Matematisk statistik
Stockholms universitet

Avvecklingsfunktionen i sjukförsäkring

Johanna Eriksson

Examensarbete 2009:2

Postal address:

Matematisk statistik
Dept. of Mathematics
Stockholms universitet
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Matematisk statistik
Stockholms universitet
Examensarbete **2009:2**,
http://www.math.su.se/matstat

Avvecklingsfunktionen i sjukförsäkring

Johanna Eriksson*

Februari 2009

Sammanfattning

För att bestämma storleken på framtida utbetalningar i sjukförsäkring och premiebefrielseförsäkring måste försäkringsbolag göra antaganden om hur länge de som blivit sjuka kommer fortsätta att vara sjuka. Till detta används den så kallade avvecklingsfunktionen. Genom bland annat Nelson-Aalenskattning har i detta arbete metoder tagits fram för att bestämma avvecklingsfunktionen både som matris och som formel. En formel med tre exponentialfunktioner visade sig vara det bästa alternativet för Folksam Livs bestånd med tanke på dess sammansättning och olika karenser. Arbetet resulterar nu i ett förslag till ett nytt avvecklingsantagande för sjukförsäkring och premiebefrielseförsäkring i Folksam Liv.

*Postal address: Matematisk statistik, Stockholms universitet, SE-106 91, Stockholm, Sweden. E-mail: johanna.eriksson@folksam.se. Handledare: Anders Björkström

Abstract

To decide the amount of future payments in disability insurance and waiver of premium, companies must make assumptions about how long those who are disabled will stay disabled. For that the so called termination function is used. The termination function is in this report estimated by the Nelson Aalen approach, both as a matrix and a formula. A formula with three exponential functions appeared to be the best option for this data set from Folksam Liv taking its specific composition and waiting time in mind. The work now results in a suggestion to a new termination assumption for disability insurance and waiver of premium in Folksam Liv.

Innehåll

1 Förord

Denna rapport utgör mitt examensarbete för en magisterexamen i matematisk statistik på Stockholms universitet. Arbetet är utfört med hjälp av datamaterial om sjukförsäkring från min arbetsplats Folksam.

Jag vill rikta ett stort tack till Mats Johansson på Folksam Liv för allt han har lärt mig och för att han alltid outtröttligt svarar så pedagogiskt och kunnigt på mina frågor. Jag har verkligen tur som får arbeta med och lära mig av dig.

Mycket tack även till min handledare Anders Björkström på Instutitionen för Matematisk statistik vid Stockholms universitet för sitt vidareförmedlande av sitt kunnande om matematisk statistik samt stöd under arbetets gång.

Ett sista tack även till alla på Folksam som bidragit med kunskap och idéer och för att ni gör mina arbetsdagar så trevliga.

2 Introduktion

Det är viktigt för alla försäkringsbolag att göra prognoser för hur mycket pengar som kommer att behöva betalas ut i framtiden. Detta för att kunna sätta rätt premier och veta hur stora summor man måste ha för att kunna betala ut det man har lovat. Det här arbetet är inriktat på betalningar som utgår på grund av sjukdom.

När man beräknar hur stora utbetalningarna till de som blivit sjuka förväntas bli behöver man bland annat göra antaganden om hur länge en sjuk person kommer fortsätta att vara sjuk. Till detta används något som kallas avvecklingsfunktionen. Detta arbete syftar till att hitta en ny avvecklingsfunktion. Men låt oss nu ta det från början:

3 Bakgrundsinformation

Begreppet "att vara sjuk"

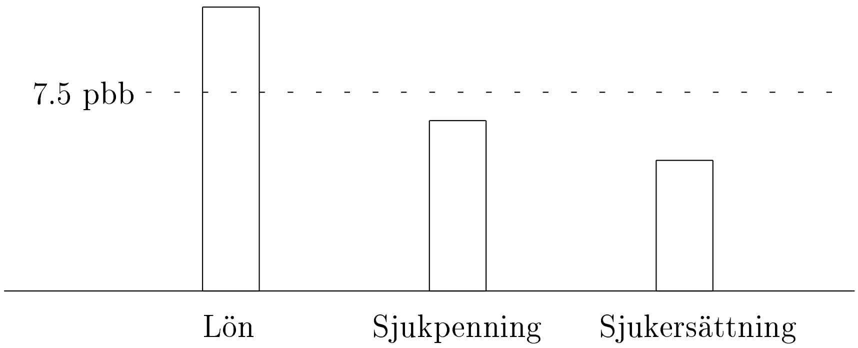
När man inom försäkring talar om "att vara sjuk" ska det inte uppfattas rent medicinskt. Det som egentligen åsyftas är nedsättning av förmågan att utföra arbete. Denna nedsättning av arbetsförmågan ska i sin tur vara orsakad av sjukdom eller olycksfall.

Man inser att sjukdom i försäkringsmedicisk mening får en subjektiv dimension eftersom två individer med samma diagnos kan ha helt olika upplevelse av hur arbetsförmågan påverkas. Vidare kan en viss sjukdom medföra att man blir helt förhindrad att utöva vissa yrken medan arbetsförmågan i ett annat yrke kan vara helt opåverkad av samma sjukdom. Eftersom sjukförsäkringens uppgift är att ersätta förlorad inkomst är det rimligt att arbetsförmågan under en skälig övergångsperiod bedöms för det arbete man hade när man insjuknade. Med tiden bör dock den sjuke undersöka om det finns något annat arbete på arbetsmarknaden som denne skulle kunna utföra. Tolkningen av sjukdomsbegreppet påverkas således av politiska faktorer eftersom det är lagarna om den allmänna sjukförsäkringen som reglerar vem som ska klassas som sjuk.

Den allmänna sjukförsäkringen

Sedan den 1/7 2008 gäller nya regler för den allmänna sjukförsäkringen i Sverige. Här ges ett sammandrag av hur vår försäkring ser ut:

Om ett sjukfall har blivit längre än 14 dagar utbetalas sjukpenning från Försäkringskassan. Försäkringskassan betalar en procentuellt beräknad ersättning för lön upp till 7.5 prisbasbelopp (pbb, 2009 betyder det ca 27000 kr/månad) vilket alltså får följden att lönedelar utöver detta inte kompenseras alls. Om sjukfallet beräknas vara bestående får man i stället sjukersättning. Sjukersättning är en lägre ersättning än sjukpenning. Se förenklad illustration av ersättningarna i figur 1.



Figur 1: Exempel på hur ersättningen från den allmänna sjukförsäkringen kan se ut för en person som tjänar mer än 7.5 pbb per år.

Den förlust man gör som sjuk är alltså det glapp som uppstår genom att Försäkringskassan bara betalar en viss procentandel av lönen i ersättning och även det man eventuellt tjänar utöver 7.5 prisbasbelopp. För att förlora mindre i inkomst om man blir sjuk kan man teckna en sjukförsäkring hos ett försäkringsbolag.

Sjukförsäkring och premiebefrielseförsäkring

Sjukförsäkring är en försäkring som ska kompensera för inkomstförlust på grund av sjukdom eller olycksfall. Den betalas därför ut med avtalade fasta månadsbelopp som beror på lönen. Man har också en avtalad slutålder för försäkringen, oftast 65 år. Ersättningen är proportionell till i hur stor grad man är arbetsoförmögen, om man exempelvis är arbetsoförmögen till 75% får man 75% av sitt månadsbelopp. Många försäkringsbolag har satt en gräns vid att den sammanlagda ersättningen högst får uppgå till 90% av lönen; man ska aldrig kunna tjäna på att vara sjuk.

Premiebefrielseförsäkring är en tilläggsförsäkring till livförsäkring som innebär att man inte behöver betala premien till sin livförsäkring om man blir sjuk. Den fungerar alltså på samma sätt som en sjukförsäkring förutom att utbetalningarna går direkt till den sjukes livförsäkring. Inga fall från premiebefrielseförsäkring ligger till grund för denna studie men en avvecklingsfunktion används även inom denna försäkringsgren.

Diskontering

Att diskontera ett framtida belopp innebär att man i beräkningarna tar hänsyn till att pengar förräntas. Genom att göra antaganden för vad räntan kommer att vara kan man beräkna hur stor del av summan man måste ha i sin ägo just nu[1].

Exempel med förenklade antaganden:
Vill man ha 100 kr om tio år och antar en årlig fast ränta på 5% ser uppställningen ut så här: $X \cdot 1.05^{10} = 100$. Alltså är $X = 100 \cdot \frac{1}{1.05^{10}} \approx 60$. Man måste alltså ha 60 kronor idag för att kunna ha 100 kronor om 10 år.

Aktsamma och betryggande antaganden

Inom försäkring måste man förutse hur saker kommer att utveckla sig i framtiden och därför använder man sig av olika antaganden. Två samlingsnamn för antaganden man gör är aktsamma respektive betryggande antaganden. Aktsamma antaganden betyder att de antaganden om bland annat framtida ränta, sjuklighet med mera man gör är realistiska, de får alltså inte innehålla några säkerhetsmarginaler. Betryggande antaganden innebär å andra sidan att man faktiskt använder sig av säkerhetsmarginaler. I vissa fall använder man sig av aktsamma antaganden, i andra av betryggande.

[1]Tidigare använde man sig av en fast ränta, så kallad högsta ränta. En relativt ny regel är att den ränta som används för diskontering ska hämtas från en räntekurva. Att använda räntekurva innebär att man diskonterar med olika räntor beroende på när i tiden utbetalningarna ska ske.

Tjänstepensionsförsäkring och övrig livförsäkring

Om en försäkring är av typen tjänstepensionsförsäkring betyder det att det är personens arbetsgivare som betalar premierna för försäkringen. All annan försäkring kallas för övrig livförsäkring. Enligt tjänstepensionsdirektivet som kom år 2006 måste försäkringsbolagen ha separat skuldtäckning för tjänstepensionsförsäkring. Detta betyder att man ekonomiskt måste hålla isär tjänstepensionsförsäkring och övrig livförsäkring. Reserven för tjänstepensionsförsäkring skall räknas med aktsamma antaganden medan reserven för övrig livförsäkring ska räknas med betryggande antaganden.

4 Problem och lösningsförslag

4.1 Premiesättning och reservberäkning

För att kunna sälja sjukförsäkring måste man ha en uppfattning om hur stora försäkringsutbetalningarna kommer vara och därmed hur stor premie man ska begära för att täcka alla kostnader. För detta behövs skattningar av hur många som kommer att utnyttja sin försäkring (insjukna) och i hur stor del dessa kommer att göra det (hur länge de kommer fortsätta vara sjuka). Insjuknandet ligger utanför ramarna för denna rapport men både för att skatta den och hur långa sjukfallen förväntas bli använder man sina tidigare erfarenheter; de uppgifter man har om hur det varit förut ligger till grund för det man tror om framtiden.

För försäkringstagarnas trygghet regleras all försäkringsverksamhet av stränga regler. Bland annat kontrolleras det löpande att varje bolag har tillräckliga medel för att kunna betala ut det de har lovat. Det kapital man måste ha till framtida utgifter kallas för reserven.

Pengarna i reserven måste investeras med eftertanke för att man inte ska riskera att vara utan likvida medel när det är dags för utbetalning. Man måste alltså både se till att ha tillräckligt stor reserv och att placera den klokt. Men att placera med låg risk innebär även ofta låg avkastning vilket gör att man ändå inte vill övervärdera sina åtaganden. Man vill ha en så liten reserv som möjligt så att överskottet kan investeras mer riskfyllt. Sammanfattningsvis kan man säga att man bör ha en så stor reserv att man klarar sina utbetalningar samtidigt som den är så liten som möjligt för chans till god avkastning.

4.2 Reservens beståndsdelar

Varje försäkringsgren måste bära sig själv ekonomiskt vilket betyder att en gren som går med vinst inte får subventionera en annan gren som går med förlust. Varje gren har därför bland annat sin egen reserv. Det görs även fler uppdelningar efter det; reserven består av

1. Reserv för kända skador

   Reserven för kända skador är reserven för de sjukfall som har anmälts och som man i nuläget betalar ersättning för.

2. Reserv för okända skador

   Reserven för okända skador är reserven för de skador som man tror redan har inträffat men som inte har hunnit anmälas till försäkringsbolaget ännu.

3. Reserv för skadebehandlingskostnader

   Reserven för skadebehandlingskostnader är summan man måste ha för att ha råd med det administrativa arbetet kring skadorna som ersätts.

Efter detta skall man som tidigare nämnts även inom varje grupp dela upp reserven på tjänstepensionsförsäkring och övrig livförsäkring.

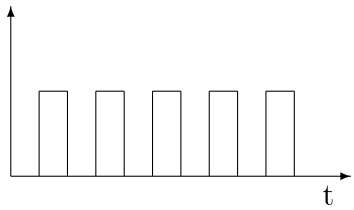
Detta arbete fokuserar på hur man får ut en så exakt reserv för kända skador som möjligt (både tjänstepensionsförsäkring och övrig livförsäkring).

4.3 Reserven för kända skador

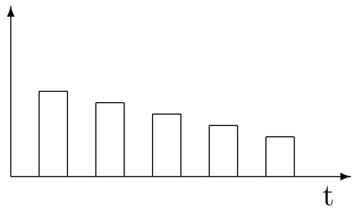
En person som är sjuk idag har en möjlig ersättning på hela sitt månadsbelopp tills försäkringen upphör (vanligtvis vid 65 års ålder). Reserven är summan av staplarna i figur 2.

Eftersom vi vet att vissa sjuka personer blir friska igen och andra dör så är det inte realistiskt att tro att en sjuk person kommer att vara sjuk ända tills försäkringen upphör. Något sorts antagande måste därför användas för hur fort sjukavvecklingen förväntas ske. Detta antagande kallas för avvecklingsfunktionen. När man räknar med att sjukfallet kan avvecklas ser diagrammet ut som i figur 3:

Slutligen bör man räkna med att man får avkastning på de pengar man har i reserven vilket betyder att man inte behöver ha hela summan från

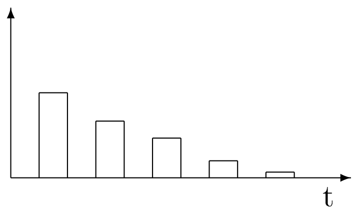


Figur 2: Möjlig ersättning



Figur 3: Förväntad ersättning

början. Genom diskontering av utbetalningarna som ligger i framtiden kan man se hur mycket av pengarna man måste ha i reserven idag, se figur 4.



Figur 4: Diskonterad förväntad ersättning

På detta sätt har vi fått reserven att bli mindre och mer realistisk. Som man kan se är avvecklingsfunktionen en viktig del i detta.

Tidigare har man beräknat reserven enligt en kontinuerlig modell. Beräkningen såg då ut så här:

$$B \cdot \frac{1}{\lambda(x,t_0)} \cdot \int_{t_0}^{z-x} \lambda(x,t) \cdot e^{-\delta(t-t_0)}\, dt$$

B är det årliga beloppet som betalas ut, x är åldern vid insjuknandet, z är slutåldern för försäkringen (oftast 65), λ är avvecklingsfunktionen, δ är den antagna ränteintensiteten som används för diskonteringen och t_0 är hur lång tid som gått sedan insjuknandet till nu.

Att räkna enligt kontinuerlig modell har varit en approximation av utbetalningar som ju i grunden är diskreta. På grund av ökad datorkraft är det nu möjligt att beräkna reserven diskret. I och med nya föreskrifter som säger att man ska använda räntekurva istället för en fast ränta har det också varit enklast att börja räkna diskret.

Reservberäkningen ser med diskret beräkning ut så här:

$$\frac{B}{12} \cdot \frac{1}{\lambda(x, t_0)} \quad \cdot \quad \sum_{n=1}^{12(z-x)} \lambda(x, t_0 + \frac{n}{12}) \cdot e^{-(\delta_n \cdot \frac{n}{12})}$$

δ_n är ränteintensiteten från punkt n i räntekurvan.

I och med övergången till diskret beräkning behöver alltså inte avvecklingsfunktionen vara analytiskt integrerbar längre.

4.4 Definition av avvecklingsfunktionen

Avvecklingsfunktionen är den betingade sannolikheten att en person kvarstår som sjuk t år efter insjuknandet givet att denne insjuknade vid åldern x år. Avvecklingsfunktionen betecknas $\lambda(x, t)$ ("lambda").

En mer matematisk definition:

Ett sjukdomsfalls totala längd antas vara ett utfall av den stokastiska variabeln T_i för person i, i=1,2,...n. Variablerna $T_1, T_2, ..., T_n$ antas vara oberoende och likafördelade och anta värden i $(0, \infty)$ med fördelningsfunktion F och täthetsfunktion f. Avvecklingsfunktionen är då

$$\lambda(x, t) = P(T > t | X = x)$$

4.5 Dagens avvecklingsfunktion

Den avvecklingsfunktion som används i någon variant av de flesta försäkringsbolagen i Sverige idag kom bland annat till för att man ville ha en funktion som var analytiskt integrerbar. Detta krävdes för att kunna räkna ut reserven när den beräknades kontinuerligt. Funktionen varierar beroende på ålder vid insjuknande, hur lång tid som gått sedan insjuknande och på om det gäller tjänstepensionsförsäkring eller övrig livförsäkring. Funktionen kan till exempel se ut så här:

$$\lambda(x,t) = a_x \cdot e^{-80t} + b_x \cdot e^{-13t} + c_x \cdot e^{-4.5t} + d_x \cdot (0.15 \cdot e^{-0.3t} + 0.85 \cdot e^{-0.01t})$$

där

$a_x = 1 - b_x - c_x - d_x,$
$b_x = 0.12,$
$c_x = 0.006 \cdot e^{0.04x}$
$d_x = 0.00065 + 0.000018 \cdot e^{0.13x}$

Det är alltså fem stycken exponentialfunktioner i avvecklingsfunktionen och man kan se det som att varje exponentialfunktion representerar en viss sorts sjukdom som har sin egen avvecklingstakt. Siffrorna 80 och så vidare är själva snabbheten i avvecklingen för just den sjukdomen. Eftersom det är exponentialfördelat är väntevärdet för den snabbaste avvecklingen $1/80 = 0.0125$ år, dvs ca 4.5 dagar. Den långsammaste avvecklingen är 0.01 vilket ger ett väntevärde på 100 år. 100 år kan verka för mycket men det är för att få rätt avtrappning på de längsta fallen, det blir rätt eftersom antagandet bara gäller fram till 65 års ålder. Se sammanställning i tabell 1.

Värde i formeln	Ungefärligt väntevärde
80	4.5 dagar
13	28 dagar
4.5	2.5 månader
0.3	3.3 år
0.01	100 år

Tabell 1: Avvecklingsfunktionens olika sjukfallslängder.

Ingen hänsyn tas alltså till om det specifika sjukfallet har en svår diagnos eller ej. Detta vore naturligtvis bra men skulle göra det hela än mer komplicerat.

Variablerna $a_x - d_x$ är sannolikheten att personen har fått just den sortens sjukdom som ger tillhörande avvecklingstakt. Det är bara dessa variabler som beror på åldern, man kan se att ju högre ålder (större x), desto mer sannolikt är det att hamna i någon av de tre sista exponentialfunktionerna med de långsammaste avvecklingstakterna, se tabell 2.

Ålder	a_x	b_x	c_x	d_x
25	0.72	0.12	0.16	0.001
40	0.58	0.12	0.30	0.004
60	0.17	0.12	0.66	0.045

Tabell 2: Sannolikheten att tillhöra de olika avvecklingstakterna beror på ålder. Ju äldre man är desto större sannolikhet att tillhöra exponentialfunktionerna med långsammare avveckling.

Idag har de tekniska förutsättningarna gjort att avvecklingsfunktionen inte behöver vara analytiskt integrerbar. En annan anledning att byta avvecklingsfunktion är att man tidigare haft mycket kortare karenser i försäkringarna. Man har haft så korta karenstider som sju dagar i många försäkringar medan den kortaste karensen i försäkringar som säljs idag är en månad. Detta påverkar naturligtvis också funktionens utseende, de flesta sjukfallen avvecklas ju i början.

4.6 Lösningsförslag

Matris

Eftersom förutsättningarna för användandet av avvecklingsfunktionen har ändrats har det nu öppnats nya möjligheter att ha en avvecklingsfunktion som ser ut på ett helt nytt sätt. Man behöver inte ens begränsa sig till att det måste vara en formel, ett annat förslag skulle istället kunna vara att använda sig av en matris. Denna matris skulle då innehålla olika värden för sannolikheten att kvarstå som sjuk beroende på valfria variabler. Detta skulle ge ett visst antal matriser, hur många beror på hur många variabler man tar med. I ett enkelt fall då bara ålder vid insjuknande och tid sedan insjuknande är med skulle det kunna se ut som i tabell 3.

Ålder |Tid	t_1	t_2	$\dots$	t_n
x_1	λ_{11}	λ_{12}	$\dots$	λ_{1n}
x_2	λ_{21}	λ_{22}	$\dots$	λ_{2n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_m	λ_{m1}	λ_{m2}	$\dots$	λ_{mn}

Tabell 3: Avvecklingen beroende på ålder vid insjuknande och tid sedan insjuknande.

Om man använder sig av en matris krävs en viss utjämning av resultatet för att undvika att sannolikheterna varierar orealistiskt på grund av slumpen. Detta för att talen λ_{ij} kommer att bygga på skattningar gjorda från ett begränsat datamaterial.

Formel

Ett annat alternativ är att även fortsättningsvis använda en formel men att anpassa den till det nya datamaterialet. Eftersom de två första exponentialfunktionerna i den nuvarande avvecklingsfunktionen gäller för riktigt korta sjukfall kan man tänka sig att en funktion med färre exponentialfunktioner skulle passa bra nu när karenserna är längre. En funktion med tre exponentialfunktioner skulle då se ut så här:

$$\lambda(x,t) = a_x \cdot e^{-p_1 \cdot\ t} + b_x \cdot e^{-p_2 \cdot\ t} + c_x \cdot e^{-p_3 \cdot\ t}$$

5 Genomförande och resultat

5.1 Datamaterialet

Det datamaterial som har använts till den här studien kommer från Folksam Livs register över sjukfall inom individuell sjukförsäkring.

Observera att den här sortens undersökning skiljer sig ifrån “vanliga” statistiska undersökningar på så sätt att man här inte gör ett urval av data. Eftersom det är avvecklingen för just Folksam Livs bestånd vi vill skatta så är det data vi har alla utfall som finns. Sedan är det ändå naturligtvis bättre ju större datamaterial man har eftersom slumpen kan inverka på så sätt att det framkommer att framtida fall inte riktigt har samma avveckling som det man nu ser.

Undersökningen täcker alla sjukfall som varit pågående från och med 1995-01-01 till och med 2007-12-31. Sjukfallen kan både ha startat tidigare och varit pågående vid undersökningens slut, så länge de har varit pågående någon gång under undersökningstiden.

Vissa observationer tillhörde försäkringar utan karensinformation, dessa är uteslutna ur undersökningen. De försäkringar som bara ger ersättning under sjukpenningtid har också tagits bort eftersom man då inte kan se om personen verkligen har tillfrisknat eller om den bara har blivit beviljad sjukersättning.

Efter denna reduktion innehåller datamaterialet 8747 observationer av sjukfall varav 5577 stycken avvecklades under undersökningsperioden. Datamaterialet består av 4707 kvinnor och 4040 män. 775 observationer tillhör tjänstepensionsförsäkring och 7972 övrig livförsäkring. För mer detaljerad information, se tabell 4.

Försäkringstyp \ Kön	Kvinnor	Män	Totalt
Tjänstepensionsförsäkring	183	592	775
Övrig livförsäkring	4524	3448	7972
Totalt	4707	4040	8747

Tabell 4: Antal sjukfall uppdelat på försäkringstyp och kön.

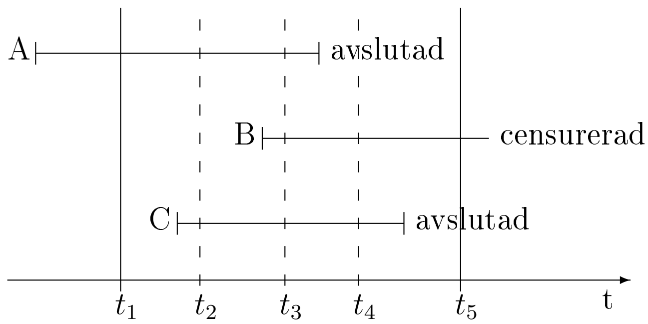
Datamaterialet innehåller mer övrig livförsäkring än tjänstepensionsförsäkring vilket gör det svårare att göra bra skattningar för tjänstepensionsförsäkring. Det skulle ha varit bra med mer data, särskilt för gruppen kvinnor med tjänstepensionsförsäkring. Ytterligare data som inte har funnits att tillgå nu kommer senare att kunna inhämtas från andra verksamheter inom Folksam. Detta material innehåller typen tjänstepensionsförsäkring så bättre studier av denna variabel kommer att kunna göras i framtiden.

Medelsjuklängden bland sjukfallen var 39 månader och medianen 17 månader vilket visar att *en* exponentialfördelning inte passar som approximation. Förklaring: Utgå från en stokastisk variabel X som är exponentialfördelad med parameter λ. Dess väntevärde är $1/\lambda$. Härled medianen genom att lösa ekvationen $P(X > x) = 1/2$, som ju ger $e^{(-\lambda x)} = 1/2$. Om man sen sätter upp kvoten mellan väntevärde och median så ser man att kvoten är ett och samma tal för alla exponentialfördelningar, nämligen $\ln 2 \approx 0.7$. Om *en* exponentialfördelning hade passat så borde alltså medianen varit ungefär $39 \cdot \ln 2 \approx 27$. Däremot kan fortfarande *flera* exponentialfunktioner, som i lösningsförslaget ovan, passa bra in på data.

Försäkringsfallen tillhör olika sorters sjukförsäkring och de kan ha olika villkor och karenser. Det är inte säkert att allt som gör skillnad för hur länge man är sjuk går att mäta och finns med i de underlag man kan ta fram men det går att se mycket: Datamaterialet innehåller information om bland annat insjuknandetidpunkt, avslutstidpunkt, avslutsorsak, kön, karens, ålder vid insjuknande, invaliditetsgrad och om det är en tjänstepensionsförsäkring eller ej.

5.2 Jämförelse av olika sjukfall

Sjukfallen som ingår i den här studien har inträffat under en lång tidsperiod. Några har hunnit hålla på i flera år medan andra precis påbörjades innan undersökningstidens slut. Låt oss titta på ett exempel, se även illustration av exemplet i figur 5.



Figur 5: Sjukfallens inträffande i tiden.

Säg att vår undersökning startar vid tiden t_1 och avslutas vid tiden t_5. Vid t_1 har sjukfall A redan pågått ett tag. Vid t_2 har även sjukfall C påbörjats och vid t_3 är alla tre sjukfallen igång. Vid t_4 har sjukfall A av någon anledning avslutats och vid t_5 har även sjukfall C avslutats. Sjukfall B hinner inte avslutas under undersökningsperioden och blir därför censurerad, vi kommer inte veta hur långt det sjukfallet egentligen blev.

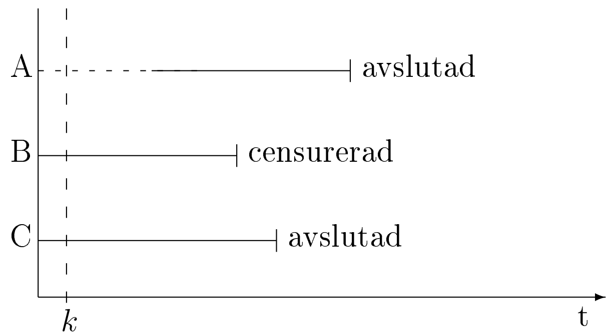
Om ett sjukfall avslutats genom tillfrisknande eller dödsfall vet man sjukfallets hela längd, annars kan man bara se en del av sjukfallet.

För att kunna jämföra sjukfallens längd måste man först och främst se till att man kan se skillnad mellan ett sjukfall som avslutats genom

1. tillfrisknande eller dödsfall

2. att undersökningsperioden tagit slut eller att individen uppnått försäkringens slutålder.

Om man inte kunde skilja på dessa skulle resultatet visa alldeles för korta sjukfall, de som avslutats av anledning 2 är ju egentligen längre än vad vi kan se. På samma sätt finns det även sjukfall som uppstod innan undersökningens början. I det fallet måste man tänka på att det kan ha funnits fler fall som uppstod samtidigt men som hann avslutas innan undersökningen började och som därför inte kan observeras. Likadant är det med den karens som försäkringen har, inga sjukfall som är kortare än karensen kommer med i studien. Har man inte det i åtanke visar resultatet för långa sjukfall. För att komma runt de här problemen utan att behöva ta bort sjukfall ur studien tittar man på något som kallas för riskfunktionen, R. Denna visar hur många av fallen som var under risk att avvecklas under en viss tidpunkt. Man börjar med att sortera om sjukfallen efter hur lång tid de pågått, samma sjukfall som innan kan nu ses på nytt sätt i figur 6.

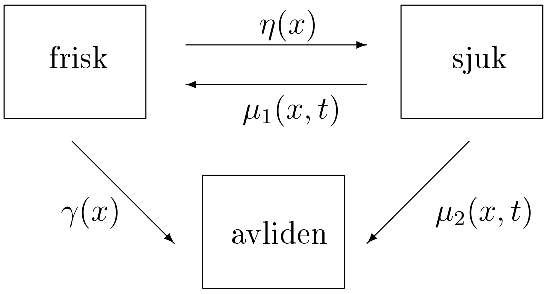


Figur 6: Jämförelse av sjukfallens längd

På det här sättet kan man se hur långa sjukfallen är i förhållande till varandra samt hålla reda på hur lång tid av sjukfallet som var under undersökningsperioden. Under tid före undersökningen (streckad linje i figuren) samt under karenstiden var fallet inte under risk att avvecklas (för hade det avvecklats då så hade vi inte fått med det i undersökningen). Inget av fallen var alltså under risk före tidpunkt k då karensen upphörde.

5.3 Skattning av avvecklingsfunktionen

Man kan se på sjukförsäkring som om människor kan befinna sig i olika tillstånd; antingen är man frisk, sjuk eller avliden. Se tillståndsfiguren; figur 7.



Figur 7: De olika tillstånden.

Från att vara frisk kan man övergå till att bli sjuk eller avliden, från sjuk kan man bli frisk igen eller avlida. Är man däremot avliden kan man inte övergå till något annat. Övergångsintensiteterna betecknas $\eta(x)$, $\mu_1(x,t)$, $\mu_2(x,t)$ respektive $\gamma(x)$. I denna studie är vi intresserade av $\mu_1(x,t)$ och $\mu_2(x,t)$, att gå från sjuk till något av de andra tillstånden. Med andra ord är vi intresserade av $\mu_1(x,t) + \mu_2(x,t) = \mu(x,t)$; avvecklingsintensiteten.

Nu vidare till sambandet mellan avvecklingsintensiteten och avvecklingsfunktionen. Förutsättningarna för definitionen av avvecklingsfunktionen lyder som skrivits ovan:

Ett sjukdomsfalls totala längd antas vara ett utfall av en stokastisk variabel , T_i, för person i, i=1,2,...n. Variablerna $T_1, T_2, ..., T_n$ antas vara oberoende och likafördelade, ta värden i $(0, \infty)$ och ha fördelningsfunktion F och täthetsfunktion f.

Eftersom avvecklingsfunktionen, $\lambda(t)$, är måttet på hur stor sannolikhet det är att *inte* ha tillfrisknat ser sambandet ut så här:

$$\lambda(t) = 1 - F(t)$$

där

$$F(t) = P(0 < T \leq t)$$

Avvecklingsintensiteten, μ, definieras som

$$\mu(t) = \lim_{h \to 0^+} \frac{1}{h}\, P(t < T \leq t + h \mid T > t).$$

$\mu(t)$ är då sannolikheten att en person avvecklas i det korta tidsintervallet $(t, t+h)$ givet att personen var sjuk vid t. Två personer antas inte kunna avvecklas vid samma tidpunkt.

Vidare får man

$$\mu(t) = \lim_{h \to 0^+} \frac{1}{h} \frac{P(T \leq t+h) - P(T \leq t)}{P(T > t)} =$$

$$= \frac{1}{P(T > t)} \lim_{h \to 0^+} \frac{P(T \leq t+h) - P(T \leq t)}{h} =$$

$$= \frac{1}{\lambda(t)} \lim_{h \to 0^+} \frac{F(t+h) - F(t)}{h} = \frac{F'(t)}{\lambda(t)}$$

Eftersom $F(t) = 1 - \lambda(t)$ och därför $F'(t) = -\lambda'(t)$ blir

$$\mu(t) = \frac{-\lambda'(t)}{\lambda(t)}$$

Sambandet $\frac{d}{dt}ln(g(t)) = \frac{g\ '(t)}{g(t)}$ ger

$$\mu(t) = -\frac{d}{dt}(ln(\lambda(t)))$$

$$\Leftrightarrow$$

$$\int_0^t -\mu(s)ds = ln(\lambda(t))$$

$$\Leftrightarrow$$

$$e^{-\int_0^t \mu(s)ds} = \lambda(t)$$

λ ser då ut så här:

$$\lambda(t) = e^{-\beta(t)}$$

där

$$\beta(t) = \int_0^t \mu(s)\ ds$$

Vi behöver alltså kunna skatta μ för att kunna skatta λ. För detta har i det här arbetet använts Nelson-Aalenskattningar. Dessa bygger på martingalteori och använder sig av antal personer som är under risk att avvecklas vid varje tidpunkt, R(t). Se appendix A för sammanfattning kring martingaler och appendix B för fördjupning om Nelson-Aalenskattning.

λ skattas på detta sätt som

$$\hat{\lambda}(t) = e^{-\sum \frac{1}{R(\tau_j)}}\ ,\quad \tau_j \leq t$$

där $\tau_1 < \tau_2 < ...$ är de tider då det skett avveckling.

På grund av att skadelängderna för de olika sjukfallen inte kan mätas exakt utan måste mätas i antal dagar inträffar det då och då att olika sjukfall har precis lika lång skadelängd. Detta kan hanteras exempelvis genom att räkna ut $\lambda(t)$ en gång för varje observation man har vid en specifik sjuklängd och sedan ta det senaste (lägsta) till den fortsatta uträkningen.

Riskfunktionen R(t) som används för uträkningen av $\lambda(t)$ måste anpassas till hur många olika förklarande variabler man vill använda sig av; endast de som tillhör samma grupp av antaganden ska räknas som under risk.

5.4 Val av variabler att uppdela efter

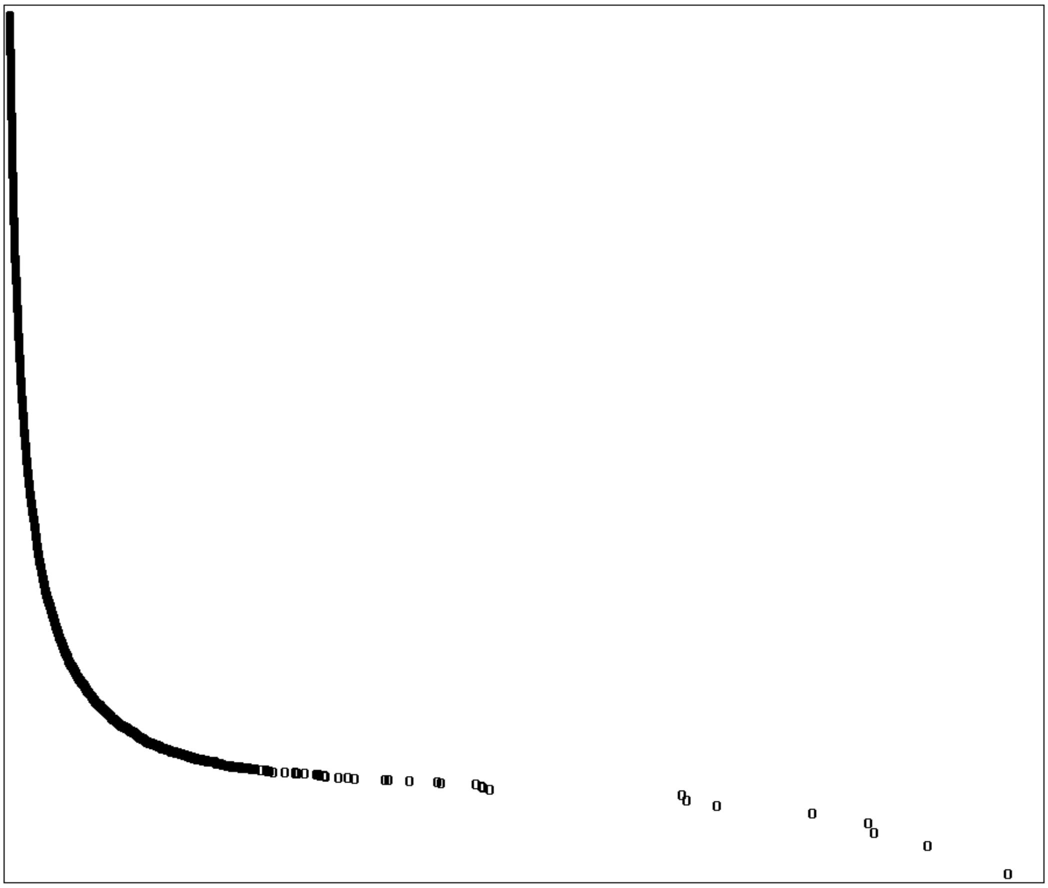
Försäkringsfallen skiljer sig åt i och med att de tillhör olika sorters sjukförsäkring och har olika villkor och karenser. Det är inte säkert att allt som ger skillnad för sjukfallslängden går att mäta och finns med i de underlag man kan ta fram. De uppgifter som finns att tillgå och som verkar vara av intresse är kön, karens, ålder vid insjuknande, invaliditetsgrad och om det är en tjänstepensionsförsäkring eller ej. Alla dessa variabler skulle kunna vara avgörande för hur länge en person är sjuk. Även om det kan finnas fler påverkande variabler så är de ovan nämnda hur som helst säkerligen de viktigaste.

Även om det finns många intressanta variabler måste man begränsa sig så att man har tillräckligt många observationer för att få säkra skattningar. För att kontrollera detta har jag tagit fram konfidensband, se appendix C för tillvägagångssätt. Konfidensbanden är nyttiga på så sätt att man lätt kan se i vilket intervall den riktiga avvecklingsfunktionen ligger (vid vald konfidensgrad). På så vis kan man se om noggrannheten är tillräcklig för det man vill uppnå. Är det inte tillräckligt bra måste man ta bort någon av sina uppdelningar så att det blir färre grupper med fler observationer i varje. Ingen undersökning om eventuell korrelation är gjord utan varje variabel är undersökt var för sig. Graferna som följer nedan har av sekretesskäl inga utskrivna axlar.

Lite om de olika variablerna:

- Skadelängd

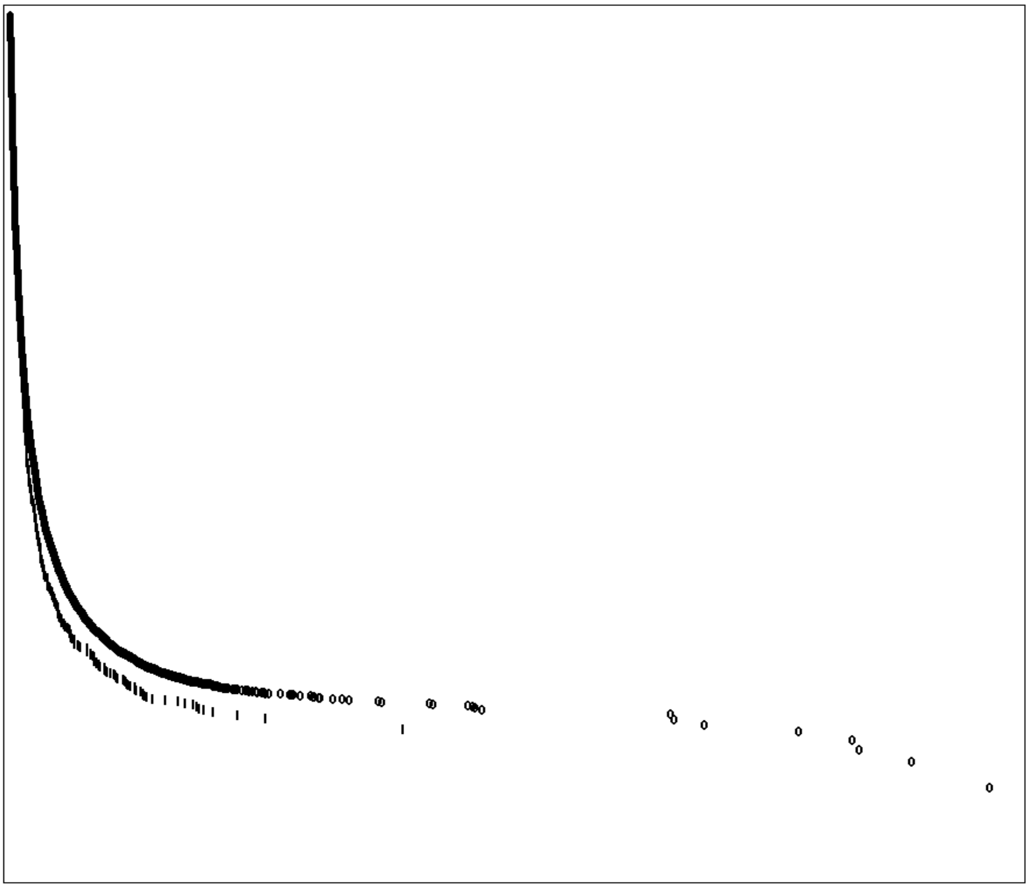
  Skadelängden, hur länge man varit sjuk, är grundläggande i den här sortens undersökning. Det är naturligt att om det bara gått en kort tid sedan insjuknande så är risken att man fortfarande är sjuk större. Däremot är antagligen avvecklingstakten större i början och mindre efter lång tid. Detta kan ses i figur 8 som att avvecklingsfunktionen hela tiden minskar (sannolikheten att vara sjuk minskar) samtidigt som lutningen på kurvan blir mindre brant (avvecklingstakten blir långsammare). För de riktigt långa sjukfallen kan man dock observera ytterligare en förändring av avvecklingstakten. Kurvans lutning blir i slutet större igen; avvecklingstakten ökar. Vad detta beror på är osäkert, ett förslag är att det är på grund av för få observationer med långa sjukfall men kanske också att dödligheten slår igenom mer för dessa eftersom personerna blivit äldre.



Figur 8: Avvecklingsfunktionens utseende vid en ökande skadelängd. Tid sedan insjuknande på x-axeln och λ på y-axeln.

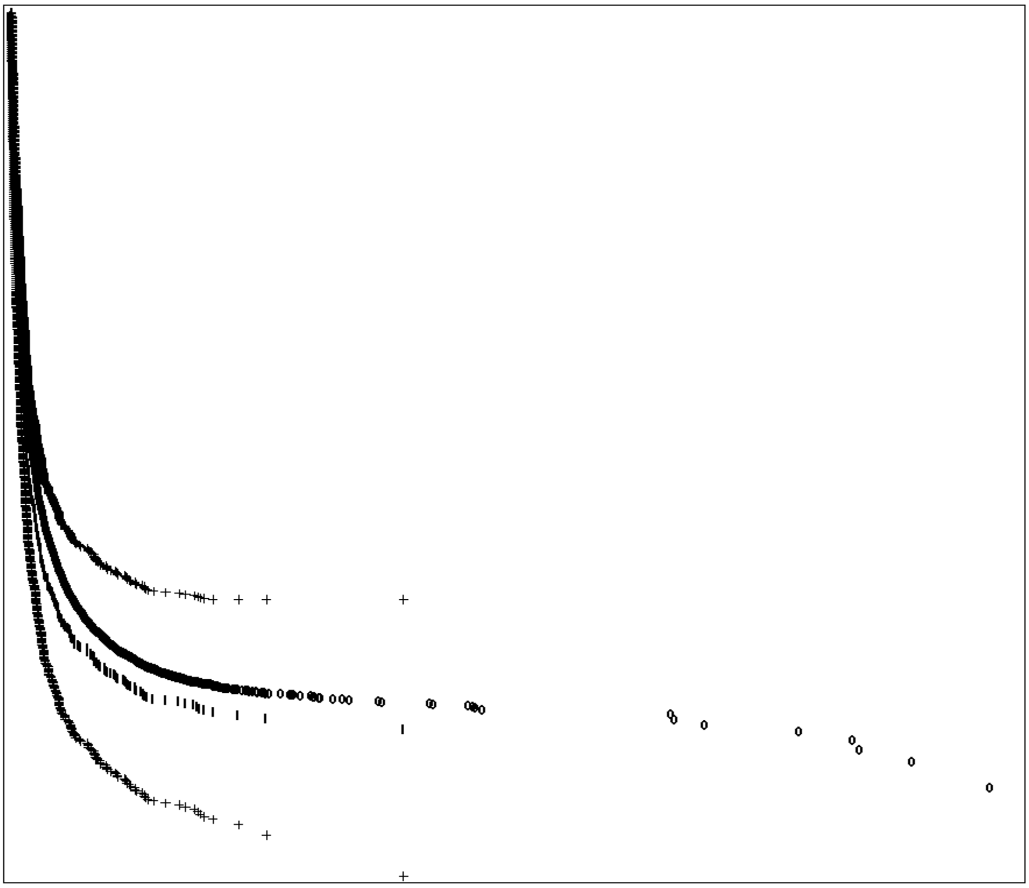
- Tjänstepension och övrig livförsäkring

Med tanke på att det är ett lagkrav att hålla isär tjänstepensionsförsäkring och övrig livförsäkring kan det kännas naturligt att låta den variabeln vara med i den utvalda modellen. I figur 9 kan man se att det även verkar finnas skillnader i sjuklängd mellan de två grupperna.



Figur 9: Avvecklingsfunktionen uppdelad på tjänstepension och övrig livförsäkring. O=Övrig livförsäkring, I=Tjänstepensionsförsäkring. Tid sedan insjuknande på x-axeln och λ på y-axeln.

Datamaterialet innehåller tyvärr mer övrig livförsäkring än tjänstepensionsförsäkring (se tabell 4 ovan) vilket gör det svårare att göra bra skattningar för tjänstepensionsförsäkring. Tittar man på konfidensbanden ser man att det inte går att vara säker på att de två grupperna verkligen skiljer sig åt, se figur 10.

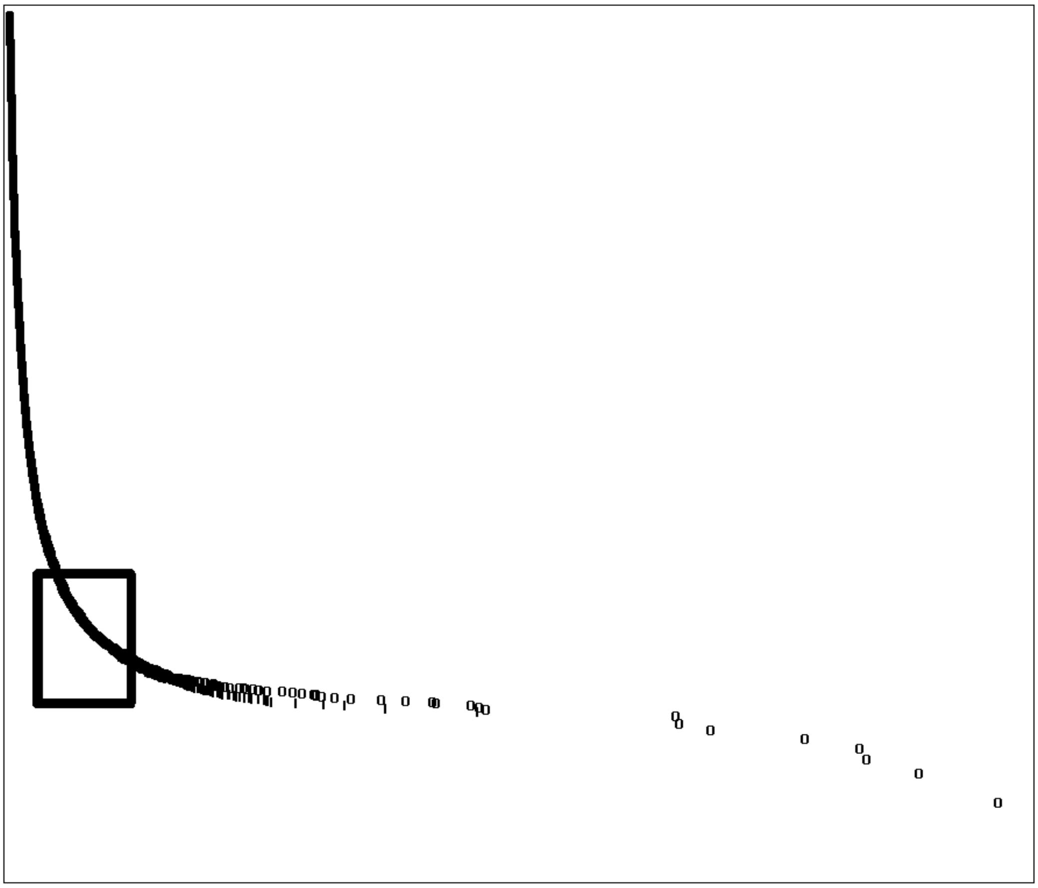


Figur 10: Konfidensband. O=Övrig livförsäkring I=Tjänstepensionsförsäkring +=Konfidensband för tjänstepensionsförsäkring. Man kan inte utesluta att tjänstepensionsförsäkringskurvan är lika med övrig liv-kurvan. Tid sedan insjuknande på x-axeln och λ på y-axeln.

Man kan naturligtvis särredovisa tjänstepensionsförsäkring och övrig livförsäkring även om man gör antagandet att de har samma avveckling, däremot skulle det rent intuitivt kännas bättre om man själv kunde tro på en skillnad mellan dessa bestånd. Det ser i data ut som en skillnad men det går inte att vara tillräckligt säker. Med endast detta datamaterial får man därför hoppa över att ta med variabeln.

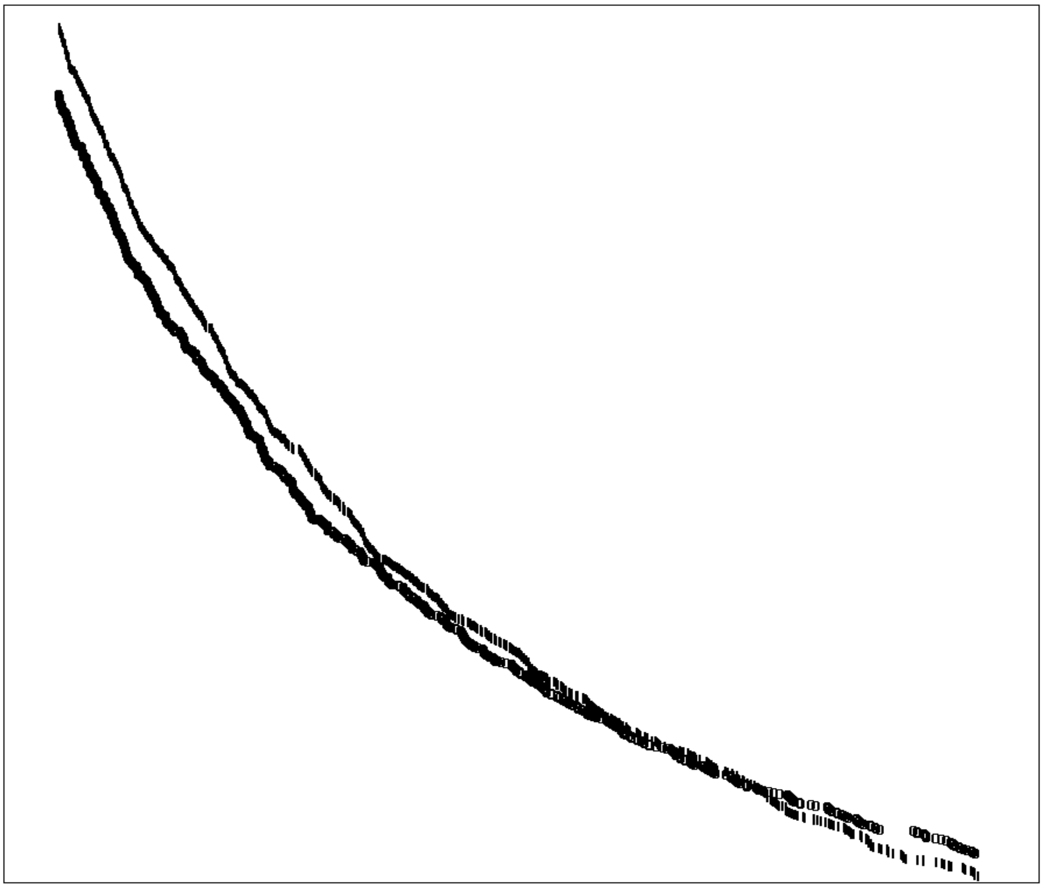
- Kön

  Skillnader mellan kön är naturligtvis viktiga att ta med om de finns men när man tittar på hela beståndet uppdelat på kön ligger de två funktionerna nära varandra. I figur 11 går det inte att se mycket skillnad mellan de två kurvorna. Den del av bilden som ligger inom rutan finns uppförstorad i figur 12. Där kan man se att funktionerna till och med korsar varandra.



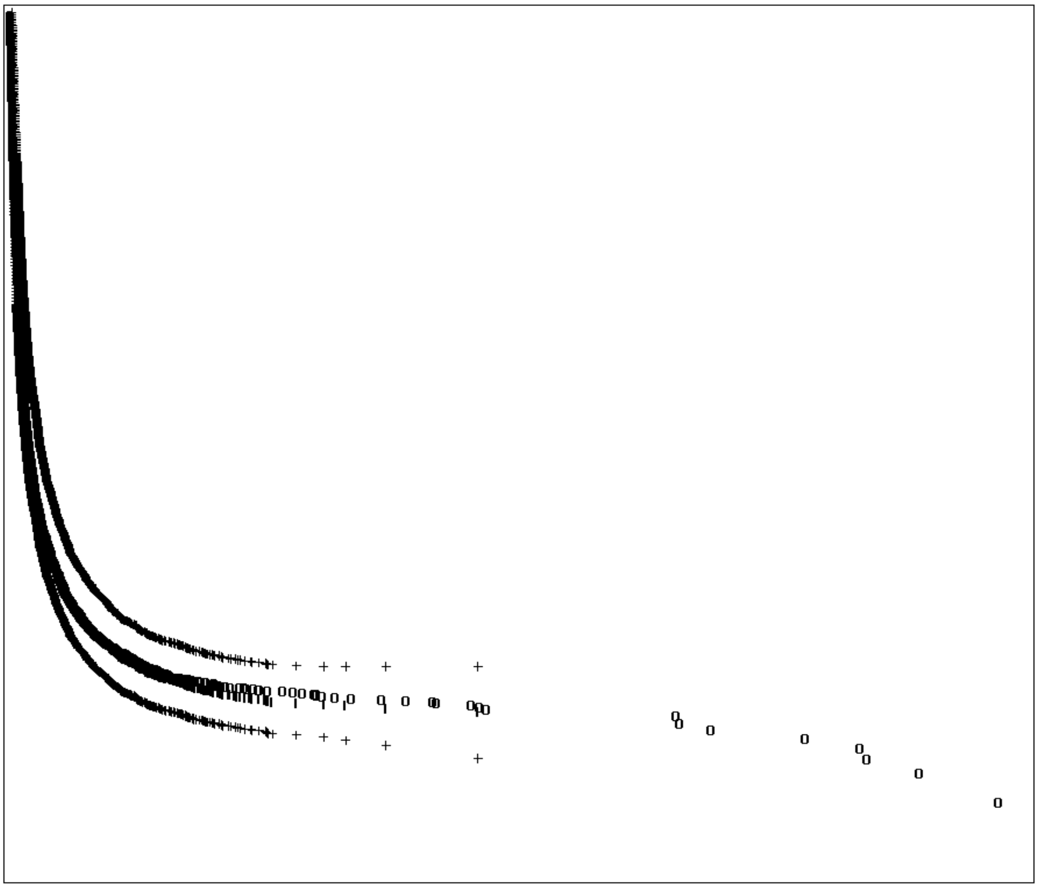
Figur 11: Avvecklingsfunktionen uppdelat på kön. I=Kvinnor, O=Män. Tid sedan insjuknande på x-axeln och λ på y-axeln. Se förstoring av rutan i nästa bild.

Skillnaderna är så små att det inte går att dra slutsatsen att de skiljer



Figur 12: Utvald del av avvecklingsfunktionen uppdelat på kön. I=Kvinnor, O=Män. Tid sedan insjuknande på x-axeln och λ på y-axeln.

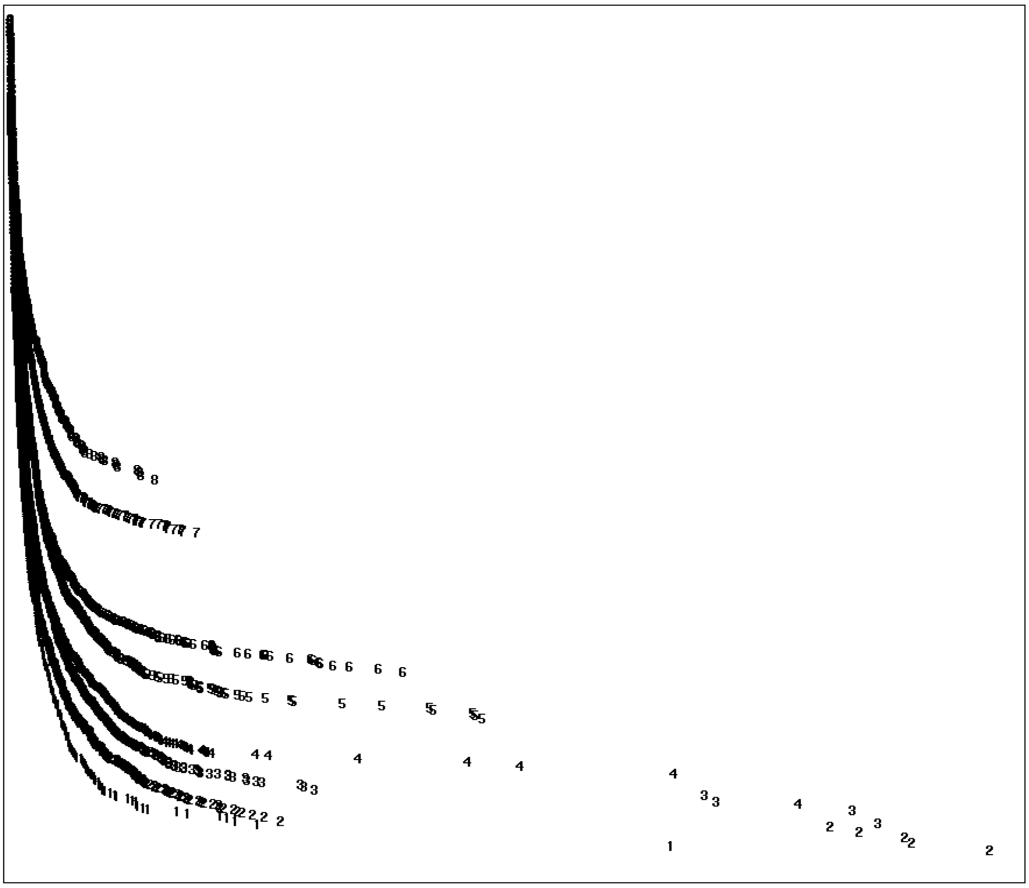
sig åt, se konfidensband för kvinnor i figur 13.



Figur 13: Konfidensband kön. I=Kvinnor, O=Män, +=konfidensband för kvinnor. Variabeln verkar inte vara meningsfull att använda. Tid sedan insjuknande på x-axeln och λ på y-axeln.

- Ålder vid insjuknandet

  Studien innehåller information om sjukfall som inträffat i åldrarna 25-65. Vilken ålder man hade när man insjuknade har tidigare använts i avvecklingsfunktionen och det verkar rimligt att anta att en äldre person har svårare att tillfriskna. Misstanken bekräftas när man tittar på datamaterialet. I figur 14 ses den stora skillnaden i avveckling mellan olika åldrar.



Figur 14: Avvecklingsfunktionen vid olika insjuknandeåldrar. Åldrarna är uppdelade i femårsintervall från 25-29 år (1) till 60-64 år (8). Tid sedan insjuknande på x-axeln och λ på y-axeln.

- Karens

Det finns många olika karenstider i försäkringarna, det beror på att de följer de nivåer som finns/funnits i andra försäkringssystem. I datamaterialet finns allt från fall med karens på sju dagar till andra med tolv månaders karens samt de med rörlig karens (R-karens) som gäller fram till beviljad sjukersättning. Inga andra fall än de som uppfyllt sin karens syns i statistiken.

Vilken karens man har på sin försäkring kan påverka avvecklingstakten. Detta beror på att en lång karens uppmuntrar till tidigt tillfrisknande, det är helt enkelt svårt att vara utan ersättning så länge som krävs. Har man däremot en kort karens är man mer benägen att vara hemma längre eftersom man klarar sig bättre ekonomiskt.

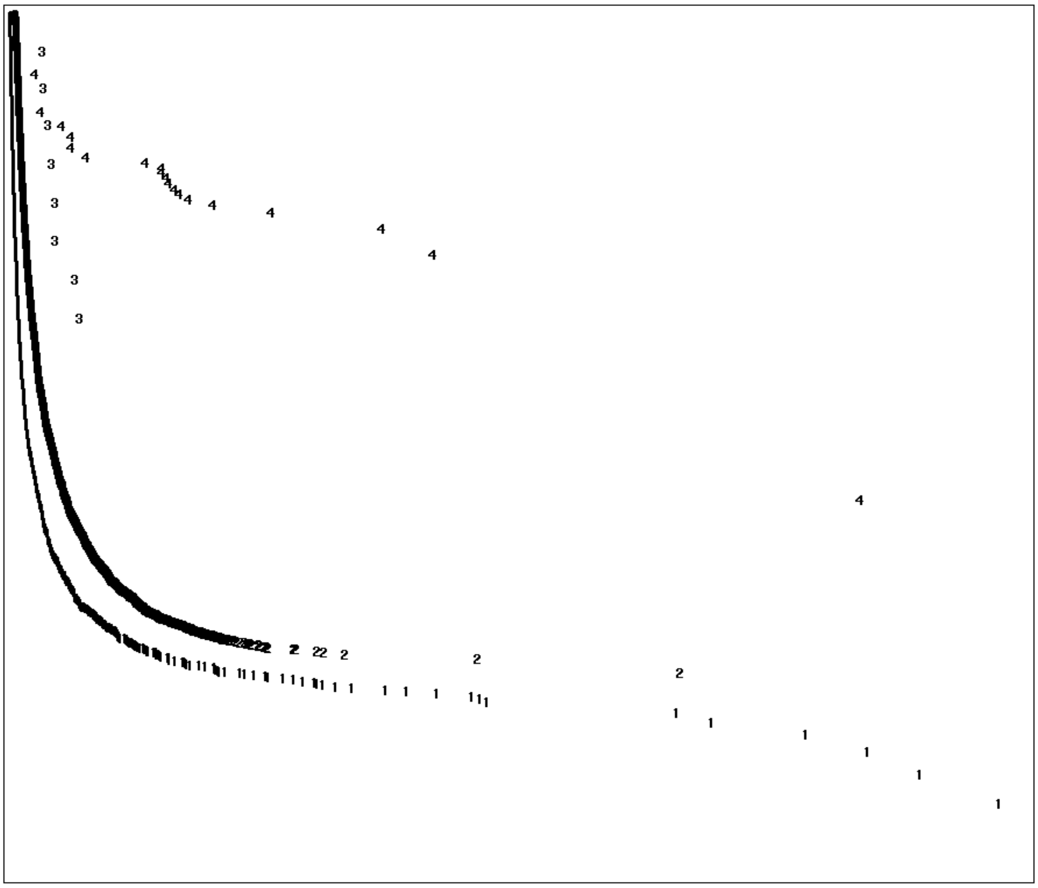
Många karenser har försvunnit på grund av att den allmänna sjukförsäkringen har ändrats. De försäkringar som fortfarande säljs har karensen en månad, tre månader, tolv månader (sällan) eller R-karens.

Att dela upp i grupper beroende på vilken karens man har på försäkringen ger tyvärr en kraftig reduktion av tillförlitligheten av skattningarna eftersom det blir för få observationer i vissa grupper. Detta kan ses i figur 15.

En annan anledning till att inte ta med de olika karenserna i modellen är också att de mer visar historien, de speglar inte karenserna till de försäkringar som säljs idag och kommer därmed att bli mer och mer oviktiga för helheten. Att bara dela upp på de grupper där det finns många försäkringar ger en skev bild eftersom de andra mer ovanliga karenserna då måste ingå i någon av dessa grupper. Slutsatsen blir alltså att det är bättre att inte använda karens som en variabel i funktionen.

- Invaliditetsgrad

Invaliditetsgraden, i hur stor del man är sjuk, räknas i fyra steg; man kan vara 25%, 50%, 75% eller 100% sjuk. I teorin borde invaliditetsgraden inverka på sannolikheten att bli helt frisk. En person som är så pass frisk att den fortfarande arbetar till någon del borde ha större sannolikhet att bli frisk än någon som är helt sjuk. I praktiken är det tyvärr inte så lätt att se dessa skillnader. I vissa villkor räknas det som om den försäkrade har blivit frisk så fort invaliditetsgraden går ner till 25% (ersättningen upphör då) medan man i andra villkor endast räknas som frisk om invaliditetsgraden gått ner till noll. Dessutom varierar invaliditetsgraden ofta väldigt mycket upp och ner för sjukfallen medan



Figur 15: Avvecklingsfunktionen indelad i olika grupper beroende på karens. $1=$ Karens <1 månad $\quad 2=1$ månad $<$ Karens ≤ 3 månader $\quad 3=$ Karens >3 månader $\quad 4=$ R-karens. Tid sedan insjuknande på x-axeln och λ på y-axeln.

man i datamaterialet bara kan se vilken den senaste invaliditetsgraden är. På grund av detta har jag tagit bort denna information och räknar alla pågående sjukfall som om de har invaliditetsgrad 100%.

När man gör skattningar blir det mer tillförlitligt ju färre uppdelningar i olika grupper man gör. Eftersom mängden data är begränsad är det viktigt att ha så få variabler och därmed grupper som möjligt. De variabler som borde vara kvar med tanke på diskussionen ovan är ålder vid insjuknandet och skadelängd. Det vore egentligen önskvärt att göra två matriser/formler, en för tjänstepensionsförsäkring och en för övrig livförsäkring. Detta går dock inte på grund av för få observationer i kategorin tjänstepensionsförsäkring.

5.5 Anpassning till data med en matris

För att få ner antalet "rutor" i matrisen behöver även variablerna ålder vid insjuknande och tid sedan insjuknande delas upp i färre grupper. Ålder delas här upp i femårsintervall från 25 till 65 års ålder, det blir 8 grupper. Tid sedan insjuknande delas för de första två åren upp i tre-månadersintervall, för år tre till nio läggs de ihop årsvis och resterande observationer får höra till gruppen "tio år eller längre". På detta sätt får man 17 olika grupper för tid sedan insjuknande. Anledningen till att det finfördelas mer i början är att det är där avvecklingsfunktionen, λ, ändras mest och även att vi där har mycket data. Detta gör att det blir $8 \cdot 17 = 136$ rutor att fylla med varsitt skattat λ. För detta är medelvärdet bland observationerna som hör till samma ruta använt.

Resultatet lämnas av sekretess-skäl ej ut men de kan se ut som i tabell 5.

Tid |Ålder	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
0-3 mån	0,85	·	0,89	0,81	·	0,89	·	0,91
3-6 mån	0,78	0,75	0,78	0,74	0,89	0,79	0,76	0,89
6-9 mån	0,68	0,7	·	0,64	0,78	0,75	0,65	0,78
9-12 mån	0,60	0,52	0,68	0,59	0,65	0,67	0,60	·
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4-5 år	·	0,40	·	0,42	0,49	·	0,42	0,53
5-6 år	0,39	0,38	0,41	0,38	·	0,41	·	·
6-7 år	0,36	·	·	0,36	0,43	0,36	·	·
7-8 år	·	0,31	0,37	·	0,40	·	·	·
8-9 år	0,32	·	·	·	·	0,33	·	·
9-10 år	0,28	0,28	·	0,30	0,37	·	·	·
över 10 år	·	·	·	·	·	0,30	·	·

Tabell 5: Skattad avvecklingsmatris.

Som man kan se är det en prick på många platser i matrisen, det betyder att det inte fanns någon observation som hörde till den gruppen. Det finns få observationer i första raden, detta beror på att många karenser inte har upphört innan tre månader. De långa skadelängderna har av naturliga skäl också få observationer. Eftersom det inte finns data nog för att fylla hela matrisen krävs det att man finner en metod att så bra som möjligt skatta vad det ska stå i de tomma rutorna. Detta görs genom interpolation och extrapolation.

Interpolation

När man vill veta ett värde *mellan* två punkter använder man sig av interpolation. I det här fallet passar exponentiell interpolation bäst.
Låt oss anta att $\lambda(t)$ antar värdet A för t=0 och B för t=1. Ett värde $\lambda(u)$, 0<u<1, fås fram genom

$$\lambda(u) = e^{u \cdot ln(B) + (1-u) \cdot ln(A)}$$

Extrapolation

När man vill veta ett värde *utanför* intervallet av punkter använder man sig av extrapolation. Ett exempel på tillvägagångssätt är att låta nästa ruta få värdet i den innan minus den procentuella minskningen mellan de två tidigare rutorna. I matematiska termer blir det:
$\lambda_{i+2} = \lambda_{i+1} \cdot (1 - (\frac{\lambda_i - \lambda_{i+1}}{\lambda_i}))$

Efter interpolation och extrapolation ser matrisen ut ungefär som i tabell 6.

Observera att det för höga åldrar ska vara tomt i rutorna för de långa skadelängderna på grund av att man endast kan mätas som sjuk fram till 65 års ålder.

Som vi såg fanns det inte tillräckligt med observationer för att direkt fylla alla rutorna. Man kan nog även misstänka att vissa av de rutor som blev ifyllda från början inte är helt tillförlitliga eftersom de kanske inte baseras på så många observationer. Ibland kan en högre åldersklass verka ha ett lägre λ trots att man till exempel i figur 14 kunde se att äldre är sjuka under längre tid. Man får då utgå från sina egna förväntningar och utjämna matrisen, se tabell 7.

När matrisen är framtagen återstår endast att anpassa reservuträkningen så att rätt värde på λ hämtas ur matrisen. En fördel med att på detta sätt samla olika värden på λ i en matris är att det kan kännas mer överblickbart, man ser lättare vad värdena egentligen är och hur de varierar.

Tid |Ålder	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
0-3 mån	0,85	0,84	0,89	0,81	0,90	0,89	0,85	0,91
3-6 mån	0,78	0,75	0,78	0,74	0,89	0,79	0,76	0,89
6-9 mån	0,68	0,7	0,72	0,64	0,78	0,75	0,65	0,78
9-12 mån	0,60	0,52	0,68	0,59	0,65	0,67	0,60	0,69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4-5 år	0,42	0,40	0,43	0,42	0,49	0,43	0,42	0,53
5-6 år	0,39	0,38	0,41	0,38	0,47	0,41	0,40	·
6-7 år	0,36	0,35	0,38	0,36	0,43	0,36	0,39	·
7-8 år	0,34	0,31	0,37	0,33	0,40	0,35	0,35	·
8-9 år	0,32	0,30	0,34	0,32	0,39	0,33	0,34	·
9-10 år	0,28	0,28	0,32	0,30	0,37	0,31	0,33	·
över 10 år	0,25	0,27	0,30	0,29	0,36	0,30	·	·

Tabell 6: Helt ifylld avvecklingsmatris.

Tid |Ålder	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
0-3 mån	0,81	0,84	0,85	0,86	0,87	0,89	0,90	0,91
3-6 mån	0,74	0,75	0,75	0,76	0,78	0,79	0,89	0,89
6-9 mån	0,64	0,65	0,68	0,72	0,75	0,77	0,78	0,78
9-12 mån	0,52	0,55	0,56	0,60	0,62	0,64	0,65	0,69
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4-5 år	0,40	0,40	0,42	0,43	0,44	0,45	0,49	0,53
5-6 år	0,38	0,39	0,39	0,40	0,41	0,42	0,47	·
6-7 år	0,35	0,37	0,38	0,39	0,39	0,40	0,46	·
7-8 år	0,31	0,33	0,34	0,34	0,35	0,39	0,44	·
8-9 år	0,30	0,32	0,33	0,33	0,34	0,38	0,41	·
9-10 år	0,28	0,28	0,31	0,32	0,33	0,37	0,40	·
över 10 år	0,25	0,27	0,29	0,30	0,31	0,35	·	·

Tabell 7: Den utjämnade och därmed färdiga avvecklingsmatrisen.

5.6 Anpassning till data med en formel

Vid skattandet av en ny formel behövs som nämnts ovan inte de två första exponentialfunktionerna. Detta beror på att dessa mest avspeglar det som händer vid så små t att de numera hamnar inom karenstiden (eftersom karenstiderna är längre nu). En funktion med tre exponentialfunktioner är alltså anpassad till datamaterialet. Formeln kan på grund av sekretess inte redovisas här men har denna form:

$$\lambda(x,t) = a_x \cdot e^{-p_1 \cdot\ t} + b_x \cdot e^{-p_2 \cdot\ t} + c_x \cdot e^{-p_3 \cdot\ t}$$

där

$$a_x = 1 - b_x - c_x,$$
$$b_x = b_1 + b_2 \cdot e^{\ b_3 x},$$
$$c_x = c_1 + c_2 \cdot e^{\ c_3 x}$$

Variablerna b_i och c_i $(i = 1, 2, 3)$ behöver inte vara positiva utan kan rent teoretiskt både vara lika med noll och negativa, så länge den sammanlagda effekten blir att de som insjuknar vid äldre ålder har långsammare avveckling.

För att få fram den mest lämpade formeln måste man ändra parametrarna p_i, b_i och c_i $(i = 1, 2, 3)$ tills den passar så bra som möjligt till datamaterialet. För detta har jag använt minsta kvadrat-metoden, se appendix D.

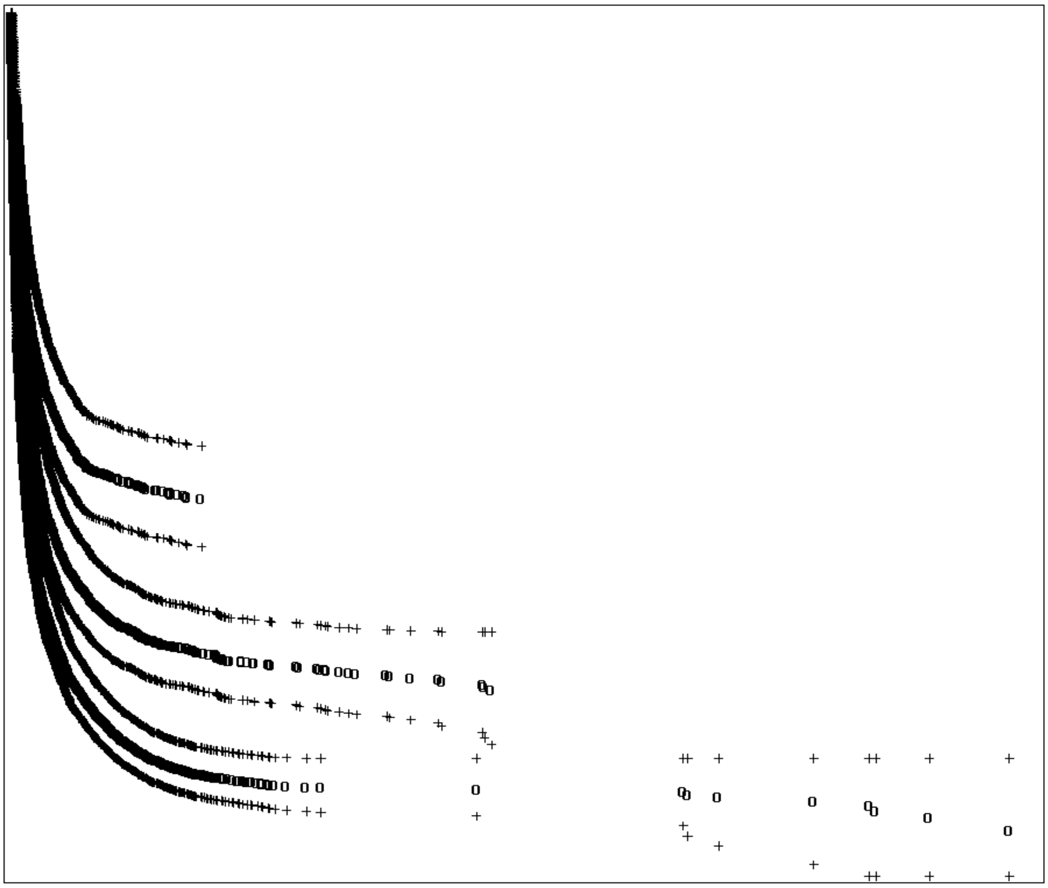
Fördelen med att använda en formel är att man får förändringar i resultatet även vid små ändringar i skadelängd och ålder vid insjuknande. Man kan dela in i färre grupper avseende till exempel ålder vid insjuknande men ändå få olika värden för alla åldrar. Detta beror på att man inte söker ut ett specifikt värde som när man har en matris utan att formeln alltid ändras med åldern (hur man än fick fram denna formel).

Så för att få med åldersberoendet utan att behöva få så osäkra skattningar kan man här dela in i färre grupper.

Om man än en gång tittar på figur 14 kan man se att det finns vissa funktioner som ligger närmare varandra än andra. Åldersgrupperna 1-4 verkar ligga i en grupp och likadant 5 och 6 respektive 7 och 8. Ska man välja färre grupper verkar dessa bra.

Resultatet av att dela in på detta sätt kan ses i figur 16. De 95%-iga Konfidensbanden visar att avvecklingsfunktionen verkligen beror på ålder. Detta är ett bra sätt att få en formel att passa datamaterialet, man får dock tänka på att alla förenklingar och sammanslagningar leder till mindre precision.

Detta verkar vara en uppdelning som både ger säkra skattningar och tar hänsyn till åldersberoendet.



Figur 16: Avvecklingsfunktionen uppdelad på tre åldrar med tillhörande konfidensband. O=avvecklingsfunktionen. +=konfidensband. Tid sedan insjuknande på x-axeln och λ på y-axeln.

6 Slutsats

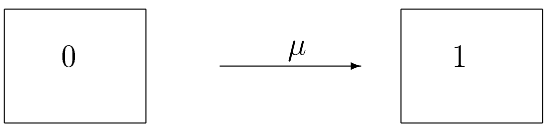
Det går att göra om avvecklingsfunktionen till både en matris och en ny formel. Med tanke på att man kan göra en formel mer känslig för små ändringar av variablerna så överväger fördelarna med att fortsätta ha en formel. Det går inte att göra två formler; en för tjänstepensionsförsäkring och en för övrig livförsäkring, på grund av för lite data. Med hjälp av data som sen ska ingå i studien kommer kanske detta att kunna göras senare. Endast en formel med variablerna ålder vid insjuknande och tid sedan insjuknande duger dock bra för att få formeln att passa datamaterialet. Variabeln ålder vid insjuknande skulle kunna delas upp i olika stora intervaller, det framkommer att i det här fallet är varianten med tre åldersgrupper bäst. Skattningarna gjorda med dennna indelning blir godtagbart säkra och ett förslag till ny avvecklingsfunktion är framtagen.

A Sammanfattande bakgrundsinformation om räkneprocesser och martingaler

En räkneprocess $\{N(t),\ 0 \leq t \leq \infty\}$ är en stokastisk process som räknar antal händelser som inträffat fram till tiden t. $N(0) = 0$ och $P(N(t) < \infty) = 1$. Vidare är processen högerkontinuerlig och varje hopp är av storleken 1. Låt F_{t-} vara historien; all information om inträffade händelser som finns att få fram till tiden t.

En multivariat räkneprocess $\{N = [N_1(t), N_2(t), ..., N_n(t)]\}$, har samma egenskaper som en enkel räkneprocess men man måste också anta att inga händelser kan inträffa samtidigt.

I vårt fall med en population av sjuka personer så kan bara en sak inträffa; personen kan avvecklas från sjukdomen. Att bli avvecklad betyder i det här fallet att man blir frisk eller dör. N(t) kan alltså anta värdena 0 (sjuk) eller 1 (avvecklad), se figur 17.



Figur 17: Övergång från sjuk till avvecklad.

μ är intensiteten med vilken den sjuke övergår till att vara avvecklad, därav kallas den avvecklingsintensiteten.

Ökningen dN(t) är antingen 0 eller 1 vid varje given tid t. För att en person ska kunna avvecklas precis vid tiden t gäller det att den inte redan har avvecklats och att observationen inte blivit censurerad innan t. Därför låter man R(t) vara en indikatorfunktion, då R(t)=1 är personen under risk och då R(t)=0 är personen inte under risk att avvecklas. Sannolikheten att person i ska avvecklas under det lilla tidsintervallet dt givet historien F_{t-} blir då

$$P(dN_i(t) = 1|F_{t-}) = \mu_i(t) \cdot R_i(t)\ dt$$

där $\mu_i(t)$ är avvecklingsintensiteten för person i. Vi antar att populationen är homogen, $\mu_1(t) = \mu_2(t) = ... = \mu_n(t)$ (alla i populationen har samma avvecklingsintensitet). Därmed får vi

$$P(dN_i(t) = 1|F_{t-}) = \mu(t) \cdot R_i(t)\ dt.$$

Låt

$$\alpha_i(t) = \mu \cdot R_i(t)$$

och

$$\alpha(t) = \alpha_1(t), \alpha_2(t), \dots, \alpha_n(t)$$

$\alpha(t)$ är då intensiteten för den multivariata räkneprocessen N.

Vi får då

$$\alpha(t) = \sum_{i=1}^{n} \mu \cdot R_i(t) = \mu \cdot R(t)$$

där R(t) är totala antalet sjuka personer under risk att avvecklas vid tiden t.

* * *

Definition av martingal

Den stokastiska processen X är en martingal om

- X är anpassad till F_{t-}
- $E[|X(t)|] < \infty$ för alla t
- $E[X(t)|F_{s-}] = X(s)$ för alla s och t där $s \leq t$.

Man kan säga att en martingal inte har någon systematisk drift.

En submartingal uppfyller de två första villkoren ovan men det tredje ändras till

- $E[X(t)|F_{s-}] \geq X(s)$ för alla s och t där $s \leq t$.

* * *

Låt nu $M(t) = N(t) - A(t)$ där $A'(t) = \alpha(t)$.

Eftersom

$$E[dN(t)|F_{t-}] = \alpha(t)\ dt$$

blir

$$E[dM(t)|F_{t-}] = 0.$$

På samma sätt ger

$$E[N(t)|F_{t-}] = A(t)$$

att

$$E[M(t)|F_{t-}] = 0.$$

M är då enligt definitionen ovan en martingal, närmare bestämt en submartingal. A(t) är kompensator till N(t), det innebär att A(t) är den enda funktion som gör $N(t) - A(t)$ till en martingal.

B Nelson-Aalenskattning

Vi vill finna en approximation till

$$\lambda(t) = e^{-\beta(t)}$$

där

$$\beta(t) = \int_0^t \mu(s)\ ds$$

Vid Nelson-Aalenskattning använder man martingalen M (se appendix A för mer information om räkneprocesser och martingaler):

N(t) är antalet personer som har tillfrisknat vid tiden t. Då är dN(t) ökningen precis vid t. Låt nu

$$dM(t) = dN(t) - \alpha(t)\ dt$$

där $\alpha(t) = \mu(t) \cdot R(t)$.

$\mu(t)$ är avvecklingsintensiteten och R(t) är antal personer under risk vid tiden t. Eftersom $E[dM(t)|F_{t-}] = 0$ enligt martingalteorin så är

$$dN(t) = \mu(t) \cdot R(t)\ dt + slump$$

och en skattning av $\mu(t)$ skulle vara $\frac{dN(t)}{R(t)}$. Skattningen av β blir då $\int_0^t \frac{1}{R(s)}\ dN(s)$. R(t) kan dock ibland vara 0 och därför använder man sig av en indikatorfunktion, J, som är ett då R(t) är positiv och noll annars. Skattningen av β blir då denna:

$$\hat{\beta}(t) = \int_0^t \frac{J(s)}{R(s)}\ dN(s)$$

Låt nu $\tau_1 < \tau_2 < ...$ vara olika hopptider för N. dN(t)=1 när t är något av dessa τ_i och noll annars. Med detta synsätt behövs inte indikatorfunktionen längre och skattningen av β kan skrivas så här:

$$\hat{\beta}(t) = \sum_{\tau_i \leq\ t} \frac{1}{R(\tau_i)}$$

Detta ger då

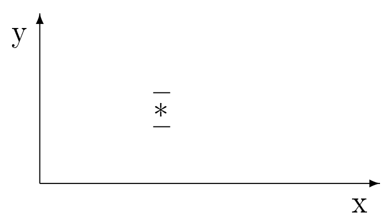
$$\hat{\lambda}(t) = e^{-\sum \frac{1}{R(\tau_i)}}\ , \quad \tau_i \leq t\ .$$

Skattningen av λ är klar.

C Konfidensintervall och konfidensband

När man får fram resultat ur datamaterial finns det alltid en risk att man fått ett skevt urval på grund av slumpen. Ju fler observationer man har desto mindre påverkar de eventuella konstiga resultaten helheten och ens slutsatser blir mer korrekta. Ett sätt att se hur bra skattningar man har är att göra konfidensintervall och konfidensband.

Om man vill se hur säker skattningen för en punkt är så kan man göra ett konfidensintervall. Ett konfidensintervall visar hur långt man måste gå från väntevärdet för en skattning för att vara säker på att man med en viss sannolikhet har inringat alla möjliga utfall för denna, se figur 18.

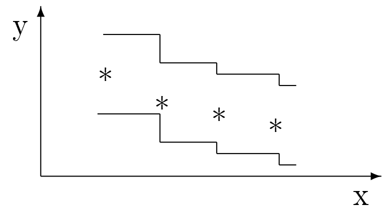


Figur 18: Konfidensintervall för en punkt. * är skattningen av y vid det specifika x-värdet och strecken ovanför och under visar osäkerheten, alltså var * med den valda konfidensgraden lika gärna skulle kunna ligga.

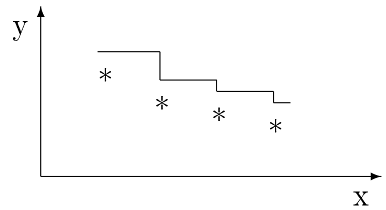
Ju fler punkter man vill göra skattningar för desto större risk att åtminstone någon hamnar utanför sitt konfidensintervall. Gör man till exempel konfidensintervall med konfidensgraden 95% ingår det ju i definitionen att ungefär var 20:e skattning har hamnat utanför konfidensintervallet. Har man flera punkter man vill ska stämma så kan man i stället använda sig av konfidensband. Konfidensbandet visar samma sak som konfidensintervallet fast för alla punkter samtidigt. För att inringa alla samtidigt som med konfidensbandet så krävs det naturligtvis att man går längre från väntevärdet i varje punkt än vid konfidensintervall, se figur 19.

Eftersom vi i det här fallet vill se till att hela funktionen hamnar innanför så är det ett konfidensband vi vill ta fram.

Konfidensbandet för avvecklingsfunktionen behöver inte vara dubbelsidigt, det räcker att se till att vi inte har för snabb avveckling. Skulle vi anta för långsam avveckling är det i enlighet med tidigare resonemang inte lika farligt. I och med att man gör konfidensbandet ensidigt får man också som bonus en lägre övre gräns, se figur 20.



Figur 19: Konfidensband för flera punkter. * är skattningarna av y vid olika x-värden och strecken ovanför och under visar osäkerheten, alltså var * med den valda konfidensgraden lika gärna skulle kunna ligga.



Figur 20: Ett ensidigt konfidensband ligger närmare skattningarna än ett dubbelsidigt.

Konfidensband för avvecklingsfunktionen tas fram genom

$$\hat{\lambda}(t) + c_\alpha \cdot \hat{\lambda}(t) \cdot \frac{1 + \hat{\hat{a}}(t)}{\sqrt{n}}$$

där n är antalet observationer och

$$\hat{\hat{a}}(t) = \sum_{i:\ \tau_i \leq\ t} \frac{n}{R(\tau_i)^2}$$

Beroende på vilken konfidensgrad man vill ha ändras c_α enligt tabell 8.

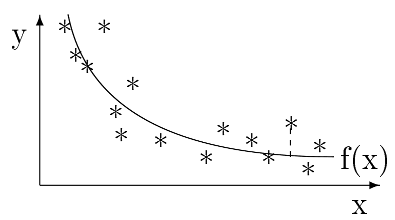
$1-\alpha$	0.95	0.975	0.995	0.9995
c_α	1.2238	1.3581	1.6276	1.9495

Tabell 8: Parametrar till ensidigt konfidensband

D Minsta kvadrat-metoden

När man anpassar en funktion till data vill man naturligtvis att funktionen som väljs är den som passar bäst till datamaterialet. Vilken som passar bäst är däremot inte självklart, det finns många sätt att mäta. Ett sätt som ofta används är minsta kvadrat-metoden.

Låt $y_1, y_2, ..., y_n$ vara observerade mätpunkter för $x_1, x_2, ..., x_n$. Ofta varierar dessa uppmätta värden beroende på till exempel slump eller mätfel. Förutom det så har man oftast inte värden för alla punkter man är intresserad av. Man vill därför med hjälp av de observerade punkter man har skatta en funktion, $f(x)$, för att kunna få fram värden för alla x. Man vill hitta den funktion som ligger närmast punkterna, därför mäter man hur långt det är från funktionen till de olika mätpunkterna, se figur 21.



Figur 21: Anpassning av funktion till observerade datapunkter

Den streckade linjen är avståndet från punkten (x_j, y_j) till funktionen $f(x_j)$. Kvadraten på detta avstånd är det som ligger till grund för vilken funktion man väljer. Den bästa funktionen anses med denna metod vara den som minimerar summan av dessa fel i kvadrat; $\sum_{i=1}^{n}(y_i - f(x_i))^2$.

Referenser

[1] Bjørk, T. (1998), *Arbitrage Theory in Continuous Time*, Oxford, Storbritannien.

[2] Ekhult, H. (1990), *Approximation av en funktion av två variabler med hjälp av en punktmatris. Tillämpning på avvecklingsfunktionen i sjukförsäkring.* Meddelande nr. 4, Försäkringstekniska Forskningsnämnden, Stockholm, Sverige.

[3] FTN, (2008), *Instruktion för rapportering till FTNs sjuklighetsundersökningar*, Försäkringstekniska Forskningsnämnden, Stockholm, Sverige.

[4] FTN, (1989-1995), *Sjuklighetsundersökningen*, Meddelande nr 1-27B, Försäkringstekniska Forskningsnämnden, Stockholm, Sverige.

[5] Miličević, M. (2002), *Long-Term Health Insurance*, Examensarbete vid Stockholms universitet, Sverige.

[6] Sandström, A. (1987), *An empirical study of t-frequencies and termination functions in long-term sickness insurance in Sweden*, Meddelande nr 71, The Research Council of Actuarial Science, Stockholm, Sverige.

[7] Sandström, A. (1990), *Long-Term Health Insurance*, Sverige Reinsurance Company 75 Years, Sverige.