concept

Frode Drevland

Rett og riktig

En gjennomgang av Statens vegvesens analysemodell

NTNU
Det skapende universitet

concept

© Concept-programmet 2005

Concept rapport nr. 1070-6

Rett og riktig

En gjennomgang av Statens Vegvesens analysemodell

Frode Drevland

ISBN 82-92506-16-0 (papirversjon)

ISBN 82-92506-17-9 (elektronisk versjon)

Sammendrag: Denne rapporten tar for seg beregningsmetodene (trinnvis kalkulasjon) som benyttes av Statens Vegvesen (SVV) ved utarbeidelse av kostnadsoverslag. Den inneholder en teoretisk gjennomgang av formelverket som benyttes og det foreslås mulige forbedringer. Videre er beregningsresultatet fra et faktisk kostnadsoverslag utført med SVVs gjeldende formelverk sammenlignet med en Monte Carlo simulering av det samme prosjektet. Det er også sett på hva følsomheten for feil er hvis man anslår feil inngangsverdier til et kostnadselement. Rapporten konkluderer med at SVVs formelverk er så godt som det kan være ut ifra de generelle formlene de er basert på, men at disse generelle formlene i seg selv kan forbedres.

Dato: *1.7.2005*

Utgiver: *Concept-programmet*
Institutt for bygg, anlegg og transport
Norges teknisk- naturvitenskapelige universitet
Høgskoleringen 7A
7491 NTNU - Trondheim

Tel. 73594640
Fax. 73597021
http://www.concept.ntnu.no

Ansvaret for informasjonen i rapportene som produseres på oppdrag fra Concept-programmet ligger hos oppdragstaker. Synspunkter og konklusjoner står for forfatternes regning og er ikke nødvendigvis sammenfallende med Concept-programmets syn.

Forord

Institutt for bygg, anlegg og transport ved NTNU har siden midt på 80-tallet arbeidet med og utviklet metoden Trinnvis kalkulasjon; en stokastisk kalkulasjonsmodell som identifiserer og eksponerer usikkerheten i estimatene og kalkyleresultatet.

Sannsynlighetsvurderingene og det matematisk/statistiske grunnlaget for tallbehandlingen er basert på en analytisk modell hvor de enkelte kalkyleelementene er antatt å ha en Erlang sannsynlighetsfordeling, og hvor den eksakte tetthetskurven fastslås ved å anslå tre punkter på kurven. Tallbehandlingen gjøres ved hjelp av et formelverk som tilpasses til hvilke punkter på kurven man velger å anslå.

I løpet av de 20 årene vi har jobbet med usikkerhetsanalyser har det tid om annet dukket opp spørsmål om valg av fordelingsfunksjoner og om hvorvidt det formelveket som benyttes er riktig og nøyaktig nok.

Av frykt for å avlede fokus fra det som vi mener er det viktigste elementet i usikkerhetsanalysen, nemlig å sikre en input som er mest mulig i samsvar med virkeligheten, har vi valgt å ligge lavt i forhold til diskusjonen om fordelingsfunksjoner og tallbehandling. Vår oppfatning har hele tiden vært at de feil man eventuelt gjør her er små i forhold til de feil man kan få ved å være ukritisk i forhold til input.

Innføring av ekstern kvalitetssikring av store statlige investeringer har imidlertid bidratt til en utvikling i retning av en viss standardisering av usikkerhetsanalysene, og derav økt fokus på datahjelpemidler og det matematisk/statistiske grunnlaget disse bygger på.

Som følge av uttrykte ønsker fra Finansdepartementet, Samferdselsdepartementet og Statens vegvesen om å gjøre en vurdering av formelverket som ligger i datahjelpemidlet til Statens vegvesen, ANSLAG, og ønsker fra flere konsulenter om å verifisere formelverket for Trinnvis kalkulasjon med Erlangfordelinger, har vi i denne rapporten gjort teoretiske utredninger rundt den analytiske modellen, og samtidig gjort sammenlikning mot resultatene fra Monte Carlo- simuleringer.

Rapportens kanskje viktigst bidrag til diskusjonen om "riktige" modeller for usikkerhetsanalyser er imidlertid de analysene som er gjort av følsomheten for feil i input. Disse følsomhetsanalysene er allmenngyldige og gjelder enten man sverger til analytiske modeller eller simuleringer.

Den direkte målgruppen for rapporten er alle i og utenfor "kvalitetssikringsregimet" som på en eller annen måte skal forholde seg til Statens vegvesens kalkulasjonsmodell ANSLAG. Vi antar den vil være av interesse også for andre som driver med usikkerhetsanalyser.

Trondheim/NTNU 09.08.2005

Kjell Austeng

(prosjektleder for Concept delprosjekt "Usikkerhetsanalyser")

Sammendrag

Frode Drevland

Rett og riktig

En gjennomgang av Statens Vegvesens analysemodell

Concept rapport nr. 1070-6

Denne rapporten tar for seg beregningsmetodene (trinnvis kalkulasjon) som benyttes av Statens Vegvesen (SVV) ved utarbeidelse av kostnadsoverslag. Den inneholder en teoretisk gjennomgang av formelverket som benyttes og det foreslås mulige forbedringer. Videre er beregningsresultatet fra et faktisk kostnadsoverslag utført med SVVs gjeldende formelverk sammenlignet med en Monte Carlo simulering av det samme prosjektet. Det er også sett på hva følsomheten for feil er hvis man anslår feil inngangsverdier til et kostnadselement. Rapporten konkluderer med at Statens Vegvesens verktøy Anslag regner omtrent så korrekt som det går an å bli innenfor dagens gjeldende formelparadigme. Feilen som gjøres på forventningsverdien er neglisjerbar med tanken på potensielle feil som følge av gale inngangsdata. Feilen på det absolutte standardavviket er noe større, men ser vi på det relative standardavviket blir også feilen her svært liten. Det er i alle fall ikke mulig å få det til å bli mer korrekte ved bruk av de standard trinnvisformlene. Dette er dog mulig ved å bruke forbedrede formler som er presentert i rapporten.

Summary

Frode Drevland

Right or wrong?

A discussion of the Norwegian Public Road Authority's method for uncertainty analysis

Concept Report no. 1070-6

This report looks into the methods of calculation (Successive Calculation) that are used by the Norwegian Public Road Authority (NPRA) in making cost estimates. It contains a theoretical discussion of the collection of formulas that are being employed and suggests possible improvements. Further on are the calculation results from an actual project, calculated using these formulas, compared to a Monte Carlo simulation of the same project. The report also looks into what the error in the results will be if the estimated input parameters of a cost element are not correct. The report concludes that the NPRAs collection of formulas are as good as the can be considering the generic formulas that they are derived from, but that these generic formulas in themselves can be improved upon. The report presents alterantive formulas that will achieve a higher level of accuracy.

Innhold

Forord 1
Sammendrag 3
Summary 5
Innhold 7
Figuroversikt 9
1. Innledning 11
2. Generelt om beregningsmetoder 13
3. Om Erlang- og Gammafordelingene 15
3.1 Gammafordelingen 15
3.2 Erlangfordelingen 16
3.3 Noen formler for en forskjøvet Gammafordeling 17
3.4 Venstreskjeve gammafordelinger 18
4. Trinnvis-formlene 21
4.1 Grunnlag for trinnvis formlene 21
4.2 Forutsetninger for vurdering av formelverket 22
4.3 Typiske skjevhetsforhold og relative standardavvik 23
4.4 Theta optimal 26
4.5 Zeta optimal 36
4.6 Dynamisk Theta og Zeta 40
4.7 Regne ut Alfa, Beta og forskyvning 43
5. Monte Carlo simulering av faktisk prosjekt 45
5.1 Om simuleringsoppsettet 46
5.2 Resultater av simuleringen 47
6. Følsomhet for feil i inngangsdata 49
6.1 Følsomhet for feil nedre anslag (n) 50
6.2 Følsomhet for feil anslag for mest sannsynlig verdi (s) 53
6.3 Følsomhet for feil øvre anslag (ø) 56
6.4 Konklusjoner om følsomhet for feil i inngangsdata 58
7. Konklusjoner 59
Litteratur 61
Vedlegg-A Hvordan finne Alfa, Beta og forskyvning fra trippelanslag 63
Vedlegg-B Begrepsdefinisjoner 67

Figuroversikt

Figur 3-1.....Variasjon av Gammafordelingens formparameter.
Øverst: Alfa = 2, Beta = 1. Nederst: Alfa = 50, Beta = 1 15
Figur 3-2.....Variasjon av Gammafordelingens skalarparameter.
Øverst: Alfa = 10, Beta = 1.
Nederst: Alfa = 10, Beta = 100.....16
Figur 3-3.....Forskyvning av Gammafordelingen
Alfa = 10, Beta = 10,
Forskyvning = 0 og 100.....16
Figur 3-4.....Gammafordelingen og dens speilede variant 18
Figur 4-1.....Fordeling av skjevhetsforhold for Statens vegvesen
og konsulenter.....24
Figur 4-2.....Typiske relative standardavvik for kostnadsanslag i Statens
vegvesen 25
Figur 4-3.....Theta optimal for 10/90 uttrykt ved Alfa 27
Figur 4-4.....Theta optimal for 10 / 90 gitt ved skjevhetsforhold.....28
Figur 4-5.....Feil på forventningsverdi for Theta mellom 0,42 og 0,43 29
Figur 4-6.....Feil på forventningsverdi ved bruk av Theta = 0,42 30
Figur 4-7.....Feil på forventningsverdi ved bruk av Theta = 0,42 ,
plottet mot Alfa.....31
Figur 4-8.....Feil på forventningsverdi ved bruk av Theta = 1 32
Figur 4-9.....Feil på forventningsverdi ved bruk av Theta = 1,
plottet mot Alfa.....32
Figur 4-10.....Theta = 1 og Theta = 0,42 plottet mot hverandre.....33
Figur 4-11.....Feil på forventningsverdi for Theta = 1 for ulike
skjevhetsforhold ved ekstreme relative standardavvik.....34
Figur 4-12.....Feil på forventningsverdi for ulike Theta verdier for ulike
skjevhetsforhold ved ekstreme relative standardavvik.....35
Figur 4-13.....Zeta optimal for 10 / 90 for ulike verdier av Alfa.....37
Figur 4-14.....Zeta optimal for 10 / 90 for ulike skjevhetsforhold 38
Figur 4-15.....Kurve som faktisk beregnes 39
Figur 4-16.....Angitt kurve.....39
Figur 4-17.....Feil på standardavviket ved bruk av Zeta = 2,53 40
Figur 4-18.....Feil på forventningsverdi ved bruk av dynamisk Theta
og Theta = 0,42.....42
Figur 4-19.....Feil på standardavvik ved bruk av dynamisk Zeta
og Zeta = 2,53.....42
Figur 5-1.....Overordnet simuleringsstruktur.....46
Figur 5-2.....Simuleringsresultat. Verdier i 1000-kr.....47
Figur 6-1.....Forventningsverdiens følsomhet for feil nedre anslag (n) for
ulike standardavvik ved Alfa = 10.....50
Figur 6-2.....Forventningsverdiens følsomhet for feil nedre anslag (n) for
ulike Alfa ved 10% standardavvik.....51
Figur 6-3.....Standardavvikets følsomhet for feil nedre anslag (n) for
ulike Alfa 52

Figur 6-4.....Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike standardavvik ved Alfa = 10..... 53
Figur 6-5.....Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa ved 10% standardavvik..... 54
Figur 6-6.....Standardavvikets følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa..... 54
Figur 6-7.....Helt like i form, men speilvendt..... 55
Figur 6-8.....Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike standardavvik ved Alfa = 10..... 56
Figur 6-9.....Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike Alfa ved 10% standardavvik..... 57
Figur 6-10.....Standardavvikets følsomhet for feil øvre anslag (ø) for ulike Alfa 57

1. Innledning

I forbindelse med ekstern kvalitetssikring av et prosjekt i Statens Vegvesen (SVV) kom ekstern kvalitetssikrer frem til et resultat som lå til 2% lavere enn det SVVs eget KO (kostnadsoverslag) tilsa, uten at det kunne pekes på noen spesielle forhold i kostnadsoverslaget som kan forklare avviket. Det ble i den sammenheng stilt spørsmål ved om det formelverket som SVV støtter seg til var riktig og nøyaktig nok. På bakgrunn av dette ble NTNU forespurt om å gjøre en utreding av validiteten av SVVs formelverk.

Vi har angrepet denne problemstillingen fra to vinkler. Vi har benyttet en Monte Carlo simulering for å kontrollregne et faktisk prosjekt og for å bruke denne metoden som fasit for svaret Statens Vegvesen kom frem til ved bruk av sin metode. I tilegg har vi gjennomført en teoretisk gjennomgang av formelverket.

En del grunnleggende statistiske begreper som er benyttet i rapporten samt begreper som introduseres i denne rapporten er forklart i Vedlegg-B.

Alle grafer i denne rapporten er generert ved hjelp av Gafa. Dette er et verktøy for å analysere gammfordelinger utviklet i forbindelse denne rapporeten. Gafa kan lastes ned fra Concepts hjemmesider (www.concept.ntnu.no).

2. Generelt om beregningsmetoder

Det finnes to prinsipielle måter å beregne stokastiske kostnadsanslag; ved bruk av analytiske metoder (for eksempel suksessiv/trinnvis kalkulasjon) eller ved bruk av simuleringsmetoder (for eksempel Monte Carlo simulering).

Ved bruk av analytiske metoder beskrives hver av kostnadene matematisk. Ligningene for hver av kostnadene regnes så summen til en nytt uttrykk som gir fordelingen av mulige utfall. Å modellere og regne en kostnadsanalyse ved bruk av eksakte matematiske sannsynlighetsfordelinger og beregningsmetoder er både meget tidkrevende og krever betydelige matematiske ferdigheter. Suksessiv kalkulasjon og formelverket som SVV benytter i dataprogrammet ANSLAG er en analytisk metode "light". I stedet for å regne eksakt benytter man lineære tilnærmingsformler som gir et svar som er tilstrekkelig nøyaktig.

Monte Carlo simulering er mer en "rå-makt" metode. I stedet for å regne gjennom kalkylen en gang regner man gjennom den fra noen hundre til flere tusen ganger. Det vil si, man får naturligvis en datamaskin til å gjøre det for seg. For hver gjennomregning trekkes en tilfeldig verdi ut fra kostnadssannsynlighetsfordelingene og benyttes i beregningen. Med andre ord for hver gjennomregning "kaster" datamaskinen en "terning" for å finne ut hvilken verdi den skal bruke for en gitt kostnad. Datamaskinen vil for hver gjennomregning ta vare på de tallene man er interessert i. Etter at simuleringen er kjørt ferdig vil den da ha et statistisk grunnlag for utarbeide en sannsynlighetsfordeling for sluttsummen og andre tall vi måtte ønske å få ut.

Begge metodene skal i prinsippet gi det samme resultatet når de arbeider med den samme innputten (struktur og data).

3. Om Erlang- og Gammafordelingene

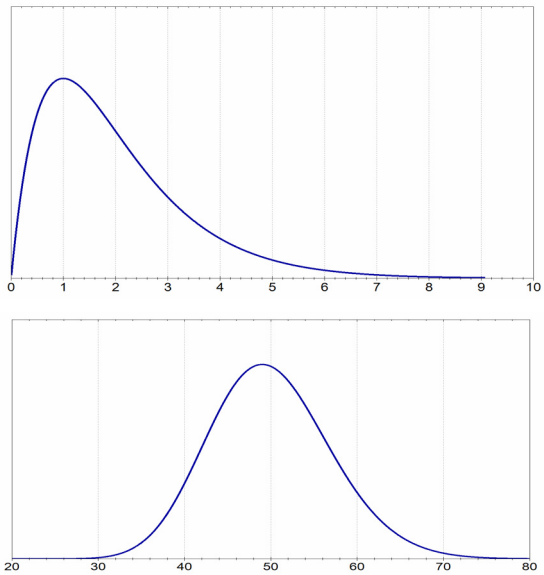
Før vi begynner gjennomgår formelverket som benyttes av Statens Vegvesen er det viktig å ha klart for seg en del ting om fordelingen de baserer seg på, gammafordelingen.

3.1 Gammafordelingen

Gammafordelingen er en høyreskjev fordeling gitt av to parametere; Alfa og Beta. Alfa avgjør kurvens form, det vil si hvor høyreskjev den er, mens Beta er en skalarparameter som avgjør tallenes størrelse.

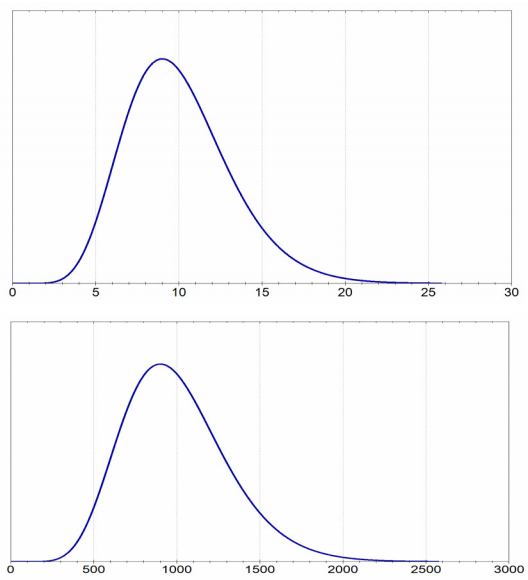
Form

Et viktig fakta om gammafordelingen er at den er per definisjon høyreskjev, det vil si at fordelingens mest sannsynlige verdi ligger til venstre for fordelingens median. Som vi kan se av den øverste kurven i figur 3-1 er fordelingen svært høyreskjev ved lave verdier av alfa. Ved større verdier av alfa vil kurven bevege seg mot en normalfordeling, som vi kan se av den nederste kurven. Hvis alfa går mot uendelig vil kurvens form gå mot en normalfordeling, men den vil alltid være litt høyreskjev selv om det er snakk om svært små verdier som det vil være umulig å se det med det blotte øye. Dette kan dog ha konsekvenser om man bruker gammafordelinger i simuleringer og enkelte analytiske metoder. Ved bruk av tilnærmingsformler av den typen Statens Vegvesen benytter er ikke dette noe problem.

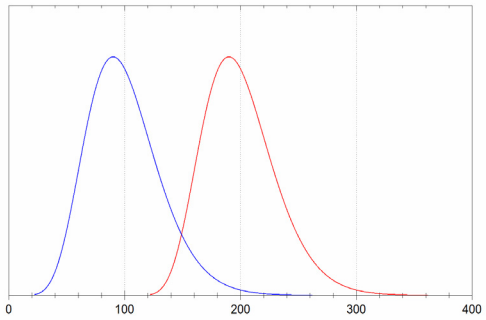


Figur 3-1 *Variasjon av Gammafordelingens formparameter. Øverst: Alfa = 2, Beta = 1. Nederst: Alfa = 50, Beta = 1*

TIPS! Ved bruk av simuleringsverktøy er det bedre å bytte ut gammafordelingen med en normalfordeling hvis man har et symmetrisk fordelt element.



Figur 3-2 *Variasjon av Gammafordelingens skalarparameter.*
Øverst: Alfa = 10, Beta = 1.
Nederst: Alfa = 10, Beta = 100



Figur 3-3 *Forskyvning av Gammafordelingen*
Alfa = 10, Beta = 10,
Forskyvning = 0 og 100

Skalering

Både Alfa og Beta bidrar til å skalere fordelingen som vi kan se i figur 3-1 til 3-4. Men at kurven blir skalert som følge av at Alfa øker er kun en bi effekt. Alfas funksjon er å bestemme kurvens form. Beta derimot bidrar kun til å skalere kurven. Som vi ser av figur 3-2 er kurvene identisk i form, men tallene på x-aksen har blitt skalert opp på den nederste kurvenmed en faktor på hundre som følge av at beta er endret fra 1 til 100.

Plassering

Gammafordelingen starter per definisjon i null. Men man kan forskyve fordelingen slik man vil ved å legge til et vilkårlig tall til alle avleste eller beregnede verdier.

Hvis vi ser disse tre parameterne i forhold til trinnvisparametrene n, s og ø beskrevet i neste kapittel, kan vi si at Beta bestemmer avstanden mellom n og ø, Alfa bestemmer hvor s ligger relativt til n og ø og forskyvningsparameteren bestemmer fordelingens nullpunkt og er altså da et tall som blir lagt til flatt til alle verdier som hentes ut av fordelingen (det vil si n, s, ø og forventningsverdi etc.).

3.2 Erlangfordelingen

Erlangfordelingen er et spesialtilfelle av gammafordelingen. Hvis vi har en gammafordeling der Alfa er et heltall får vi Erlangfordelingen. Erlangfordelingens formparameter kalles k i stedet for Alfa. Denne har den fordelen at den har noe enklere matematisk formel. De opprinnelig Trinnvis-formlene som er omtalt i neste kapitel er basert på Erlangfordelingen. Det synes å ha vært flere ulike grunner til at man opprinnelig valgte å benytte Erlangfordelingen, men slik trinnvis-formlene brukes per i dag er det Gamma-fordelingen som er den korrekte fordelingen å forholde seg til, så Gamma-fordelingen blir brukt i det følgende i denne rapporten.

3.3 Noen formler for en forskjøvet Gammafordeling

Parametre

α - kontinuerlig formparameter $\alpha > 0$

β - kontinuerlig skalarparameter

φ - forskyvningsparameter

Forventingsverdi

$\beta > 0 \; E(x) = \alpha\beta + \varphi$

Varians

$\sigma^2 = \alpha\beta^2$

Mode (Mest sannsynlige verdi)

$s = \beta(\alpha - 1) + \varphi$

Tetthetsfunksjon

$$f(x) = \begin{cases} \dfrac{\beta^{\alpha}(x - \varphi)^{\alpha - 1} e^{-x - \varphi/\beta}}{\Gamma(\alpha)} & \text{hvis } (x - \varphi) > 0 \\ 0 & \text{ellers} \end{cases}$$

hvor $\Gamma(\alpha)$ er Gammafunksjonen definert som $\Gamma(\alpha) = \int_0^{\infty} t^{(\alpha - 1)} e^{-t} dt$ for alle reelle tall $\alpha > 0$

Kummulativ tetthetsfunksjon

$$F(x) = \begin{cases} \dfrac{\Gamma(\alpha, (x - \varphi)/\beta)}{\Gamma(\alpha)} & \text{hvis } (x - \varphi) > 0 \\ 0 & \text{ellers} \end{cases}$$

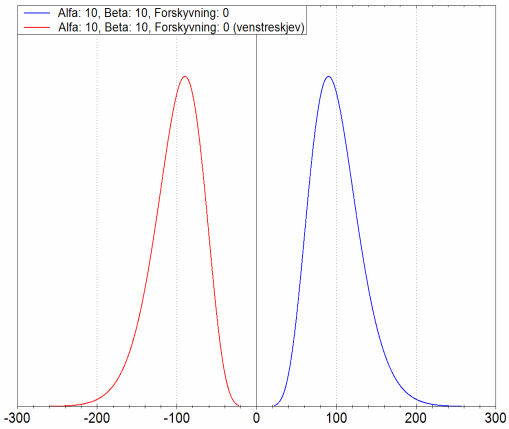
hvor Γ er den Inkomplette gammafunksjonen definert som $\Gamma(z, a) = \int_a^{\infty} t^{(z - 1)} e^{2-t} dt$ for alle reelle tall $z>0$

3.4 Venstreskjeve gammafordelinger

Gammafordelingen er per definisjon høyreskjev, men det vil ofte hende at man ønsker å beregne et kostnadselement man antar har en venstreskjev fordeling. Dette avsnittet tar for seg hvordan ved å benytte en speilet gammafordeling kan oppnå dette. Dette er plassert her for å ha teorien om gammafordelingen samlet på et sted, men det forutsetter at man er kjent med konseptet av trippelanslag og trinnvisformlene som omtales i neste kapittel.

Definisjonen av en venstreskjev gammafordeling

En uforskjøvet gammafordeling går fra null til uendelig. Vi definerer en venstreskjev fordeling til å være den speilede varianten av dette, det vil si en fordeling som går fra null til minus uendelig.Figur 3-4 viser eksempel på en vanlig høyreskjev gammafordeling og en speilet venstreskjev variant.



Figur 3-4 *Gammafordelingen og dens speilede variant*

På parametersiden definerer vi skalarparameteren Beta til å være negativ for venstreskjeve kurver. Dette gjør at vi kan bruke formlene for forventningsverdi og varians som beskrevet i faktaramma på forrige side.

For å kunne bruke formlen for tetthetsfunksjonene må vi gjøre et par triks. Vi bruker formelen for tetthetsfunksjonene som normalt, men vi snur fortegnet på Beta og x i formelen. For den vanlige tetthetsfunksjonen er et vilkårlig kurvepunkt for tetthetsfunksjonen for en venstreskjev kurve gitt som (x, f(-x)) der Beta i f(x) er positiv. Når det gjelder den kumulative tetthetsfunksjonen må man i tilegg snu på resultatet ellers vil S-kurven gå gal vei. Et vilkårlig punkt på den kumulative tetthetskurven er da gitt ved. (x, 1-f(-x)) der Beta i f(x) er positiv.

Transformasjon av trippelanslag

De færreste vil ha behov for å befatte seg med formlene fra forrige avsnitt for speiling av tetthetsfunksjonene, men det kan være tilfeller der man har behov for speile trippelanslag. Hvis man bruker de lineære tilnærmingsformlene kjent fra trinnvis kalkulasjon er ikke venstreskjeve trippelanslag noe problem, men hvis man skulle ønske å bruke mer nøyaktige analytiske modeller eller simulering står man ovenfor et problem. Gammafordelingen er per definisjon høyreskjev. Men dette betyr ikke at det er umulig å få beregnet verdier for et venstreskjevt anslag. Man må bare speile fordelingen slik at den fremstår som høyreskjev og trikse litt. I det følgende er en oppskrift for hvordan man går fram.

N = Nedre anslag, S = mest Sannsynlig Anslag, Ø = Øvre Anslag

Ved et venstre skjevt anslag N,S,Ø gjøres følgende:

1. Snu fortegnene på alle tre anslagene.

2. Snu på rekkefølgen

Summert opp:

Transformert N,S og Ø blir i forhold til det opprinnelige trippelanslagt uttrykt som

$N_{Trans} = - Ø$

$S_{Trans} = - S$

$Ø_{Trans} = - N$

3. Regn ut forventningsverdien av trippelanslaget snu fortegnet på resultatet

$E(x) = - E(x)_{Trans}$

Eksempel på utregning av en venstreskjev fordeling

Trippelanslag: 60-100-110

N_{Trans} = -110

S_{Trans}= -S= -100

$Ø_{Trans}$= -Ø = -60

Nytt trippelanslag = -110, -100, -60

$E(x)_{Trans}$ =

$E(x) = - E(x)_{Trans} = -(-87{,}6) = 87{,}6$

Regner ut opprinnelig trippelanslag direkte:

$$E(x) = \frac{-110 + (0,42(-100)) - 60}{2,42} = 87,6$$

Som vi ser av beregningsresultatene gir metodene for speiling samme resultatet som å regne ut verdien direkte ved hjelp av trinnvisformelen for forventningsverdien. Tilfeller hvor det er nødvendig å gjøre denne øvelsen er beskrevet senere i rapporten.

4. Trinnvis-formlene

4.1 Grunnlag for trinnvis formlene

De opprinnelige formlene for trinnvis eller suksessiv kalkulasjon ble utarbeidet av Steen Lichtenberg på 70-tallet. Ifølge Lichtenberg kan sannsynlighetsfordelingen til et kostnadselement i et prosjekt uttrykkes ved hjelp av en Erlangfordeling. Formlene Lictenberg utarbeidet var basert på at man hadde k = 10 for fordelingen. I sine bøker skriver han at dette representerer den mest typiske skjevheten. Videre skriver han at selv om den faktiske skjevheten skulle være noe annet så ville formlene gi en tilstrekkelig god tilnærming, og at det selv med rimelig ekstreme forskjeller i skjevhet mellom formelens ideal og inngangsdata, vil man kun få et par promille feil for forventningsverdien. I bøkene hans er det så vidt vi kan finne ikke ført form for bevis for dette eller på andre måter dokumentert.

Formlene (se boksen til høyre) tar utgangspunkt i at man gjør et trippelanslag for et en kostnad der man anslår verdiene for 1% og 99 % kvartilene, samt mest sannsynlige verdi.

Trinnvis metoden og Lichtenbergs formler ble første gang brukt her i landet for ca 25 år siden. Rundt omkring årtusenskiftet gikk man etter hvert over til å anslå verdien for 10% og 90% kvantilene i stedet for 1% og 99% kvantilene. Reviderte versjoner av formlene dukket opp i forbindelse med denne overgangen. Ulike aktører kom opp med ulike varianter. Felles for dem alle er at de har hatt ingen eller manglende teoretisk forankring. Den mest riktige revisjonen av formlene som har vært i bruk er de som er vist midt i rammen til høyre.

Vi ser at forskjellen mellom formlene for forventningsverdien mellom 1/99 og 10/90 ligger i hva vektingen av mest sannsynlige verdi skal være. På generell form kan vi skrive disse formlene som visst nederst i rammen til høyre. Denne "vektfaktoren" har av Kjell Austeng ved NTNU blitt gitt navnet Theta. På tilsvarende vis innfører vi her Zeta som navn for standardavvikets delingsfaktor.

Trinnvis formler

Opprinnelige for 1/99

$$E = \frac{n + 2,95s + ø}{4,95}$$

$$\sigma = \frac{ø - n}{4,6}$$

Revidert for 10/90

$$E = \frac{n + 0,42s + ø}{2,42}$$

$$\sigma = \frac{ø - n}{2,53}$$

Generelle formler

$$E = \frac{n + \theta s + ø}{\theta}$$

$$\sigma = \frac{ø - n}{\zeta}$$

Parametre

E = forventningsverdi

σ =standardavvik

n = nedre anslag

s = mest sannsynlige verdi

ø = øvre anslag

θ =Theta
Vekting av sannsynlig verdi ved utregning av forventingsverdi

ζ =Zeta
Delingsfaktor ved utregning av standardavviket

4.2 Forutsetninger for vurdering av formelverket

Det er noen viktige forutsetninger vi må ha klart for oss over når det vi skal se på hvor godt formelverket for Trinnvis kalkulasjon er.

1. Det finnes en virkelig sannsynlighetsfordeling for kostnadselementet som anslås.
2. Denne virkelige sannsynlighetsfordelingen er gitt ved en gammafordeling, eventuelt forskjøvet og/eller speilet
3. De tallene man legger inn i formlene stemmer med den virkelige sannsynlighetsfordelingen.

Om disse forutsetningene er holdbare i alle sammenhenger er heller tvilsomt. Spesielt når det gjelder det siste punktet. Men disse forutsetningene er nå engang hva som ligger til grunn for formelverket.

Forutsetningen 2, om at den virkelige sannsynlighetsfordelingen er gitt ved en gammafordeling, trenger vi for å ha en fasit å vurdere formelverket opp mot. Det er ikke dermed sagt at virkeligheten faktisk forholder seg til en Gammafordeling.

Hvis man skal gjøre en vurdering av hvor godt formelverket er uten at man legger disse til grunn vil være en oppgave av gigantiske dimensjoner.

I kapittel 6 ser vi nærmere på hva som skjer hvis den siste forutsetningen viser seg å ikkeholde.

4.3 Typiske skjevhetsforhold og relative standardavvik

Vi skal videre i denne rapporten se på hvor stor feilene kan bli som følge av feil i formlene og feil i inngangsdata. Hvor store disse feilene blir avhenger i stor grad av to faktorer, hvor skjeve fordelingene er og hvor store relative standardavvik de har. Resultatene senere i rapporten presenteres i stor grad i form avkurver plottet mot disse to parameterene. Det er derfor på sin plass at vi først ser på hva som er typiske verdier for disse slik at vi har en referanse å forholde oss til.

For å kartlegge disse typiske verdiene hentet vi inn tre ulike usikkerhetsanalyser fra Statens vegvesen og et tilsvarende antall fra hver av tre konsulentfirmaer som driver med ekstern kvalitetssikring. Alle analyser er utført av ulike prosessledere og resurrsgrupper. Vi gikk inn i hver analyse og så på skjevhetsforholdet og relativt standardavvik for hvert enkelt trippelanslag og laget en statistikk på grunnlag av dette. Resultatet presenteres nedenfor.

Skjevhetsforhold

Kurvens grad av skjevhet kan uttrykkes ved bruk av Gammafordelingens formparameter Alfa. Men dette er en ikke-lineær parameter, og derfor ikke særlig enkel å forholde seg til. Vi innfører derfor i denne rapporten begrepet skjevhetsforhold.

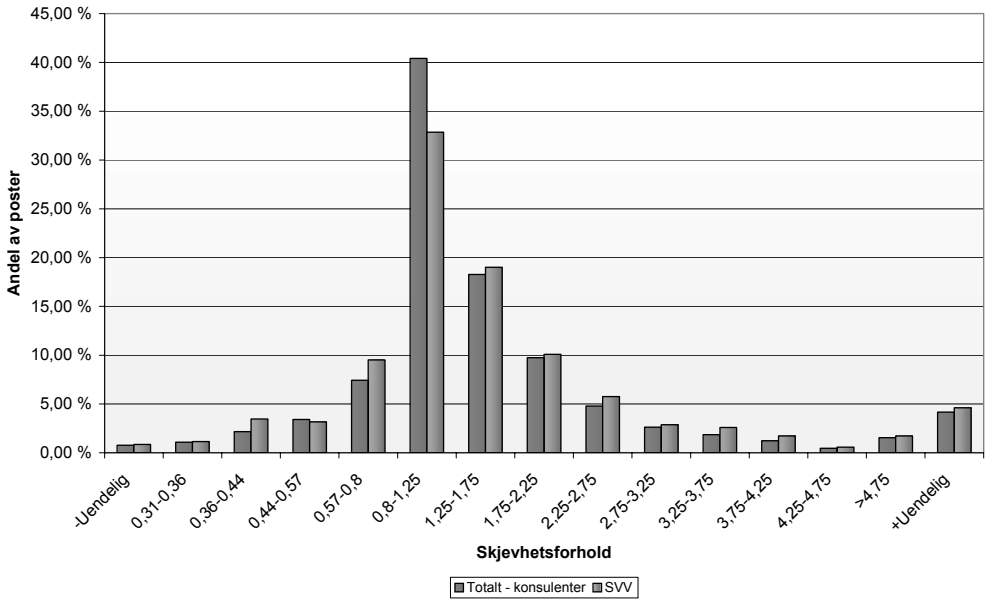
Vi definerer skjevhetsforholdet til å være differansen mellom øvre verdi og mest sannsynlige verdi delt på differansen av nedre og mest sannsynlige verdi:

$$Skf = \frac{\text{ø-s}}{\text{s-n}}$$

Merk at slik vi har definert skjevhetsforholdet er det lineært for høyreskjeve fordelinger, mens den for venstreskjeve fordelinger blir ikke-lineært.

Skjevhetsforhold er interessant for oss av to grunner. For det første er formlene vi benytter lineære tilnærmingsformler. Som vi skal nærmer på senere i rapporten, er disse formlene tilpasset et gitt skjevhetsforhold, og jo mer vi avviker fra dette jo større blir feilen.

Den andre grunnen til at skjevhetsforholdet er interessant for oss er når vi skal se på feil i inngangparameterene. En høysreskjev kurve har en lang hale til høyre. Jo skjevere kurven er jo lengre er halen (i forhold til resten av kurven). En lang hale betyr at det er større avstand mellom prosentkvantilene ute på halen. Det vil si å gå fra 80 %-kvantilet til 90%-kvantilet har mye mer å si i kroner å øre enn for en symmetrisk kurve. Det betyr at konsekvensen av å anslå feil prsoentkvantil blir mye større.



Figur 4-1 *Fordeling av skjevhetsforhold for Statens vegvesen og konsulenter*

Figur 4-1 viser statistikken som ble utarbeidet for postenes skjevhetsforhold. Tallene som er lavere enn null på x-aksen kan virke litt merkelige. Det som er viktig å huske på her er at disse representerer venstreskjeve kurver og skjevhetsforholdet da ikke er lineært stigende. Grunnen til at ikke mer runde tall ble valgt for kategoriseringen er at et betydelig antall av trippelanslagene i grunnlagsmaterialet hadde runde skjevhetsforhold, og det ville gitt et statisk feil bilde å brukt runde tall på avgrensningene.

Vi kan regne med at 70-75% av alle trippelanslag vil ha et skjevhetsforhold mellom 0,57 og 1,75, det vil ikke særlig skjeve. Bare et fåtall vil ha skjevhetsforhold større enn fem. Man skal likevel være oppmerksom på er at dette fåtallet poster som har store skjevhetsforhold som regel er store poster i en analyse. Det er typisk der hvor man har nedre og mest sannsynlig verdi tilnærmet lik at man ser store skjevhetsforhold. Dette forekommer først og fremst på postene for indre og ytte påvirkningsfaktorer og andre prosentpåslagsposter som totalt sett utgjør en stor andel av forventningsverdien for et kostnadsanslag.

Relativt standardavvik

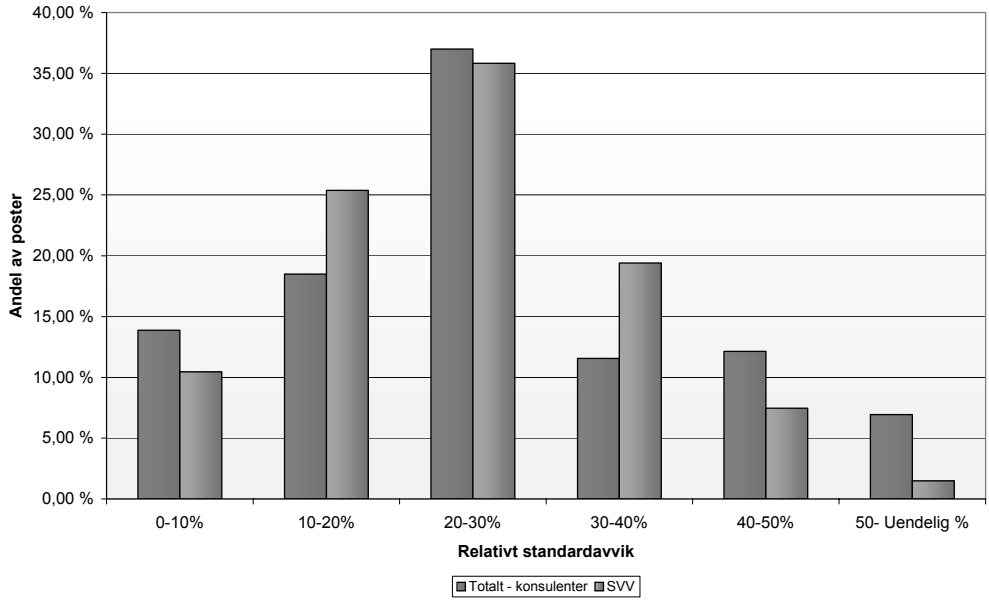
Relativt standardavvik er et mye brukt begrep. Det er definert som standardavviket delt forventningsverdien. Eller sagt med andre ord, hvor stor spredningen er i forhold til forventningsverdien.

$$\sigma_{rel} = \frac{\sigma}{E(x)}$$

Det relative standardavviket har mye si for hvor stor feilene blir. Det kommer at vi benytter en *forskjøvet* Gammafordeling. Det vi i si at benytter en fast tallverdi pluss en Gammafordeling. For standardavviket er denne tallverdien uinteressant. Standardavviket sier kun noe om spredningen på verdiene, ikke noe om hvor store de er. Men for forventningsverdien er det absolutt interessant. Når man forskyver fordelingen langs x-aksen blir alle verdier hentet ut

fra kurven tilsvarende større. Hvis vi for forskyver kurven med 10, så øker verdeien for 10 %-kvantilet med 10, og likeså for alle andre kvantiler samt mest sannsynlige verdi og forventningsverdien. Trinnvisformelen for forventningsverdien anslår veriden direkte for en forksjøvet gammafordeling. Men vi kunne like gjerne ha delt opp beregningen i to trinn. Først regnet ut forventingsverdien av en uforskjøvet Gammafordeling og deretter lagt forskyningstilegget. Resulatet hadde blitt det samme. Men det er klart at i dette tilfellet må eventuelle feil i formlene slå til på i utregningen av forventningsverdien for Gammafordelingen, det andre er jo bare å legge til en gitt tallverdi. Det er også klart jo større denne "sikre " tallverdien er, jo mindre blir en eventuell bergegningsfeil i forhold til den totale forventningsveriden(med forskyvning). Samtidig vet vi at dette sikre tallet ikke har noe å si for standardavviket, så om vi øker denne tallveriden øker vi altså forventingsverdien men ikke standardavviket. Dette får da naturlig nok den følgen at det relative standardavviket synker. Poenget her er da at konskekvensen av gjøre feil med lave relative standardavvik er mye lavere enn ved høye som følge av at det ligger et "sikkert" tall i bunn.

Figur 4-2 viser hvor stort det relative standardavviket vil være statistisk sett for en post i et kostnadsoverslag utført av Statens vegvesen. Man kan kan ha i bakhodet når man leser denne rapporten at det typsike relative standardavviket ligger på ca. 25 %.



Figur 4-2 *Typiske relative standardavvik for kostnadsanslag i Statens vegvesen*

4.4 Theta optimal

Theta er, som beskrevet i forrige avsnitt, en vektfaktor som angir hvor mye s (mest sannsynlige verdi) skal vektlegges ved beregningen av forventingsverdien for en sannsynlighetsfordeling. Theta optimal er den verdien av Theta som ved bruk av suksessivformlene vil gi en forventningsverdi lik den matematiske korrekte verdien for sannsynlighetsfordelingen. Hva Theta optimal vil være for en gitt sannsynlighetsfordelingen vil avhenge av både hvilke kvantil det er som benyttes når man anslår nedre og øvre verdier i trippelanslaget og hvilken alfa sannsynlighetsfordelingen har. På bakgrunn av den generelle formelen for forventingsverdien visst i forrige avsnitt kan vi ved hjelp av kjente formler for gammafordelingen sette opp et generelt uttrykk hva Theta optimal vil være.

Utledning av generell formel

Vi antar at forventingsverdien av en gammefordeling kan finnes ved hjelp av en lineær tilnærmingsformel på formen:

1) $$E(x) = \frac{n+\theta s+ø}{2+\theta}$$

Der *E(x)* er forventningsverdien, *n* er nedre anslag, *s* er mest sannsynlige verdi og *ø* er øvre anslag.

Noen formler som gjelder gammafordelingen er:

Mode eller mest sannsynlige verdi

2) $$s = \beta(\alpha-1)$$

Forventningsverdi:

3) $$E(x) = \alpha\beta$$

Ved å skrive om 2) får vi et uttrykk for Beta 4):

4) $$\beta = \frac{s}{\alpha-1}$$

Vi setter så 4) inn i 3) og får et uttrykk for forventningsverdien uttrykt ved s og Alfa 5):

5) $$E(x) = \alpha\beta = \frac{\alpha s}{\alpha-1}$$

Vi setter så den nye uttrykket 5) for forventningsverdien likt vår antatte formel 1) og får ut et uttrykk for Theta gitt ved n, s, ø og Alfa.

6) $$\frac{n+\theta s+ø}{2+\theta} = \frac{\alpha s}{\alpha-1}$$

7) $$\theta = \frac{\alpha n-n+\alpha ø-ø-2\alpha s}{s}$$

Dette uttrykket 7) for Theta er generelt, i neste avsnitt går vi videre og bruker dette uttrykket til å finne hva Theta optimal vil være for ulike alfa ved bruk av 10/90 kvantilene.

Theta optimal for 10/90

Med utgangspunkt i den generelle formelen for Theta optimal fra forrige avsnitt kan vi regne ut hva Theta optimal vil være ved bruk av spesifikke prosentkvantiler. I dette avsnitte ser vi på hva Theta optimal vil være ved bruk av 10/90 kvantilene som basis for øvre og nedre verdier i et trippelanslag.

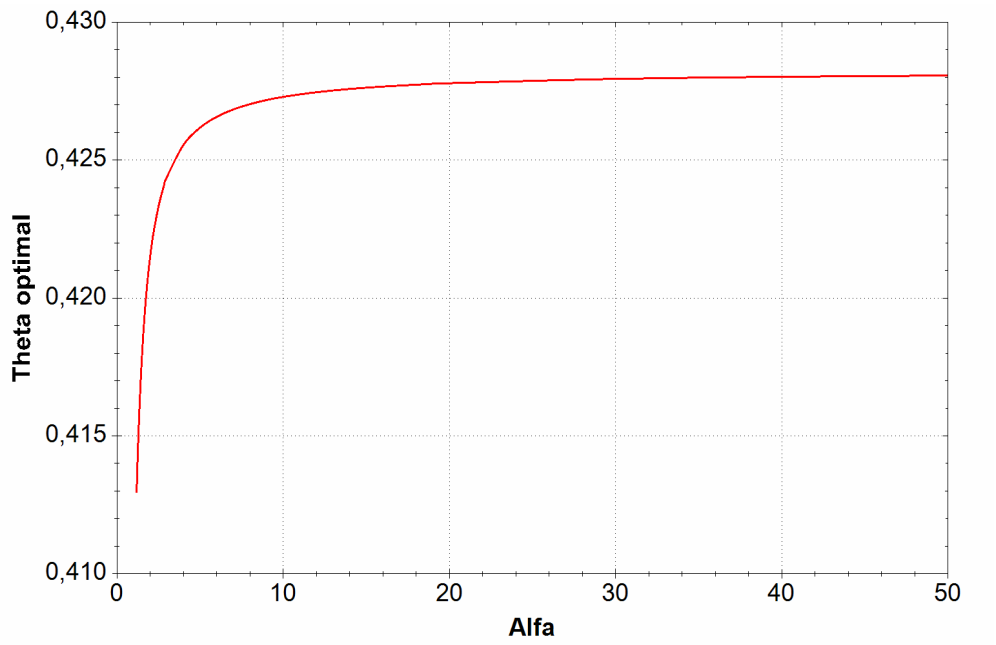
Den generelle formelen for Theta optimal er:

7) $$\theta = \frac{\alpha n - n + \alpha ø - ø - 2\alpha s}{s}$$

Vi ser at i formelen for Theta optimal er alfa en parameter. Som beskrevet tidligere på side 15 beskriver Alfa kurvens form. Det dette betyr er at Theta vil være avhengig av hvor skjev kurven er. Det betyr at det ikke er nok å regne ut en verdi for Theta optimal, vi er nødt til å regne ut en serie verdier for å se hvor mye Theta optimal varier med skjevheten.

En gammafordeling bestemmes entydig ved hjelp av en tre parametere. Dette kan være Alfa, Beta og Forskyvning som beskrevet i avsnitt 3.1, men det er også gjøres ved å fastslå tre punkter på kurven som for eksempel n, s og ø eller kombinasjon av disse to måtene. For å generere de tallsett med verdiene vi trenger (Alfa, n, s og ø) er det enklest å benytte Alfa, Beta og Forskyvning og regne ut n, s, og ø fra disse.

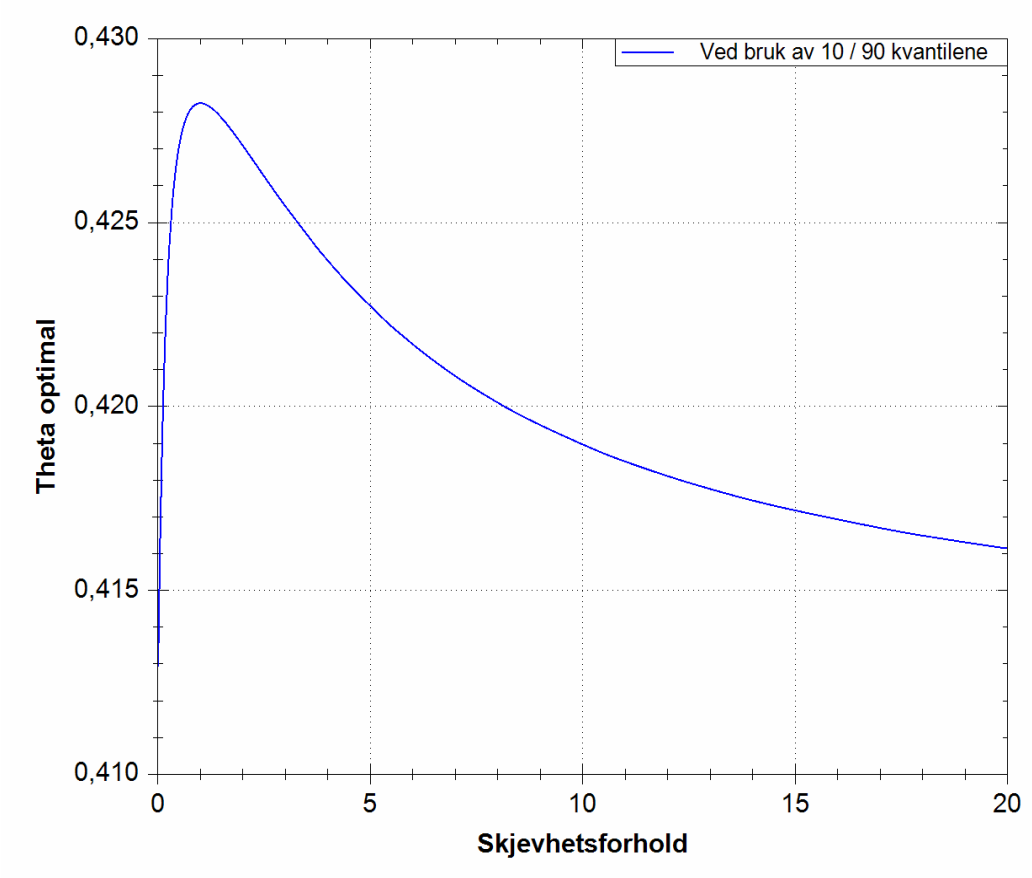
Vi gjør dette ved at Beta settes lik 10, Forskyvning settes lik 0 og Alfa varieres. Verdi for s regnes ut ved hjelp av tidligere beskrevet formel 2). n og ø regnes ut ved å ta den inverse funksjonen for gammafordelingen av Alfa, Beta og den aktuelle prosentkvantilen (10 og 90). (for eksempel funksjon GAMMAINV i Excel.) Resultatet blir da som vist i figur 4-3.



Figur 4-3 *Theta optimal for 10/90 uttrykt ved Alfa*

Vi ser av figuren at det er svært liten variasjon av hva som er optimal Theta verdi for ulike verdier av Alfa. For Alfa tilnærmet lik 1 vil optimal Theta være 0,408. Mens det for Alfa lik 50 vil optimal Theta være 0,428. Altså en forskjell på 0,02.

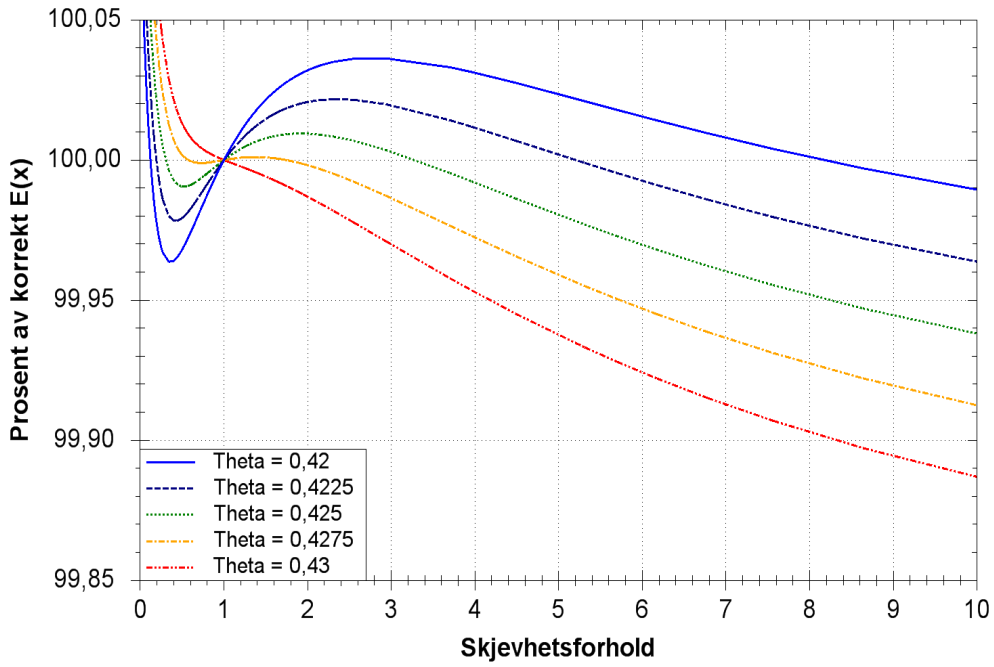
De færreste har noe særlig forhold til Alfa, spesielt med tanke på at kurvens form endres ikke-lineært med Alfa. Det er nok enklere for de fleste å forholde seg til skjevhetsforholdet. Vi definerer skjevhetsforholdet til å være differansen mellom øvre verdi og mest sannsynlige verdi delt på differansen av nedre og mest sannsynlige verdi ((ø-s)/(n-s)). Figur 4-4 viser hva optimal Theta vil være for et gitt skjevhetsforhold.



Figur 4-4 *Theta optimal for 10 / 90 gitt ved skjevhetsforhold*

Figur 4-4 viser hva Theta optimal vil være for skjevhetsforhold både større og mindre enn en. Altså både høyre og venstreskjeve fordelinger. Merk at verdien for Theta optimal blir akkurat det samme for en venstreskjev fordeling, men som følge av at den blir plottet mot en ulik skala ser det forskjellig ut på kurven. Skjevhetsforholdet for en venstreskjev kurve er likt 1/skjevhetsforholdet for en høyreskjev kurve.

Det vi kan konkludere ut fra dette er at ved bruk av trinnvis formlene for 10/90 må Theta ligge et sted mellom 0,42 og 0,43 for de skjevhetsforholdene som vi kan regne med at det er realistisk at vi kommer over i praksis. Figur 4-5 viser hva ulike Theta verdier gir av feil på forventingsverdien for ulike skjevhetsforhold. Vi ser at en Theta verdi på rundt 0,4225 gir minst feil ved lave skjevhetsforhold. En Theta verdi på 0,425 er antageligvis bedre egnet i praksis der man må regne med at inngangsdataene har en viss skjevhet. I alle fall er her kun snakk om et par tidels promille i forskjell på forventningsverdien.



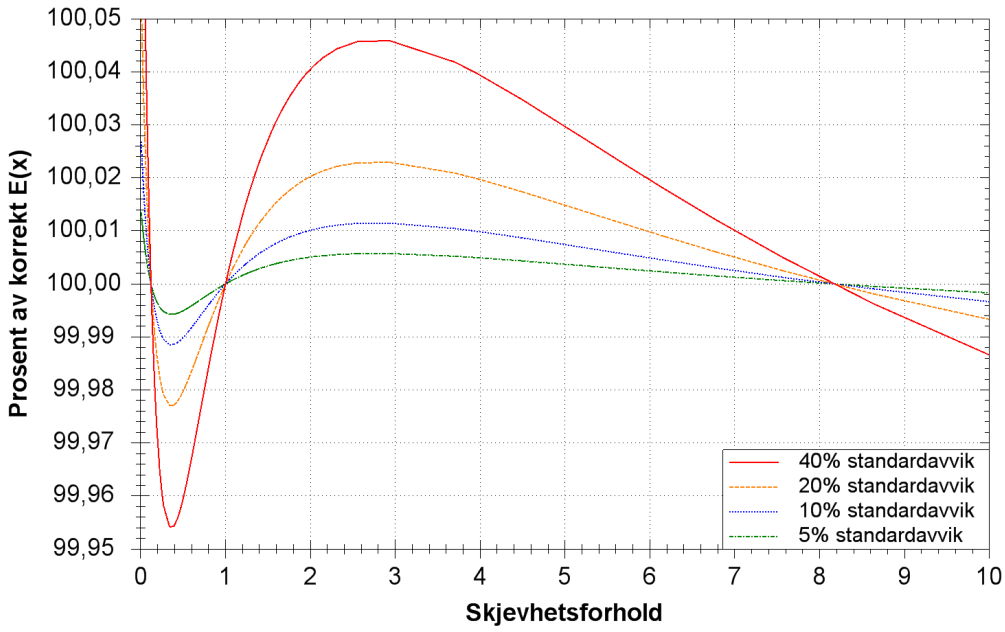
Figur 4-5 *Feil på forventningsverdi for Theta mellom 0,42 og 0,43*

I stedet for å bruke statiske verdier for Theta vil en bedre metode i alle fall være å bruke dynamiske verdier for Theta, noe som er beskrevet senere i avsnitt 4.3.

Feil på forventningsverdi ved bruk av Theta = 0,42

Statens Vegvesen benytter i sitt beregningsprogram, Anslag, en verdi for Theta på 0,42. Figur 4-6 viser hvor stort avviket på forventningsverdien blir ved bruke av denne Theta-verdien for ulike skjevhetsforholdstall. Som vi ser av figuren er maks avvik mindre enn en halv promille for trippelanslag med standardavvik på 40 %, det vil si for tilfeller der standardavvikets størrelse er 40 % av forventningsverdiens størrelse. Relativt standardavvik høyere enn dette er uvanlig at man finner på vanlige kalkyleposter På generelle forhold (indre og ytre påvirkningsfaktorer) derimot kan man finne relative standardavvik som er betraktelig høyere som igjen gir større feil på forventningsverdien. I verste fall kan man ved bruk av Theta = 0,42 få en feil på tre promille.

Feilen på forventingsverdien kunne blitt mindre ved å finjustere Theta som visst i forrige avsnitt, men feilen her er uansett å betrakte som neglisjerbart med tanke på den usikkerheten som ligger i inngangsdataene. (Se kapittel 6).



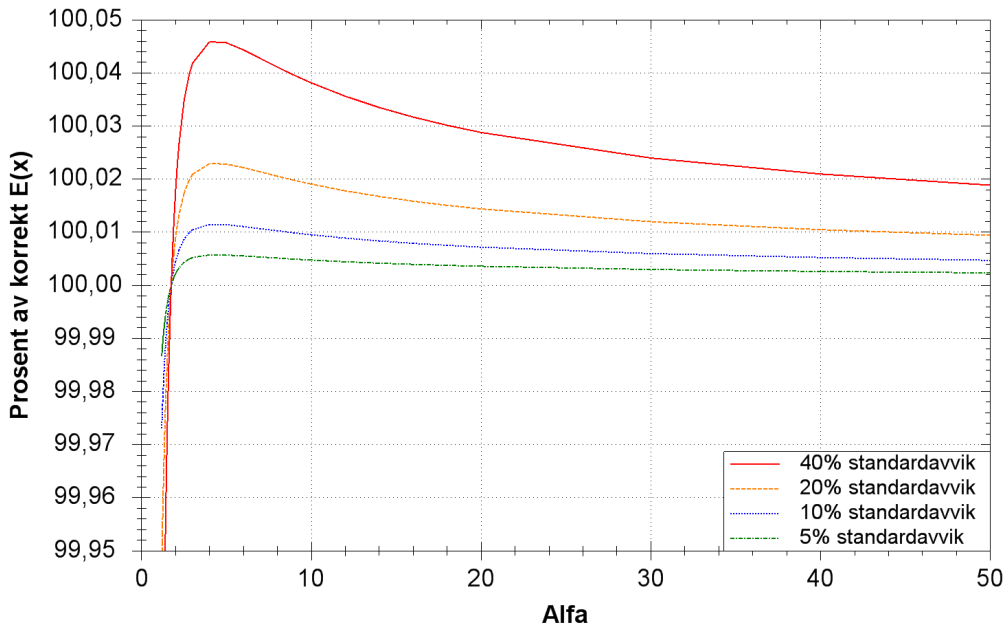
Figur 4-6 *Feil på forventningsverdi ved bruk av Theta = 0,42*

Figur 4-6 har tegnet inn flere kurver for ulike relative standardavvik og vi ser at jo større det relative standardavviket er jo større blir feilen. Det relative standardavviket avhenger i stor grad av fordelingens forskyvningsparameter. Det relativet standardavviker er standardavviket, som er uavhengig av forskyvingen, delt på forventningsverdien som er avhengig av forskyvningen, noe som gjør at jo større forskyvningsparameteren er jo mindre blir det relative standardavviket.

Vi kan se på dette som om at et kostnadsanslag har en deterministisk del og en stokastisk del. Der forskyvningen utgjør den deterministiske delen og en uforskjøvet gammafordeling gitt ved Alfa og Beta utgjør den stokastiske. Det er da naturlig at jo større den deterministiske delen er, jo mindre blir den totale feilen som følge av at den stokastiske delen, og eventuelle feil der, får mindre og si for helheten.

Grunnen til at kurvene i figur 4-6 får den formen den får her er at Theta = 0,42 er optimalveriden for en svært skjev fordeling med et skjevhetsforhold på 8 (se figur 4-5) Dette gjør at forventingsverdien blir mest korrekt jo nærmere man er dette skjevhetsforholdet. Vi ser også at feilen avtar når skjevhetsforholdet går ned mot 1. Dette skyldes at fordelingen da etter hvert blir så symmetrisk at mest sannsynlig verdi ikke får noe å si. Et skjevhetsforhold på 1 betyr at vi har en normalfordeling og forventningsverdien vil være lik snittet av nedre og øvre anslag.

Figur 4-7 viser det samme som figur 4-6, men med feilen på forventningsverdien plottet mot Alfa fremfor skjevhetsforholdet. Grunnen til at utslaget på y-aksen er større på denne figuren enn på den forrige er at figur 4-7 er plottet fra og med Alfa = 1,18. Dette tilsvarer et skjevhetsforhold på 158,65, Noe som er ca 16 ganger større enn hva som er vist i figur 4-6.

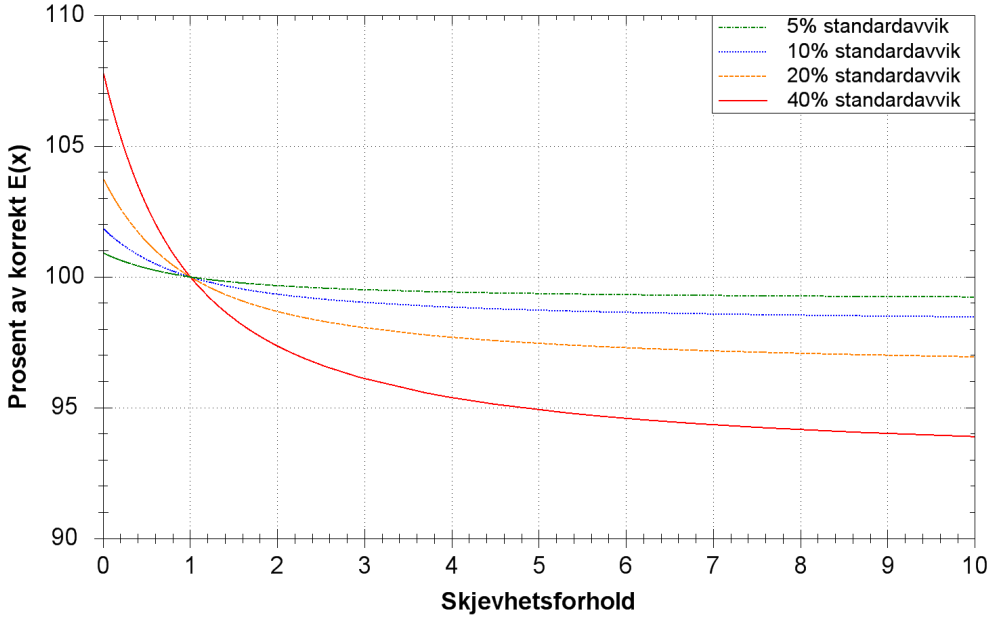


Figur 4-7 *Feil på forventningsverdi ved bruk av Theta = 0,42 , plottet mot Alfa*

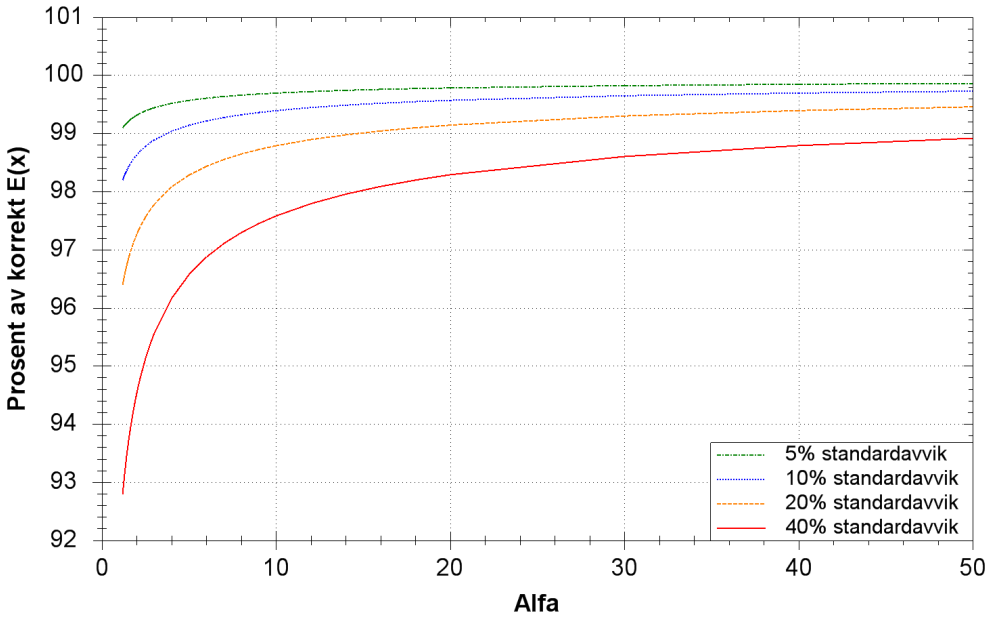
Feil på forventningsverdi ved bruk av for stor Theta

Bruken av Theta = 1 i Trinnvis-formlene er ikke spesielt utbredt. Men vi har sett nærmere på den som et eksempel på hva det medfører å bruke en Theta-verdi som er større enn hva vi tidligere har bestemt Theta optimal til å være.

Figur 4-8 viser hva feilen blir ved bruk av Theta = 1. Som vi ser av kurvene blir feilene betydelige. Feilene er her opp mot seks prosent i de mest ekstreme tilfellene der det for Theta = 0,42 maks var en halv promille. Hvis vi beveger oss lengre bort fra Theta optimal vil feile bli større. For eksempel ville en kurve plottet for Theta = 2,95 gitt oss opp mot 14 % feil. (2,95 er Theta som brukes når input er 1 % kvantilet - mest sannsynlig verdi - 99 % kvantilet).



Figur 4-8 *Feil på forventningsverdi ved bruk av Theta = 1*

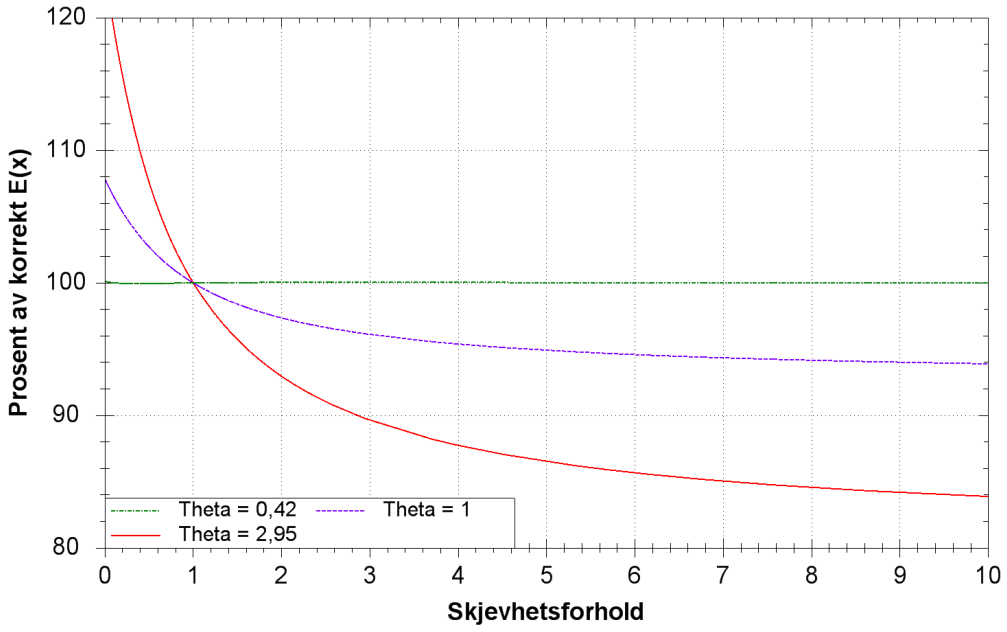


Figur 4-9 *Feil på forventningsverdi ved bruk av Theta = 1, plottet mot Alfa*

Gap mellom Theta = 1, Theta = 0,42 og Theta = 2,95

Figur 14 viser forskjellen i feil for Theta = 1, Theta = 0,42 og Theta = 2,95 plottet mot hverandre. Vi ser at når disse kurvene plottes inn på samme skala blir feilkurven for Theta = 0,42 tilnærmet rettlinjet i forhold til feilkurvene for Theta = 1 og Theta = 2,95.

Denne figuren forklarer delvis et fenomen man har observert i Statens Vegvesen. Ved overgang fra Anslag versjon 3.0.3 (som hadde feil formel lagt inn med Theta = 2,95) til versjon 3.0.5 har man sett at forventningsverdiene på kostnadsoverslagene har gått opp rundt 2 %, men i enkelte tilfeller mye mer. For eksempel økte kostnadsoverslaget for "Rv 659 Nordøyvegen" med hele 8,6 % kun som følge av at man gikk fra den ene til den andre versjonen av Anslag. Det viste seg her at en god del av trippelanslagene for dette prosjektet var svært høyreskjeve. Og som vi ser av figuren blir avviket større jo større skjevheten er.



Figur 4-10 *Theta = 1 og Theta = 0,42 plottet mot hverandre*

Men dette var ikke hele forklaringen på hvorfor feilen ble såpass stor som 8,6 %. Selv om enkelte av trippelanslagene var svært høyreskjeve, var det også mange som var normalfordelte eller bare litt skjeve. Den gjennomsnittlige skjevheten på kalkylepostene var ikke alene nok til å gi et såpass stort utslag. Hvorfor ble da feilen såpass stor? Svaret ligger i multiplikasjonsposter. Hvis vi har en post som har en enhetspris og en mengde som begge har en feil i forventingsverdien på 2 % vil posten ha en feil på 4,04 %.

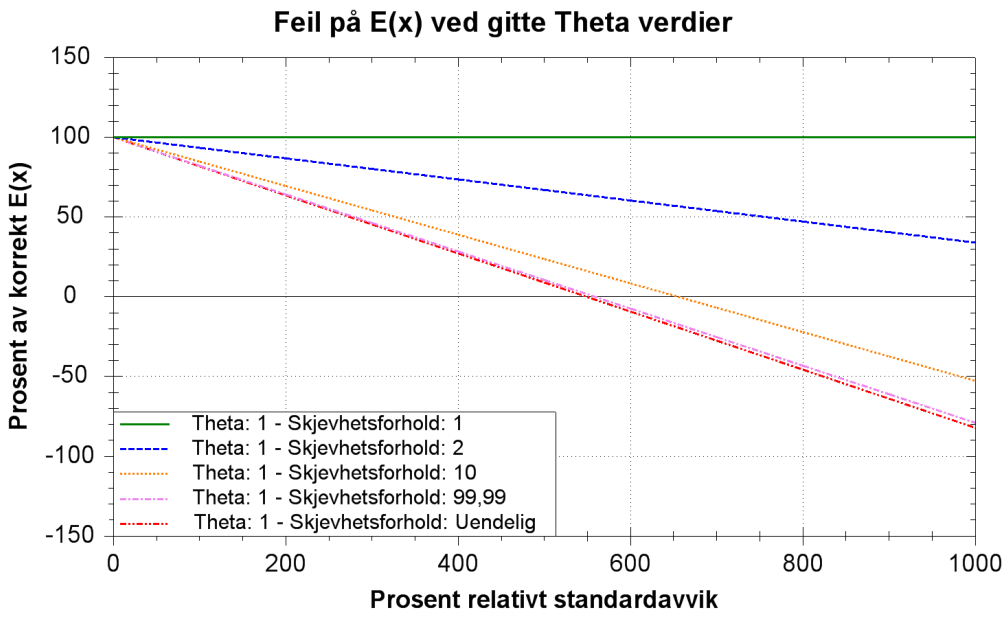
Særlig tydelig blir effekten av multiplisering av feil på indre og ytre påvirkningsfaktorer, spesielt fordi feilen der i utgangspunktet som regel er mye større.

Feil på indre og ytre påvirkningsfaktorer ved bruk av uriktig Theta

De feilkurvene vi har sett på hittil har vært for standardavvik opp til 40 %. Som tidligere nevnt er det uvanlig at vi finner relative standardavvik større enn det for vanlige kalkyleposter. For indre og ytre påvirkningsfaktorer er det derimot meget vanlig å ha relative standardavvik som ligger langt over dette. Trippelanslagene for indre og ytre påvirkningsfaktorer angis vanligvis som påslagsfaktorer omkring 1,0, men beregningsmessig ligger de sentrert rundt 0. Noe som betyr at vi får en forventingsverdi i nærheten av 0 og en relativt sett stor spredning omkring denne.

Figur 4-11 viser hva feilen på forventningsverdien blir for ulike skjevhetsforhold ved bruk av Theta = 1 for ekstremt høye relative standardavvik. Vi ser at ved et relativt standardavvik på 0 er feilen på forventingsverdien lik 0 og at den deretter øker lineært med at det relative standardavviket øker. En situasjon der man har et relativt standardavvik på null tilsvarer at man har et fullstendig deterministisk anslag, ergo er det irrelevant hvor godt formlene regner ut forventningsverdien av den stokastiske delen.

Vi ser også av figur 4-11 at feilen på forventningsverdien øker ikke-lineært med skjevhetsforholdet. For eksempel er det en prosentvis mindre økning av feilen på forventingsverdien ved å gå fra et skjevhetsforhold på 2 til et på 10, enn det er å gå fra et skjevhetsforhold på 10 til et på 99,99. Dette stemmer overens de foregående figurene som er "omvendt" plottet. Det vil si de har en kurve for hvert relative standardavvik og skjevhetsforholdet langs x-aksen.



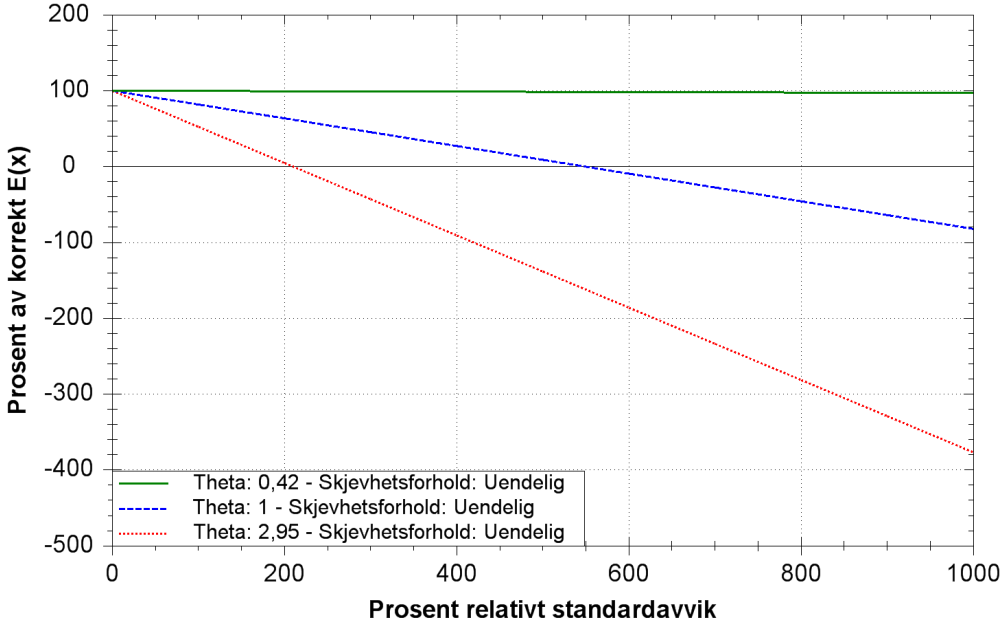
Figur 4-11 *Feil på forventningsverdi for Theta = 1 for ulike skjevhetsforhold ved ekstreme relative standardavvik*

Vi ser av figur 4-11 at det i ved svært høye skjevhetsforhold og relative standardavvik faktisk er mulig at man får en forventningsverdi som er såpass mye lavere enn det den skulle ha vært at den ender opp med motsatt fortegn. Det skal riktignok et veldig sært trippelanslag til for

å komme i denne situasjonen. For eksempel vil trippelanlaget -10, -10, 20 gi et uendelig stort skjevhetsforhold og et relativt standardavvik på bortimot 500 %.

Den største feilen man i praksis kan regne med å møte er når mest sannsynlig verdi er lik 0. Trinnvisformlene for Theta = 1 og Theta = 2,95 har en den interessante egenskapen at de for tilfeller der mest sannsynlig verdi er lik 0 bestandig gir en forventningsverdi som kun er henholdsvis 80,7 % og 48,9 % av det den korrekte verdien for høyreskjeve trippelanslag.

Figur 4-12 viser de mest ekstreme kurvene for alle de tre ulike verdiene av Theta som er omtalt her. Vi ser at ved 1000 % relativt standardavvik ligger forventningsverdien på ca -380 % av korrekt for Theta = 2,95, ca -80% for Theta = 1 og ca 97% for Theta = 0,42. Det vi kan si ut i fra dette er at selv om denne situasjonen som sagt er svært lite sannsynlig i praksis viser den at theta = 0,42 gir en formel som er svært robust selv ved ekstreme trippelanslag.



Figur 4-12 *Feil på forventningsverdi for ulike Theta verdier for ulike skjevhetsforhold ved ekstreme relative standardavvik*

4.5 Zeta optimal

Zeta optimal er den verdien som må settes inn under brøkstreken i Trinnvisformlene for standardavviket for at den beregnede verdien for standardavviket skal bli matematisk korrekt. Tilsvarende som for Theta kan vi også finne et uttrykk for Zeta optimal ved hjelp av kjente formler for gammafordelingen.

Utredning av generell formel

Varians av en gammafordeling:

8) $\sigma^2 = \alpha\beta^2$

Standardavvik av en gammafordeling

9) $\sigma = \sqrt{\alpha\beta^2}$

Generalisert trinnvisformel for standardavvik

10) $\sigma = \frac{ø - n}{\zeta}$

Hvis vi setter sammen 9 og 10 og snur litt på uttrykket får vi:

11) $\zeta = \frac{ø - n}{\sqrt{\alpha\beta^2}}$

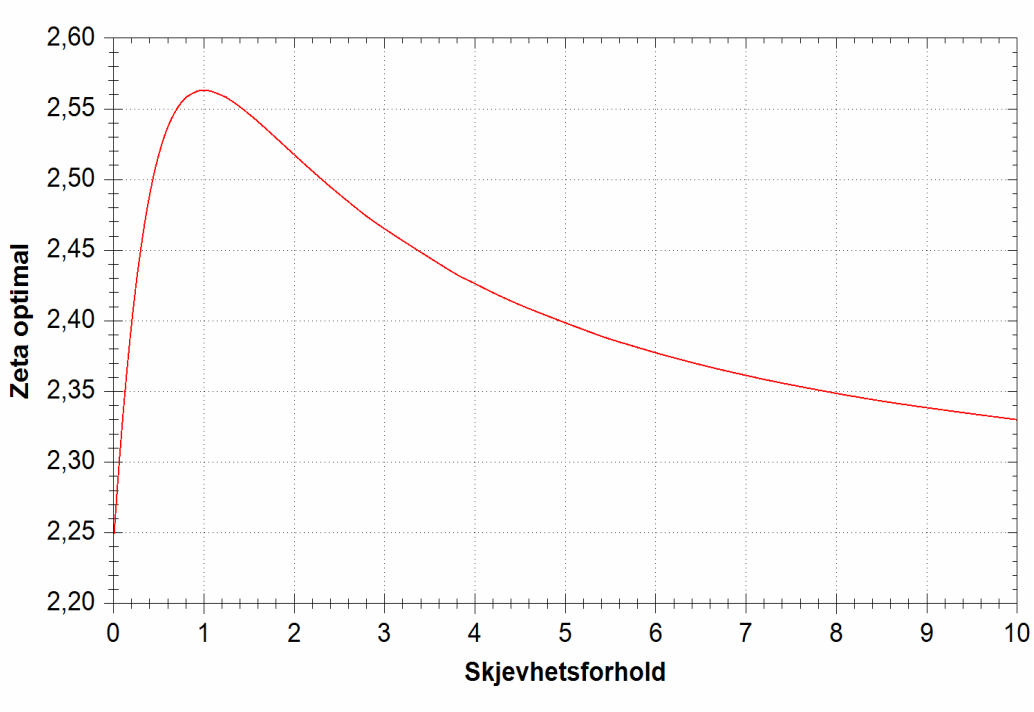
Dette er det generelle uttrykket for Zeta optimal. I neste avsnitt ser vi på hva Zeta optimal bør være ved bruk av 10/90 kvantilene.

Zeta optimal for 10/90

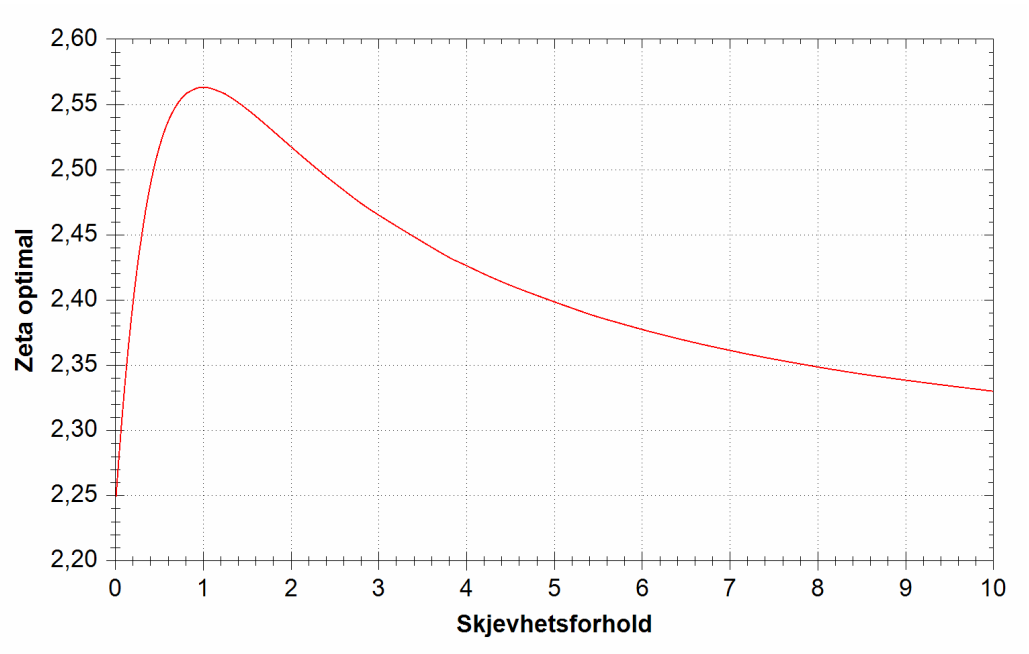
Det generelle utrykket for Zeta optimal er:

11) $$\zeta = \frac{ø - n}{\sqrt{\alpha\beta^2}}$$

Vi tar utgangspunkt i tallsettene generert tidligere for beregningen av Theta optimal og regner ut verdier for Zeta optimal for 10/90 ved hjelp av disse. Dette gir oss resultatet vist i Figur 17 og Figur 18.



Figur 4-13 *Zeta optimal for 10 / 90 for ulike verdier av Alfa*



Figur 4-14 *Zeta optimal for 10 / 90 for ulike skjevhetsforhold*

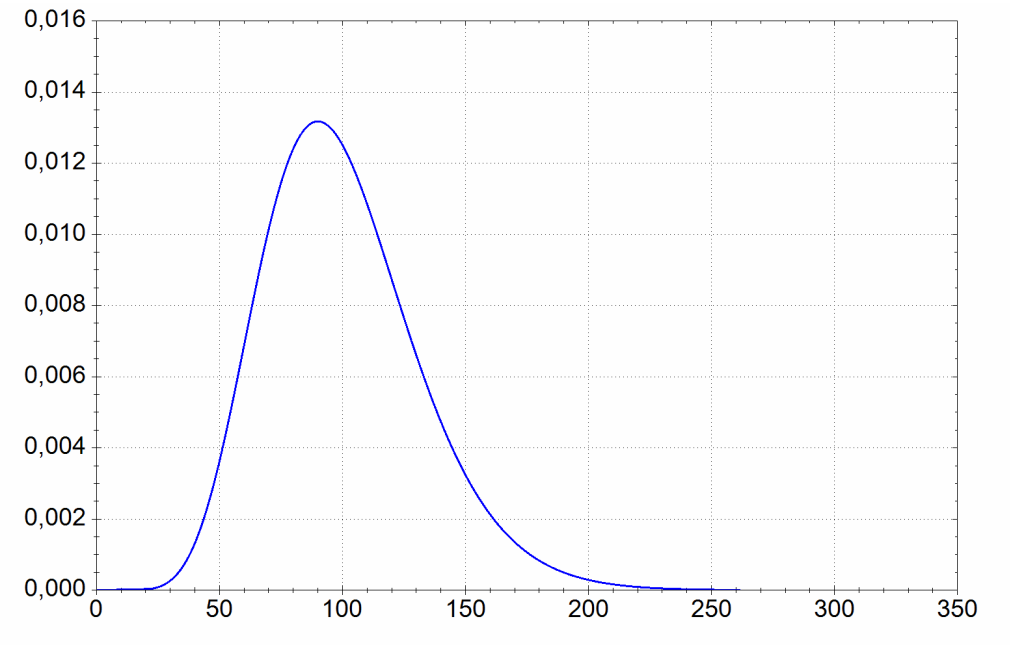
Vi ser her at spennet i Zeta optimal vist i kurvene på forrige side er mye større enn for Theta optimal vist i kurvene i avsnitt 4.2.2. Grunnen til dette er at mest sannsynlig verdi, s, ikke inngår i trinnvisformelen for standardavviket.

10) $$\sigma = \frac{ø - n}{\zeta}$$

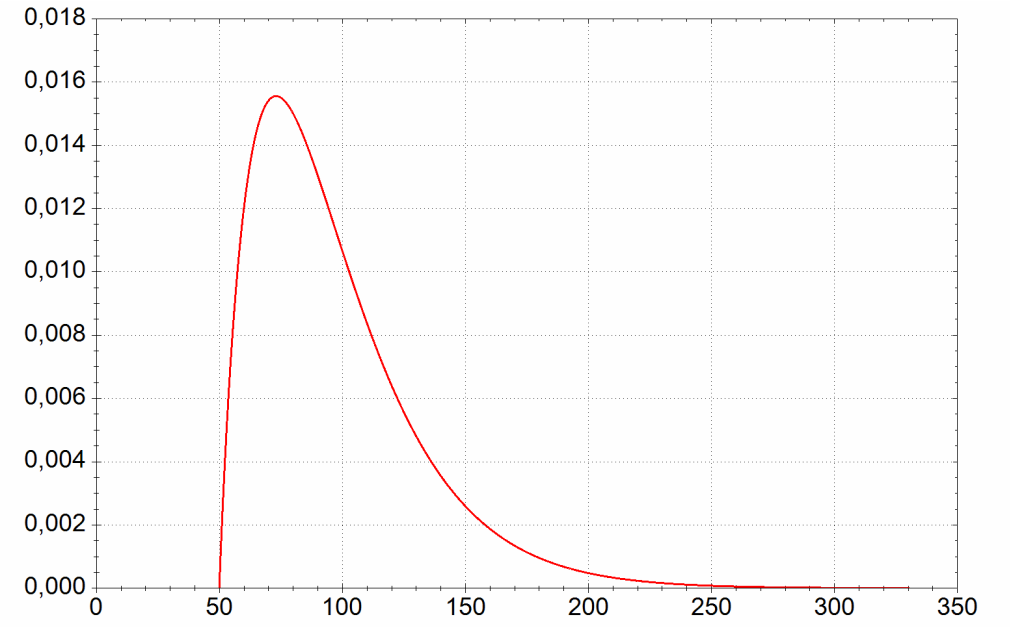
Dette betyr en gitt Zeta verdi tilsier et gitt skjevhetsforhold. Zeta = 2,53 er den som er vanlige å bruke ved anslag av 10/90 kvantilene. Denne verdien er korrekt for et skjevhetsforhold på 1,87 (Alfa = 10), så vi kan si at formelen er finstilt for dette skjevhetsforholdet. Dette betyr i praksis at man forskyver mest sannsynlige verdi til 41,26 %-kvantilet, som er hvor mest sannsynlig ligger for gammafordelinger med Alfa =10.

Følgen av at Zeta ikke tas hensyn til i trinnvisformlen for standardavvik

La oss anta at vi skal beregne standardavviket for fordelingen vist i Figur 20. Denne fordelingen har et skjevhetsforhold på 6,17 (Alfa=2). Ved bruk av formel 11 med Zeta = 2,53 vil fordelingen som vi faktisk beregner standardavviket for være den som er visst i figur 4-15.



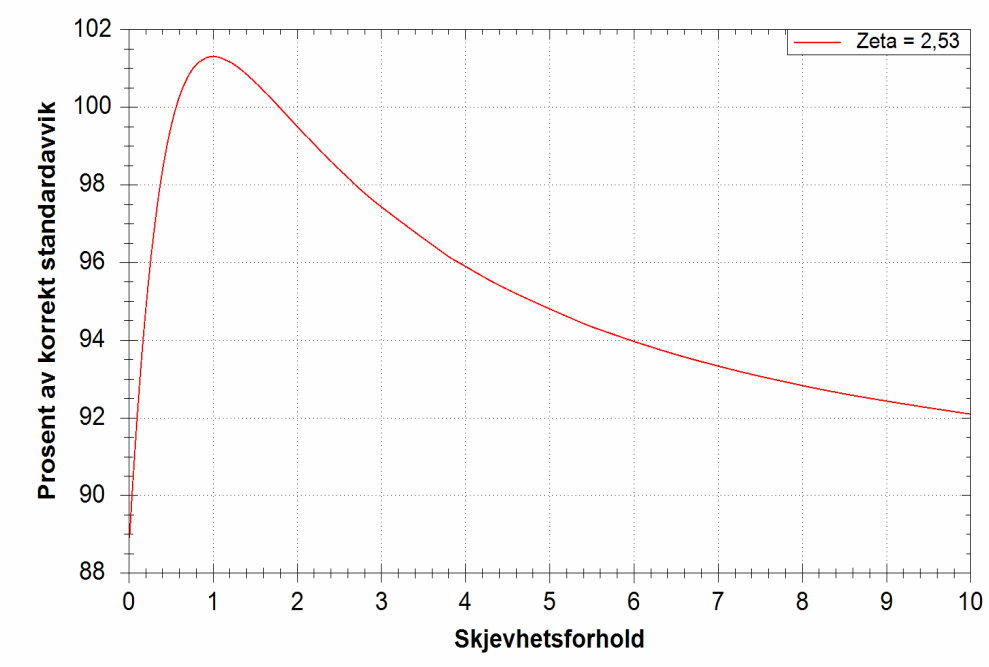
Figur 4-15 *Kurve som faktisk beregnes*



Figur 4-16 *Angitt kurve*

I tilfellet over vil det beregnede standardavviket være 6 % for lavt. Figur 4-17 viser hva feilen blir for ulike skjevhetsforhold ved bruk av zeta = 2,53. Hvis vi begrenser oss til de skjevhetsforholdene vi vanligvis kan forvente å finne i en usikkerhetskalkyle, mellom 0,5 og 4, ser vi at feilen er på opp mot 4%. En annen ting vi kan se er at symmetriske fordelte trippelanslag gir en feil på bortimot 1,5% på standardavviket.

Vi ser at kurven at fordelinger som er mer symmetriske enn "idealet" vil får for høye standardavvik, mens de som er mer skjeve får for lave beregnede standardavvik.



Figur 4-17 *Feil på standardavviket ved bruk av Zeta = 2,53*

Det vi kan konkludere med er at trinnvisformelen for standardavviket er mye mer unøyaktig enn formelen for forventningsverdi, og i motsetning til Theta er det ikke mulig å finne noen god allmenngyldig verdi for Zeta, skal man bruke en statisk verdi for Zeta må man ta et valg og finjustere formelen for et gitt skjevhetsforhold. Det vil dog være en mye bedre løsning å benytte dynamiske verdier for Zeta som presentert i neste kapittel.

4.6 Dynamisk Theta og Zeta

Trinnvis-formlene benytter seg av statiske verdier for Theta og Zeta. Som vi har sett er variasjonene i hva som er optimale verdier for disse for ulike alfa svært små når det gjelder Theta og ikke full så små når det gjelder Zeta. Men formlene gir i alle falle en rimelig god tilnærming til korrekte verdier med tanke på den usikkerheten som ligger i inngangsdataene. Hvis man ønsker en større grad av nøyaktighet vil en mulighet være å bruke dynamisk Theta og Zeta. Det vil si at man ved utregning av et trippelanslag bruker de optimale verdiene for Theta og Zeta for det aktuelle skjevhetsforholdet. Dette krever at det er mulig å beregne disse på en grei måte. En grei måte vil si at det enkelt og greit kan implementeres i dataprogrammer for beregning av usikkerhetsanalyser. Å skulle bruke en slik metode i forbindelse med håndregning vil være dødfødt.

Hvis vi legger en trendlinje på den høyreskjeve delen av kurvene vist i figur 4-3 'Theta optimal for 10/90 uttrykt ved Alfa" på side 27 og i figur 4-13 'Zeta optimal for 10 / 90 for ulike verdier av Alfa" på side 37 vil vi se at de er tilnærmet parabelformet. Vi antar da at vi kan uttrykke optimal Theta og Zeta ved hjelp av polynomuttrykk av skjevhetsforholdet.

Hvis *Phi* er skjevhetsforholdet

$$\phi = \frac{ø - s}{s - n} \qquad ø - s > s - n$$

$$\phi = \frac{s - n}{ø - s} \qquad ø - s < s - n$$

$$\phi = 1 \qquad ø - s = s - n$$

så vil et generelt uttrykk for Theta og Zeta være:

$$a\phi^3 + b\phi^2 + c\phi + d$$

der *a*, *b* og *c* er konstanter. Disse bestemmes ved å gjennomføre regresjon på datasettet til kurvene i figur 4-3 og figur 4-13. Ved å gjøre dette får vi følgende formler for dynamisk Theta og Zeta.

$$\theta = 0,0000039753126 28\phi^3 + (0,000009459489819\phi^2) - 0,001602543687\phi + 0,4301263456$$

$$\zeta = -0,0001693580287\phi^3 + 0,005991846477\phi^2 - 0,07392170812\phi + 2,640746721$$

Disse formlene vil ikke gi korrekte verdier for vært høye skjevhetsforhold (over ca 20). I og med at dette sjelden forekommer kan man i disse tilfellene gå over på en gitt forhåndsdefinert verdi. Det vil riktignok si at ved høye skjevhetsforhold vil få noen unøyaktigheter. Men for det første er så høye skjevhetsforhold svært sjeldne og for det andre vil det fortsatt gi et mer korrekt svar enn dagens formler gir i og med at man kan velge en Theta-verdi optimalisert for et smalere område.

Ved bruk av disse formlene finnes to spesialtilfelle som må håndteres. Det ene er når nedre eller øvre anslag er lik mest sannsynlig anslag og det andre er når alle tre anslagene er like. Begge disse to tilfellene vil føre til feil i et dataprogram i og med at de forårsaker en divisjon med null hvis man prøver å regne ut skjevhetsforholdet for fordelingen.

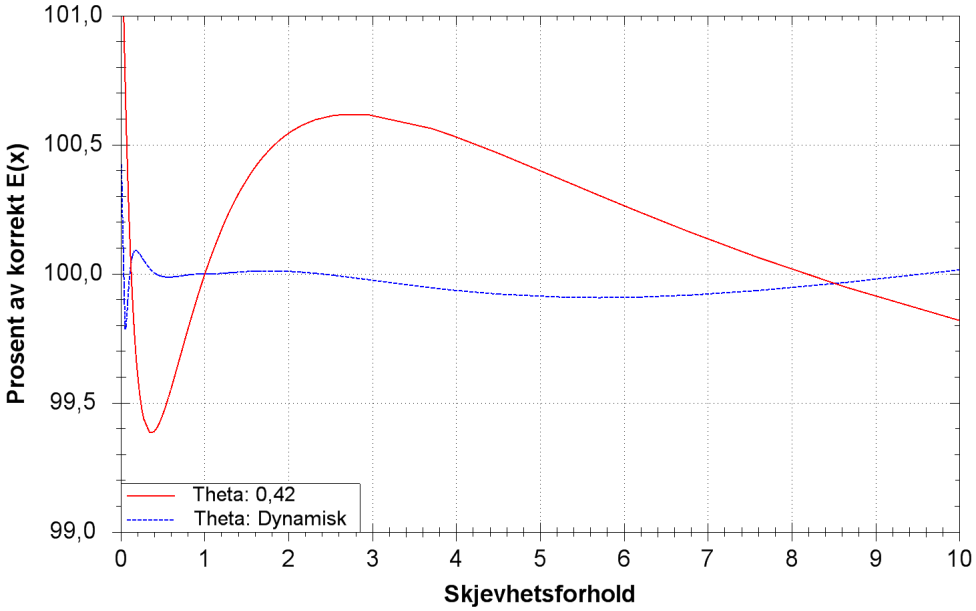
Tilfellet der nedre eller øvre anslag er lik mest sannsynlig anslag tilsvarer et uendelig stort skjevhetsforhold. For dette tilfellet er Theta-optimal lik 4,85 og Zeta optimal er lik 2,24.

For tilfellet der alle der alle tre anslagene er like spiller det ingen rolle hvilke Theta- og Zeta-verdier en bruker. Resultatet blir det samme uansett.

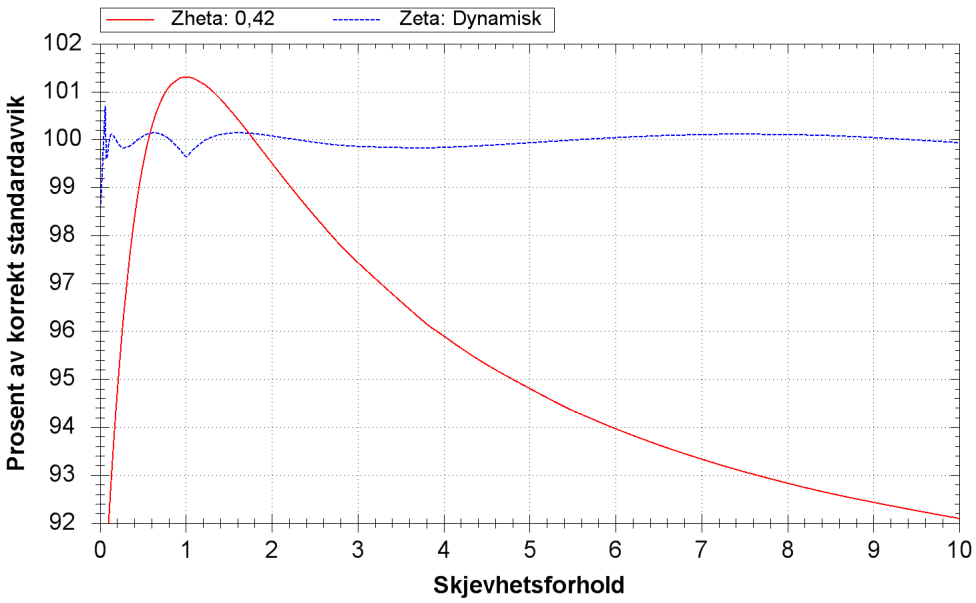
Alt i alt er det derfor ingenting i veien for at dette ikke kan implementeres i et dataprogram på en slik måte at det blir abstrahert vekk for brukeren slik at det ikke blir mer komplisert å vurdere eller gi innputt til analysen.

Å bruke disse formlene for å beregne en egen Theta og Zeta for hvert trippelanslag vil et langt mer nøyaktig resultat enn bruk av dagens formler. Figur 4-18 og 4-19 viser henholdsvis

sammenligninger av hva feilen av forventningsverdien og standardavviket blir for de gamle formlene og den nye dynamiske varianten



Figur 4-18 *Feil på forventningsverdi ved bruk av dynamisk Theta og Theta = 0,42*



Figur 4-19 *Feil på standardavvik ved bruk av dynamisk Zeta og Zeta = 2,53*

4.7 Regne ut Alfa, Beta og forskyvning

En alternativ og mer nøyaktig metode til å bruke Trinnvis-formlene til å regne ut forventingsverdi og standardavvik for et trippelanslag er å bruke Gammafordelingens egne formler for dette (se avsnitt 3.1 'Gammafordelingen" på side 15). Dette krever at vi først klarer å omsette et trippelanslag til Gammafordelingens parametere Alfa, Beta og forskyvning. Dette er i utgangspunktet ikke en enkel sak og er i praksis umulig å gjøre uten å ta datamaskiner til hjelp. For å muliggjøre dette har vi utviklet et sett algoritmer som beregner disse parametrene ut i fra et gitt trippelanslag.. Disse ligger vedlagt i Vedlegg-A.

Et problem ved bruk av denne metoden er at den er betydelig mye tregere enn å benytte enten de vanlige Trinnvis-formlene eller de dynamiske variantene. Den er dog ikke tregere enn at en vanlig datamaskin klarer å regne ut disse parametrene på et brøkdels sekund. Så når det gjelder beregning av en enkeltpost i et kalkyleprogram vil det ikke være noen merkbar forskjell for brukeren, men det kan by på problemer når man skal regne i hop postene, multipliserer inn faktorer etc. Dette vil være avhengig av hvordan programmet er bygd opp. I dette ligger også en annen ulempe med denne metoden kontra dynamisk utregning av Theta og Zeta. Den er mye vanskeligere å implementere. Selv med formlene Vedlegg-A. tilgjengelig er det ikke tilrådelig å prøve å implementere dette med mindre man enten har svært gode kunnskaper om Gammafordelingen og de nødvendige statistiske beregningsmetodene eller et hav av tid for å sette seg inn i dette.

Selv om det er visse utfordringer i å implementere denne metoden bør man absolutt vurdere den ved utvikling av kalkulasjonsprogrammer. Ikke bare er den så å si 100 % nøyaktig, med denne metoden kan også potensielt ivareta skjevheten i inngangsdataene ved summering av poster.

5. Monte Carlo simulering av faktisk prosjekt

Forutsatt at prosjektet er korrekt modellert og at inndata er korrekte vil resultatet av en Monte Carlo simulering konvergere mot virkeligheten når simuleringen kjøres med tilstrekkelig antall iterasjoner (gjennomregninger) i henhold til Strongs lov om store tall (Law og Kelton 2000).

I og med at vi kan forvente et korrekt svar ut av en Monte Carlo simulering valgte vi å bruke denne metoden for å kontrollregne en faktisk prosjektkalkyle gjennomført av Statens Vegvesen.

Problemet med å bruke Monte Carlo simulering til en slik kontrollregning er at det ikke er en triviell sak å omsette inndata fra en Trinnvis-kalkyle til parametere som et simuleringsprogram kan ta imot. Oss bekjent er det ingen simuleringsprogrammer som lar en sette inn et trippelanslag på den formen som benyttes av Statens Vegvesen og andre. Det vil si der man angir mest sannsynlige verdi samt en øvre og nedre verdi ved gitte prosentkvantiler.

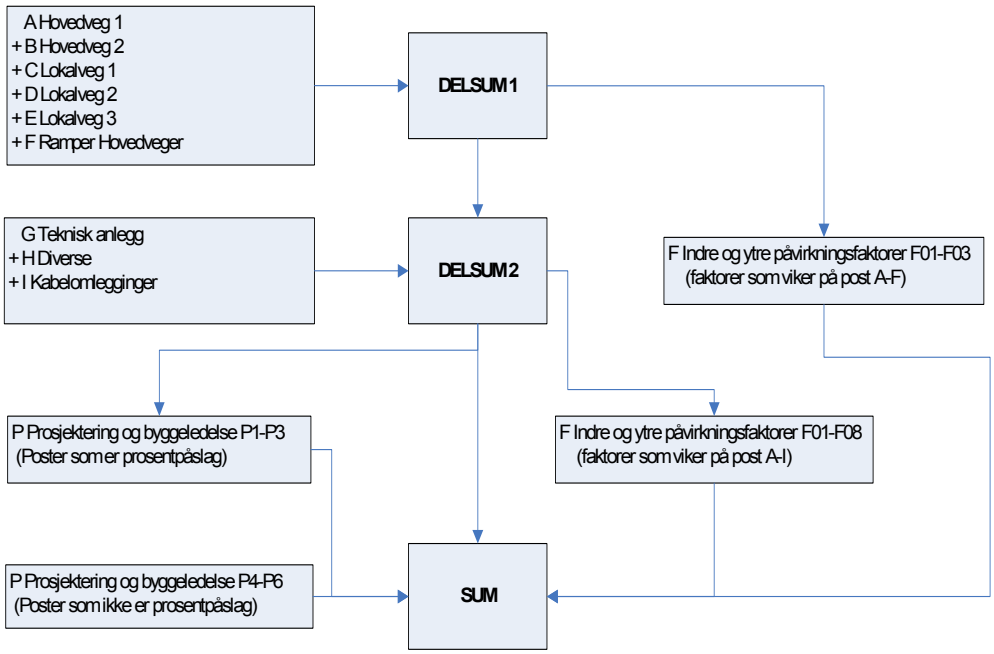
I @Risk og Crystall Ball har man valget mellom å bruke gammafordelingens normale parametere Alfa, Beta og Forskyvning (kalt Shift i engelskspråklige programmer) eller bytte ut en eller flere av disse med en vilkårlig prosentkvantil, men det er ikke mulig på noe vis å direkte legge inn mest sannsynlige verdi. I Definitive Scenario virker det ut ifra dialogene for innlegging som om den midtre verdien man legger inn er mest sannsynlige verdi. Men hvis man leser programmets dokumentasjon fremgår det at det faktisk er 50% kvantilen som legges inn. Sistnevnte program har for så vidt heller ikke mulighet for å benytte gammafordelingen, man er nødt til å benytte en lognormalfordeling. En annen svakhet med disse simuleringsprogrammene er at de heller ikke lar en direkte benytte venstreskjeve gammafordelinger.

Vi kom oss rundt disse to manglene ved å transformere de venstreskjeve fordelingene og transformere tilbake resultatet som beskrevet i på side 19 i kapittel 3 og benytte algortimen beskrevet i Vedlegg-A for å finne Alfa.

5.1 Om simuleringsoppsettet

I simuleringen valgte vi å begrense oss til å se på hva sum og standardavvik ble for hele prosjektet. Grunnen til dette var at det er umulig å få ut sammenlignbare svar på delsummer. Ved bruk av trinnvismetoden blir deler av påvirkningsfaktorenes standardavvik ført tilbake til posten de virker på. Dette er for å anskueliggjøre hvor årsaken til denne usikkerheten ligger. Denne tilbakeføringen er ikke mulig å få til ved simulering.

Figur 5-1 viser en skisse av simuleringsstrukturene. En del av postene i kalkylene beregnes på grunnlag av verdien av andre poster, dette gjaldt her noen av postene for byggherrekostander og faktorene for generelle forhold. Felles for disse var at de virket enten alle på post A til F eller på post A til I. To delsummer ble satt opp slik at disse postene enkelt kunne beregnes ved å multiplisere mot en sekkepost og ikke en mengde av enkeltposter.



Figur 5-1 *Overordnet simuleringsstruktur*

En simulering vil som nevnt gi et helt nøyaktig resultat av den gitte innputten forutsatt at man kjører den med tilstrekkelig antall iterasjoner. 100% nøyaktighet vil dog kreve svært mange iterasjoner. Vanligvis vil man nøye seg med en unøyaktighet på noen promille. I og med at vi skulle bruke tallene fra denne simuleringen som en fasit valgte vi å kjøre et relativt høyt antall iterasjoner. Simuleringen ble kjørt 2 ganger med ulikt såtall[1] for å kontrollere nøyaktigheten. Hver gjennomkjøring bestod av 100 000 iterasjoner.

Simuleringen ble modellert og gjennomført ved bruk av @Risk fra Palisade Decision Tools.

1. Et såtall er hva tilfeldig-tall-generatoren i et simuleringsprogram tar utgangspunkt i. Samme såtall vil gi samme strøm av tilfeldige tall (og dermed helt identisk resultat) hvis simuleringen kjøres på nytt

5.2 Resultater av simuleringen

Figur 5-2 viser resultatet av simulering og en sammenligning av Statens Vegvesens tall opp mot dette. Som vi ser av tabellen er forventningsverdien beregnet av SVV er så å si identisk med simuleringsresultatet. Avviket her ligger i samme størrelsesorden som avviket mellom de to simuleringsgjennomføringene. Standardavviket derimot har et mye større avvik på litt over 6 prosent. Dette stemmer bra med det vi har sett på tidligere om trinnvisformlenes nøyaktighet i kapittel 4. Konklusjonen er at formlene for forventningsverdien er gode, mens formlene for standardavviket blir noe unøyaktige som følge av at de ikke tar hensyn til mest sannsynlig verdi fra trippelanslaget.

Figur 5-2 *Simuleringsresultat. Verdier i 1000-kr*

	Sim 1	Sim 2	Avvik mellom simuleringene	Snitt av sim	SVVs tall	Avvik
Forventnings-verdi	1 559 142	1 559 282	0,01 %	1 559 212	1 558 921	-0,02 %
Standardavvik	107 282	107 767	0,45 %	107 524	100 890	-6,17 %
Relativt standardavvik	6,88 %	6,91 %	0,44 %	6,90 %	6,47 %	-6,15 %

6. Følsomhet for feil i inngangsdata

Det vi har sett på så langt er hvorvidt selve formelverket er korrekt under forutsetning om at alle inngangsdata er korrekt. Men hva om de ikke er det? Hva om vi bommer når vi anslår en verdi? Hva skjer da?

For å finne ut av det gjennomførte vi følsomhetsanalyser for trippelanslag. Måten dette er gjort på er ved at to av verdiene i trippelanslaget er holdt konstant mens den tredje varieres.

De to andre forutsetningene som er gitt i 4.1 antas fortsatt å være gyldige. Det vil si vi antar at kostnadselementet vi anslår har en virkelig sannsynlighetsfordeling, som er ukjent for oss, men som kan uttrykkes ved hjelp av en gammafordeling.

Hvor stort feilen blir på forventningsverdien er til dels avhengig av den virkelige kurvens skjevhet og av hvor stor spredning det er i verdiene i forhold til forventningsverdien, altså det relative standardavviket til den virkelig kurven.

For standardavviket er det kun den virkelige kurvens skjevhet som har noe å si.

Grafene som er visst i det følgende er for høyreskjeve fordelinger. For venstreskjeve kurver vil det bli de samme grafene bortsett fra at de vil være speilet, utslaget på y-aksen vil gå andre veien og kurvene for n vil være kurven for ø og vice versa.

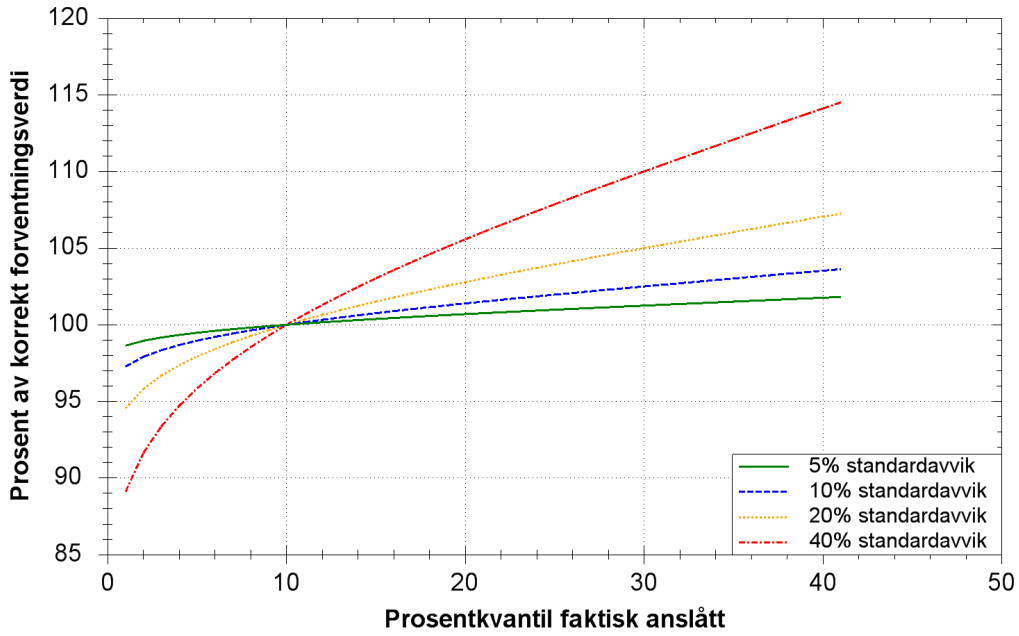
Det er tatt utgangspunkt i trippelanslag ved bruk av 10/90.

Vi har kun sett på følsomheten for feil inngangsdata for enkeltelementer. Det vil si der det er angitt verdier som standardavvik etc. er det verdiene for det enkelte element som er angitt, ikke totalt av en kostnadskalkyle.

6.1 Følsomhet for feil nedre anslag (n)

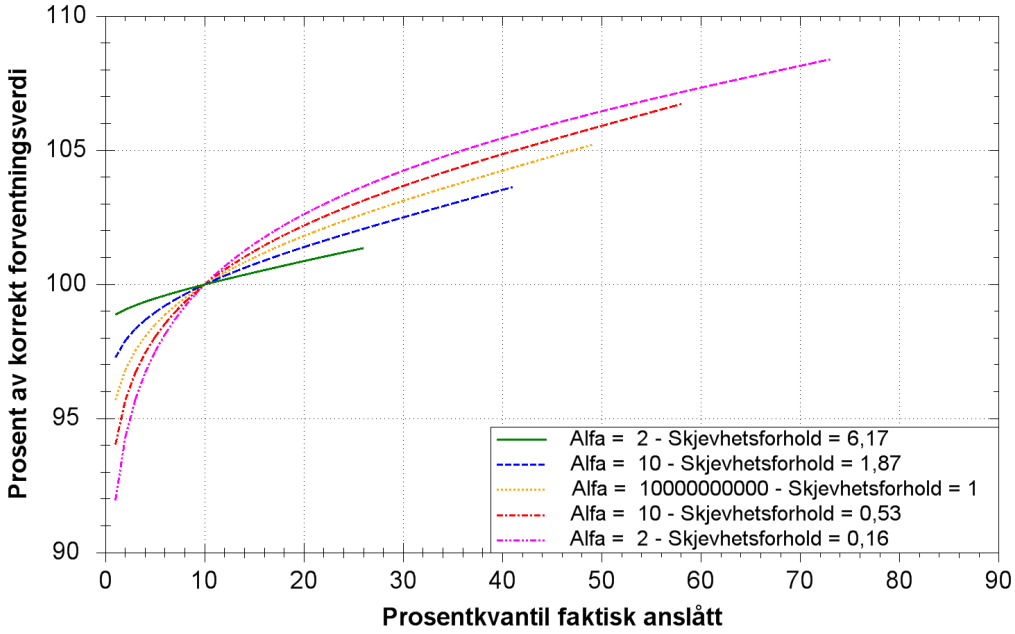
Figur 6-1 viser hvor stor feilen på forventningsverdien blir i prosent ved feil nedre anslag n. Den anslåtte n er her uttrykt ved hvilket prosentkvantil på den virkelige fordelingen vi faktisk anslår. Verdien som viser å x-aksen er altså "virkelighetens prosentkvantiler". Det er forutsatt at vi ikke kan anslå verdier som er utenfor den virkelige fordelingen. I og med at en høyreskjev gammafordeling er åpen på høyre side er dette en rimelig grei forutsetning. 100% kvantilet for fordelingen svarer til en verdi som er uendelig stor, og det er vel ganske så usannsynlig at vi anslår noe som er større enn dette. Den virkelige fordelingen vil dog ha et absolutt minimumspunkt, men i praksis er det ikke spesielt sannsynlig at vi anslår en nedre verdi som er hinsides mye lavere en hva den virkelige verdien er.

Følsomheten for n er i figuren vist fra 1%-kvantilet til 41%-kvantilet. Grunnen til den øvre grensen er at mest sannsynlig verdi, som vi forutsetter at vi har anslått korrekt, for en gammafordeling med Alfa = 10 ligger på 41,26%-kvantilet, Vi går her ut fra at vi ikke anslår nedre verdi n til å være større enn mest sannsynlige verdi (selv om dette er mulig ved ekstremt høyreskjeve fordelinger).



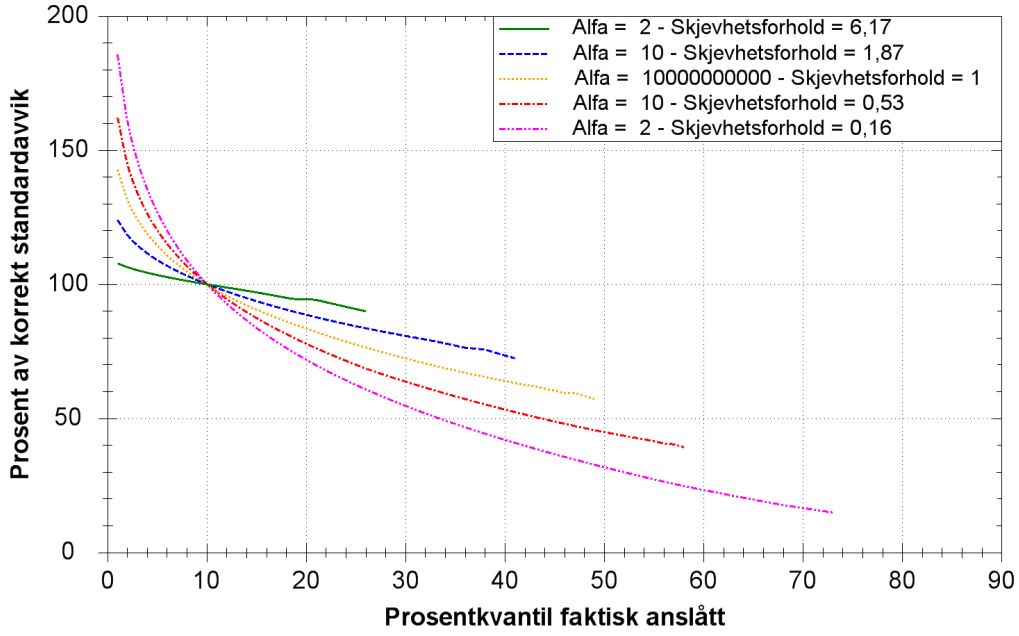
Figur 6-1 *Forventningsverdiens følsomhet for feil nedre anslag (n) for ulike standardavvik ved Alfa = 10*

Figur 26 viser hva feilen blir på forventingsverdi for ulike Alfa/skjevhetsforhold ved 10% standardavvik. Vi ser her at feilen på forventningsverdien blir større jo mer venstreskjev den virkelige fordelingen er. Dette kommer av at n her i utgangspunktet ligger lengre unna kurvens tyndgepunkt (tenk vektstangprinsippet) og at hvert prosentkvantil utgjør et større hopp i absolutte verdier. Å anslå n feil for svært høyreskjeve kurver gir små utslag, mens det for svært venstreskjeve kurver gir meget store utslag.



Figur 6-2 *Forventningsverdiens følsomhet for feil nedre anslag (n) for ulike Alfa ved 10% standardavvik*

TIPS! Jo skjevere fordelingen er, jo større er konsekvensen av å bomme. Hvis man har et skjevt kostnadselement er det svært viktig å anslå den mest ekstreme verdien riktig.



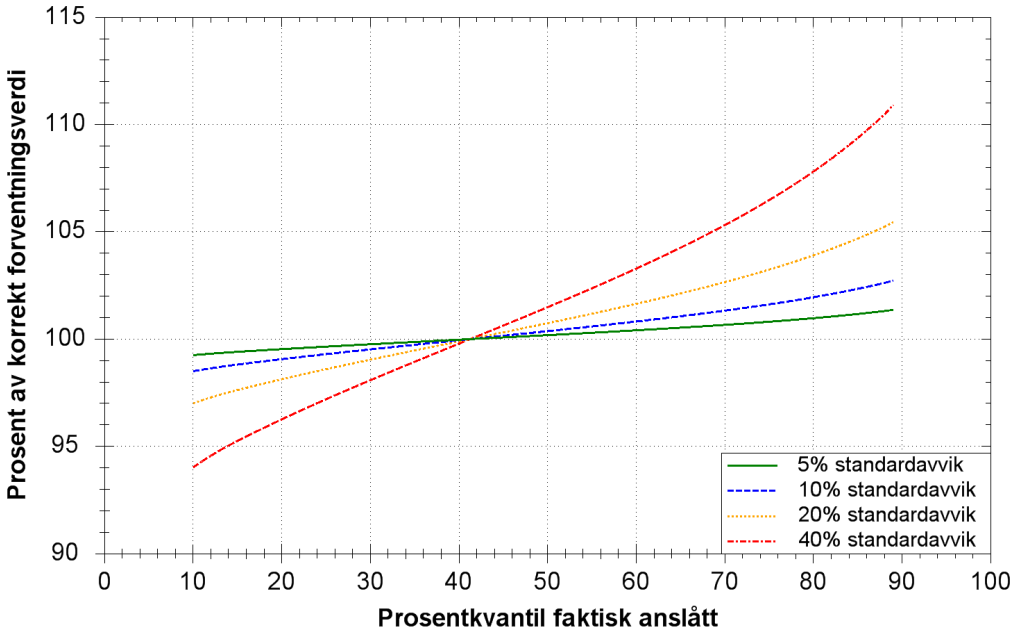
Figur 6-3 *Standardavvikets følsomhet for feil nedre anslag (n) for ulike Alfa*

Figur 6-3 viser standardavvikets følsomhet for feil nedre anslag. Vi ser her at hvis vi anslår n for lavt øker standardavviket. Dette er naturlig i og med at vi med å gjøre dette øker avstanden fra n til ø. Feilen på standardavviket blir også størst for venstreskjeve fordelinger på grunn av at hvert prosentkvantil representerer et større gap i absolutte størrelser for disse.

I PRAKSIS: Anslår du nedre verdi for lavt vil forventningsverdien gå ned og standardavviket vil øke

6.2 Følsomhet for feil anslag for mest sannsynlig verdi (s)

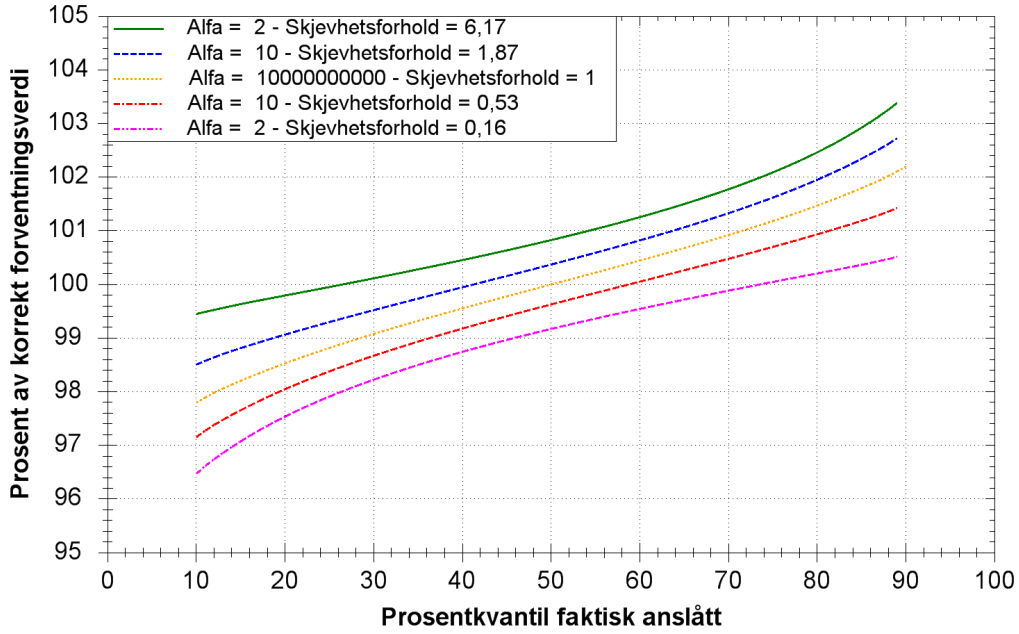
Figur 6-4 viser følsomheten for feil anslag av mest sannsynlig verdi for ulike relative standardavvik ved Alfa = 10. I og med at vi antar at vi har anslått nedre og øvre verdi korrekt er kurven avgrenset ved 10% og 90% kvantilene. Vi ser at feilen man gjør ved å anslå feil her er mye mindre enn hvis man anslår nedre verdi feil. Og som vi skal senere gjelder det samme i forhold til anslag av øvre verdi. Konsekvensen av å bomme på mest sannsynlig verdi er altså mye mindre enn å bomme på ytterpunktene



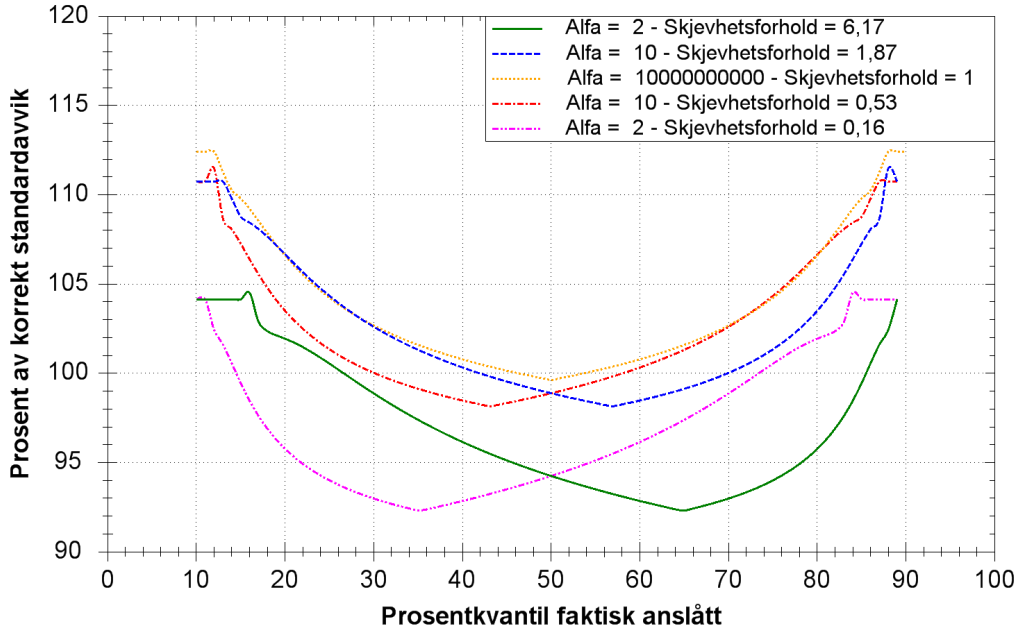
Figur 6-4 *Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike standardavvik ved Alfa = 10*

TIPS! Anslå ytterpunktene først før du anslår mest sannsynlig verdi.

Figur 6-5 viser forventingsverdiens følsomhet for feil mest sannsynlig verdi. I og med at alfaverdien bestemmer i hvilket prosentkvantil mest sannsynlig verdi ligger krysser kurvene 100% linja ved ulike punkter. Det samme gjelder for Figur 6-5.

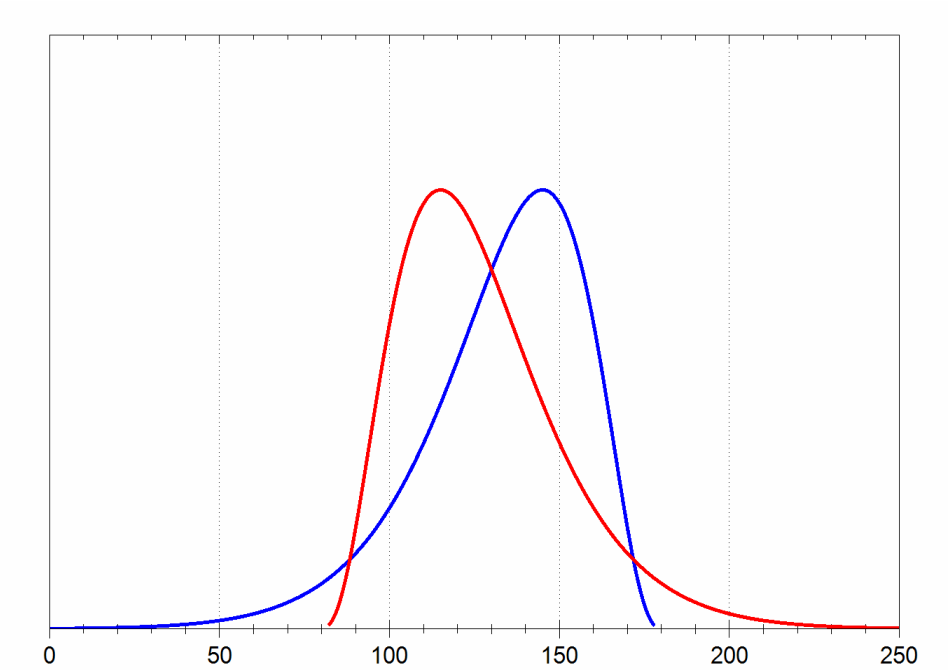


Figur 6-5 *Forventningsverdiens følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa ved 10% standardavvik*



Figur 6-6 *Standardavvikets følsomhet for feil mest sannsynlig anslag (s) for ulike Alfa*

Det man også kan observere for figur 6-6 er at standardavviket er korrekt på to punkter. Standardavviket blir korrekt hvis man anslår mest sannsynlig verdi til å ligge i det korrekte prosentkvantilet, men også hvis man anslår verdien til å ligge i 100% minus det korrekte kvantilet, med andre ord en kurve som er helt lik i formen, men speilvendt, forhold til den korrekte som vist i Figur 6-7. Hvis vi ser på formlene for standardavviket fra avsnitt 3.1 og definisjonen av en venstreskjev gammafordelingpå side 18 i kapittel 3 ser vi lett hvorfor standardavviket her bli det samme.



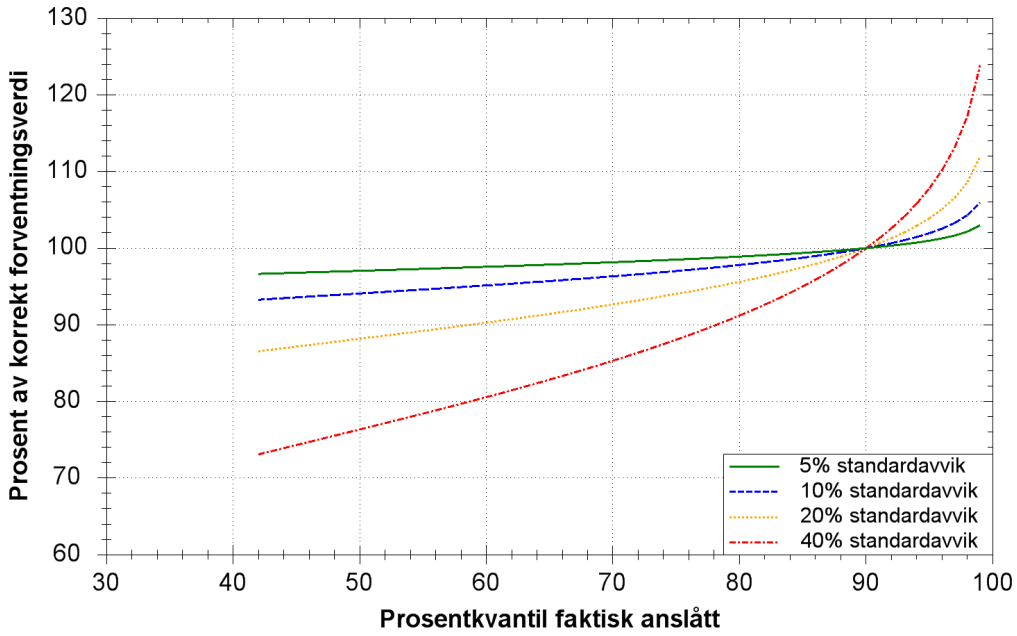
Figur 6-7 *Helt like i form, men speilvendt*

$\sigma = \sqrt{\alpha\beta'^2}$ og vi definerer en venstreskjev gammafordeling til å ha negativ Beta.

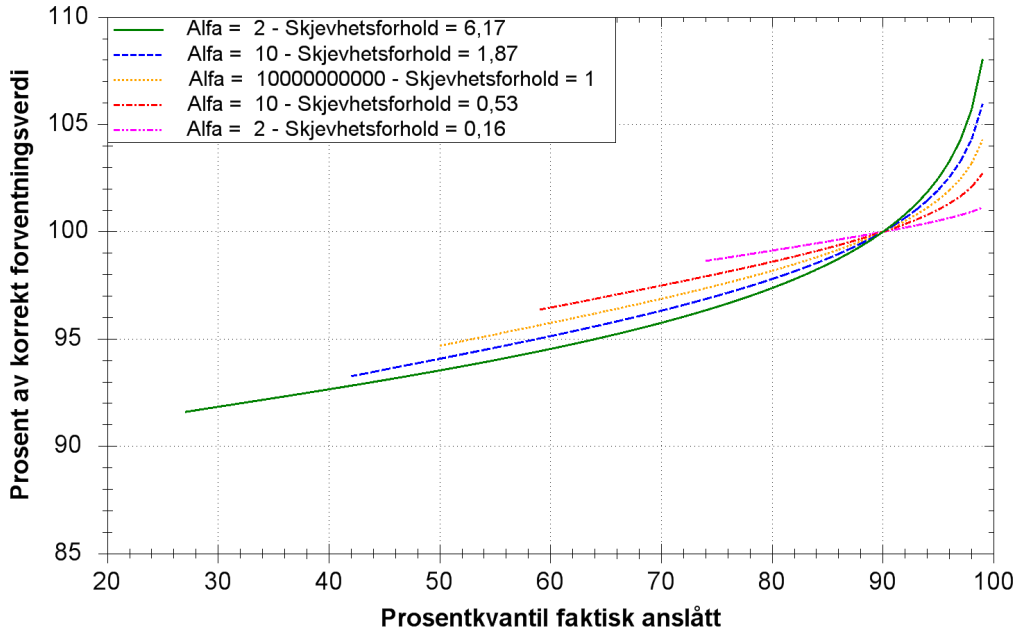
I og med at Beta kvadreres i formlene får ikke det negative fortegnet noe å si og standardavviket blir det samme. (Merk at det ikke blir slik for forventningsverdien for her kvadreres ikke Beta og forskyvningsparameteren som også spiller inn vil ikke være den samme.

6.3 Følsomhet for feil øvre anslag (ø)

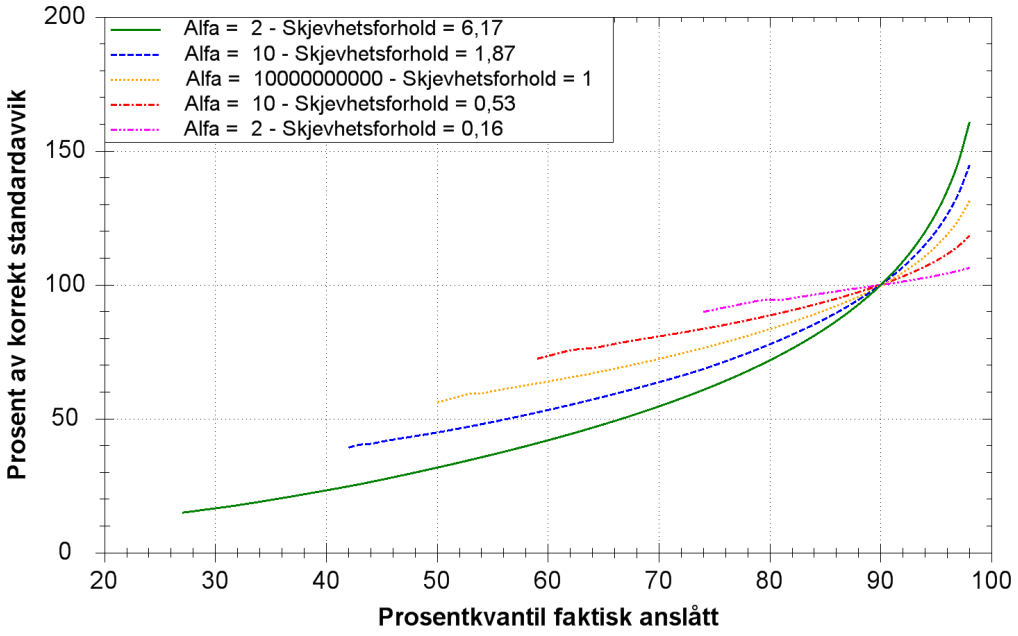
Figur 6-8, 6-9 og 6-10 viser det samme for feil øvre anslag som figurene som ble vist i avsnitt 6.1 for feil nedre anslag. Utslagene blir større for øvre anslag i og med at vi i har tatt utgangspunkt i en høyreskjev fordeling med Alfa = 10. Hadde vi valgt en venstreskjev fordeling hadde resultatet blitt motsatt. Det vil si kurvene blir like i formen, men speilet om 50%-kvantilet og 100%-korrekt linjen.



Figur 6-8 *Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike standardavvik ved Alfa = 10*



Figur 6-9 *Forventningsverdiens følsomhet for feil øvre anslag (ø) for ulike Alfa ved 10% standardavvik*



Figur 6-10 *Standardavvikets følsomhet for feil øvre anslag (ø) for ulike Alfa*

6.4 Konklusjoner om følsomhet for feil i inngangsdata

Noe av det vi har sett her er ikke direkte overraskende. Jo større den relative usikkerheten er på den virkelige fordelingen, jo større blir konsekvensen av å anslå feil. Det som kanskje er mest interessant her er å se hvor liten konsekvensen av å anslå feil mest sannsynlig verdi er i forhold til å anslå ytterpunktene feil. Dett understreker viktigheten av å i en analysesammenheng å først bestemme disse før man går over til å bestemme den mest sannsynlige verdien.

Det vi også kan se er at feil anslag av inngangsverdiene fører til langt større feil i resultatene enn det som finnes av unøyaktigheter i Statens Vegvesens formelverk. Trinnsvis-formelen for forventningsverdien er såpass nøyaktig at enhver feil i inngangsdataene vil gi større utslag enn en eventuell feil som følge av formelen. Formlene for standardavviket er som sagt mer unøyaktig og feilen der er avhengig av hvor skjev fordelingen er. For de skjevhetsforholdene man typisk finner i praksis i usikkerhetsanalyser blir feilene på standardavviket større enn hva den potensielle formelfeilene er hvis vi plasser et anslag 1 prosentkvantil feil. For eksempel, hvis vi anslår 10 %-kvantilet til å være verdien som faktisk er 11%-kvantilet, vil feilen på forventningsverdien være større enn en eventuell feil som følge av unøyaktighet i formlene. (Dette gjelder forutsatt at vi bruker en korrekt verdi for Theta.)

7. Konklusjoner

Statens Vegvesens verktøy Anslag regner omtrent så korrekt som det går an å bli innenfor dagens gjeldende formelparadigme. Feilen som gjøres på forventningsverdien er neglisjerbar med tanken på potensielle feil som følge av gale inngangsdata. Feilen på det absolutte standardavviket er noe større, men ser vi på det relative standardavviket blir også feilen her svært liten. Det er i alle fall ikke mulig å få det til å bli mer korrekte ved bruk av de standard trinnvisformlene.

Vi har sett på et par alternative metoder her som man bør vurdere når man i fremtiden skal oppgradere eller bytte ut analyseverktøyene. Dynamisk utregning av Theta og Zeta verdier gir en stor forbedring i formlenes nøyaktighet og er enkelt nok til at det ikke vil by på noen problemer å implementere metoden i et dataverktøy. Den andre alternative metoden er å regne ut alfa, beta og forskyvning fra trippelanslaget og så regne med nøyaktige analytiske formler. Denne metoden er den mest nøyaktige av de analytiske metodene som er mulig å implementere i et dataprogram, men den er betydelig vanskeligere å implementere. Monte Carlo simulering er selvsagt et alternativ til å benytte analytiske formler. Å vurdere analytiske metoder kontra simuleringer har vært utenfor mandatet, men det som kan sies her om Monte Carlo simuleringer er at så lenge struktur og innputt er rett, og såfremt man kjører simuleringen med tilstrekkelig antall iterasjoner, vil resultat som kommer ut være tilnærmet helt korrekt.

Den kanskje viktigste konklusjonen vi kan trekke ut i fra dette arbeidet er hvor avgjørende det er å arbeide systematisk og godt med å få frem rett innputt (både tallverdier og struktur). Feil man gjør her vil stort sett bestandig overskygge eventuelle feil som kommer av formlene.

Litteratur

Austeng, Kjell & Hugsted, Reidar (1995): *Trinnvis kalkulasjon*. Institutt for bygg- og anleggsteknikk, NTH.

Kreyszig, Erwin (1999): *Advanced engineering mathematics*, 8th edition. John Wiley & Sons. ISBN 0-471-15496-2

Law, Averill M. & Kelton, David W. (2000): *Simulation Modelling and Analysis, third edition*. McGraw-Hill Higher Education. ISBN 0-07-059292-6

Lichtenberg, Steen. (1978): *PROJEKT PLANLEGGING - i en foranderlig verden*. Polyteknisk Forlag Lyngby.

Palisade Corporation. (2002): *A Concise Summary of @RISK Probability Distribution Functions*. Palisade Corporation.

Spiegel, Murray R. & Stephens, Larry J. (1999): *Schaum's Outline of Theory and Problems of Statistics Third Edition*. McGraw-Hill. ISBN 0-07-060281-6

Walpole, Ronald E, Myers, Raymond H. & Myers, Sharon L. (1998): *Probability and Statistics for Engineers and Scientists*. Prentice Hall. ISBN 0-13-095246-X

Vedlegg-A Hvordan finne Alfa, Beta og forskyvning fra trippelanslag

For å kunne benytte algoritmene som er oppgitt her må man tilgang på en funksjon for å regne ut den inverse kumulative tetthetsfunksjonen til en gammafordeling, slik som funksjonen GAMMAINV i Excel. Denne er dog ikke særlig egnet den ikke klarer verdier for alfa større enn ca 300. Eller finnes kildekode til funksjoner som regner ut den inverse kumulative tetthetsfunksjonen til en gammafordeling fritt tilgjengelig på internett.

Denne algoritmen klare å regne ut alfa for skjevhetsforhold opp til litt over 13 000 og takler i tilegg spesialtilfellet der n = s =eller s = ø. Algoritmen takler ikke tilfellet der n = s = ø da dette tilfellet ikke gir noen mening for en gammafordeling. Algoritmen forutsetter en høyreskjev kurve. Hvis man har en venstreskjev kurve må tallene transformeres først. Se "Transformasjon av trippelanslag" på side 19.

Algoritme for å finne Alfa

Inngangsparametre: n, s og ø (nedre, mest sannsynlig og øvre anslag)

```
1.  Hvis n = s eller ø = s returner alfa = 1,1563

2.  Regn ut skjevhetsforholdet SFnsø fra n, s og ø.

3.  Hvis SFnsø = 1 returner alfa = 1 000 000 000 000. Se
    merknad 1

4.  Hvis SFnsø < 1, SFnsø = 1/SFnsø

5.  Gjør et grovt anslag på alfa. Se merknad 2

6.  Sett alfa = 10, beta = 10

7.  Regn ut skjevhetsforholdet SFab fra alfa og beta. Se
    merknad 3

8.  Regn ut den interne forholdet mellom
    skjevhetsforholdene. i = SFab / SFnsø

9.  Hvis i = 1 ± 0,0001 avslutt og returner alfa

10. Hvis vi antall iterasjoner hittil er større enn 50
    reduser i. Se merknad 4

11. Hvis i > 1 alfa = alfa * i
    Gå tilbake til punkt 3

12. Hvis i < 1 alfa = alfa / (2 - i)

13. Gå tilbake til punkt 3.
```

Merknad 1

Alfa bør i dette tilfellet settes så stor som mulig. Hvor stor den bør settes til vil være avhengig av datasystemet og hva verdien skal brukes til i senere bergninger. Det bunner ut i er at man skal sette den så stor som mulig slik at den ikke forårsaker overflow feil andre steder i programmet.

Merknad 2

Formel for å gjøre grovt anslag av alfa:

$$\text{alfa} = \frac{3978883,4}{f^{100}} - \frac{57,7483922}{e^{f^2}} + \frac{404,931744}{e^f} + 5,04650656$$

f i formelen er lik skjevhetsforholdet SFnsø i algoritmen.

Merknad 3

Se egen algoritme.

Merknad 4

Når vi har en veldig skjev kurve vil alfa være lav og små endringer i alfa representerer store endringer i skjevhetsforholdet. For å klare å få algoritmen til å konvergere må vi da redusere i for å få til mindre endringer av alfa.

i reduseres i første omgang etter hva den hittil beregnede alfa ligger på som følger:

$$\alpha < 2 \qquad i = \sqrt{i}$$

$$\alpha < 1,75 \qquad i = \sqrt[4]{i}$$

$$\alpha < 1,25 \qquad i = \sqrt[8]{i}$$

$$\alpha < 1,2 \qquad i = \sqrt[16]{i}$$

I enkelte tilfeller vil heller ikke denne reduksjonen være tilstrekkelig til å få algoritmen til å konvergere, vi må i så fall redusere i enda mer. Hvis i fortsatt ikke har konvergert når vi har nådd 5000 iterasjoner setter vi $i = \sqrt[16]{i}$

Dette kommer i tilegg til den reduksjonen vi eventuelt har gjort tidligere som følge av lav alfa. Hvis algoritmen fortsatt ikke har konvergert når vi har nådd 10 000 iterasjoner reduser vi igjen $i = \sqrt[4]{i}$

Også denne gangen i tilegg til eventuelt andre reduksjoner som er gjort.

Regne ut skjevhetsforholdet fra alfa og beta

Denne algoritmen kan brukes til å finne skjevhetsforhold for en gammafordelingen der man kjenner alfa og beta.

Inngangsparametre: alfa, beta, nkvan, økvan der nkvan og økvan er kvantilene vi benytter for nedre og øvre anslag.

```
1.   Regn ut s ved s=beta * (alfa-1)

2.   Regn ut n ved å ta den inverse kummulative
     tetthetsfunksjonen for gammafordelingen for nkvan, alfa
     og beta.

3.   Regn ut ø ved å ta den inverse kummulative
     tetthetsfunksjonen for gammafordelingen for økvan, alfa
     og beta.

4.   Regn ut skjevhetsforholdet som normalt,
     skjevhetsforhold = (ø-s)/(s-n)
```

Algoritme for å finne beta

Denne algoritmen baserer seg på at for en gitt alfa eller skjevhetsforhold, vil en beta-verdi representere en spesifikk avstand mellom to prosentkvantiler.

Denne algoritmen forutsetter en høyreskjev kurve. Hvis man har en venstreskjev kurve må tallene transformeres først. Se "Transformasjon av trippelanslag" på side 19. Hvis man benytter definisjonen av at venstreskjeve gammafordelinger har negativ beta må man i tilegg endre fortegn på beta etter å ha kjørt algoritmen som den står.

Inngangsparametere: alfa, n, ø, n-kvantil, ø-kvantil

```
1.   Regn ut hva n og ø vil være for den beregnede beta ved
     bruk av den inverse kumulative tetthetsfunksjonen for de
     aktuelle prosentkvantilene. Vi kaller disse verdiene xn
     og xø.

2.   Regn ut i som følger: i = (ø-n)/(xø-xn)

3.   Hvis i <1,0001 og i>0,9999 avslutt og returner beta

4.   Hvis i>1 beta = beta * i. Gå tilbake til 1.

5.   Hvis beta = beta / (2-i). Gå tilbake til 1.
```

Algoritme for å finne forskyvning

Forutsatt at man vet alfa og beta for en gammafordeling kan man regne ut hva forskyvningen vil være ved hjelp av et hvilket som helst punkt på kurven som man vet x-verdi og prosentkvantil for.

Inngangsparameter: alfa, beta, x-verdi, kvantil.

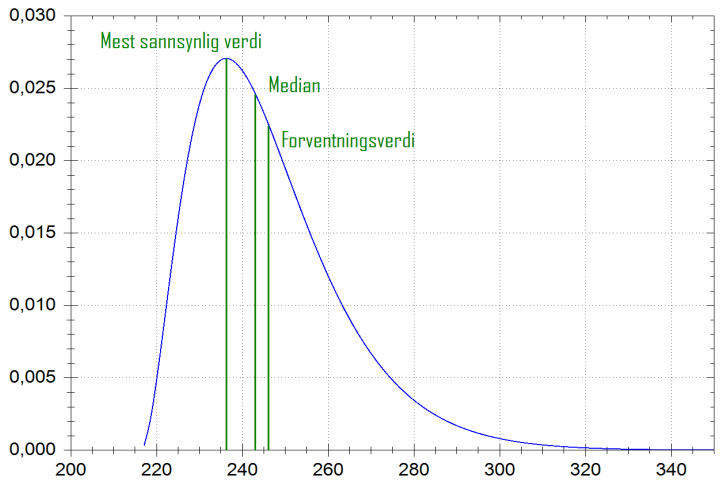
```
1.   Regn ut x-verdien er for det oppgitte kvantilet for en
     uforskjøvet gammafordeling.

2.   Returner resultatet fra 1 minus x-verdi
```

Vedlegg-B Begrepsdefinisjoner

Gjennomsnitt - mål på sentraltendens

Et gjennomsnitt er en verdi som er typisk, eller representativ, for et sett med data. Siden slike typiske verdier tenderer til å ligge sentralt innen et sett med data ordnet etter størrelse, gjennomsnitt kalles også mål på sentraltendens.

Flere ulike typer gjennomsnitt kan defineres. De mest vanlig er den aritmetiske middelverdien (arithmetic mean), medianen (median) og mest sannsynlig verdi (mode)



Forventningsverdi

Et av målene for en sannsynlighetsfordelings sentraltendens. Forventningsverdien representerer kurvearealets tyngdepunkt. Benevnes vanligvis matematisk som E(x).

Mode - Mest sannsynlig verdi

Mode eller mest sannsynlig verdi er et av målene for en sannsynlighetsfordelings sentraltendens. Moden representerer den verdien som det er størst sannsynlighet for at vil opptre, det vil si toppunktet i en sannsynlighetsfordeling.

Median

Et av målene for en sannsynlighetsfordelings sentraltendens. Medianen representer den verdien som er midt i fordelingens sannsynlighetsmasse. Det vil si at medianen er den verdien hvor det er like stor sannsynlighet at man får en større eller mindre verdi. For en kontinuerlig sannsynlighetsfordeling vil det si det punktet på kurven der halvparten av arealet til tetthetskurven ligger til venstre og den andre halvparten av arealet ligger til høyre. For symmetriske fordelinger er denne lik forventingsverdien og den mest sannsynlige verdien (mode)

Varians

Variansen til en sannsynlighetsfordeling er et mål på statistisk spredning. Den indikerer hvor langt fra forventningsverdien verdiene i fordelingen er. For en ikke-kontinuerlig fordeling kan vi definerer det som "gjennomsnittet av kvadratet av avstanden av hvert datapunkt fra standardavviket. Benevnes matematisk som σ^2.

Standardavvik

Standardavviket er det vanligste målet for statistisk spredning. Det er definert som kvadratroten av variansen. Standardavviket er definert som dette for å gi oss et mål på spredningen som er et ikke-negativt tall og som har denne samme enheten som verdien den gir spredning om. Benevnes matematisk som σ.

Relativt standardavvik

Relativt standardavvik sier hvor stort standardavviket er i forhold til forventningsverdien. Relativt standardavvik = standardavvik/forventningsverdi.

Kvantil og prosentkvantil

Kvantiler er en type beliggenhetsmål. De angir hvor på sannsynlighetsskalaen en verdi befinner seg. Hvis vi deler opp en sannsynlighetsfordeling i et vilkårlig antall like store (i absolutte verdier langs x-aksen) deler, kaller vi hvert av grensepunktene til disse av delene et kvantil. Deler vi fordelingen opp i 100 like store deler kaller vi det prosentkvantiler. Et gitt prosentkvantil P representer den X-verdien på kurven det er P% sannsynlighet at ikke overstiges.

Trippelanslag

Tre anslåtte verdier for et kostnadselement, faktor etc. som til sammen bestemmer sannsynlighetsfordeling for det aktuelle kostnadselement faktor etc. Normalt sett anslås en nedre eller best tenkelig verdi og en øvre eller best tenkelig verdi i tilegg til den verdien man anser som mest sannsynlig vil opptre. Nedre og øvre anslag anslås som regel i motsatte prosentkvantiler, det vil for eksempel si 1 og 99 kvantilene eller 10 og 99.

Nedre anslag N

Nedre verdi som anslås i et trippelanslag. Mest vanlig å definere denne verdien til å ligge i 1 eller 10 prosentkvantilet.

Mest sannsynlig anslag S

Midtpunktet som anslås i et trippelanslag. Se forøvrig mode

Øvre anslag Ø

Øvre verdi som anslås i et trippelanslag. Mest vanlig å definere denne verdien til å ligge i 99 eller 90 prosentkvantilet.

Høyre- og venstreskjevhet

En høyreskjev sannsynlighetsfordeling har en hale mot høyre. Fordelingens mest sannsynlig verdi, eller mode, er da mindre enn medianen mens forventningsverdien er større enn medianen. For en venstreskjev fordeling er det motsatt, den har en hale mot venstre og fordelingens mest sannsynlige verdi er større enn medianen mens forventningsverdien er mindre enn medianen. For en helt symmetrisk kurve vil alle disse tre verdiene være like.

Skjevhetsforhold

Begrepet innføres i denne rapporten for å si hvor høyre- eller venstreskjev en gammafordeling er. Skjevhetsforholdet er definert som avstanden mellom øvre og mest sannsynlig anslag delt på avstanden mellom nedre og mest sannsynlig anslag. Det vil si at et skjevhetsforhold større enn én tilsier en høyreskjev kurve

Alfa (α)

Alfa er en ikke-lineær parameter som bestemmer gammafordelingens form (eller skjevhet). En lav Alfa tilsier en skjev kurve, mens en uendelig stor Alfa tilsier en kurve som er identisk med normalfordelingen bortsatt fra at den ikke er åpen i den ene enden..

Beta (β)

Beta er parameteren som skalerer gammafordelingen.

Forskyvning

En gammafordeling står per definisjon fast i null, men ved å forskyve den kan man sette dette minimumspunktet til hvor man vil.

Theta (θ)

Theta er vektfaktoren som benyttes på mest sannsynlig verdi i Trinnvis-formelen for forventningsverdien.

$$\frac{n + \theta s + ø}{2 + \theta}$$

Zeta (ζ)

Zeta er delingsfaktoren som benyttes i Trinnvis-formelen for standardavviket.

$$\frac{ø \angle n}{\zeta}$$