psi(v)\right] = \int_v^{v+K}\psi(u)du - \int_v^{v+K}\left[1-F(u)\right]du - \int_v^{v+K}\int_{-\infty}^{u}\psi(u-y)dF(y)du \qquad (20)$$

(idet venstre side ikke generelt kan settes lik $B\int_v^{v+K}\psi'(u)du$.

Det er forøvrig fristende å dividere (19) med du og la $du \to 0$. Da fremkommer en differensial-integralligning for $\psi(u)$. Men bemerk at denne grenseovergang fører til vanskeligheter idet den deriverte ikke er entydig gitt over alt. Se eksempel nedenfor.)

Hvis $B = 0$, dividerer vi i (18) med dt, integrerer fra v til $v+K$, og (20) fremkommer med $B = 0$. (20) gjelder derfor for alle B.

I (20) lar vi nå $K \to \infty$, benytter at $\psi(u) \leq e^{-Ru}$ (og skifter betegnelsene u og v). Vi får da

$$B\psi(u) = \int_u^{\infty}\left[1-F(v)\right]dv + \int_u^{\infty}\int_{-\infty}^{v}\psi(v-y)dF(y)dv - \int_u^{\infty}\psi(v)dv \qquad (21)$$

Vi tar for oss det nest siste ledd i (21) og betegner det med J. Ved betraktning av fig.1 ser vi at

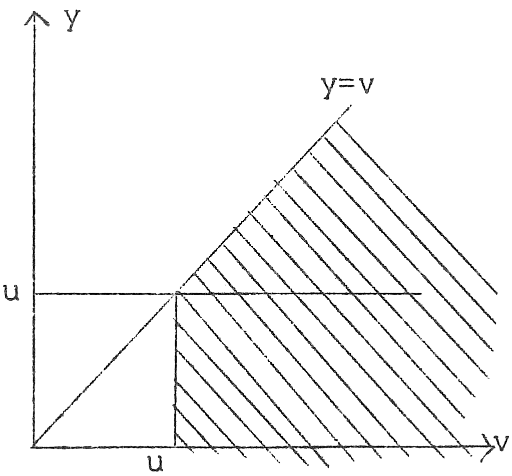


Fig.1.

$$J = \int_{-\infty}^{u}\int_{u}^{\infty}\psi(v-y)dvdF(y) + \int_{u}^{\infty}\int_{y}^{\infty}\psi(v-y)dvdF(y),$$

(idet integrasjonen foregår over det skraverte område. Merk at øvre grense i integrasjonsintervallene er inkludert, nedre grense ikke). Vi innfører $w = v-y$ som ny integrasjonsvariabel i stedet for v, og ombytter igjen integrasjonsrekkefølgen (fig.2).

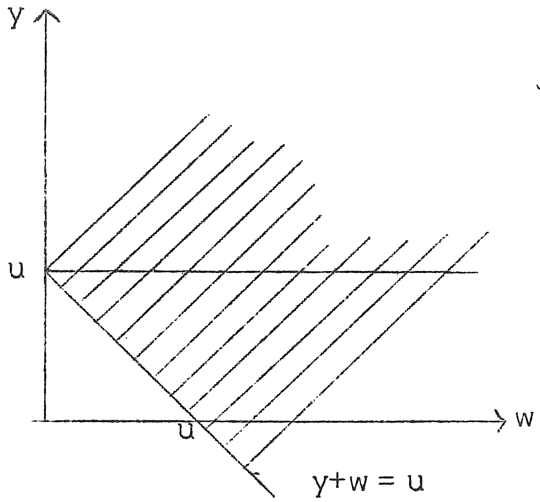


Fig.2.

$$J = \int_{-\infty}^{u}\int_{u-y}^{\infty}\psi(w)dwdF(y) + \int_{u}^{\infty}\int_{o}^{\infty}\psi(w)dwdF(y)$$

$$= \int_{o}^{\infty}\int_{-\infty}^{+\infty}\psi(w)dF(y)dw - \int_{o}^{\infty}\int_{-\infty}^{u-w}\psi(w)dF(y)dw$$

$$= \int_{o}^{\infty}\psi(w)dw - \int_{o}^{\infty}\psi(w)F(u-w)dw$$

$$= \int_{o}^{\infty}\psi(w)\left[1-F(u-w)\right]dw$$

Altså blir ligning (21)

$$B\psi(u) = \int_{u}^{\infty}\left[1-F(v)\right]dv + \int_{o}^{\infty}\psi(v)\left[F_o(u-v)-F(u-v)\right]dv \tag{22}$$

som gjelder for alle $u \geq 0$. Av (19) følger at <u>for $B \neq 0$ er</u>

$\psi(u)$ kontinuerlig for $u > 0$. Derimot er $u = 0$, som vi skal se senere, vanligvis et diskontinuitetspunkt.

Vi skal se på spesialtilfellet med bare positive erstatninger. Det er utvilsomt det tilfelle som er av størst betydning i praksis. (Men det utelukker altså situasjoner i livrenteinstitusjoner hvor premiereserve frigis til fordel for selskapet ved død.)

Vi antar altså at $F(0) = 0$. Det medfører at $F(y)$ er kontinuerlig for $y = 0$. Videre er $B > p_1 > 0$.

Integralligningen (22) kan da skrives

$$B\psi(u) = \int_u^\infty [1-F(v)]dv + \int_0^u \psi(u-v)\,[1-F(v)]dv \tag{23}$$

Setter vi her $u = 0$, får vi

$$B\psi(0) = \int_0^\infty [1-F(v)]dv$$

Vi finner ved delvis integrasjon og bruk av (14) (under den forutsetning at $p(a) < \infty$ for en $a > 0$)

$$\int_0^\infty [1-F(v)]dv = \int_0^\infty v dF(v) = p_1 \tag{24}$$

Herav den viktige formel

$$\underline{\psi(0) = \frac{p_1}{B}} \tag{25}$$

Vi ser at siden $\psi(u) = 1$ for $u < 0$, vil $\psi(u)$ være diskontinuerlig i origo når $B \neq p_1$.

Eksempel 1: Vi skal se på det spesialtilfellet hvor den sammensatte Poissonprosess reduserer seg til en enkel Poissonprosess, dvs. det tilfelle hvor den eventuelle erstatning alltid er 1. Vi har da

$$F(y) = 0 \quad \text{når} \quad y < 1, \quad F(1) = 1.$$

Vi finner av (22)

$$B\psi(u) = 1-u+\int_0^u \psi(v)dv, \quad \text{når} \quad u \leq 1 \tag{26}$$

$$B\psi(u) = \int_{u-1}^u \psi(v)dv, \quad \text{når} \quad u \geq 1 \tag{27}$$

Da følger av (26) at den deriverte $\psi'(u)$ eksisterer for $0 \leq u < 1$, og vi har derfor i dette intervall

$$B\psi'(u) = -1+\psi(u)$$

På den annen side har vi direkte av (26) ved å sette $u = 0$ i (26)

$$\psi(0) = \frac{1}{B} \quad \text{hvor} \quad B > p_1 = 1.$$

Herav finner vi ved løsning av differensialligningen

$$\psi(u) = 1-(1-\frac{1}{B})e^{\frac{u}{B}} \quad 0 \leq u \leq 1. \tag{28}$$

Da $\psi(u)$ er kontinuerlig for $u > 0$, følger av (27) at $\psi(u)$ er deriverbar for $u > 1$. Vi har derfor for $u > 1$

$$B\psi'(u) = \psi(u) - \psi(u-1)$$

Anta nå $1 < u < 2$. Dette medfører $u-1 < 1$. Derfor er $\psi(u-1)$ gitt ved (28), og vi får differensialligningen

$$B\psi'(u) = \psi(u) - 1 + (1-\tfrac{1}{B})e^{\frac{u-1}{B}}$$

Løses denne differensialligning med initialbetingelsen (ifølge (28))

$$\psi(1) = 1-(1-\tfrac{1}{B})e^{\frac{1}{B}}$$

fåes

$$\underline{\psi(u) = 1 + \tfrac{1}{B}(1-\tfrac{1}{B})(u-1)e^{\frac{u-1}{B}} - (1-\tfrac{1}{B})e^{\frac{u}{B}}} \qquad 1 \leq u \leq 2 \qquad (29)$$

Ved derivasjon av $\psi(u)$ ser vi at vi får forskjellig resultat for $\psi'(1)$ etter som man bruker (28) eller (29). Det viser at ligningen (19), oppfattet som en differensialligning, generelt ikke kan være riktig.

Vi finner forøvrig

$$\psi(u) = 1-(1-\tfrac{1}{B}) \sum_{j=0}^{\nu-1} \frac{1}{j!} \left(\frac{j-u}{B}\right)^j e^{\frac{u-j}{B}} \quad \text{for } \nu-1 \leq u \leq \nu \qquad (30)$$

Eksempel 2: La F være gitt ved

$$dF = e^{-y}dy \quad \text{for } y \geq 0, \quad dF = 0 \quad \text{for } y < 0$$

dvs. $F(y) = 1-e^{-y}$. (23) blir da:

$$B\psi(u) = \int_u^{\infty} e^{-v}dv + \int_0^u \psi(u-v)e^{-v}dv = e^{-u}+e^{-u}\int_0^u \psi(t)e^{t}dt$$

(idet vi har innført $t = u-v$). Setter $\chi(u) = e^u\psi(u)$, som gir

$$B\chi(u) = 1 + \int_0^u \chi(t)dt$$

Herav $B\chi'(u) = \chi(u)$.

Denne differensialligning har løsningen $\chi(u) = Ce^{\frac{u}{B}}$, som gir

$$\psi(u) = Ce^{-\frac{B-1}{B}u}$$

Konstanten C bestemmes av (25), idet

$$\psi(0) = \frac{p_1}{B} = \frac{1}{B}\int_0^{\infty} ydF = \frac{1}{B}$$

Altså

$$\underline{\psi(u) = \frac{1}{B}e^{-\frac{B-1}{B}u}}$$

Vi vender tilbake til den generelle situasjon med $F(0) = 0$.

Anta at vi vil beregne ψ for ekvidistante verdier av u. Etter eventuell skalatransformasjon kan vi la dem være $u = 0,1,2,\ldots$. Vi setter da tilnærmet, idet vi anvender trapesregelen på siste integral i (23)

$$B\psi(u) = \int_u^{\infty}\left[1-F(v)\right]dv + \frac{1}{2}\left[\psi(u)+2\sum_{v=1}^{u-1}\psi(u-v)\left[1-F(v)\right]+\psi(0)(1-F(u))\right] \tag{31}$$

eller ved (25)

$$(B-\tfrac{1}{2})\psi(u) = \int_u^\infty \left[1-F(v)\right] dv + \frac{p_1(1-F(u))}{2B} + \sum_{v=1}^{u-1} \psi(u-v)\left[1-F(v)\right] \tag{32}$$

for $u = 1,2,\ldots,$ idet summen faller bort for $u = 1$. Vi finner altså til suksessiv beregning av $\psi(1), \psi(2), \ldots$

$$(B-\tfrac{1}{2})\psi(1) = \int_1^\infty \left[1-F(v)\right] dv + \frac{p_1(1-F(1))}{2B}$$

$$(B-\tfrac{1}{2})\psi(2) = \int_2^\infty \left[1-F(v)\right] dv + \frac{p_1(1-F(2))}{2B} + \psi(1)\left[1-F(1)\right]$$

osv.

I enkelte tilfelle kan det, som vist i eksempel 2, angis en eksakt eksplisit løsning av (23). I eksemplet var $dF = e^{-y}dy$. La mer generelt

$$dF = \sum_{n=1}^{N} h_n \beta_n e^{-\beta_n y} dy \tag{33}$$

hvor $\beta_1 < \beta_2 < \ldots \beta_N$; $h_n > 0$ og $\sum h_n = 1$.

Da har vi

$$\psi(u) = \sum_1^N \frac{B-p_1}{p'(R_n)-B} e^{-R_n u} \tag{34}$$

hvor $R_1, R_2, \ldots, R_N$ er røttene i ligningen

$$1+Bs-\sum_{1}^{N}\frac{h_n\beta_n}{\beta_n-s}=0 \qquad (35)$$

og

$$p'(R)=\sum_{1}^{\infty}\frac{h_\nu\beta_\nu}{(\beta_\nu-R)^2},\quad p_1=\sum_{1}^{\infty}\frac{h_\nu}{\beta_\nu} \qquad (36)$$

Bemerk at (35) kan skrives $P_N(s) = 0$, hvor P_N er et N-te grads polynom. Det kan bevises at denne ligning har N reelle positive røtter (når $B > p_1$).

En nærliggende generell fremgangsmåte for numerisk løsning av (23) med en vilkårlig F, er nå åpenbart å tilnærme F med et uttrykk av formen (33), såfremt dette gir forsvarlig approksimasjon. Man bestemmer da først alle β_n og h_n best mulig, og lar $\psi(u)$ være bestemt ved (34).

I.4. Matematisk appendiks.

A. Kumulantene.

I dette appendiks vil vi for flere formål trenge relasjoner mellom deriverte og logaritmisk deriverte av funksjoner. La $w(x) = \log v(x)$ idet $v(x) > 0$. Ved ν gangers derivasjon av $v'(x) = w'(x)v(x)$, finner vi følgende rekursjonsformel:

$$v^{(\nu+1)}(x)=\sum_{j=0}^{\nu}\binom{\nu}{j}w^{(j+1)}(x)v^{(\nu-j)}(x) \qquad (1)$$

Herav kan vi suksessivt uttrykke de deriverte ved de logaritmisk deriverte og vice versa. Vi finner $v' = w'v = w'e^w$,

$v'' = w'v'+w''v = (w'^2+w'')e^w$, $v^{(3)} = w'v''+2w''v'+w^{(3)}v = \left[w'^3+3w'w''+w^{(3)}\right]e^w$, osv. Omvendt $w' = \frac{1}{v}v'$, $w'' = \frac{1}{v}\left[v''-w'v'\right] = \frac{v''}{v} - \frac{v'^2}{v^2}$, $w^{(3)} = \frac{1}{v}\left[v^{(3)}-w'v''-2w''v'\right] = \frac{v^{(3)}}{v} - 3\,\frac{v'v''}{v^2} + 2\,\frac{v'^3}{v^3}$, osv.

Omfattende formler over sammenhengen mellom deriverte og logaritmisk deriverte finnes i Maurice G. Kendall: The Advanced Theory of Statistics. London 1945, p.62-63, formel 3.29 og 3.33. Kendalls κ_ν svarer til nærværende $w^{(\nu)}$ og Kendalls μ'_ν svarer til nærværende $\frac{v^{(\nu)}}{v}$.

La X være en stokastisk variabel med kumulativ sannsynlighetsfunksjon $F(x)$. Vi kaller

$$p(s) = \int_{-\infty}^{+\infty} e^{xs}dF(x) = Ee^{Xs} \tag{2}$$

den momentgenererende funksjon (mgf.) for den variable X. Vi vet at $p(s)$ eksisterer iallfall for s rent imaginær, dvs. $s = it$, og at $p(it) = \varphi(t)$ er den karakteristiske funksjon som definerer F entydig.

Såfremt det ν-te moment p_ν eksisterer, har vi

$$p_\nu = EX^\nu = p^{(\nu)}(0) \tag{3}$$

(Om nødvendig må derivasjonen foretas i rent imaginær retning, dvs. $p^{(\nu)}(0) = \frac{d^\nu}{d(it)^\nu}\, p(it)\Big|_{t=0}$.)

Vi kaller størrelsen

$$\kappa_\nu = \kappa_\nu(X) = \frac{d^\nu}{ds^\nu} \log p(s)\Big|_{s=0} \tag{4}$$

ν-te kumulant for X. Sammenhengen mellom momenter og kumulanter er gitt ved (1) med $w^{(0)} = 0$, $w^{(\nu)} = \kappa_\nu$, $v^{(0)} = 1$, $v^{(\nu)} = p_\nu$.

Vi skal finne kumulantene for $aX+b$ uttrykt ved kumulantene for X. Den mgf. for $aX+b$ er

$$\bar{p}(s) = Ee^{(aX+b)s} = e^{bs}p(as)$$

dvs.

$$\log \bar{p}(s) = bs + \log p(as)$$

Ved å derivere og sette $s = 0$, får vi

$$\kappa_1(aX+b) = a\kappa_1(X)+b,\ \kappa_\nu(aX+b) = a^\nu \kappa_\nu(X);\ \nu \geqq 2. \tag{5}$$

Setter vi spesielt $a = 1$, $b = -\kappa_1 = -EX$, ser vi at kumulantene for $X-EX$ er $\kappa_1(X-EX) = 0$, $\kappa_\nu(X-EX) = \kappa_\nu$ for $\nu \geqq 2$. Momentene for $X-EX$ er sentralmomentene

$$\lambda_\nu = E(X-EX)^\nu \quad ;\ \nu = 2,3,\ldots \tag{6}$$

mens $\lambda_1(X) = EX$ og $\lambda_1(X-EX) = 0$.

For sammenhengen mellom kumulanter og momenter for $X-EX$ finner vi av (1), idet vi benytter at $\lambda_1(X-EX) = \kappa_1(X-EX) = 0$ og $\lambda_0 = 1$

$$\lambda_{\nu+1} = \sum_{j=1}^{\nu-2} \binom{\nu}{j}\kappa_{j+1}\lambda_{\nu-j} + \kappa_{\nu+1} \tag{7}$$

hvor summen i (7) bortfaller når $\nu < 3$.

Herav spesielt

$$\begin{aligned} &\varkappa_1 = EX, \ \varkappa_2 = \lambda_2, \ \varkappa_3 = \lambda_3, \ \varkappa_4 = \lambda_4 - 3\lambda_2^2, \\ &\varkappa_5 = \lambda_5 - 10\lambda_2\lambda_3, \ \varkappa_6 = \lambda_6 - 15\lambda_2\lambda_4 - 10\lambda_3^2 + 30\lambda_2^3 \end{aligned} \tag{8}$$

Det er altså en en-entydig sammenheng mellom kumulanter og (sentral-)momenter. Forsåvidt synes det likegyldig enten vi benytter momenter eller kumulanter til å karakterisere sannsynlighetsfordelinger. Når kumulantene likevel av og til foretrekkes er det fordi de har noen bekvemme matematiske egenskaper. Først og fremst er de <u>additive</u> på samme måte som variansen er det. Vi vet at såfremt $X_1, X_2, \ldots, X_n$ er <u>uavhengige</u>, er $\operatorname{var} \sum X_j = \sum \operatorname{var} X_j$. Vi har tilsvarende for enhver ν

$$\varkappa_\nu\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \varkappa_\nu(X_j) \tag{9}$$

som er <u>fundamentalegenskapen ved kumulanter</u>.

Vi skal bevise (9) og antar at de uavhengige stokastiske variable $X_1, X_2, \ldots, X_n$ har mgf. henholdsvis $p_1(s), \ldots, p_n(s)$. Mgf. for $X = \sum X_j$ betegner vi med $p(s)$. Vi finner

$$p(s) = Ee^{Xs} = Ee^{s\sum X_j} = \prod_{j=1}^{n} Ee^{X_j s} = p_1(s) \ldots p_n(s)$$

Altså

$$\log p(s) = \sum_{j=1}^{n} \log p_j(s)$$

Herav følger (9) ved å derivere ν ganger og sette $s = 0$.
Bemerk at fordi $\varkappa_\nu = \lambda_\nu$ for $\nu = 1,2,3$, har vi

$$\lambda_\nu(\Sigma X_j) = \sum_j \lambda_\nu(X_j) \quad \text{for} \quad \nu = 1,2,3,$$

mens dette ikke gjelder generelt for $\nu = 4,5,\ldots$. Det er derfor først og fremst når høyere momenter enn tredje moment er involvert at det kan være en fordel å arbeide med kumulantene istedenfor sentralmomentene. Skal vi således finne $\lambda_4(\Sigma X_j)$, kan man først finne $\varkappa_4(X_j)$ og var X_j, og dernest $\lambda_4(\Sigma X_j) = \varkappa_4(\Sigma X_j)+3(\lambda_2(\Sigma X_j))^2 = \Sigma\varkappa_4(X_j)+3(\sum_j \text{var } X_j)^2$.

Eks.1. Hvis X er normal (ξ,σ), finner vi for mgf. $p(s) = e^{\xi s+\frac{\sigma^2}{2}s^2}$ og herav finner vi direkte av (4)

$$\varkappa_1 = \xi, \quad \varkappa_2 = \sigma^2, \quad \varkappa_\nu = 0; \quad \nu > 2.$$

Eks.2. Hvis X er kji-kvadratfordelt med ρ frihetsgrader, finner vi direkte av (4)

$$\varkappa_\nu = \rho(\nu-1)!2^{\nu-1}$$

Spesielt $\varkappa_1 = \rho$, $\varkappa_2 = 2\rho$, $\varkappa_3 = 8\rho$, $\varkappa_4 = 48\rho$, $\varkappa_5 = 384\rho$, $\varkappa_6 = 3840\rho$. Herav for sentralmomentene ved (8)
$\lambda_1 = \rho$, $\lambda_2 = 2\rho$, $\lambda_3 = 8\rho$, $\lambda_4 = 48\rho+12\rho^2$, $\lambda_5 = 384\rho+160\rho^2$.

Eks.3. Hvis X er Poissonfordelt med parameter λ finner vi direkte av (4)

$$\varkappa_\nu = \lambda$$

Herav $\lambda_1 = \lambda_2 = \lambda_3 = \lambda, \lambda_4 = \lambda + 3\lambda^2, \lambda_5 = \lambda + 10\lambda^2,$
$\lambda_6 = \lambda + 25\lambda^2 + 30\lambda^3.$

Eks. 4. Hvis X er antall forsøk med kjennetegn A i en Bernoulli forsøksrekke med n forsøk, og $\pi = \Pr(A)$, finner vi $p(s) = (\pi e^s + 1 - \pi)^n$, og altså ved (4)

$$\varkappa_\nu = n\pi \frac{d^{\nu-1}}{ds^{\nu-1}} \frac{1}{\pi + (1-\pi)e^{-s}}\Big|_{s=0}$$

Herav spesielt

$\varkappa_1 = n\pi, \varkappa_2 = n\pi(1-\pi), \varkappa_3 = n\pi(1-\pi)(1-2\pi),$
$\varkappa_4 = n\pi(1-\pi)[1-6\pi(1-\pi)]$

De standardiserte kumulanter γ_ν er kumulantene for den standardiserte variabel, $\frac{X-EX}{\sigma}$ hvor $\sigma^2 = \operatorname{var} X$. Vi har i følge (5)

$$\gamma_1 = 0, \gamma_2 = 1, \quad \gamma_\nu = \frac{\varkappa_\nu}{\sigma^\nu} = \frac{\varkappa_\nu}{\varkappa_2^{\frac{\nu}{2}}} \; ; \nu > 2 \tag{10}$$

B. Hermite-polynomer. Ved derivasjon av $e^{-\frac{x^2}{2}}$ finner vi

$$\frac{d}{dx} e^{-\frac{x^2}{2}} = -xe^{-\frac{x^2}{2}}, \quad \frac{d^2}{dx^2} e^{-\frac{x^2}{2}} = (x^2-1)e^{-\frac{x^2}{2}},$$

$$\frac{d^3}{dx^3} e^{-\frac{x^2}{2}} = -(x^3-3x)e^{-\frac{x^2}{2}}$$

Vi skriver generelt

$$\frac{d^n}{dx^n} e^{-\frac{x^2}{2}} = (-1)^n H_n(x) e^{-\frac{x^2}{2}} \tag{11}$$

som definisjonsligning for $H_n(x)$. Vi finner ved derivasjon av (11)

$$\frac{d^{n+1}}{dx^{n+1}} e^{-\frac{x^2}{2}} = \left[(-1)^n H_n'(x) + (-1)^{n+1} x H_n(x)\right] e^{-\frac{x^2}{2}}$$

Altså

$$H_{n+1}(x) = xH_n(x) - H_n'(x) \tag{12}$$

Det følger lett ved induksjon at alle $H_n(x)$ er polynomer, de såkalte <u>Hermite-polynomer</u>. Vi finner spesielt

$$\begin{gathered} H_o(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2-1, \quad H_3(x) = x^3-3x, \\ H_4(x) = x^4-6x^2+3, \quad H_5(x) = x^5-10x^3+15x, \quad H_6(x) = x^6-15x^4+45x^2-15 \end{gathered} \tag{13}$$

[Ved å derivere

$$\frac{d}{dx} e^{-\frac{x^2}{2}} = -xe^{-\frac{x^2}{2}}$$

n ganger mhp. x, fåes

$$\frac{d^{n+1}}{dx^{n+1}} e^{-\frac{x^2}{2}} = -x \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} - \binom{n}{1} \frac{d^{n-1}}{dx^{n-1}} e^{-\frac{x^2}{2}}$$

dvs.

$$H_{n+1}(x) = xH_n(x)-nH_{n-1}(x) \tag{14}$$

Av (12) og (14) finnes $H_n'(x) = nH_{n-1}(x)$, eller

$$H_{n+1}'(x) = (n+1)H_n(x) \tag{15}$$

Hvis vi nå deriverer (12) og benytter (15) finner vi

$$H_n''(x)-xH_n'(x)+nH_n(x) = 0 \tag{16}$$

som er en differensialligning for $H_n(x)$. Av (12) og (13) ser vi at vi kan skrive

$$H_n(x) = x^n+a_{1,n}x^{n-2}+a_{2,n}x^{n-4}+\ldots+a_{s,n}x^{n-2s}+\ldots \tag{17}$$

Innsettes dette i (16), fåes

$$(n-2s+1)(n-2s+2)a_{s-1,n}+2sa_{s,n} = 0$$

Herav

$$a_{s,n} = \frac{(-1)^s}{2^s}\,\frac{n!}{(n-2s)!s!} \tag{18}$$

Ved Taylors formel har vi

$$e^{-\frac{1}{2}(x+h)^2} = \sum_{n=0}^{\infty} \frac{h^n}{n!} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}} = \sum_{n=0}^{\infty} \frac{(-h)^n}{n!} H_n(x) e^{-\frac{x^2}{2}}$$

Ved å sette $h = -s$, får vi

$$e^{-\frac{s^2}{2} + sx} = \sum_{n=0}^{\infty} H_n(x) \frac{s^n}{n!} \tag{19}$$

Setter vi $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, ser vi at (11) kan skrives

$$g^{(n)}(x) = (-1)^n H_n(x) g(x) \tag{20}$$

Ved delvis integrasjon finner vi derfor

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) g(x) dx = (-1)^n \int_{-\infty}^{+\infty} H_m(x) g^{(n)}(x) dx =$$

$$= (-1)^{n-1} \int_{-\infty}^{+\infty} H_m' g^{(n-1)}(x) dx$$

Ved gjentatt delvis integrasjon finner vi sluttelig

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) g(x) dx = \int_{-\infty}^{+\infty} H_m^{(n)}(x) g(x) dx$$

idet vi forutsetter $m \leq n$. Men nå er $H_m^{(n)}(x) = 0$ hvis $m < n$ og $H_n^{(n)}(x) = n!$ Altså

$$\int_{-\infty}^{+\infty} H_m(x) H_n(x) g(x) dx = \begin{cases} n! & \text{hvis } m = n \\ 0 & \text{hvis } m \neq n \end{cases} \tag{21}$$

Altså er polynomene $\frac{1}{\sqrt{n!}} H_n(x)$; $n = 0,1,2,\ldots$ ortogonale relativt til $g(x)$. Multipliserer vi (19) med $g^{(m)}(x)$, integrerer fra $-\infty$ til $+\infty$ og benytter (21), får vi

$$\int_{-\infty}^{+\infty} e^{sx} g^{(m)}(x)dx = (-s)^m e^{\frac{s^2}{2}} \tag{22}$$

C. Gram-Charlier-rekken.

La X være en stokastisk variabel med forventing 0 og varians 1 og med (standardiserte) kumulanter $\gamma_1, \gamma_2, \gamma_3, \ldots$ Vi skal undersøke muligheten for å approksimere sannsynlighetstettheten for X med et uttrykk av formen

$$f(x) = \sum_{n=0}^{R} \frac{c_n}{n!} g^{(n)}(x) = g(x) \sum_{n=0}^{R} (-1)^n \frac{c_n}{n!} H_n(x) \tag{23}$$

hvor $g(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$.

La oss multiplisere (23) med $H_m(x)$, integrere fra $-\infty$ til $+\infty$ og benytte (21). Da får vi

$$c_m = (-1)^m \int_{-\infty}^{+\infty} H_m(x)f(x)dx \tag{24}$$

Herav, og av (13) ser vi at vi kan uttrykke c_m ved momentene

$$p_n = \int_{-\infty}^{+\infty} x^n f(x)dx$$

Koeffisientene c_m kan imidlertid enklest uttrykkes ved kumulantene. Mgf for X er i følge (23)

$$p(s) = \int_{-\infty}^{+\infty} e^{sx} f(x)dx = \sum_{n=0}^{R} \frac{c_n}{n!} \int_{-\infty}^{+\infty} e^{sx} g^{(n)}(x)dx$$

Altså ved (22)

$$v = e^{-\frac{s^2}{2}} p(s) = \sum_{n=0}^{R} \frac{c_n}{n!}(-s)^n \tag{25}$$

ν-te deriverte av v for $s = 0$ er altså, ifølge uttrykket på høyre side lik $(-1)^\nu c_\nu$; mens den ν-te deriverte av $w = \log v$, ifølge venstre side er lik 0 for $\nu = 1$ og 2 og lik γ_ν for $\nu \geq 3$. Innsettes dette i (1) får vi

$c_1 = c_2 = 0$,

$$c_{\nu+1} = \sum_{j=2}^{\nu-3} (-1)^{j+1} \binom{\nu}{j} \gamma_{j+1} c_{\nu-j} + (-1)^{\nu+1} \gamma_{\nu+1} \tag{26}$$

(idet summen faller bort når $\nu - 3 < 2$, dvs. $\nu < 5$). Vi finner spesielt

$$\begin{aligned}
&c_0 = 1, \quad c_1 = c_2 = 0, \quad c_3 = -\gamma_3, \quad c_4 = \gamma_4, \quad c_5 = -\gamma_5, \\
&c_6 = \gamma_6 + 10\gamma_3^2, \quad c_7 = -(\gamma_7 + 35\gamma_3\gamma_4), \quad c_8 = \gamma_8 + 56\gamma_3\gamma_5 + 35\gamma_4^2, \\
&c_9 = -(\gamma_9 + 84\gamma_3\gamma_6 + 126\gamma_4\gamma_5 + 280\gamma_3^3)
\end{aligned} \tag{27}$$

Dermed er koeffisientene i rekken (23) bestemt.

D. Edgeworth rekke.

La $Y_1, Y_2, \ldots, Y_N$ være uavhengige og identisk fordelte med forventning 0 og varians 1 samt med kumulanter γ_ν'. Vi skal approksimere sannsynlighetsfordelingen for den standardiserte sum

$$X = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} Y_j \tag{28}$$

Vi finner ved (5) og (9) for kumulantene for X

$$\gamma_1 = 0, \qquad \gamma_\nu = \frac{\gamma_\nu'}{N^{\frac{\nu}{2}-1}} \tag{29}$$

Vi innsetter dette i (23) med c_n gitt ved (27), ordner leddene etter stigende potenser av $N^{-\frac{1}{2}}$ og sløyfer ledd av høyere orden enn $N^{-\frac{3}{2}}$

$$f(x) = g(x)\left\{1 + N^{-\frac{1}{2}} \frac{\gamma_3'}{6} H_3(x) + N^{-1}\left[\frac{\gamma_4'}{24} H_4(x) + \frac{\gamma_3'^2}{72} H_6(x)\right] + \right.$$

$$\left. + N^{-\frac{3}{2}}\left[\frac{\gamma_5'}{120} H_5(x) + \frac{\gamma_3'\gamma_4'}{144} H_7(x) + \frac{\gamma_3'^3}{1296} H_9(x)\right]\right\} \tag{30}$$

Alt etter hvor stor nøyaktighet som ønskes, kan ledd av orden $N^{-\frac{1}{2}}$, N^{-1}, $N^{-\frac{3}{2}}$ tas med. Også generelt når X ikke er en standardisert sum kan det lønne seg å trekke inn leddene i Gram-Charliers rekke i rekkefølgen som antydet ved (30). Derved fremkommer Edgeworth rekke for sannsynlighetstettheten for en standardisert variabel med kumulanter $\gamma_3, \gamma_4, \ldots$.

$$f(x) = g(x)\left\{1 + \frac{\gamma_3}{6} H_3(x) + \left[\frac{\gamma_4}{24} H_4(x) + \frac{\gamma_3^2}{72} H_6(x)\right] + \right.$$
$$\left. + \left[\frac{\gamma_5}{120} H_5(x) + \frac{\gamma_3\gamma_4}{144} H_7(x) + \frac{\gamma_3^3}{1296} H_9(x)\right]\right\} \tag{30}$$

idet hakeparentesene samler ledd som hører sammen, og som enten alle bør tas med, eller ikke tas med ved approksimeringene.

For å finne approksimasjonsformler for den kumulative sannsynlighetsfunksjon $F(x)$ for X, benytter vi oss av (20) som fører til

$$\int_{-\infty}^{x} g(t)H_\nu(t)dt = -H_{\nu-1}(x)g(x) = (-1)^\nu G^{(\nu)}(x) \tag{31}$$

hvor $G(x)$ er det gaussiske integral. Herav de viktige formler

$$\underline{F(x) = G(x) - \frac{\gamma_3}{6} G^{(3)}(x) + \frac{\gamma_4}{24} G^{(4)}(x) + \frac{\gamma_3^2}{72} G^{(6)}(x)} \tag{32}$$

eller

$$\underline{F(x) = G(x) - g(x)\left\{\frac{\gamma_3}{6} H_2(x) + \frac{\gamma_4}{24} H_3(x) + \frac{\gamma_3^2}{72} H_5(x)\right\}} \tag{33}$$

En kortfattet oversiktstabell over $G^{(n)}(x)$ for $n = 0,1,2,3,4,5,6,7$; $x = 0.0\ (0.1)\ 4.0$ finnes i Harald Cramer: Mathematical methods of Statistics, Uppsala 1945, p.557. Ved. nøyaktigere beregninger bør en benytte (33) med

$$H_2(x) = x^2-1, \quad H_3(x) = x(x^2-3), \quad H_5(x) = [x^2(x^2-10)+15]x \tag{34}$$

sammen med nøyaktige tabeller over $g(x)$ og $G(x)$. (34) vil som regel være den mest bekvemme form for numeriske regninger. For omfattende tabelleringer med samme γ_3 og γ_4 vil det lønne seg å skrive ut hakeparentesen i (33) som et polynom, og så omforme dette tilsvarende (34), altså

$$a_o x^5 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = \left\{\left[\left[a_o x^2 + a_1\right] x + a_2\right] x + a_3\right\} x + a_4 \tag{35}$$

Eksempel: Hvis Y er kji-kvadratfordelt med 2 frihetsgrader har vi

$$\Pr(Y \leq y) = 1 - e^{-\frac{y}{2}} = G(x) - \frac{1}{3}G^{(3)}(x) + \frac{1}{4}G^{(4)}(x) + \frac{1}{18}G^{(6)}(x)$$

hvor

$$x = \frac{y-2}{2}$$

Hvis Y er kji-kvadratfordelt med 4 frihetsgrader har vi

$$\Pr(Y \leq y) = 1 - \left(\frac{y}{2} + 1\right)e^{-\frac{y}{2}} = G(x) - 0.2357G^{(3)}(x) + 0.125G^{(4)}(x) + $$
$$+ 0.0278G^{(6)}(x)$$

hvor

$$x = \frac{y-4}{2.8284}$$

Eksempel 2: Se stensilet "Noen forsikringsmatematiske emner" kapitel III. Beregning av ruinsannsynligheten ved aggregering av polisene. Kumulantene for overskuddene, uttrykt ved kumulantene κ_ν for $e^{-\delta T}$, er

$$\kappa_\nu(F_a) = \frac{N}{\delta^\nu}\kappa_\nu;\quad \kappa(F_A) = (-1)^\nu N\kappa_\nu;\quad \kappa(F_{A/a}) = (-1)^\nu N(\tfrac{\rho}{\delta}+1)^\nu\kappa_\nu$$

De standardiserte kumulanter γ_ν for overskuddene blir like store i de 3 situasjoner og er gitt ved

$$\gamma_\nu = (-1)^\nu \frac{\kappa_\nu}{N^{\frac{\nu}{2}-1}\sigma^\nu};\ \nu \geqq 2$$

der $\sigma^2 = \kappa_2$.

II. RISIKOVURDERING AV LIVSFORSIKRINGSVIRKSOMHETEN

II.1. Lundbergs makromodell for vurdering av risikoen ved en åpen forsikringsbestand.

La oss se litt på hva som skjer i et vanlig privat forsikringsselskap - enten det nå er et skade - eller livsforsikringsselskap.

På et tidspunkt t har selskapet N_t poliser i sin portefølje. For en vilkårlig av disse poliser, la oss si nr. i, er sannsynligheten for at krav skal oppstå i tidsintervallet t til $t+dt$ lik $\lambda_i(t)dt$. Avhengig av de nærmere omstendigheter i forbindelse med forsikringstilfellets inntreden er den forsikringssum som skal utbetales forskjellig. La sannsynligheten være $F_{it}(y)$ for at forsikringssummen Y er mindre enn eller lik y, gitt at krav oppstår på polisen. Vi vil foreløpig forutsette at det ikke er noen sammenheng mellom kravene for de N_t poliser; mer presist: De N_t polisers skjebner i intervallet $(t,t+dt)$ er stokastisk uavhengige.

Vi vil tenke oss at den forsikrede på tidspunktet t må betale premie for den forsikring som han har i tidsintervallet $(t,t+dt)$. Nettopremie for polise nr. i vil i så fall være

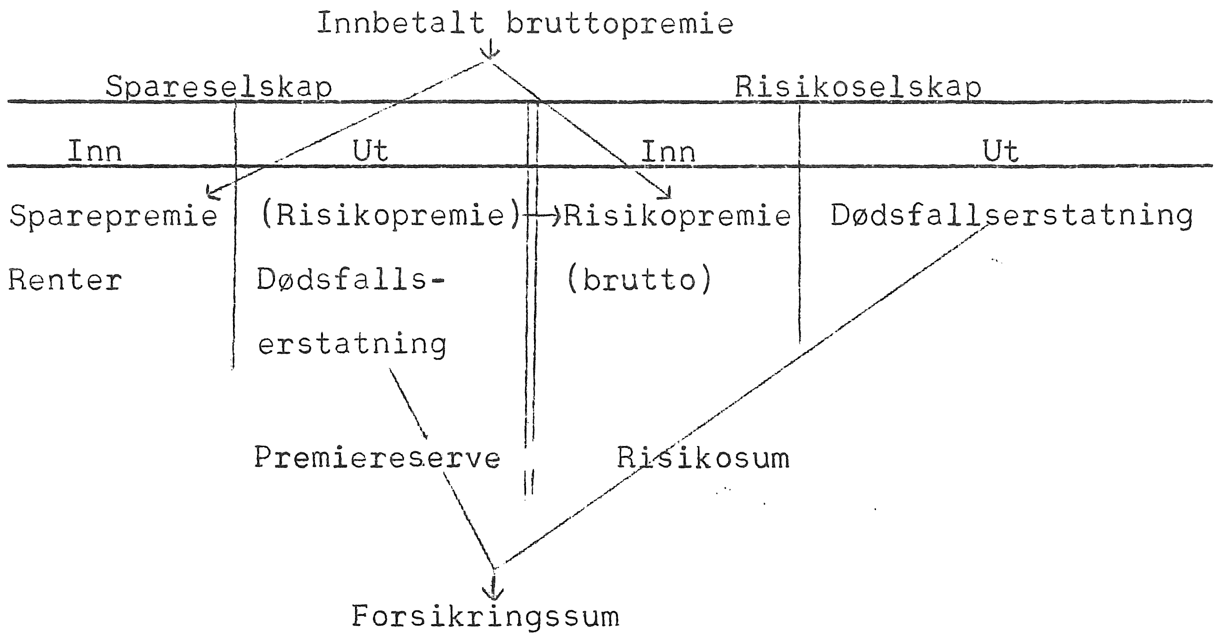
$$\pi_i(t)dt = \lambda_i(t)dt \int_{-\infty}^{+\infty} ydF_{it}(y) \tag{1}$$

(Muligheten for negativ forsikringssum vil bli kommentert senere). Dette er nettorisikopremien. La bruttorisikopremien være $B_i(t)dt$. Denne premien er ment å skulle dekke forsikringen i intervallet t til $t+dt$. Vi vil anta at den fremkommer ved et konstant prosentvis tillegg til nettopremien, altså $B_i(t) = (1+\alpha)\pi_i(t)$. Nå vet vi at i alminnelighet foregår ikke den faktiske premiebetaling ifølge avtalen på denne måten. I skadeforsikring vil det riktignok ofte være tilnærmet riktig å beskrive premiebetalingen som ovenfor, særlig hvis premieterminene er korte, men i livsforsikring vet vi at den faktiske premiebetaling ofte er en helt annen. Premie erlegges på visse tidspunkter til dekning av aktuell og fremtidig risiko. Resultatet er at det er kortere eller lengre tidsrom hvor der overhodet ikke betales premie. Beløpet til dekning av bruttorisikopremien for disse perioder er innbetalt tidligere. Når avtalt premie overstiger bruttorisikopremien i vedkommende tidsintervall vil det overskytende bli lagt til side til dekning av senere risikopremier. Resultatet er den velkjente fondsopplegning i form av premiereserve. Enhver forsikret har en premiereserve "på konto" i forsikringsselskapet. Den effektive forsikringssum som er gjenstand for forsikring er derfor forsikringssummen ifølge kontrakten minus den forsikredes premiereserve, dvs. <u>risikosummen</u>. I prinsippet oppstår fonds også i skadeforsikring (p.g.a. diskret premiebetaling). De vil være av mindre størrelsesorden, fordi sjansen for krav er konstant og avtalt premie justeres ettersom avtalt forsikringssum varierer.

Vi skal oppfatte disse fonds som de forsikredes oppsparte midler som egentlig ikke vedrører selskapets risikoforretning. Selskapene har på ethvert tidspunkt "rett" til å ta av dette fond til reduksjon av forsikringssummen og til utredning av risikopremien. Selskapet oppfører fondet som et passivum og det er <u>ruinert hvis det ikke har aktiva til dekning av dette passivum</u>.

Vi kan derfor når det gjelder vurdering av risikoen forbundet med usikkerheten ved kravene og forsikringssummens størrelse, arbeide med det skjema som er beskrevet ovenfor. Vi betrakter m.a.o. risikoforretningen isolert og <u>holder sparepremiene og premiereserven utenfor</u>. Vi ser bare på risikopremiene og risikosummen. I denne risikoforretning oppstår naturligvis ingen premiereserve. Differensen mellom tidligere innbetalte tariffrisikopremier og utbetalt risikosum går til et <u>sikkerhetsfond</u> (bonusfond), eller det utbetales direkte som aksjeutbytte eller som bonus til de forsikrede.

Vi tenker oss altså selskapet oppdelt i et spareselskap og et risikoselskap, skjematisk fremstilt som nedenfor:



(I virkeligheten er selskapet først konkurs når det ikke har disponible midler til å avsette premiereserve og til å dekke risikoselskapets erstatningsutbetalinger. Det "totale" selskap kan således manipulere slik at (noen av) midlene til risikoselskapet anvendes i spareselskapet, eller omvendt.

Forsikringsselskapets totale risiko er <u>ikke</u> lik summen av risikoene for de to forretninger. Man kan f.eks. tjene så godt på spareselskapet at overrenten som oppnås der, virker som en støtpute som <u>demper</u> risikoen i risikoselskapet.)

I dette avsnitt vil vi imidlertid se helt bort fra spareselskapet og bare studere risikoselskapet. Vi vil betrakte dette som konkurs hvis det ikke kan dekke sine spesielle forpliktelser uten hjelp "utenfra", dvs. fra spareselskapet.

Sannsynligheten for et krav på selskapet i tidsintervallet t til $t+dt$ er åpenbart

$$1 - \prod_{i=1}^{N_t} (1-\lambda_i(t)dt) = dt \sum_{i=1}^{N_t} \lambda_i(t)+o(dt^2) = \lambda_t dt+o(dt^2) \qquad (2)$$

idet vi har innført

$$\lambda_t = \sum_{i=1}^{N_t} \lambda_i(t) \qquad (3)$$

Sannsynligheten for krav på polise nr. i og nr. j i intervallet er åpenbart $\lambda_i(t)\cdot\lambda_j(t)dt^2$. Sannsynligheten for mer enn et krav i et intervall av lengde dt er altså av størrelsesorden dt^2 eller mindre. Med en nøyaktighet av denne størrelsesorden kan vi altså betrakte krav på de enkelte poliser i intervallet t til $t+dt$ som hverandre utelukkende begivenheter.

Den betingede kumulative sannsynlighet for at kravet i intervallet t til $t+dt$ er mindre eller lik y gitt at krav oppstår i intervallet, blir etter I.1.(7)

$$F_t(y) = \frac{1}{\lambda_t} \sum_{i=1}^{N_t} \lambda_i(t)F_{it}(y) \qquad (4)$$

I skadeforsikring vil det beløp som blir utbetalt på den enkelte polise være avhengig av skadens omfang. Fordelingen gitt ved $F_{it}(y)$ vil derfor ha en viss spredning. Spredningen i fordelingen gitt ved $F_t(y)$ vil vanligvis være større fordi risikosummen også vil avhenge av hvilken polise det dreier seg om.

I livsforsikring er $\lambda_i(t)$ en dødsintensitet, og sannsynlighetsmassen definert ved $F_{it}(y)$ er konsentrert i ett punkt, svarende til at kravets størrelse er gitt, f.eks. lik differensen mellom forsikringssum ifølge kontrakten og premiereserven for polisen. Spredningen i fordelingen $F_t(y)$ vil derfor i sin helhet bero på variasjonen i risikosummen fra polise til polise.

(En unntagelse er invalidepensjonsforsikring hvor man vil ha en viss spredning i fordelingen, idet invalidepensjonens størrelse avhenger av graden av invaliditet. Her vil derfor $F_{it}(y)$ ikke være ettpunktsfordeling.)

La ellers erstatningen for polise nr. i på tidspunkt t med sikkerhet være S_{it}, slik at

$$F_{it}(y) = \begin{cases} 0 & \text{for} \quad y < S_{it} \\ 1 & \text{for} \quad y \geq S_{it} \end{cases}$$

Da har vi:

$$F_t(y) = \frac{1}{\lambda_t} \sum_{\{i \mid S_{it} \leq y\}} \lambda_i(t)$$

Selv om F_{it} er en ettpunktsfordeling behøver altså ikke det samme gjelde for F_t.

Bemerk at ved en løpende livrente vil risikosummen være lik minus premiereserven, dvs. en negativ erstatning.

En eventuell enkepensjon kan bekvemmest oppfattes som to poliser, med avgangsintensiteter lik henholdsvis mannens og hustruens dødsintensiteter. La V være premiereserven når begge lever og V_1 premiereserven hvis mannen er død og hustruen lever. Risikosummene for de to poliser er da henholdsvis $V_1 - V$ og $-V$.

La oss vende tilbake til vårt generelle skjema. Total nettorisikopremie i intervallet $(t, t+dt)$ blir ifølge (1)

$$\pi(t)dt = dt \sum_i \pi_i(t) = dt \sum \lambda_i(t) \int_{-\infty}^{+\infty} ydF_{it}(y)$$

som kombinert med (4) gir

$$\pi(t) = \lambda_t \int_{-\infty}^{+\infty} ydF_t(y) \tag{5}$$

Total innbetalt bruttorisikopremie i intervallet (t,t+dt) blir

$$B(t)dt = dt \sum_i B_i(t) = (1+\alpha)\pi(t)dt \tag{6}$$

Det er selskapets totale drift som interesserer oss, ikke de enkelte poliser. Derfor har vi foretatt en aggregering av bestandens enkeltpoliser og oppnådd følgende data for intervallet (t,t+dt):

Sannsynligheten for krav i tidsintervallet er $\lambda_t dt$.

Den betingede kumulative fordeling for erstatningen, gitt at krav oppstår ved tidspunktet t, er $F_t(y)$.

Den totale nettorisikopremie er $\pi(t)dt$, hvor

$$\pi(t) = \lambda_t \int_{-\infty}^{+\infty} ydF_t(y)$$

og den totale bruttorisikopremie er $B(t) = (1+\alpha)\pi(t)$.

Vi forutsatte ovenfor at de enkelte polisers skjebne i intervallet (t,t+dt) var stokastisk uavhengige. Dette var grunnlaget for avledning av λ_t gitt ved (3). Nå er det imidlertid klart at vi uten videre kan la forutsetningen om uavhengigheten mellom polisenes skjebne falle. I stedet kan vi eksplisitt postulere at det er en sannsynlighet på $\lambda_t dt$ for krav i intervallet $(t,t+dt)$ uten å bekymre oss om hvorledes denne sannsynlighet kan avledes av sannsynligheter knyttet til enkeltpolisene. Dermed behøver vi heller ikke bekymre oss med distinksjonen mellom enkeltpolisene, som kan være uklar både i skade- og livsforsikring. (Det er jo under alle omstendigheter klart at det vi har kalt "polise" ikke kan identifiseres med den enhet det utstedes polisebrev på.)

Vi vil nå innføre en ny forutsetning i vårt system som ikke kan avledes av våre forutsetninger om de enkelte poliser. <u>Vi vil forutsette at erstatningsbeløpene og antall krav for forskjellige ikke-overlappende tidsintervaller er stokastisk uavhengige</u>. Spesielt vil altså det totale erstatningsbeløp i et intervall $(t,t+dt)$ være uavhengig av hva som er utbetalt på poliser før tidspunktet t. Ved en første refleksjon kan det synes som om dette er en svært lite realistisk forutsetning. I livs- eller brannforsikring opphører ofte en polise å eksistere når erstatning finner sted. Disse polisers skjebne etter tidspunktet t avhenger derfor i høy grad av hva som er skjedd før t. Betydningen av dette moment blir imidlertid uvesentlig når vi istedenfor å betrakte den enkelte polise, ser på erstatningsbeløpene under ett i et normalt selskap med mange poliser og med stadig nytegning. Hvor mange og hvor store erstatninger som vil bli utbetalt et gitt år vil først og fremst avhenge av bestandens sammensetning. Hva som skjedde foregående år av erstatningsutbetalinger har relativt liten innflytelse på bestanden, og dermed på årets erstatninger. Over et lengere tidsrom bakover i tid vil det være nytegninger som i første rekke får betydning for bestanden i dag, mens de fortidige erstatninger vil være av relativt liten betydning.

Til slutt skal vi forutsette at fordelingen F_t for enkelterstatningene er uavhengig av tiden, vi skriver $F_t(y) = F(y)$. Dette betyr at erstatningsbeløpene for de poliser på hvilke det faktisk oppstår krav, har en fordeling som ikke varierer med tiden. Men bemerk at forutsetningen ikke utelukker at totalerstatningen over forskjellige tidsrom varierer sterkt p.g.a. variasjonene i antall krav. Forutsetningen er naturligvis innskrenkende; men vær oppmerksom på at den godt kan være oppfylt selv om erstatningsfordelingen for de enkelte poliser varierer med tiden.

(Med vår tids inflasjon kan denne forutsetning virke meget urealistisk, idet erstatningsbeløpene i nominelle kroner vil ha en tendens til å øke i størrelse med tiden. Hvis man imidlertid tenker seg et indeks som reduserer (deflaterer) de nominelle erstatningsbeløp til "faste" kroner, virker det rimeligere at fordelingen for de deflaterte erstatningsbeløp er uavhengig av tiden.)

Den statistiske bestemmelse av $F(y)$ kan tenkes å skje ved å ta opp statistikk over de fortidige erstatninger i tilfellene hvor krav har oppstått. En hyppighetsfordeling for disse erstatninger gir et tilnærmet estimat av fordelingen F. Vi kan naturligvis ikke i sin alminnelighet estimere F statistisk ved å lage en hyppighetsfordeling over de hypotetiske erstatninger som ville ha blitt utbetalt i tilfelle krav på de forskjellige poliser i bestanden. En slik estimering ville hvile på den forutsetning at sannsynligheten for krav var like stor for alle poliser. Derimot kan F bestemmes statistisk av bestandstallene på en annen måte, som vi skal komme tilbake til senere.

Vi ser at forutsetningene vi har gjort om forsikringsbestanden har ledet oss til den sammensatte Poissonprosess som beskrevet i I.1. Det er dette som er Lundbergs makromodell for forsikringsvirksomheten. Som i I.1 kan kravintensiteten λ_t formelt elimineres ved å innføre effektiv tid.* Modellen inneholder da to elementer F og α. På grunnlag av modellen kan man nå bestemme den kumulative sannsynlighetsfunksjon $F(x;t)$ for samlet erstatning $X(t)$ i et tidsrom av lengde t. Med en gitt sikkerhetsreserve u til å starte med, kan man også beregne $\psi(u)$ som er sannsynligheten for ruin i fremtiden under forutsetning av at hele det fremtidige overskudd går til sikkerhetsfondet. Herav kan risikoen ved forsikringsvirksomheten vurderes.

*) Premieintensiteten $B(t)$ vil avhenge av tiden bare via λ_t, og blir derfor konstant når man innfører effektiv tid. Vi har nemlig

$$B(t) = (1+\alpha)\pi(t) = (1+\alpha)\lambda_t \int y\,dF_t(y) = (1+\alpha)\lambda_t p_1$$

Innbetalt premie i et effektivt tidsintervall av lengde τ blir da

$$\int_0^\tau B(t)dt = (1+\alpha)p_1 \int_0^\tau \lambda_t dt = (1+\alpha)p_1\tau$$

Altså er premieintensiteten etter transformasjon til effektiv tid τ, lik $(1+\alpha)p_1$.

Det er klart at risikoen ved forsikringsvirksomhet på et gitt tidspunkt må vurderes på grunnlag av de tre grunnelementer i Lundbergs modell : F, α, u. Risikoen kan endres ved å variere disse tre størrelser. Den blir åpenbart mindre hvis α eller u øker. Den kan også minskes ved å skjære av den øvre "hale" av erstatningsfordelingen F, f.eks. ved å reassurere all overskytende erstatning over et gitt beløp. Det er altså naturlig å velge en funksjon av F, α, u som mål for risikoen. Mange mål for totalrisikoen kan naturligvis komme på tale. Ett av dem er

$$\psi(u) = \psi_{F,\alpha}(u)$$

Størrelsen angir sannsynligheten for ruin en eller annen gang, men sier intet om når ruin skal inntreffe. Selv om denne størrelse var nær 1, kan dette alene ikke taes som bevis for at selskapet er i faresonen. Beregningen av ψ hviler dessuten på den forutsetning at hele overskuddet under enhver omstendighet legges tilside. Slik vet vi at det ikke foregår i et selskap i praksis. Hvis selskapet over visse perioder har relativt betydelige overskudd, vil deler av dette bli utdelt som bonus, aksjeutbytte e.l.

ψ må derfor ikke oppfattes som noe mer enn et måltall som muligens kan brukes til å sammenligne risikoen i forskjellige situasjoner. Hvis f.eks. selskapet nedsetter sikkerhetstillegget til nettopremiene fra α til $\alpha - \delta$, kan dette kompenseres ved å øke sikkerhetsreserven fra u til $u+\Delta$, såfremt $\psi_{F,\alpha-\delta}(u+\Delta) = \psi_{F,\alpha}(u)$. Hvis selskapet innfører en reassuranseordning som endrer fordelingen for erstatningene fra F til G, kan man minske sikkerhetstillegget med δ uten å endre risikoen, såfremt $\psi_{G,\alpha-\delta}(u) = \psi_{F,\alpha}(u)$. Alt dette forutsetter naturligvis at ψ aksepteres som mål for risikoen.

II.2. Stopptap-reassuranse.

En stopptap reassuransekontrakt mellom cedenten (dvs. den som blir reassurert) og reassurandøren går ut på følgende: Hvis dødsfallserstatningene i et gitt tidsintervall overstiger et visst beløp, skal reassurandøren betale en bestemt sum til cedenten. For denne reassuranse må cedenten erlegge en reassuransepremie ved periodens begynnelse.

Vi skal se på beregning av reassuransepremien ved en spesiell slags stopptapkontrakt. Vi vil gå ut fra at hvis dødsfallserstatningene + premiereserveavsetningen for perioden overstiger rente + premieinntektene, skal reassurandøren godtgjøre hele det underskudd som derved oppstår.[x)]

Cedenten kan representere en åpen eller lukket bestand av poliser. Det vanlige vil naturligvis være at bestanden er åpen, og det er den situasjon vi vil betrakte. Vi må da åpenbart

x) Beregning av reassuransepremie etter den her beskrevne metode er gjennomført i en konkret situasjon av aktuar Magnus Hegranes. Vi følger her stort sett hans tankegang.

gjøre en antagelse om tilgangen og den frivillige avgang i kontraktperioden. En enkel måte å gjøre det på er å anta at bestanden er stasjonær, dvs. at til enhver tid i perioden er antall personer i bestanden i alder $(x,x+dx)$ lik $\lambda_x dx$ for $z \leq x \leq \omega$. Vi antar altså at de λ_{40} 40-åringer på et gitt tidspunkt vil om, f.eks. 2 år, være erstattet med λ_{40} nye 40-åringer som utgjør de λ_{38} 38-åringer som var 38 år for to år, siden med fradrag av de som er døde eller frivillig avgåtte og med tillegg av de som er trådt inn. En spesiell måte å oppnå stasjonæritet på, er som bekjent å anta at dødeligheten holder seg konstant, at all inntredelse skjer i laveste alder z, at denne har holdt seg konstant de siste $\omega - z$ år, og at ingen frivillig avgang er mulig.

Stasjonæritet over lengre tidsintervaller er sjelden til stede, men over så korte kontraktsperioder som det er tale om ved reassuranse (1 eller 5 år) kan det likevel være tilnærmet riktig å gjøre forutsetningen.

La aldersintervallet (z,ω) være delt i delintervallene $I_1, I_2, \ldots$ med midtpunkter henholdsvis $x_1, x_2, \ldots$. Antall personer i aldersintervall I_j er da

$$L_j = \int_{I_j} \lambda_x dx \qquad (1)$$

til enhver tid i perioden. Antall døde i aldersintervallet I_j i perioden kaller vi D_j. Dette er altså antall personer som ved døden hadde alder i intervallet I_j, de er ikke utgått av antall levende L_j ved periodens begynnelse (eller på noe bestemt annet tidspunkt).

Vi antar at dødeligheten i bestanden kan angis ved dødsintensiteten μ_x. Som en tilnærmelse vil vi tenke oss at alle personer i aldersintervall I_j har alder x_j. Sannsynligheten for at det skal dø en person i aldersintervall I_j i løpet av et tidsintervall av lengde dt er

$$1-(1-\mu_{x_j}dt)^{L_j} \approx L_j\mu_{x_j}dt \tag{2}$$

Kontraktsperiodens lengde er T. Antall døde D_j i kontraktsperioden i aldersintervall I_j er altså Poissonfordelt med parameter $L_j\mu_{x_j}T$. Vi har følgelig for sannsynlighetsfordelingen for D_j:

$$\Pr(D_j = d) = \frac{(L_j\mu_{x_j}T)^d}{d!}\, e^{-L_j\mu_{x_j}T} \tag{3}$$

Videre for forventing, varians og kumulanter

$$ED_j = \operatorname{var} D_j = \varkappa_\nu(D_j) = L_j\mu_{x_j}T \tag{4}$$

Det er utvilsomt ikke helt teoretisk tilfredsstillende på den ene side å forutsette at L_j-ene er faste gjennom tiden, uten tilfeldige variasjoner, og på den annen side å forutsette at dødsfallene D_j er stokastiske variable. Hvis man eksplisitt lager en noenlunde realistisk modell for tilgangen, vil også L_j-ene være stokastiske. Hva vi - sterkt forenklet - forutsetter, er at tilgangen varierer nøyaktig i takt med dødeligheten, slik at L_j-ene blir konstante. Dermed forutsetter vi også konstant premiereserve $\sum V_{x_j}L_j$ gjennom kontraktsperioden, se ligning (11) nedenfor. For korte kontrakts-perioder vil denne mangel på realisme formodentlig ikke være vesentlig, så lenge formålet er å beregne en stopp-tap reassuransepremie for dekning av reassuranseerstatning ved en ugunstig tilfeldig variasjon i dødeligheten.

I første omgang vil vi se på en situasjon som er spesielt enkel å behandle. Polisebestanden består av enkle livsforsikringer, alle med samme forsikringssum S og samme tegningsalder z. Tilgangen består altså bare av z-åringer og den antas konstant, mens det ikke er noen frivillig avgang. Forsikringen betales med en kontinuert premie i forsikrings-tiden

$$\pi(1+\varepsilon) = S\left(\frac{1}{\bar{a}_z} - \delta\right)(1+\varepsilon) \tag{5}$$

idet renteintensiteten er δ og sikkerhetstillegget til nettopremien er $100\,\varepsilon\,\%$.

Vi antar at den totale fondsavsetning for en person i alder x er nettopremiereserven

$$SV_x = S\left(1 - \frac{\bar{a}_x}{\bar{a}_z}\right) \tag{6}$$

Premiereserven i bestanden holder seg åpenbart konstant lik

$$S \sum V_{x_j} L_j \tag{7}$$

Samlet dødsfallserstatning i løpet av kontraktperioden er

$$S \sum D_j \qquad (8)$$

Anta som en tilnærmelse at utbetalingene faller jevnt over perioden, slik at det pr. tidsenhet utbetales $\frac{S}{T} \sum D_j$.

Cedentens regnskap for perioden ser nå slik ut, når W betegner kontantverdien av overskuddet ved periodens begynnelse:

Inntekter		Utgifter	
Sikkerhetsfond	$ue^{\delta T}$	Dødsfalls-erstatninger med rentetap	$\bar{s}_{\overline{T}\rceil} \frac{S}{T} \sum D_j$
Gammel reserve med renter	$e^{\delta T} S \sum V_{x_j} L_j$	Reassuranse-premie	$e^{\delta T} P'$
Premier med renter	$\bar{s}_{\overline{T}\rceil} \pi (1+\varepsilon) \sum L_j$	Ny reserve	$S \sum V_{x_j} L_j$
		Overskudd	$We^{\delta T}$

hvor $\bar{s}_{\overline{T}\rceil} = \frac{1}{\delta}(e^{\delta T}-1)$.

Av samlet dødsfallserstatning kan som kjent noe tas av fondsmidlene, mens resten må utredes av de risikopremier som er innbetalt i perioden:

$$S \sum D_j = S \sum D_j V_{x_j} + \sum D_j R_{x_j} \qquad (9)$$

hvor $R_x = S(1-V_x)$ er risikosummen for en person i alder x. Nettopremieinntekten med renter spaltes i spareandelen og risikoandelen:

$$\bar{s}_{\overline{T}\rceil} \pi \sum L_j = \bar{s}_{\overline{T}\rceil} \sum L_j \pi_{x_j}^{sp} + \bar{s}_{\overline{T}\rceil} \sum L_j \pi_{x_j}^{ri} \qquad (10)$$

hvor π_x^{sp} og π_x^{ri} er henholdsvis spare- og risikopremien for en x-åring.

Sparepremiene oppsamles i reservefondet. Disse premier pluss det gamle fondet med renter må derfor være akkurat nok til å dekke ny reserve samt premiereserveutbetalingene til de døde. Vi har derfor

$$\bar{s}_{\overline{T}|} \sum L_j \pi_{x_j}^{sp} + e^{\delta T} S \sum V_{x_j} L_j = S \sum V_{x_j} L_j + \bar{s}_{\overline{T}|} \frac{S}{T} \sum D_j V_{x_j} \qquad (11)$$

Denne ligning må være tilnærmet riktig likegyldig hvorledes dødsfallene forløper. (Ligningen inneholder utvilsomt en liten selvmotsigelse, idet bare siste ledd er stokastisk. Se imidlertid kommentaren etter ligning (4) ovenfor.) Trekker vi ut de to ledd på venstre side av (11) på inntektssiden og de to ledd på høyre side på utgiftssiden, og diskonteres alle beløp til regnskapsperiodens begynnelse, fåes følgende regnskap:

Inntekter		Utgifter	
Sikkerhetsfond	u	Risikosumut-betalinger	$\frac{1}{T} \bar{a}_{\overline{T}|} \sum D_j R_{x_j}$
Sikkerhets-bidrag	$\bar{a}_{\overline{T}|} \pi \varepsilon \sum L_j$	Reassuranse-premie	P'
Risikopremie-inntekt	$\bar{a}_{\overline{T}|} \sum L_j \pi_{x_j}^{ri}$	Overskudd	W

Vi innfører betegnelsene

$$\begin{aligned} K &= u + \bar{a}_{\overline{T}|} (\pi\varepsilon \sum L_j + \sum L_j \pi_{x_j}^{ri}) \\ c &= \frac{1}{T} \bar{a}_{\overline{T}|} \end{aligned} \qquad (12)$$

og får

$$W = K-c \sum D_j R_{x_j} - P' \tag{13}$$

Begge størrelsene (12) lar seg lett beregne.

Reassurandøren må altså betale et beløp med kontantverdi $-W$ til cedenten hvis $W < 0$, men intet hvis $W \geqq 0$.

For den stokastiske variable W har vi nå ifølge (4) følgende forventning, varians og kumulanter:

$$\begin{aligned} \rho &= EW = K-cT \sum R_{x_j} L_j \mu_{x_j} - P' \\ \sigma^2 &= \operatorname{var} W = c^2 T \sum R_{x_j}^2 L_j \mu_{x_j} \\ \varkappa_\nu(W) &= (-1)^\nu c^\nu T \sum R_{x_j}^\nu L_j \mu_{x_j} \end{aligned} \tag{14}$$

De standardiserte kumulanter for W er derfor

$$\gamma_\nu = \frac{(-1)^\nu \quad \sum R_{x_j}^\nu L_j \mu_{x_j}}{T^{\frac{\nu}{2}-1} \left(\sum R_{x_j}^2 L_j \mu_{x_j} \right)^{\frac{\nu}{2}}} \quad ; \nu \geqq 3 \tag{15}$$

Den kumulative sannsynlighetsfunksjon for W er derfor tilnærmet

$$\Pr(W \leqq w) = G\left(\frac{w-\rho}{\sigma}\right) - \frac{\gamma_3}{6} G^{(3)}\left(\frac{w-\rho}{\sigma}\right) + \frac{\gamma_4}{24} G^{(4)}\left(\frac{w-\rho}{\sigma}\right) + \frac{\gamma_3^2}{72} G^{(6)}\left(\frac{w-\rho}{\sigma}\right) =$$

$$= H\left(\frac{w-\rho}{\sigma}\right) = G\left(\frac{w-\rho}{\sigma}\right) + H_0\left(\frac{w-\rho}{\sigma}\right) \tag{16}$$

idet definisjonen av H_0 følger av denne ligning. (Merk at H

er absolutt kontinuerlig, selv om fordelingen for W ikke behøver være det.)

Sannsynligheten for at cedenten skal motta noe fra reassurandøren ved periodens slutt er åpenbart

$$\Pr(W \leq 0) = H(-\tfrac{\rho}{\sigma}) \tag{17}$$

Nettoreassuransepremien P er forventningsverdien av det beløp som cedenten mottar, altså

$$\begin{aligned} P &= -\int_{-\infty}^{0} w\, d_w H(\tfrac{w-\rho}{\sigma}) = -\rho H(-\tfrac{\rho}{\sigma}) - \sigma \int_{-\infty}^{-\frac{\rho}{\sigma}} t\, dH(t) = \\ &= -\rho G(-\tfrac{\rho}{\sigma}) - \sigma \int_{-\infty}^{-\frac{\rho}{\sigma}} t\, dG(t) - \rho H_0(-\tfrac{\rho}{\sigma}) - \sigma \int_{-\infty}^{-\frac{\rho}{\sigma}} t\, dH_0(t) \end{aligned} \tag{18}$$

Nå har vi

$$\int_{-\infty}^{x} t\, dG(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} t e^{-\frac{t^2}{2}}\, dt = -g(x),$$

mens de to siste ledd i (18) ved delvis integrasjon blir lik

$$\sigma \int_{-\infty}^{-\frac{\rho}{\sigma}} H_0(t)\, dt$$

Innsettes uttrykket for H_0 (ifølge (16)) i dette integral og benyttes

$$\int_{-\infty}^{c} G^{(\nu)}(t)\, dt = G^{(\nu-1)}(c) = g^{(\nu-2)}(c)$$

fåes

$$\frac{P}{\sigma} = -\frac{\rho}{\sigma}\left[1-G\left(\frac{\rho}{\sigma}\right)\right]+g\left(\frac{\rho}{\sigma}\right)+\frac{\gamma_3}{6}\,g'\left(\frac{\rho}{\sigma}\right)+\frac{\gamma_4}{24}\,g^{(2)}\left(\frac{\rho}{\sigma}\right)+\frac{\gamma_3^2}{72}\,g^{(4)}\left(\frac{\rho}{\sigma}\right) \qquad (19)$$

hvor $\frac{\rho}{\sigma} = \frac{K-cT\sum R_{x_j} L_j \mu_{x_j}}{\sigma} - \frac{P'}{\sigma}$.

(19) sammen med (14) og (15) er grunnlaget for bestemmelsen av P og P'.

Hvis vi spesielt setter $P' = P$, bestemmer (19) en ren netto-reassuransepremie. Vanligvis må man imidlertid regne med tillegg til nettopremien P på grunn av at reassurandøren løper annen risiko i forbindelse med kontrakten. Man setter derfor

$$P' = P(1+\Delta) \qquad (20)$$

hvor Δ er skjønnsmessig ansatt. (19) og (20) bestemmer da P og dermed også P'.

Premien P er altså rasjonelt begrunnet ved risikoen for underskudd på grunn av tilfeldige variasjoner i dødeligheten, mens det ikke er gitt noen regel for fastsettelsen av tillegget $P\Delta$. Hvis ovenstående beregning skal ha noen interesse må situasjonen derfor være slik at de tilfeldige variasjoner i dødeligheten gir noe grunnlag for bekymring, dvs. at cedentbestanden er liten (f.eks. en liten "kasse"). Hvis cedentbestanden er stor, f.eks. et stort livsforsikringsselskap, vil risikoen for underskudd på grunn av tilfeldige variasjoner i dødeligheten være helt ubetydelig , mens kommersiell risiko og risikoen for systematisk endring i dødeligheten (av epidemisk og sekulær art) vil være det vesentlige. Da vil det vesentlige problem ved premieansettelsen bestå i å finne $P\Delta$, og det gir ovenstående utledninger intet grunnlag for.

Vår forutsetning er altså at vi har en relativt liten bestand som skal reassureres.

Vi skal nå se på en litt mer komplisert situasjon enn den vi betraktet ovenfor. Vi oppgir forutsetningen om at tegningsalder og forsikringssum er den samme for alle poliser, men forøvrig antar vi fremdeles at cedentens bestand består av livsvarige livsforsikringer. Reassuransekontrakten er som tidligere beskrevet, og vi forutsetter som ovenfor at bestanden er stasjonær.

I den enkle situasjon vi hittil har betraktet var det mulig å skrive samlet risikoutbetaling som

$$R = \sum D_j R_{x_j} \qquad (21)$$

(se ligning (13)) hvor alle R_{x_j} er konstanter mens D_j er Poissonfordelt. Hvis nå tegningsalder og forsikringssum varierer for poliser med samme aktuelle alder, vil dette også være tilfellet med udekket risiko. De D_j personer vil derfor utgå med forskjellige risikosummer, og (21) er ikke lenger riktig.

La oss betrakte de L personer i bestanden. La deres aldre være $y_1, y_2, \ldots, y_L$ og deres udekkede risici $z_1, z_2, \ldots, z_L$. Sannsynligheten for at et krav skal oppstå i løpet av $(t,t+dt)$ er

$$dt \sum_{i=1}^{L} \mu_{y_i}, \tag{22}$$

mens den betingede sannsynlighet for at det er person nr. i, med alder y_i som gjør krav gjeldende, gitt at krav oppstår, er

$$q_i = \mu_{y_i} / \sum_{i=1}^{L} \mu_{y_i} \tag{23}$$

La $Z(t)$ betegne størrelsen av selskapets utbetalingsforpliktelser på tidspunktet t. Forutsetningsvis er:

$$Pr(Z(t) \leq z \mid \text{krav fra person nr. } i) = G_{(i)}(z) \quad \text{for alle} \quad t > 0$$

og

$$G_{(i)}(z) = \begin{cases} 0 & \text{for} \quad z < z_i \\ 1 & \text{for} \quad z \geq z_i \end{cases}$$

Vi har for oss en heterogen bestand inndelt i L risikogrupper med kravintensiteter henholdsvis $\mu_{y_1},\ldots,\mu_{y_2}$ og med erstatningsfordelinger henholdsvis $G_{(1)},\ldots,G_{(L)}$. Det følger av I.1 at

$$\begin{aligned} F(z) &= Pr(Z(t) \leq z \mid \text{krav}) \\ &= \sum_{i=1}^{L} q_i G_{(i)}(z) \\ &= \frac{\sum_{z_i \leq z} \mu_{y_i}}{\sum_{i=1}^{L} \mu_{y_i}} \end{aligned}$$

La oss nå gjøre samme betraktning for aldersgruppe j hvor det er L_j personer. Da kan vi som en tilnærmelse sette alle aldre y_i like store, nemlig lik x_j. Vi finner da av

(24) for <u>den betingede kumulative sannsynlighetsfunksjon for kravets størrelse i aldersgruppe j</u>:

$$F_j(z) = L_{zj}/L_j \tag{25}$$

hvor

L_{zj} = antall risikosummer i aldersgruppe j som er $\leq z$.

<u>$F_j(z)$ er lik den kumulative hyppighetsfordeling for risikosummene i aldersgruppe j</u>. (Det er altså her tale om <u>en eksakt</u> relasjon, <u>ikke</u> den vanlige tilnærmete likhet mellom hyppighetsfordeling og sannsynlighetsfordeling.) Statistisk kan alle F_j bestemmes enten ved å oppta fullstendig statistikk over risikosummene i aldersgruppene, eller ved å oppta samme statistikk for et tilfeldig utvalg fra hver aldersgruppe.

La nå samlet utbetalt risikosum i aldersgruppe j i perioden være R_j. Denne størrelse har altså en sammensatt Poissonfordeling med kravsintensitet $\mu_{x_j} L_j$ og med sannsynlighetsfordeling for erstatningens størrelse $F_j(z)$. Herav følger at kumulantene for R_j er gitt ved

$$\varkappa(R_j) = T\mu_{x_j} L_j p_{\nu j} = T\mu_{x_j}\alpha_{\nu j} \tag{26}$$

hvor

$$p_{\nu j} = \int z^{\nu} dF_j(z) = \frac{1}{L_j}\sum_{i=1}^{L_j} Z_{ij}^{\nu} = \frac{1}{L_j}\alpha_{\nu j} \tag{27}$$

hvor $Z_{1,j},\ldots,Z_{L_j,j}$ er risikosummene i j-te aldersgruppe.

Nå er overskuddet for perioden analogt med (13) gitt ved

$$W = K - c\sum R_j - P' \tag{28}$$

hvor

$$K = u + \bar{a}_{\overline{T|}}(\varepsilon \sum \pi_i + \sum \pi_i^{ri}) \tag{29}$$

$\sum \pi_i$ er summen av nettopremiene og $\sum \pi_i^{ri}$ er summen av risikopremiene for alle L personer. c har samme mening som før.

Vi finner herav

$$\rho = \varkappa_1(W) = EW = K - P' - cT\sum \mu_{x_j} \alpha_{1_j} \tag{30}$$

$$\varkappa_\nu(W) = (-c)^\nu \, T \sum \mu_{x_j} \alpha_{\nu_j} \tag{31}$$

hvor

$$\alpha_{\nu_j} = \sum_{i=1}^{L_j} Z_{ij}^\nu \tag{32}$$

Spesielt er

$$\sigma^2 = \operatorname{var} W = \varkappa_2(W) \tag{33}$$

De standardiserte kumulanter for W blir derfor

$$\gamma_\nu = \frac{(-1)^\nu}{T^{\frac{\nu}{2}-1}} \frac{\sum \mu_{x_j} \alpha_{\nu_j}}{(\sum \mu_{x_j} \alpha_{2_j})^{\frac{\nu}{2}}} \tag{34}$$

Sannsynlighetsfordelingen for W er derfor gitt ved (16) med ρ, σ og γ_ν bestemt ved (30), (33) og (34). P og P' bestemmes som før av (19).

I praktisk forsikring fremkommer ikke tariffpremie ved å gjøre et konstant prosentvis tillegg til nettopremien. Det vanlige er at man legger de forskjellige grunnlagselementer til den sikre side og beregner premien ut fra dette. Det gjelder også μ_x som er av spesiell interesse i denne sammenheng, siden det er usikkerheten på grunn av tilfeldige variasjoner i dødeligheten som vi beskjeftiger oss med. Dette reiser ingen nye matematiske problemer av betydning.

Spørsmålet oppstår imidlertid hvilken μ_x vi skal benytte i formlene (30) - (33). <u>Det er åpenbart at vi ikke må benytte grunnlagsdødeligheten</u>. Den har jo sikkerhetsmargin, og det er denne sikkerhetsmargin som gir grunnlag for overskuddsdannelsen. Hvis vi derfor ved beregning av reassuransepremien P' går ut fra beregningsgrunnlagets μ_x, vil det si at vi regner med at dette er den faktiske dødelighet, dvs. ingen overskuddsdannelse. Det er da klart at premien P' vil bli urimelig høy. Vi må derfor konstruere et nytt dødelighetsgrunnlag for beregningen av P' som ligger nær opp til det faktiske, kanskje med en ubetydelig sikkerhetsmargin. Det er i denne forbindelse viktig å være oppmerksom på at P' er meget følsom overfor variasjoner i μ_x.