

Teknisk-naturvetenskapliga
fakulteten

Examensarbete i Matematisk statistik
D-nivå, 20 poäng

Analys av fiberlängdsfördelning hos gran (*Picea Abies*) och tall (*Pinus Sylvestris*)

Hilma Johansson

Institutionen för matematik
och matematisk statistik
Umeå universitet
901 87 UMEÅ

Analys av fiberlängdsfördelning hos gran (*Picea Abies*) och tall (*Pinus Sylvestris*)

Hilma Johansson

Examensarbete i Matematisk statistik
D-nivå, 20 poäng

December 2007

Hilma Johansson: *Analys av fiberlängdsfördelning hos gran (Picea Abies) och tall (Pinus Sylvestris)*
Institutionen för matematik och matematisk statistik, Umeå Universitet, 901 87 Umeå

©Hilma Johansson, 2007

Detta arbete har skrivits för avläggande av magisterexamen i matematisk statistik.

Arbetet har gjorts på D-nivå och omfattar 20 poäng.

Handledare: Sara de Luna

Examinator: Lennart Bondesson

Hilma Johansson: *Analysis of fibre length distributions for spruce (Picea Abies) and pine (Pinus Sylvestris)*
Department of Mathematics and Mathematical Statistics, Umeå University, 901 87 Umeå

© Hilma Johansson, 2007

This is a thesis for the degree of Master of Science in Mathematical Statistics.

Supervisor: Sara de Luna

Examiner: Lennart Bondesson

Sammanfattning

Fiberlängderna i träd styr viktiga egenskaper som t.ex. styrka och jämnhet hos papper och andra träprodukter. Inom skogsforskningen är man intresserad av att ta reda på fiberlängder i stående träd för att kunna utvärdera olika behandlingar och försök utan att såga ner träden. Bl.a. har Mörling m.fl. (2003) utvecklat metoder för att skatta fiberlängder i stående träd utan att skada dessa. Metoderna bygger på att man känner till vilken typ av fördelning fiberlängderna följer. I denna rapport undersöks fiberlängder från gran och tall tagna med 30 mm tändsticksformade prov med syfte att hitta en typ av fördelning som är flexibel nog att beskriva olika fiberlängdsfördelningar. Rapporten bygger vidare på Per Livs examensarbete (2005) där fem typer av fördelningar undersöktes på data från gran. Hans resultat var att de fördelningar han testade var för symmetriska för att anpassa sig väl till data.

Här undersöktes tre andra fördelningar vars parametrar skattades med hjälp av maximum likelihoodmetoden och jämfördes med Livs fördelningar. Det visade sig att den generaliserade gammafördelningen anpassade sig mycket bra till data från både gran och tall. Ett avsnitt ägnas därför åt ett exempel på test av skogliga hypoteser med hjälp av denna fördelning.

Abstract

Tree fiber lengths controls properties like strength and smoothness in paper and other wood products. For forest researchers it is of interest to be able to determine the fiber lengths in standing trees. Mörling et al. (2003), among others, have developed methods to estimate fiber length distributions in standing trees without damaging them. The methods require knowledge of the type of distribution that fiber lengths follow. In this thesis fiber lengths from spruce and pine are examined from 30mm high sticks. The purpose is to find a type of distribution that is flexible enough to capture many different kinds of fiber length distributions observed in the trees. The thesis continues the diploma work of Per Liv (2005) where five types of distributions were studied on fiber length data from spruce. He concluded that the distributions he tested were too symmetric to fit well to the data.

Here three other distributions were tested whose parameters were estimated using the maximum likelihood method and compared to the distributions from Liv's work. The generalized gamma distribution turned out to describe fiber length data very well from both spruce and pine. Therefore one chapter is devoted to examples of discussions of hypothesis from forestry using this distribution.

Förord

Ett stort tack till mina handledare Sara Sjöstedt-de Luna och Tommy Mörling som gav mig möjligheten att göra detta examensarbete. Trots många svårigheter längs vägen, inte minst min flytt, tycker jag att det har varit roligt och intressant att lära av er och att ni har varit mycket generösa med er tid.

Ett tack också till Per Liv som tillhandahållit sina programmeringsfiler och förklarat sitt examensarbete för mig för att underlätta mitt arbete.

Med detta examensarbete avslutar jag mina studier i matematisk statistik vid Umeå universitet och jag vill därför passa på att även tacka alla jag kommit i kontakt med på Institutionen för matematik och matematisk statistik. Jag har alltid blivit vänligt bemött av er, både som student och som anställd, och jag har många trevliga minnen med mig i bagaget när jag nu gått ut i arbetslivet.

Stockholm, november 2007

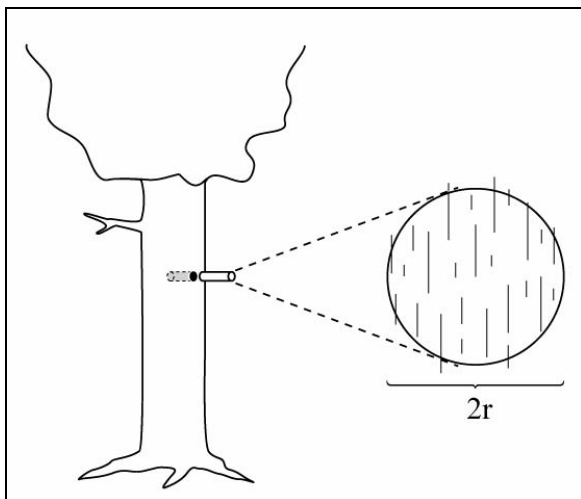
Hilma Johansson

1 Introduktion

Ett trädets fiberlängder beskriver viktiga egenskaper för såväl sågade varor som för massa- och pappersprodukter. Exempel på dessa egenskaper är styrka och jämnhet i papper (Zobel & van Buijtenen 1989). Om man känner till fördelningen för fiberlängder i ett träd kan man anpassa användandet därefter. Att utveckla metoder att skatta dessa fördelningar för stående träd är därför värdefullt för skogsforskningen. De ska vara enkla, kostnadseffektiva och skonsamma för trädet. Det skulle också kunna vara till hjälp inom skogsforskningen för att utvärdera effekter av genetisk förädling och skogskötselåtgärder i försöksträd utan att såga ner dem.

Med fibrer i denna uppsats menas trakeider som utgör ca 90 % av cellerna i ett träd (Ilvessalo-Pfäffli 1995). De kan vara mellan 1-7 mm och är ungefär hundra gånger längre än de är breda. Trakeider förser trädet med näring och vatten samt stabiliserar stammen så att trädet kan stå emot yttre påfrestningar som vind och snötryck. Märgstråleceller är en annan typ av näringslagrande celler som ligger radiellt i trädet och tillsammans med andra mindre vanliga celltyper utgör de den resterande andelen.

Mörling m.fl. (2003), Svensson m.fl. (2006) och Svensson (2007) har utvecklat metoder för att skatta fiberlängdsfördelningar i stående träd, som inte skadar trädet nämnvärt och är kostnadseffektiva. Metoderna baseras på vedprov i form av 5 mm borrhävar. Fiberlängderna i borrhävarna mäts sedan med hjälp av Kajaani FiberLab som är en automatisk optisk fiberanalysator. De föreslagna metoderna bygger på att man vet vilken typ av fördelning som fiberlängder följer. Det är då viktigt att hitta rätt typ av fördelning som är anpassningsbar nog att klara av olika egenskaper hos fiberlängderna både inom och mellan träd. T.ex. finns det studier som visar att fiberlängderna ökar med ökat avstånd från märgen inom ett träd. Sedan avtar ökningen ju närmare barken man kommer (Atmer & Thörnqvist 1982). I detta arbete kommer vi att studera vilken typ av fiberlängdsfördelningar som verkar vara rimlig att ansätta för tall och gran för att fånga upp bl.a. sådan radiell fiberlängdsvariation som beskrivs ovan.



Figur 1. Exempel på borrhävarprovtagning i ett stående träd.

Sedan länge har vedprover från stående träd i form av 5 mm borrhävar använts som metod för att mäta olika mått i ett stående träd, t.ex. stamdiametertillväxt, veddensitet och mikrofibrillvinkel. I figur 1 ses ett exempel på hur borrhävarna tas ur det stående trädet och hur fibrerna ligger vertikalt i stammen. Eftersom fibrerna kan vara så långa som 6 mm blir det problem med metoder där man

använder en 5 mm borrhärna då alltför många fibrer kapas vid provtagningen. Dessutom begränsas den maximala längden på en fiber till 5 mm och datamaterialet blir överrepresenterat av korta fibrer. Vedprov med mindre andel kapade fibrer kan erhållas via tändsticksformade vedprov, ca 30 mm höga. Nackdelen med tändsticksformade prov är att trädet måste fällas för att proverna ska kunna tas. Detta är inte optimalt då man vill kunna uttala sig om fiberlängderna i stående träd, men är användbart då vi vill undersöka vilken typ av fördelning som fiberlängder följer. När man väl vet det kan man använda Svenssons med fleras (2006) metoder för att på ett icke-destruktivt sätt ta reda på fiberlängdsfördelningar. Liv (2005) undersökte i sitt examensarbete hur väl lognormal-, trunkerad normal-, gamma-, Weibull- och betafördelning stämmer överens med granfiberfördelningar genom att använda tändsticksformade prov. Slutsatsen från Livs arbete var att ingen av fördelningarna var flexibel nog att kunna anpassa sig till data som uppvisar både negativ och positiv skevhet.

Syftet med denna rapport är att hitta en typ av fördelning som är flexibel nog att anpassa sig väl till olika slags fiberlängdsfördelningar för tall och gran och som kan fånga mått som skevheten på ett tillfredsställande sätt. Vi går därför vidare och undersöker tre andra fördelningar, Johnson Sb, segmenterad och generaliserad gamma, för både gran- och tallfibrer. Vi jämför dessa med de fördelningar Liv undersökte i sitt examensarbete dels för att se hur hans fördelningar anpassar sig på talldata och dels hur de står sig mot de tre nya, förhoppningsvis mer anpassningsbara, fördelningarna på data från både gran och tall. För att minimera antalet kapade fibrer och få ett mer representativt datamaterial använder vi oss av en 30 mm tändsticksformad provsticka. Målet är dock att hitta en fördelning som är så pass bra att man i framtiden kan använda metoder med 5 mm borrhärnor vid provtagning. På så sätt kommer vi att kunna uttala oss om ett träds fiberegenskaper utan att nämnvärt skada det, medan det fortfarande växer i skogen.

I avsnitt 2 berättar vi mer om datamaterialet och hur vi bearbetat det för att kunna analysera det på bästa sätt. Avsnitt 3 förklarar dels ML-skattningar generellt och går sedan in djupare på respektive fördelnings skattningar. I avsnitt 4 jämförs fördelningarna på data från gran medan vi i avsnitt 5 gör samma sak på data från tall. Dessa två kapitel indikerar att generaliserad gammafördelning är den fördelning som bäst anpassar sig till fiberlängderna. Avsnitt 6 illustrerar hur man skulle kunna gå till väga då man studerar skogliga frågeställningar/hypoteser kring fiberlängdsfördelningar, i detta fall rörande de talldata som vi har. Fiberlängdsfördelningarna är här de uppskattade generaliserade gammafördelningarna. Testerna ska ses som ett exempel på framtida användningsområde för fördelningen. Rapporten avslutas med diskussion i avsnitt 7.

2 Datamaterialet

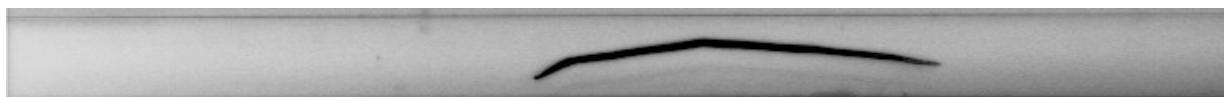
För att undersöka vilken typ av fördelning som bäst passar fiberlängder har vi haft tillgång till data från 7 tallar (*Pinus Sylvestris*) och 3 granar (*Picea abies* (L) Karst). Bland de tallar som undersökts är fem från södra Sverige, två snabbväxande (hädanefter kallade *Snabb1* och *Snabb2*) och två långsamt växande träd (*Långsam1* och *Långsam2*). Den femte tallen från södra Sverige kallas *Söder*. Två tallar från Norge finns också med i studien, där den ena har vuxit skyddat (*Skyddat*) och den andra i ett vindutsatt område (*Utsatt*). De tre granar som användes i Per Livs examensarbete (Liv, 2005) kallas här *träd 1*, *2* och *3*. Granarna kommer från ett ca 80-årigt skogsbestånd i Vindelns försökspark i norra Sverige. Från tallarna har prover tagits både vid mörgen (position 1), vid femtonde årsringen (position 2) och vid barken (position 3). Data från gran togs vid två positioner, en vid mörgen och en vid barken.

Avsnitt 2.1 – 2.4 följer i stort den presentation som gjordes av Liv (2005).

2.1 Mätning och bedömning

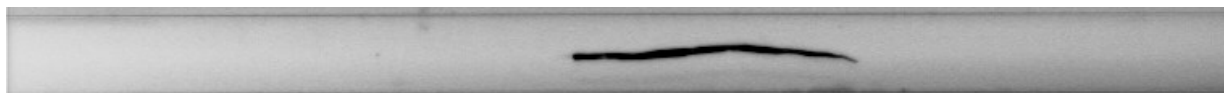
Det datamaterialet som har använts i detta arbete bestod av mätvärden från Kajaani FiberLab 3.5 (Metso Automation). De vedprov som togs från träden är tagna i brösthöjd (1,3 m) och formade som stickor och ca 30 mm höga longitudinellt. Radiellt varierade proven från ca 2 till 8 mm beroende på årsringsbredd. Varje prov omfattade minst en årsring och vid breda årsringar blev således provbitarna breda radiellt. Variationen i tangentiell ledd var betydligt mindre, ca 5-7 mm. Proven löstes upp i en 1:1 lösning väteperoxid och ättiksyra i ca 24 timmar vid 70° C. Därefter sköljdes och skakades proverna i vatten för att separera fibrerna. Innehållet späddes sedan med vatten till ca 2,5 liter och ur varje lösning togs sedan ett 25 ml delprov som analyserades med Kajaani FiberLab.

Kajaani FiberLab mäter flera hundra fibrer i minuten och är därför väldigt effektiv. Apparaturen klarar dock inte automatiskt att skilja på kapade och okapade fibrer och klarar inte heller att bedöma om fibrer klumpat ihop sig eller böjt sig så att mätningarna blir felaktiga. För varje fiber tar dock Kajaani FiberLab en bild. Dessa bilder har vi sparat och använt oss av för att i efterhand manuellt undersöka varje fiber med hjälp av ett webbaserat program och sälla bort de fibrer som av anledningarna ovan inte är lämpliga att ha med i det slutgiltiga datamaterialet. Programmet har uppdaterats efter inrådan av Liv och är uppbyggt så att man får se bilder på fibrerna med alternativ på hur man vill bedöma dem. De hela fibrerna är de som sedan ligger till grund för de stickprov som skattningarna av fiberlängdsfördelningarna är gjorda på. Nedan visas exempel på de fyra bedömningsalternativen – okapad, kapad, på annat sätt felaktig eller ihopklumpad.



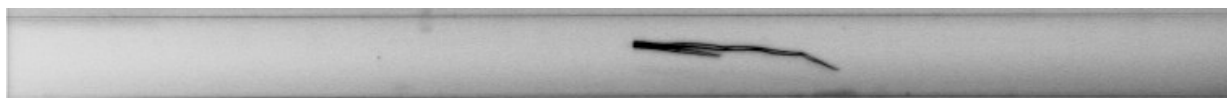
Figur 2. En okapad fiber med längd 2,6 mm.

I figur 2 ser man tydligt att ändarna på fibern smalnar av vilket tyder på att den inte är kapad. Om den ena eller båda ändarna är avhuggna bedöms fibern som kapad, se figur 3.



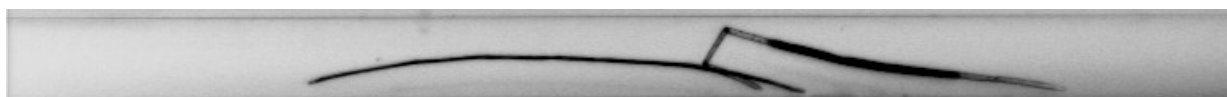
Figur 3. En fiber som bedöms som kapad eftersom den vänstra änden slutar tvärt medan den högra avsmalnar.

Fibern mäts både i skugglängd och i konturlängd. Det förstnämnda är fiberns projektion medan det sistnämnda följer fiberns krökning. Om fibern är vikt är det lätt hänt att mätningarna blir fel, som i figur 4.



Figur 4. En fiber som bedöms som felaktig p.g.a. mätfel. Skugglängden mäts till 1,25 mm och konturlängden till 1,3 mm, fast det är tydligt att det borde vara större skillnad mellan de två mätvärdena.

Om två fibrer klumpat ihop, som i figur 5 nedan, så tror Kajaani FiberLab att det bara är en fiber och mäter därför den totala längden.



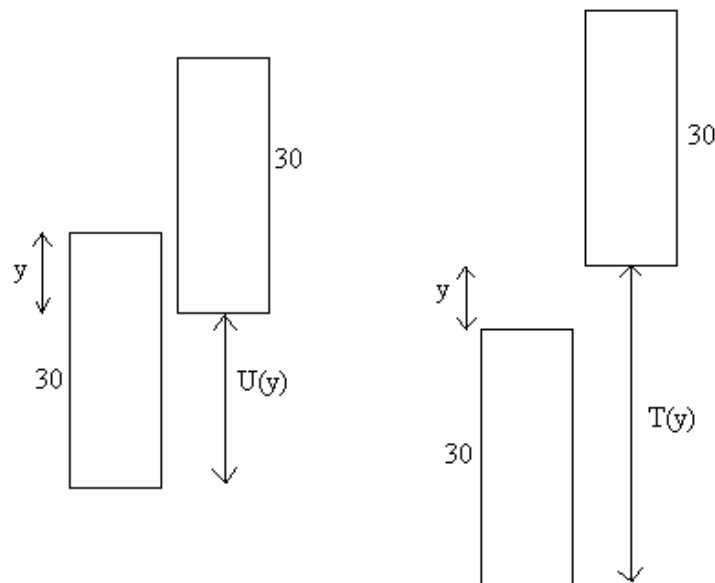
Figur 5. Två fibrer som klumpat ihop sig.

Eftersom det är svårt att avgöra om korta fibrer är kapade eller inte och dessutom vilka slags fibrer de är tas alla mätvärden mindre än 0,5 mm bort. Med denna trunkering garanterar vi att datamaterialet bara innehåller trakeider då mörkstråleceller är mellan 0,07 och 0,16 mm för tall (Nyrén & Back 1959a) och mellan 0,12 och 0,20 mm för gran (Nyrén & Back 1960).

2.2 Kapning av fibrer

En andel av fibrerna kommer att kapas vid provtagning och sannolikheten att detta inträffar, givet att en del av fibern kommer med i provet, är en funktion av fiberns längd och kan härledas geometriskt. Att en fiber kapas innebär att en av de vertikalt växande fibrerna klipps av horisontellt vid provtagningen. Genom att anta att provet tas från en slumpmässigt vald punkt på den intressanta delen av trädet kan man studera provtagningen i en dimension. Vi vill veta sannolikheten att en fiber av längd $Z=y$ inte kapas givet att en del av den kommit med i stickan. För att inte kapas måste provbitens underkant vara placerad i höjdläge inom sträckan $U(y)$, som har längden $u(y) = 30 - y$, se figur 6. Sträckan $T(y)$, i vilken den undre kanten på provet kan placeras så att åtminstone en del av fibern av längd y kommer med i provet har längden $t(y) = 30 + y$, jfr figur 5. Om vi väljer att beteckna händelsen att åtminstone en del av fibern kommit med i stickan som ST betyder det att den sökta sannolikheten blir

$$P(\text{okapad} \mid Z = y, ST) = \frac{u(y)}{t(y)} = \frac{30 - y}{30 + y}. \quad (1)$$



Figur 6. Till vänster en fiber med längd $Z=y$ samt gränsvallen för vilka positioner i höjddled den 30 mm höga tändsticksformade provbiten kan placeras på så att fibern kommer att ingå okapad. Till höger visas gränsvallen så att åtminstone en del av samma fiber kommer att ingå.

2.3 Uttuning av datamaterialet

I avsnitt 2.1. har vi beskrivit hur vi tar fram datamaterial bestående av okapade fibrer från stickan (trunkerade vid 0,5 mm). Detta är dock inte ett representativt data för den sanna fiberlängdsfördelningen i trädet. Sannolikheten att en fiber kapas vid provtagning ökar med dess längd och detta måste vi kompensera för. Med en lämplig uttuning kan vi få representativa data för att skatta fördelningen för den sanna längden av fibrer som åtminstone med någon del träffar stickan. Via den kan vi sedan komma åt fiberlängdsfördelningen i trädet. Vi måste därför också ta reda på hur fördelningarna för okapade fibrer i provet, sanna längder för de fibrer som åtminstone med någon del träffar provet, och sanna fördelningen för det stående trädets fibrer hänger samman. Nedan reder vi ut kopplingen mellan fiberlängdsfördelningen för de (sanna längderna av de) fibrer som åtminstone med någon del träffar provet och den för okapade fibrerna i provet. I avsnitt 2.4 tar vi reda på hur detta kan anknytas till fiberlängdsfördelningen i det stående trädet.

Låt Z vara den sanna längden av en fiber i trädet. Låt vidare Y vara den sanna längden av en fiber som (åtminstone med någon del) kommit med i provet och dess frekvensfunktion vara $f_Y(y)$. Låt sedan X , med frekvensfunktion $f_X(x)$ vara längden av en slumpmässigt vald okapad fiber som kommit med i provet. För ett litet intervall $(y, y + \Delta y)$ vet vi att

$$P(Z \in (y, y + \Delta y) | \text{okapad}, ST) = P(X \in (y, y + \Delta y)) \approx f_X(y)\Delta y. \quad (2)$$

Vi kan med hjälp av Bayes sats visa att

$$P(Z \in (y, y + \Delta y) | \text{okapad}, ST) = \frac{P(\text{okapad} | Z \in (y, y + \Delta y), ST)P(Z \in (y, y + \Delta y) | ST)}{P(\text{okapad} | ST)}$$

vilket betyder att

$$P(Z \in (y, y + \Delta y) | ST) = \frac{P(Z \in (y, y + \Delta y) | \text{okapad}, ST)P(\text{okapad} | ST)}{P(\text{okapad} | Z \in (y, y + \Delta y), ST)} \quad (3)$$

Tillsammans ger (2) och (3) oss att

$$\begin{aligned} f_Y(y)\Delta y &\approx P(Y \in (y, y + \Delta y)) = P(Z \in (y, y + \Delta y) | ST) \\ &\approx \frac{f_X(y)\Delta y \cdot P(\text{okapad} | ST)}{P(\text{okapad} | Z \in (y, y + \Delta y), ST)}, \end{aligned} \quad (4)$$

där $P(\text{okapad} | ST) = k$ är en konstant. Med (1) kan vi nu se att

$$P(\text{okapad} | Z \in (y, y + \Delta y), ST) \approx \frac{30 - y}{30 + y}.$$

Om vi går tillbaka och tittar på (4) får vi att

$$f_Y(y) = k \cdot f_X(y) \frac{30 + y}{30 - y}. \quad (5)$$

Ett histogram över de okapade fibrerna i provstickan ger oss en bild av $f_X(y)$. För att få en bild av fördelningen $f_Y(y)$ utfördes en uttunning av det första histogrammet med hjälp av (5) på så sätt att vi väljer ett k_{\max} så att

$$k_{\max} = P(\text{okapad} | Z = y_{\max}, ST) = \frac{30 - y_{\max}}{30 + y_{\max}},$$

där y_{\max} är längden av den längsta hela fibern. Från (4) får vi då att $f_Y(y)$ är proportionell mot $H(y)f_X(y)$ enligt

$$f_Y(y) \propto \frac{k_{\max} f_X(y)}{P(\text{okapad} | Z \in (y, y + \Delta y), ST)} \approx H(y) f_X(y),$$

där

$$H(y) = \frac{30 - y_{\max}}{30 + y_{\max}} \cdot \frac{30 + y}{30 - y}.$$

Kvoten ovan ligger mellan 0 och 1 för alla y utom för $y = y_{\max}$ där den är 1. Vi kan tunna ut observationerna från histogrammet av $f_X(y)$ och få en rättvis bild av $f_Y(y)$ genom att behålla observationsvärdet y med sannolikhet $H(y)$, eftersom $f_Y(y)$ är proportionell mot $H(y)f_X(y)$.

Detta görs praktiskt genom att för varje mätvärde, t_i , i stickprovet, slumpa fram ett tal $U \in U(0,1)$ och behålla mätvärdet t_i om $U < H(t_i)$. Utgående från de uttunnade histogrammen skattades sedan $f_Y(y)$ med ett antal olika fördelningstyper via ML-metoden för att se vilken fördelning som bäst passar.

2.4 Skillnad mellan provets och trädets fiberlängdsfördelning

Svensson med flera har utvecklat metoder för att skatta fördelningen för fiberlängden i det stående trädet med ett stickprov taget med en 5 mm borrkärna (se t ex Svensson, 2004). Längre fibrer har större sannolikhet att träffas av borrkärnan vilket resulterar i att borrkärnans fiberlängder inte kommer från samma fördelning som trädets. En längdbias som skiljer fördelningarna åt tillkommer och innebär att stickproven innehåller en för hög andel långa fibrer. Man kan minska längdbiasen genom att använda en 30 mm hög provsticka. Nedan visar vi sambandet mellan trädets fiberlängdsfördelning $f_Z(y)$ och den sanna fiberlängdsfördelningen för fibrer som åtminstone med någon del träffar provstickan, $f_Y(y)$.

Om vi antar att fibrerna är slumpmässigt packade i ett träd oavsett längd och att vi tar provet på en slumpmässigt vald position (inom det intressanta området) så är sannolikheten att en fiber av längd y , som åtminstone träffat provet med någon del, proportionell mot $30 + y$, (jämför med avsnitt 2.2).

Alltså är

$$P(ST | Z = y) \propto 30 + y.$$

Bayes sats ger

$$\begin{aligned} f_Y(y)\Delta y &\approx P(Z \in (y, y + \Delta y) | ST) \approx \frac{P(ST | Z = y)P(Z \in (y, y + \Delta y))}{P(ST)} \\ &\propto (30 + y)f_Z(y)\Delta y, \end{aligned}$$

eftersom $P(ST)$ är konstant. Alltså blir

$$f_Y(y) = d(30 + y)f_Z(y) \quad (6)$$

där d är en konstant som vi kan beräkna genom att inse att

$$1 = \int_0^{\infty} f_Y(y)dy = d \int_0^{\infty} (30 + y)f_Z(y)dy = d(30 + \mu) \quad (7)$$

där $\mu = E(Z)$. Från (6) och (7) får vi att

$$f_Y(y) = \frac{30 + y}{30 + \mu} f_Z(y),$$

där μ är väntevärdet för en fibers längd i trädet. Man kan ta reda på μ via fördelningen för Y genom

$$1 = \int_0^{\infty} f_Z(y) dy = (30 + \mu) \int_0^{\infty} \frac{1}{30 + y} f_Y(y) dy,$$

vilket medför att

$$\mu = \left(\int_0^{\infty} \frac{1}{30 + y} f_Y(y) dy \right)^{-1} - 30.$$

Liv (2005) visar i sitt examensarbete dels ovanstående samt att fördelningarna mellan fiberlängderna för de fibrer som åtminstone med någon del träffar provstickan och fiberlängdsfördelningarna i trädet är mycket lika och att det därför är acceptabelt att använda $f_Y(y)$ som en skattning av $f_Z(y)$.

3 ML-skattningar

En maximum likelihoodskattning $\hat{\theta} = (\theta_1, \dots, \theta_n)$ av en fördelnings parametrar $\theta = (\theta_1, \dots, \theta_n)$ är den skattning som maximerar den så kallade likelihoodfunktionen (se nedan). Det betyder att skattningen ger den maximerade sannolikheten för att få det utfall vi har fått. Som exempel ska vi ju i denna rapport hitta parametrar till olika fördelningar som gör att frekvensfunktionerna $f_Y(y | \theta)$ anpassar sig så bra som möjligt till datamaterialen beskrivna i avsnitt 2.3. Med andra ord ska vi alltså hitta de parametrar som maximerar utfallssannolikheten d.v.s. maximerar likelihoodfunktionen hörande till respektive fördelning. Likelihoodfunktionen ges av

$$L(\theta | y) = \prod_{i=1}^n f_Y(y_i | \theta)$$

där vi har n stycken oberoende observationer $y = (y_1, \dots, y_n) \in f_Y(y)$ och vi vill skatta parametrarna $\theta = (\theta_1, \dots, \theta_n)$. För att förenkla maximeringen kan man välja att maximera logaritmen av likelihoodfunktionen

$$l(\theta | y) = \log(L(\theta | y)).$$

Eftersom vi enbart är intresserade av trakeider (se avsnitt 2.1) och därför har tagit bort alla fibrer som är kortare än 0,5 mm från datamaterialet observerar vi endast fördelningen $f_Y(y | \theta)$ trunkerad vid konstanten $r = 0,5$ mm. Vår likelihood kommer därmed att se något annorlunda ut. Låt den sanna längden på en fiber som med någon del kommer med i provet vara Y , med frekvensfunktion $f_Y(y | \theta)$. Om då T är, den från r , trunkerade fördelningen för Y blir fördelningsfunktionen för T

$$F_T(t) = P(T \leq t) = P(Y \leq t | Y \geq r) = \frac{P(r \leq Y \leq t)}{P(Y \geq r)} = \frac{F_Y(t) - F_Y(r)}{1 - F_Y(r)}, \text{ då } t \geq r$$

Frekvensfunktionen för T blir

$$f_T(t) = \frac{d}{dt} F_T(t) = \begin{cases} \frac{f_Y(t)}{P(Y \geq r)}, & t \geq r \\ 0, & t < r. \end{cases} \quad (8)$$

Loglikelihooden för den trunkerade fördelningen T blir då, när vi observerar t_1, \dots, t_n ,

$$\begin{aligned} l(\theta | t) &= \log \left(\prod_{i=1}^n \frac{f_Y(t_i | \theta)}{P(Y \geq r)} \right) = \sum_{i=1}^n \log(f_Y(t_i | \theta)) - n \log(P_\theta(Y \geq r)) \\ &= \sum_{i=1}^n \log(f_Y(t_i | \theta)) - n \log \left(\int_r^\infty f_Y(y | \theta) dy \right). \end{aligned} \quad (9)$$

Det går inte att finna analytiska uttryck för maximum av loglikelihooden för de fördelningar vi studerar i detta arbete. Detta beror främst på termen $n \log(P_\theta(Y \geq r))$ i ekvation (9). Därför görs detta

numeriskt med hjälp av funktionen *fminsearch* i Matlab. Den minimerar en funktion utifrån dess okända parametrar och den negativa likelihoodfunktionen blir då det vi vill minimera. En mer detaljerad förklaring av *fminsearch* finns i hjälpavsnittet till Matlab Version 7.2.0.232 (R2006a).

Per Liv (2005) studerade följande möjliga fiberlängdsfördelningar för Y för gran: trunkerad normal-, lognormal-, gamma-, Weibull- och betafördelning. I detta arbete kommer vi att studera dem tillsammans med tre nya förslag på fördelningar, nämligen Johnson's Sb-fördelning, generaliserad gammafördelning samt segmenterade fördelningar. De tre nya fördelningarna presenteras nedan.

3.1 Johnson Sb-fördelning

Om en variabel $W \in N(\mu, \sigma^2)$ så är $Y = \frac{ke^W}{1+e^W}$ Johnson Sb-fördelad, $0 < Y < k$, där k är en konstant (Johnson, Kotz & Balakrishnan 1994). Man har fördelningsfunktionen

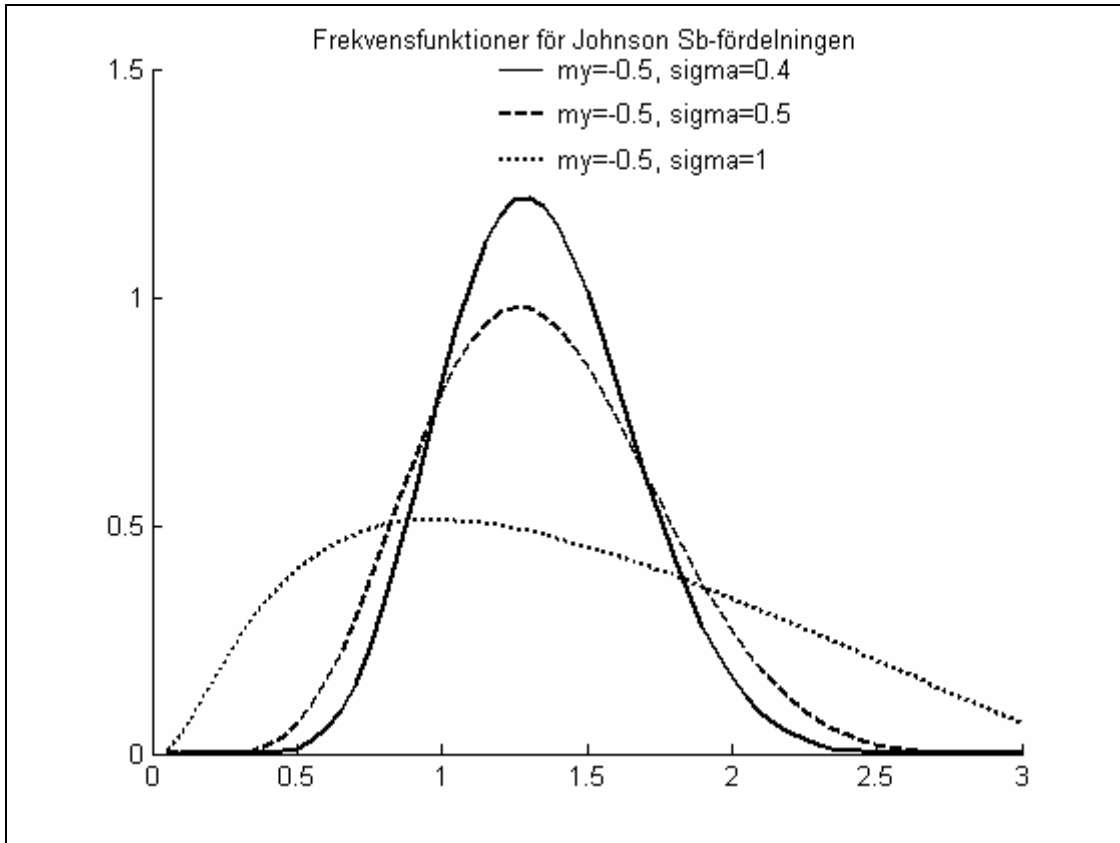
$$F_Y(y | \mu, \sigma) = P(Y \leq y) = P\left(\frac{ke^W}{1+e^W} \leq y\right) = P\left(W \leq \ln\left(\frac{y}{k-y}\right)\right) = F_W\left(\ln\left(\frac{y}{k-y}\right)\right),$$

där $F_W(w) = \int_{-\infty}^w \frac{1}{\sigma\sqrt{2\pi}} e^{-(u-\mu)^2/2\sigma^2} du$.

Genom att derivera med avseende på y får vi frekvensfunktionen

$$f_Y(y | \mu, \sigma) = \frac{k}{y(k-y)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\ln\left(\frac{y}{k-y}\right) - \mu\right)^2 / 2\sigma^2}, \quad 0 < Y < k.$$

Johnson Sb-fördelningen bygger på en normalfördelning och det syns tydligt i figur 7. Den har dessutom en förmåga att vara ickesymmetrisk som man ser på den prickiga linjen. Notera att fördelningen har en övre gräns för fiberns längder, k . Eftersom $0 < t_i < k, \forall i$ har vi valt att sätta $k = t_{\max} + 0,5$ för att garantera att $k > t_i, \forall i$.



Figur 7. Frekvensfunktioner för Johnson Sb-fördelning, där $k=y_{\max}+0,5=3,5$.

Loglikelihooden, bortsett från additiva konstanter, för de vid konstanten r trunkerade variablerna t_1, \dots, t_n blir enligt (9)

$$l(\mu, \sigma | t) = -\sum_{i=1}^n (\log(t_i) - \log(k - t_i)) - \frac{1}{2\sigma^2} \sum_{i=1}^n (\log(t_i) - \log(k - t_i) - \mu)^2 - n \log(\sigma) - n \log(1 - F_y(0,5 | \mu, \sigma)).$$

3.2 Segmenterad fördelning

En segmenterad fördelning är en sammanslagning av två eller flera fördelningar. Här har vi valt att konstruera två olika segmenterade fördelningar. En som består av en lognormalfördelning till värdet θ där den övergår till en Weibullfördelning. Den andra segmenterade fördelningen består av två Weibullfördelningar, även den med parameterskifte i θ . Frekvensfunktionen för en segmenterad fördelning är

$$f(y) = \begin{cases} \frac{f_1(y)}{b} & y \leq \theta \\ \frac{af_2(y)}{b} & y > \theta \end{cases}$$

där f_1 och f_2 är frekvensfunktioner av de fördelningar vi vill sammanfoga,

$$a = \frac{f_1(\theta)}{f_2(\theta)} \text{ och } b = F_1(\theta) + a(1 - F_2(\theta)) \text{ (Cao \& Wu 2007).}$$

Detta innebär att fördelningsfunktionen ser ut som

$$F(y) = \begin{cases} \frac{F_1(y)}{b} & y \leq \theta \\ \frac{F_1(\theta)}{b} + \frac{a}{b}(F_2(y) - F_2(\theta)) & y > \theta \end{cases}$$

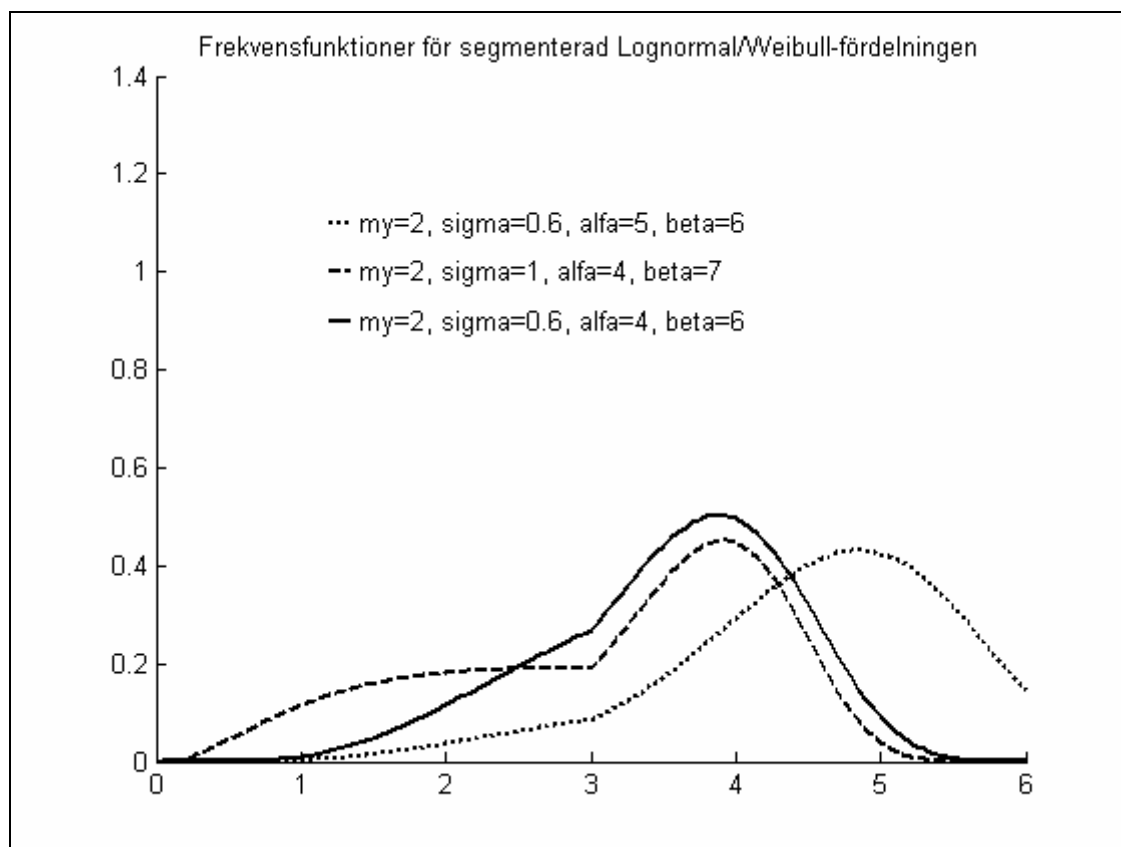
På grund av svårigheter att hitta ML-skattningar med fem okända parametrar så valde vi att sätta θ till medianen i respektive stickprov.

De fördelningar vi är intresserade av att använda är lognormalfördelning med frekvensfunktion

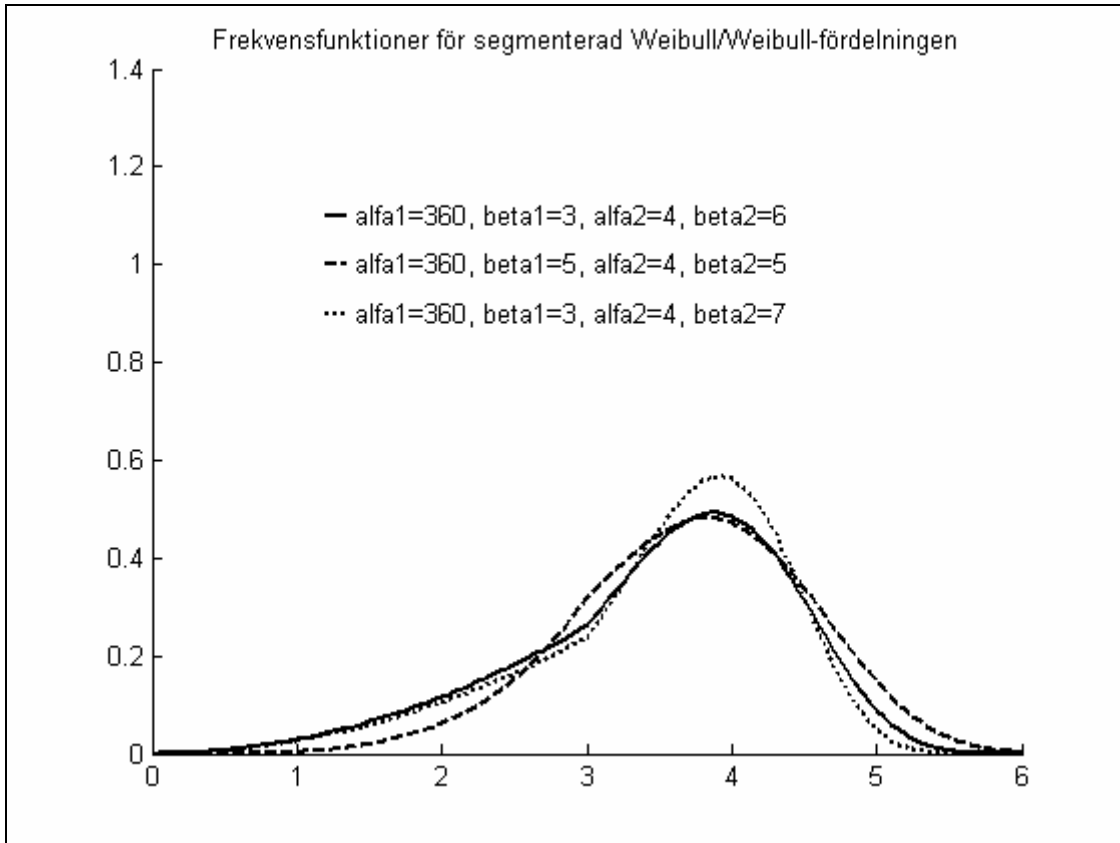
$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}$$

och Weibullfördelning med frekvensfunktion

$$f(y) = \beta\alpha^{-\beta} y^{\beta-1} e^{-\left(\frac{y}{\alpha}\right)^\beta}.$$



Figur 8. Frekvensfunktioner för segmenterad lognormal/Weibull-fördelning.



Figur 9. Frekvensfunktioner för segmenterad Weibull/Weibull-fördelning.

Som man ser i både figur 8 och 9 kan de segmenterade fördelningarna ta hänsyn till ovanliga skillnader, som orsakar olika lutningar, i fiberlängder i och med sin brytning i medianen. Fördelningarna har ett ibland ovanligt utseende med tvära brytningar i medianen.

Eftersom två olika segmenterade fördelningar provats finns det två loglikelihoodfunktioner (som beräknas enligt (9)).

1. När $f_1(y)$ är lognormalfördelad och $f_2(y)$ Weibullfördelad.

$$l(\mu, \sigma, \alpha, \beta | t) = -n \log(b) + n_m \log(a) - \sum_{i=1}^{n_m} \log(t_i) - n_m \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n_m} (\log(t_i) - \mu)^2 + (n - n_m) \log(\beta) - \beta(n - n_m) \log(\alpha) + (\beta - 1) \sum_{i=n_m+1}^n \log(t_i) - \sum_{i=n_m+1}^n \left(\frac{t_i}{\alpha} \right)^\beta - n \log(1 - F(0,5)).$$

2. När $f_1(y)$ och $f_2(y)$ båda är Weibullfördelade.

$$l(\alpha_1, \alpha_2, \beta_1, \beta_2 | t) = -n \log(b) + n_m \log(a) + n_m \log(\beta_1) - n_m \beta_1 \log(\alpha_1) + (\beta_1 - 1) \sum_{i=1}^{n_m} \log(t_{(i)}) - \sum_{i=1}^{n_m} \left(\frac{t_{(i)}}{\alpha_1} \right)^{\beta_1} +$$

$$(n - n_m) \log(\beta_2) - \beta_2 (n - n_m) \log(\alpha_2) + (\beta_2 - 1) \sum_{i=n_m+1}^n \log(t_{(i)}) - \sum_{i=n_m+1}^n \left(\frac{t_{(i)}}{\alpha_2} \right)^{\beta_2} - n \log(1 - F(0,5)).$$

n_m är antalet observationer som är mindre eller lika med medianen. Alltså är de ordnade variablerna $t_{(1)}, \dots, t_{(m)}$ alla observationer som är mindre eller lika med medianen.

3.3 Generaliserad gammafördelning

Om W är gammafördelad med formparameter α och skalparameter β tillhör $Y = \beta \left(\frac{W}{\beta} \right)^{1/c}$ en

generaliserad gammafördelning med parametrarna α , β och c (Johnson, Kotz & Balakrishnan 1994). Fördelningsfunktionen kan härledas från den för en gammafördelning.

$$F_Y(y) = P(Y \leq y) = P\left(\beta \left(\frac{W}{\beta} \right)^{1/c} \leq y \right) = P\left(W \leq \beta \left(\frac{y}{\beta} \right)^c \right) = F_W\left(\beta \left(\frac{y}{\beta} \right)^c \right),$$

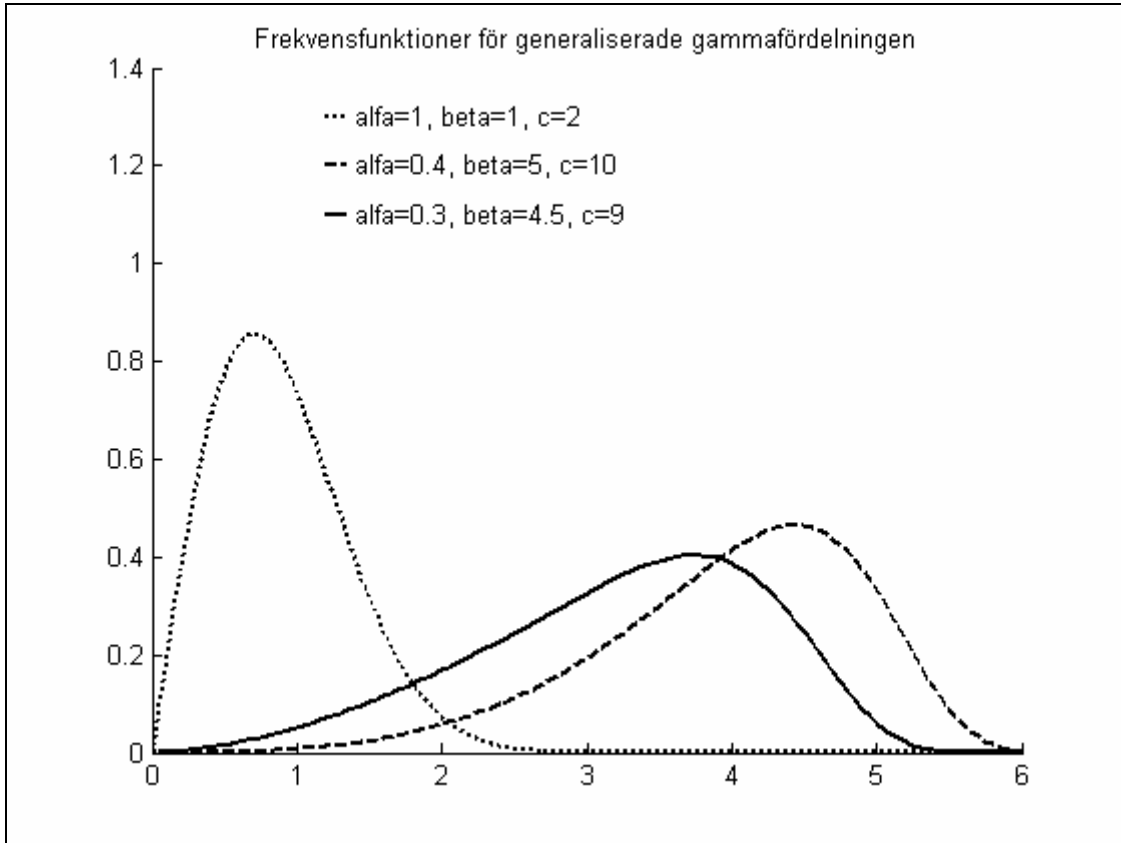
där $F_W(w) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^w t^{\alpha-1} e^{-\frac{t}{\beta}} dt$ och $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

Frekvensfunktionen blir då

$$f_Y(y) = c \left(\frac{y}{\beta} \right)^{c-1} f_W\left(\beta \left(\frac{y}{\beta} \right)^c \right) = \frac{c}{\Gamma(\alpha)} \frac{y^{c\alpha-1}}{\beta^{c\alpha}} e^{-\left(\frac{y}{\beta} \right)^c}, \quad y > 0 \quad (10)$$

eftersom

$$f_W(w) = \frac{1}{\beta^\alpha \Gamma(\alpha)} w^{\alpha-1} e^{-\frac{w}{\beta}}.$$



Figur 10. Frekvensfunktioner för generaliserade gammafördelning.

I figur 10 ser vi att en generaliserad gammafördelning kan vara både negativt och positivt skev.

Loglikelihooden för de vid konstanten $r=0,5$ trunkerade variablerna t_1, \dots, t_n blir enligt (9)

$$l(\alpha, \beta, c | t) = n \log(c) - n \log(\Gamma(\alpha)) - n \log(\beta) + (\alpha c - 1) \sum_{i=1}^n \log\left(\frac{t_i}{\beta}\right) - \sum_{i=1}^n \left(\frac{t_i}{\beta}\right)^c - n \log(1 - F_Y(0,5))$$

För att överkomma problemen med maximering av likelihoodfunktionen använder man ofta följande parametrisering,

$$\mu = \ln(\beta) + \frac{1}{c} \ln\left(\frac{1}{\lambda^2}\right), \quad \sigma = \frac{1}{c\sqrt{\alpha}} \quad \text{och} \quad \lambda = \frac{1}{\sqrt{\alpha}}.$$

Den generaliserade gammafördelningen kan vid speciella parametervärden uppföra sig som både exponential-, lognormal-, Weibull- och gammafördelningen. För att få fördelningen att bete sig som en Weibullfördelning (med parametrarna α och β) sätter vi $\lambda = 1$, $\alpha = 1/\sigma$ och $\beta = \ln(\mu)$. För en exponentialfördelning gäller att $\lambda = 1$ och $\sigma = 1$ medan $\lambda = 0$ ger en lognormalfördelning. Gammafördelningen erhålls genom att sätta $\lambda = \sigma$. Detta faktum, att generaliserad gamma kan uppföra sig som många kända fördelningar, gör att den anpassar sig bättre än många andra fördelningar. Den kan vara skev åt båda hållen men även symmetrisk vilket ger den en stor fördel.

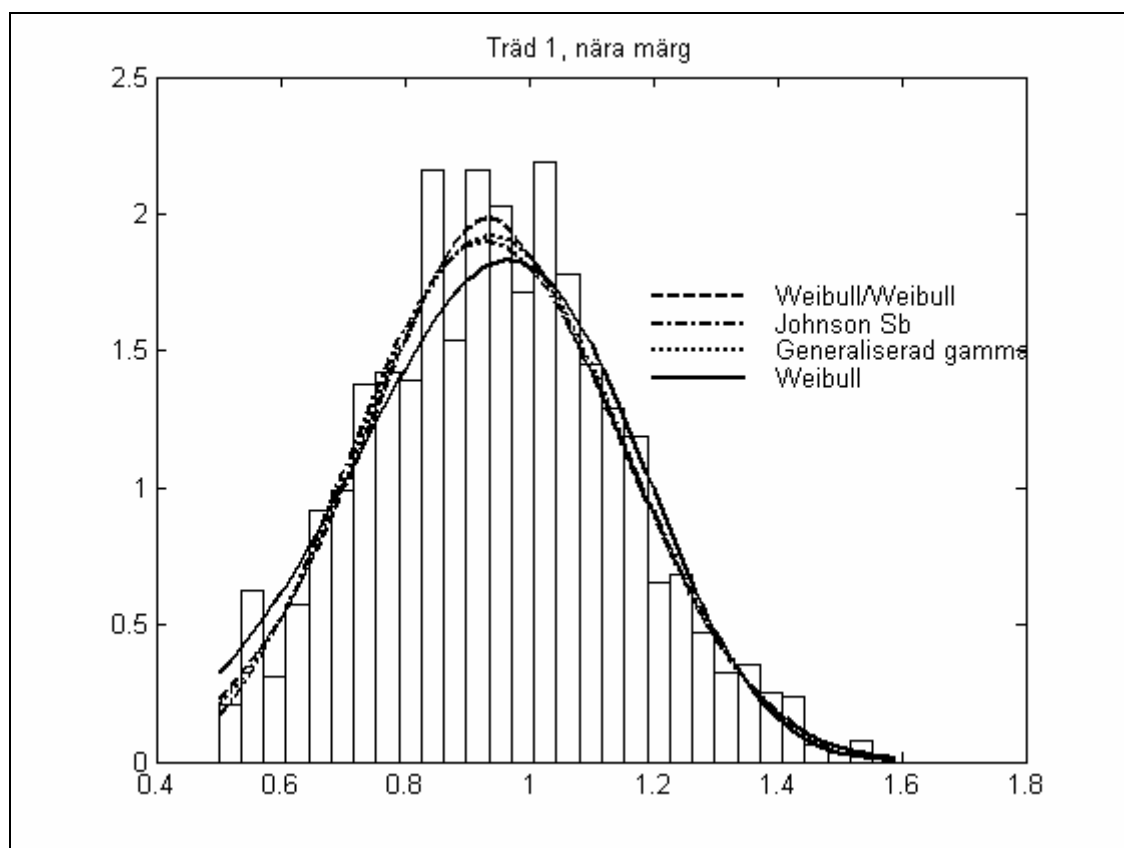
4 Fiberlängdsfördelningar för gran

Liv (2005) visar i sitt examensarbete att Weibullfördelningen var en av de fördelningar som fungerade bäst på grandata från både bark och märke. Den jämförs i detta kapitel med Johnson Sb-fördelning, generaliserad gamma och segmenterad Weibullfördelning på data från gran. Detta för att ta reda på om de nya fördelningarna ger en bättre eller sämre anpassning än de gamla. Segmenterad lognormal/Weibull-fördelning har också testats men visat sig vara sämre än den andra segmenterade fördelningen så den tas inte upp i delkapitlet nedan.

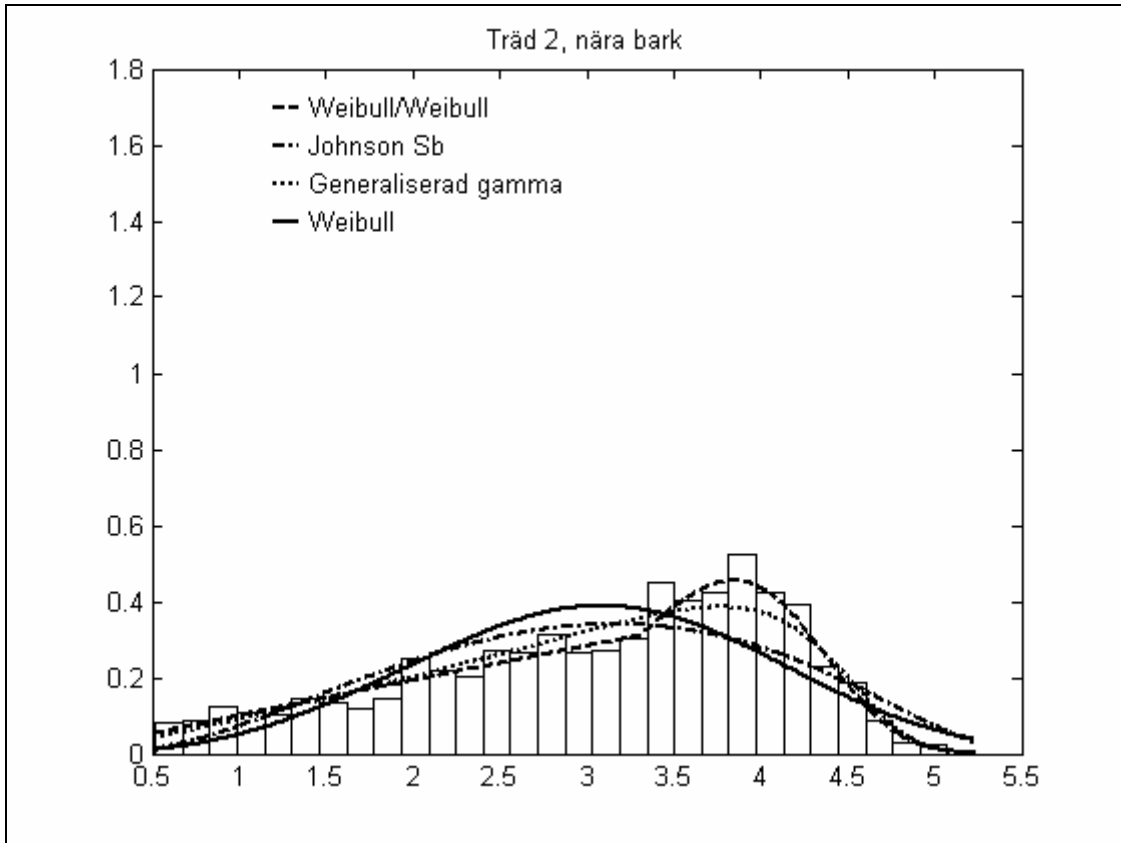
I detta avsnitt jämför vi fiberlängdsfördelningar för de tre granar som även studerades av Liv (2005). I avsnitt 4.1 visas histogram och anpassade frekvensfunktioner för de fyra undersökta fördelningarna. I avsnitt 4.2 utförs ett χ^2 -test för fördelningarna.

4.1 Histogram med anpassade frekvensfunktioner

Alla tre träd har testats både vid märke och bark och visar alla samma sak. Därför visas bara de tydligaste exemplen, träd 1 vid märke och träd 2 vid bark, i histogrammen nedan.



Figur 11. Jämförelse mellan Weibull-, Johnson Sb-, generaliserad gamma- och segmenterad Weibull/Weibull-fördelning på märke-data från gran.



Figur 12. Jämförelse mellan Weibull-, Johnson Sb-, generaliserad gamma- och segmenterad Weibull/Weibull-fördelning på barkdata från gran.

Det syns tydligt i figur 11 och 12 att Weibullfördelningen och Johnson Sb-fördelningen är alltför symmetriska för datamaterialet. Segmenterad Weibull/Weibull och generaliserad gammafördelning däremot anpassar sig bra efter data i och med att de är lätta att få skeva. Samma resultat fås om man undersöker de övriga två granarna.

4.2 χ^2 -test

För att ta reda på vilken fördelning som statistiskt fungerar mest korrekt utför vi ett χ^2 -test. Säg att vi har observerat y_1, \dots, y_n av slumpvariabeln Y och vill undersöka om Y tillhör fördelningen f . Detta gör vi genom att dela utfallsrummet för Y i k stycken klasser och observera hur många observationer som hamnar i varje klass. Antalet observationer i klass i döps till z_i och sannolikheten att värdet på en variabel tillhörande f ska ligga i klass i är p_i uträknad under nollhypotesen.

Det vi testar är

$$H_0: Y \text{ kommer från fördelningen } f$$

mot

$$H_1: Y \text{ kommer ej från } f.$$

Teststatistikan blir då

$$Q = \sum_{i=1}^k \frac{(z_i - np_i)^2}{np_i}$$

och är, under nollhypotesen, approximativt χ^2 -fördelad med $k-r-1$ frihetsgrader, där k är antalet klasser och r är antalet skattade parametrar till f . Vi kan inte förkasta nollhypotesen om $Q < \chi_{0,05}^2(k-r-1)$.

Eftersom våra olika fördelningar alla har olika antal skattade parametrar omvandlar vi resultatet till p-värden istället för att lättare kunna jämföra fördelningarna med varandra. Vi kommer då att kunna förkasta nollhypotesen om p-värdet är mindre än 0,05.

Tabell 1. Resultat från χ^2 -test där k är antalet intervall som utfallsrummet är uppdelat i, r är antalet skattade parametrar och n är antalet fibrer i provet. p Under respektive fördelning visas teststatistikan Q med p-värdena inom parentes.

Träd	k	Antal fibrer i provet, n	Weibull, $r=2$ (p-värde)	Johnson Sb, $r=2$ (p-värde)	Segmenterad Weibull/Weibull, $r=4$ (p-värde)	Generaliserad gamma, $r=3$ (p-värde)
Gran 1, bark	59	1702	219 (0)	223 (0)	136 (0)	128 (0)
Gran 2, bark	72	1486	361 (0)	269 (0)	74 (0,26)	102 (0,01)
Gran 3, bark	76	1617	216 (0)	230 (0)	90 (0,06)	98 (0,02)
Gran 1, mäg	78	1923	165 (0)	161 (0)	158 (0)	156 (0)
Gran2, mäg	10 5	4401	492 (0)	512 (0)	490 (0)	485 (0)
Gran 3, mäg	79	2304	178 (0)	271 (0)	144 (0)	166 (0)

Vi kan förkasta nollhypotesen för alla utom för segmenterad Weibull/Weibull vid två av proverna från barkdata. I övrigt är generaliserad gamma den enda av övriga fördelningar som uppvisar ett något större p-värde. Den segmenterade fördelningen och generaliserad gammafördelning är i alla sex fallen bättre än de två andra fördelningarna. Vanligtvis är det så att ju färre observationer man har desto större p-värde tenderar man att få. Detta kan man se här då de p-värden som är lite större uppkommer i de två proverna med minst observationer i.

5 Fiberlängdsfördelningar för tall

I detta avsnitt analyseras data från sju tallar där fem är från södra Sverige och två från Norge. Några är långsamt växande, andra snabbväxande och vissa är vindutsatta. Prover är tagna från tre positioner i varje träd – vid mörgen, vid femtonde årsringen och vid barken.

5.1 Kapade och korta fibrer

För att kontrollera om proven från de sju olika träden verkar uppvisa samma egenskaper beräknades antalet fibrer under 0,5 mm och antalet kapade fibrer, från mätdata givet av Kajaani FiberLab, för att kunna jämföra andelarna mellan träden och dess tre olika positioner (här kallade mörge, mellan och bark).

En likvärdig analys kunde inte utföras på data från gran då den programvara vi använde till tall inte fanns tillgänglig då.

Tabell 2. Andelen korta och kapade fibrer i proven.

Träd och position	Totalt antal fibrer i provet	Antalet fibrer under 0,5 mm (% av totala antalet fibrer i provet)	Antalet återstående fibrer	Antal kapade av återstående (%)
Snabb1, mörge	1348	275 (20)	1073	687 (64)
Snabb1, mellan	1914	569 (30)	1345	776 (58)
Snabb1, bark	1643	322 (20)	1321	788 (60)
Långsam1, mörge	968	119 (12)	849	547 (64)
Långsam1, mellan	1229	130 (11)	1099	657 (60)
Långsam1, bark	910	159 (17)	751	456 (61)
Söder, mörge	948	162 (17)	786	504 (64)
Söder, mellan	1741	168 (10)	1573	988 (63)
Söder, bark	1262	184 (15)	1078	674 (63)
Snabb2, mörge	2274	475 (21)	1799	1084 (60)
Snabb2, mellan	2022	252 (12)	1770	959 (54)
Snabb2, bark	1515	271 (18)	1244	796 (64)
Långsam2, mörge	1703	566 (33)	1137	892 (78)
Långsam2, mellan	1830	327 (18)	1503	982 (65)
Långsam2, bark	1094	117 (11)	977	673 (69)
Skyddat, mörge	1461	188 (13)	1273	836 (66)
Skyddat, mellan	1918	362 (19)	1556	919 (59)
Skyddat, bark	1623	327 (20)	1296	881 (68)
Utsatt, mörge	1321	205 (16)	1116	716 (64)

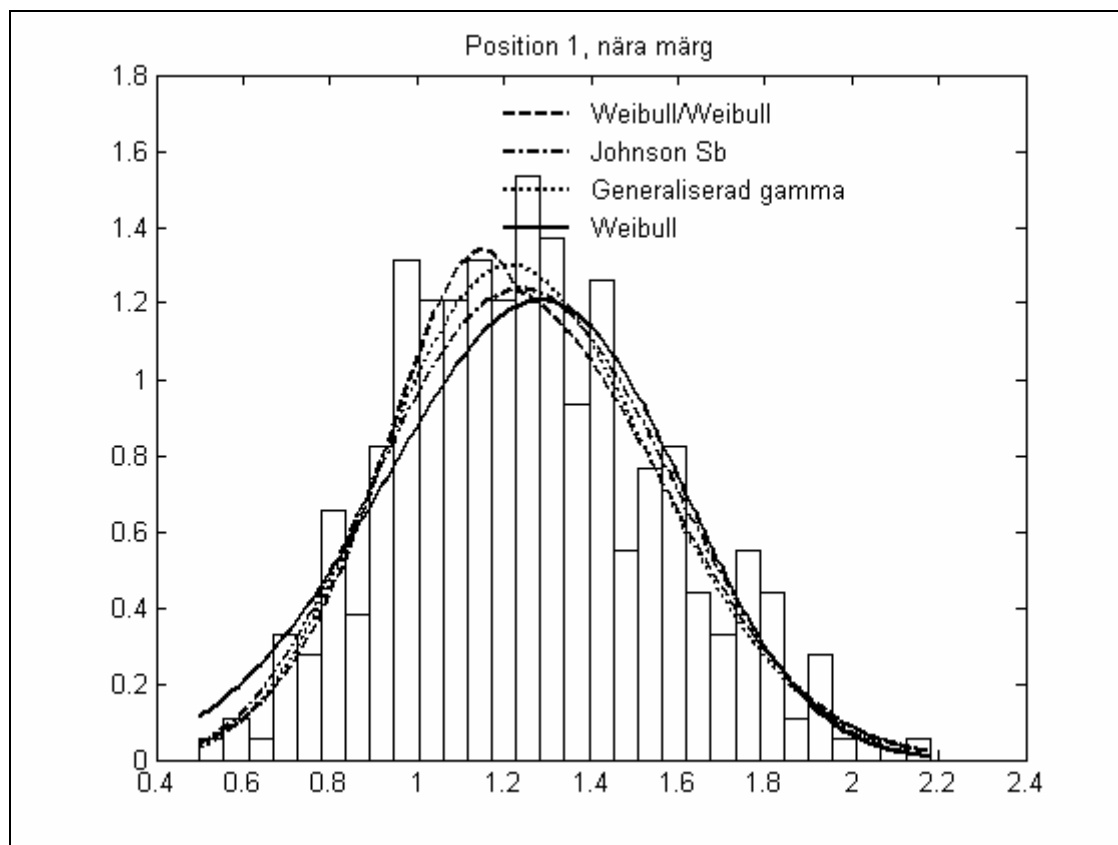
Utsatt, mellan	1755	515 (29)	1240	687 (55)
Utsatt, bark	1218	109 (9)	1109	684 (62)

Tabell 2 visar att det i genomsnitt är 63 % kapade fibrer med standardavvikelse 5,11 %. Antalet kapade fibrer är nästan lika för alla träd och ingen skillnad mellan olika positioner verkar finnas. Andelen korta fibrer i proven är kring 18 % med standardavvikelse 6,58 % och inte heller här verkar det vara någon skillnad mellan positionerna.

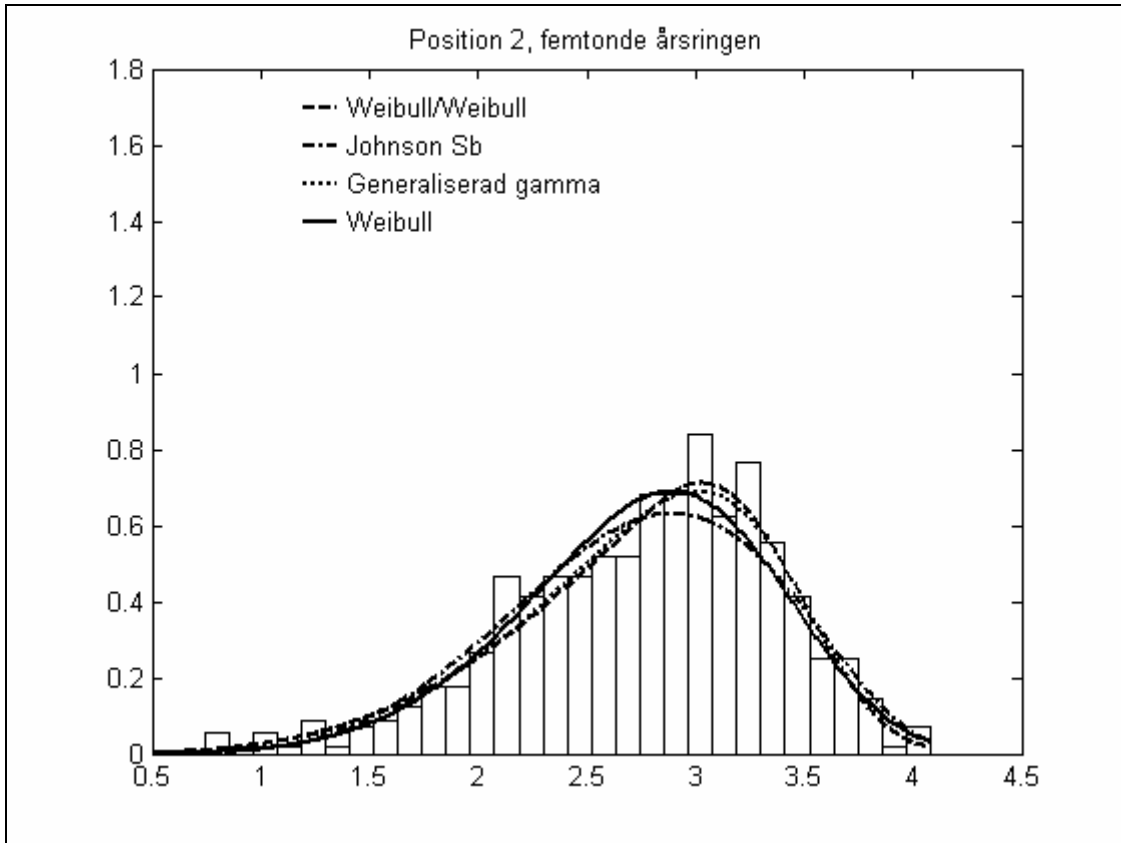
5.2 Histogram med anpassade frekvensfunktioner

Nedan visas ett histogram vardera för respektive position. Alla träd har ritats upp och undersökts men de tydligaste exemplen har valts ut för att lättare visa på resultatet.

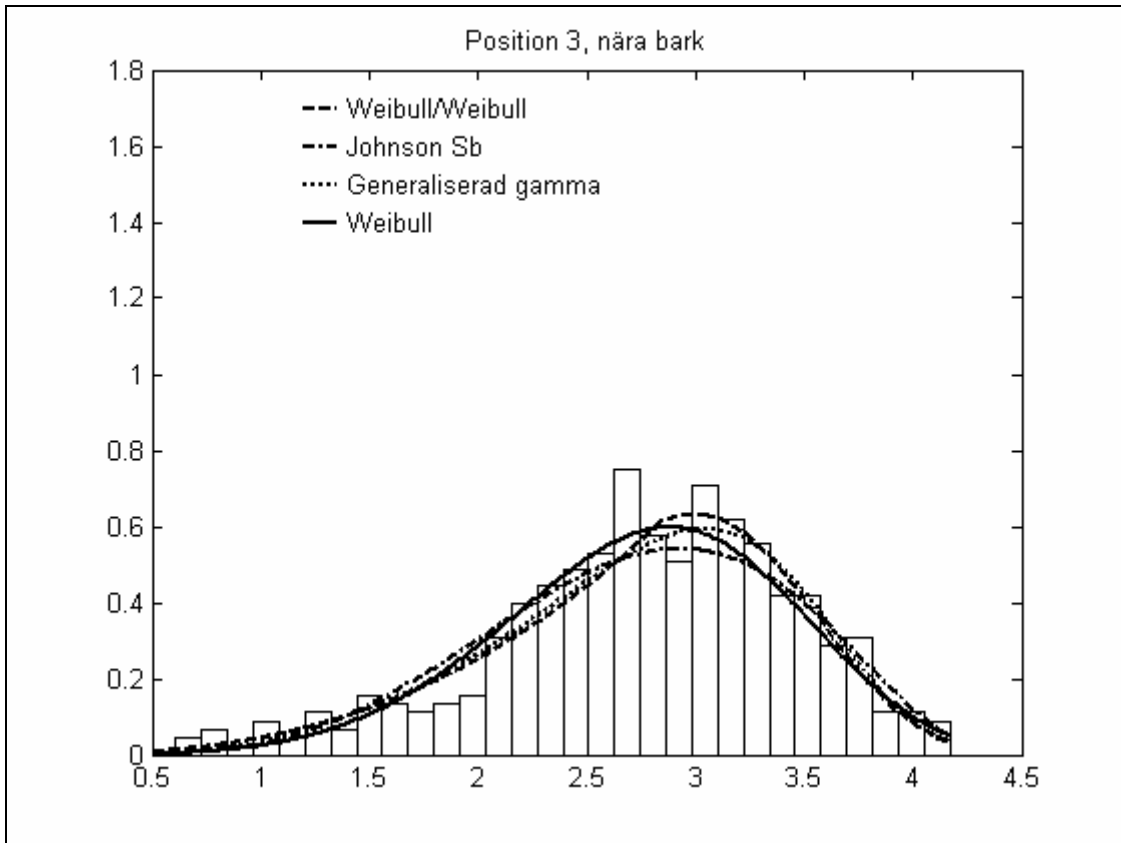
Det syns i både figur 13, 14 och 15 att Weibullfördelningen och Johnson Sb-fördelningen är för symmetriska för att anpassa sig väl till datamaterialet. Den segmenterade fördelningen har rätt skevhet i alla tre fallen men har en ganska skarp kant vid parameterbytet. Den generaliserade gammafördelningen är den som både har rätt skevhet och är följsam vilket passar bäst i det här fallet. Vi kunde ju även se i avsnitt 4.1 att generaliserad gamma även fungerar bra på data av gran vilket är en styrka.



Figur 13. Jämförelse mellan Weibull-, Johnson Sb-, generaliserad gamma- och segmenterad Weibull/Weibull-fördelning på data från tall Snabb1, nära märke.



Figur 14. Jämförelse mellan Weibull-, Johnson Sb-, generaliserad gamma- och segmenterad Weibull/Weibull-fördelning på data från tall Söder, vid femtonde årsringen.



Figur 15. Jämförelse mellan Weibull-, Johnson Sb-, generaliserad gamma- och segmenterad Weibull/Weibull-fördelning på data från tall Snabb2, nära bark.

De tre träden i figurerna har ju valts ut för att de tydligt visar det vi kommit fram till ovan men tilläggas ska att alla träd analyserats och resultatet har varit entydigt – generaliserad gamma är den fördelning som anpassats sig bäst till datamaterialet. Därför väljer vi att enbart koncentrera oss på denna fördelning i följande avsnitt för att kunna gå in lite djupare i analysen.

5.3 Sammanfattande mått

I detta avsnitt studerar vi hur väl de sammanfattande måtten (väntevärde, varians och skevhet) för generaliserad gamma överrensstämmer med de empiriska måtten (medelvärde, stickprovsvarians och stickprovsskevheter) från de bedömda, trunkeerade och uttunnade proven (hädanefter kallade proven). På grund av att data är trunkeerad vid 0,5 mm får vi jämföra de empiriska måtten med de sammanfattande måtten från en generaliserad gammafördelning trunkeerad vid 0,5. Vi gör detta för tre träd vid samtliga positioner och träden har valts ut slumpmässigt.

Vi vill alltså först ta reda på $E(T)$ där T är en trunkeerad generaliserad gammafördelad variabel. Vi har att

$$f_T(t) = \frac{f_Y(t)}{P(Y \geq 0,5)}, t \geq 0,5 \text{ vilket medför}$$

$$E(T) = \frac{1}{P(Y \geq 0,5)} \left[E(Y) - \int_0^{0,5} t f_Y(t) dt \right]$$

$$Var(T) = E(T^2) - (E(T))^2$$

$$E((T - E(T))^3) = E(T^3) - 3E(T)E(T^2) + 3(E(T))^2 E(T) - (E(T))^3.$$

Vi kan också visa att

$$E(T^2) = \frac{1}{P(Y \geq 0,5)} \left[E(Y^2) - \int_0^{0,5} t^2 f_Y(t) dt \right]$$

$$E(T^3) = \frac{1}{P(Y \geq 0,5)} \left[E(Y^3) - \int_0^{0,5} t^3 f_Y(t) dt \right].$$

Vi kan med hjälp av (10), i avsnitt 3.3, räkna ut att

$$E(Y^i) = \int_0^{\infty} y^i f_Y(y) dy = \frac{\beta^i \cdot \Gamma\left(\frac{\alpha c + i}{c}\right)}{\Gamma(\alpha)} \quad (11)$$

Vi kan även lösa $\int_0^{0,5} t f_T(t) dt$, $\int_0^{0,5} t^2 f_Y(t) dt$ och $\int_0^{0,5} t^3 f_Y(t) dt$ numeriskt när vi känner α , β och c .

Vi kan nu räkna ut väntevärdet för den trunkeerade generaliserade gammafördelningen som anpassar sig bäst till respektive träd och position genom att ersätta α , β och c med ML-skattningarna (se avsnitt 3.3)

från respektive dataset i formlerna ovan. Detta jämförs sedan med medelvärdet, $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$, för de observerade fiberlängderna t_1, \dots, t_n för respektive träd och position. De tre träd som undersöks är samma som i figurerna i avsnitt 5.1.

Tabell 3. Väntevärde för anpassad trunkerad generaliserad gammafördelning och medelvärde för respektive prov.

Träd och position	α	β	c	Väntevärde $E(T)$	Medelvärde \bar{t}
Snabb1, nära märe	4,161 0	0,641 9	2,0373	1,26	1,26
Snabb1, mellan	0,539 3	2,809 7	6,5138	2,24	2,24
Snabb1, nära bark	0,678 2	3,570 1	6,1256	3,00	3,00
Söder, nära märe	1,728 7	1,478 8	3,1209	1,65	1,65
Söder, mellan	0,511 4	3,353 6	8,5677	2,76	2,76
Söder, nära bark	0,280 8	4,122 2	14,321 5	3,27	3,28
Snabb2, nära märe	1,019 9	1,735 8	3,6567	1,59	1,59
Snabb2, mellan	0,495 4	3,071 3	6,9421	2,42	2,42
Snabb2, nära bark	0,463 2	3,483 6	7,9767	2,76	2,77

Den trunkerade fördelningen ger ett väldigt bra resultat om man studerar medelvärdena. De stämmer på två decimaler i nästan alla fallen. För att få ytterligare kunskap om hur bra generaliserad gammafördelning fungerar studerar vi standardavvikelsen.

Standardavvikelsen $\sigma = \sqrt{Var(T)}$ från generaliserad gamma jämförs med standardavvikelsen

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (t_i - \bar{t})^2}$$

för respektive prov.

Tabell 4. Standardavvikelse för anpassad generaliserad gammafördelning och standardavvikelse för respektive prov.

Träd, position	Standardavvikelse σ	Stickprovsstandardavvikelse s
Snabb1, nära märe	0,3039	0,3044
Snabb1, mellan	0,5913	0,5926
Snabb1, nära bark	0,7282	0,7266
Söder, nära märe	0,4242	0,4254
Söder, mellan	0,6021	0,6032
Söder, nära bark	0,7053	0,7095

Snabb2, nära märe	0,4621	0,4609
Snabb2, mellan	0,6404	0,6404
Snabb2, nära bark	0,6840	0,6859

Även här verkar anpassningarna ligga nära de riktiga värdena för proven. Generaliserad gamma skattar alltså både medelvärde och standardavvikelse på ett tillfredsställande sätt för vårt data från tall. Det återstår att se om den klarar av att skatta skevheten på ett bra sätt. Såsom det ser ut i histogrammen (figur 12, 13 och 14) är skevheten positiv nära märe och negativ för de två andra positionerna.

Skevheten från trunkerad generaliserad gamma, $E((T - E(T))^3)$, jämförs med provens skevhet,

$$\frac{1}{n} \sum_{i=1}^n (t_i - \bar{t})^3 .$$

Tabell 5. Skevhet för anpassad generaliserad gammafördelning och skevhet för respektive prov.

Träd, position	Anpassad skevhet	Stickprovsskevhet
Snabb1, nära märe	0,008	0,008
Snabb1, mellan	-0,070	-0,092
Snabb1, nära bark	-0,139	-0,130
Söder, nära märe	0,006	0,005
Söder, mellan	-0,123	-0,135
Söder, nära bark	-0,274	-0,383
Snabb2, nära märe	0,007	0,003
Snabb2, mellan	-0,098	-0,100
Snabb2, nära bark	-0,155	-0,194

Även fast värdena från trunkerad generaliserad gamma inte stämmer helt med de från proven så klarar fördelningen av att få rätt tecken på skevheten - positiv nära märe och negativ vid femtonde årsringen och vid barken. Något som många andra fördelningar har misslyckats med att åstadkomma.

6 Exempel på skogliga tillämpningar för fiberlängders uppförande

Som vi visade i avsnitt 4 och 5 är generaliserad gammafördelning flexibel och verkar vara lämplig att använda till att approximera fiberlängdsfördelningar, i både gran och tall - även då de växer under olika betingelser som t.ex. växthastighet, läge och utsatthet. Låt oss nu förutsätta att generaliserad gammafördelning tillräckligt väl beskriver fiberlängdsfördelningarna i gran och tall. Det innebär då att via metoder som Svensson (2007) föreslagit kan man utifrån borrhärdor, som innehåller många kapade fibrer, ändå få information om stående träd fiberlängdsfördelningar och se hur de förändras under olika betingelser.

I detta avsnitt ska vi se exempel på hur man skulle kunna nyttja detta för att studera olika skogliga fenomen med avseende på fiberlängder. Vi väljer att studera talldata som vi i detta fall bestämt fiberlängdsfördelningen för utifrån stickor. Vi antar att de skattade generaliserade gammafördelningarna tillräckligt väl beskriver fiberlängdsfördelningarna i träden och använder oss av dem i fortsättningen. De sammanfattande måtten väntevärde, standardavvikelse och skevhet kommer att studeras för några olika situationer.

Vi är nu inte intresserade av den trunkerade fördelningen utan vill kunna uttala oss om fibrer kortare än 0,5 mm också. Därför behöver vi beräkna väntevärde, varians och skevhet för slumpvariabeln Y .

Vi kan utifrån (11) och de skattade generaliserade gammafördelningarna beräkna $E(Y)$,

$$Var(Y) = E(Y^2) - (E(Y))^2$$

och

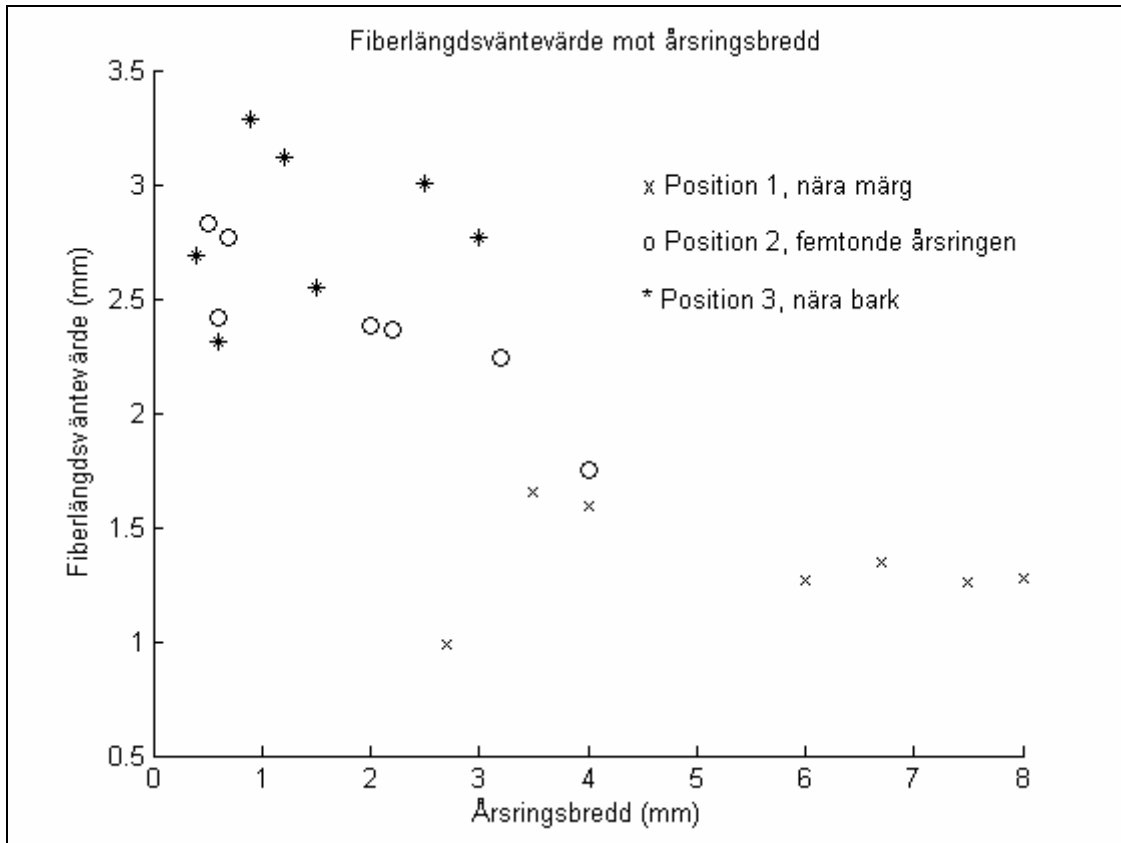
$$E((Y - E(Y))^3) = E(Y^3) - 3E(Y)E(Y^2) + 3(E(Y))^2 E(Y) - (E(Y))^3,$$

där parametrarna α , β och c , ersätts med respektive ML-skattning.

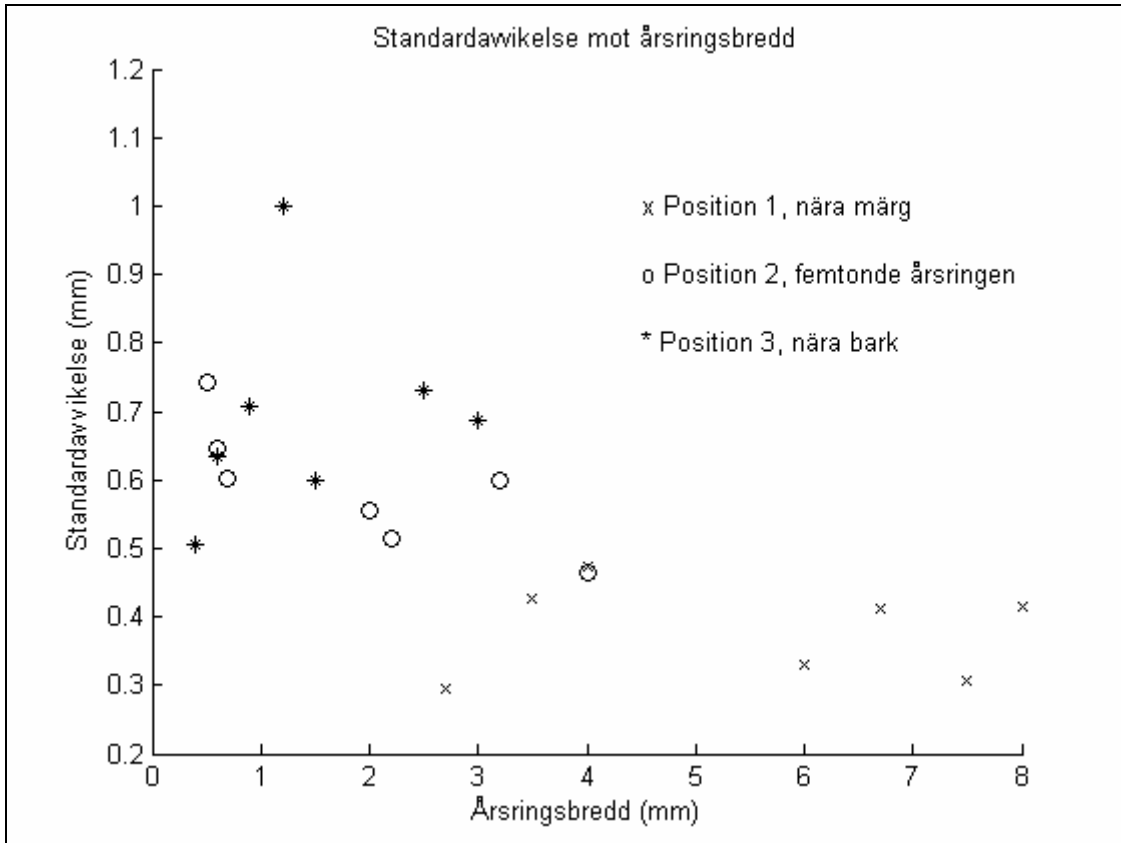
Det är intressant att studera hur fiberlängdernas väntevärde, standardavvikelse och skevhet förändras beroende på om man är nära märke eller nära bark. Vi ska samtidigt se på hur det samvarierar med årsringsbredd. Bredden på ett träd årsringar varierar beroende på olika faktorer. Till exempel så tenderar breda årsringar att indikera snabbväxande träd och gynnsamma växtförhållanden medan smala årsringar kan betyda att de växt under kargare förhållanden. I figurerna 16-18 visas skattade väntevärden, standardavvikelser, respektive skevheter för de sju tallar som ingår i denna studie. De tre separata positionerna (nära märke, vid femtonde årsringen samt nära bark) finns angivna och årsringsbredden för respektive kombination angiven. I figurerna ser vi att de bredaste årsringarna verkar finnas nära märke medan de smalaste finns närmast barken.

Figurerna 16 och 17 visar att medelfiberlängd och standardavvikelse i princip fördubblas nära bark jämfört med nära märke. Inom position 2, vid årsring femton, syns också det tydliga mönstret av att medelfiberlängd och standardavvikelse ökar med minskad årsringsbredd. Nära märke och nära bark verkar det inte vara så stor skillnad i medellängd och standardavvikelse med avseende på ändring i årsringsbredd. Att fibrer tenderar att vara kortare nära märke än nära bark är något som man vet om sedan tidigare (Atmer & Thörnqvist 1982, Nilsson 2004). Tendensen att breda årsringar har kortare fiberlängder än smala årsringar är också allmänt vedertaget trots att det finns studier som inte har hittat något samband mellan fiberlängd och årsringsbredd (Zobel & van Buijtenen 1989, Nilsson 2004).

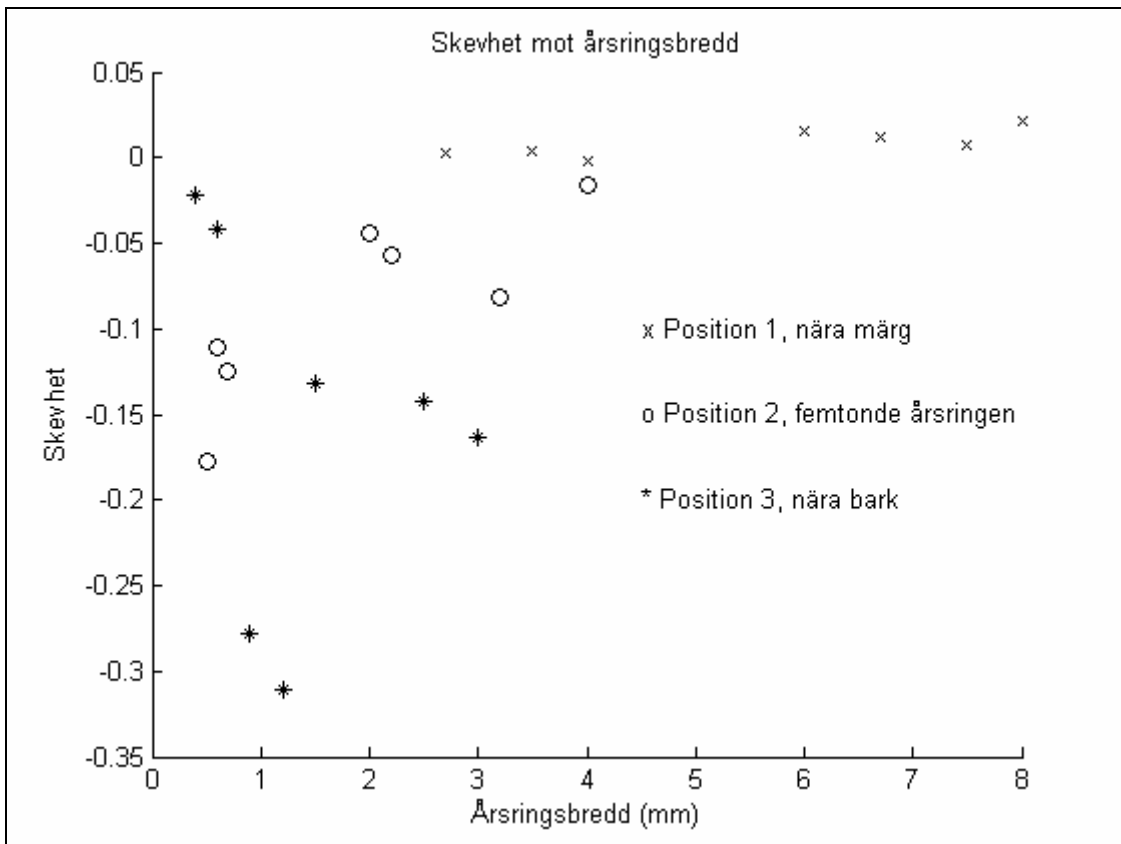
När det gäller skevhet så tenderar den att vara mycket liten eller positiv (då årsringsbredden går över 4 mm) nära märke och annars negativ. En negativ skevhet betyder att man har en längre svans nedåt mot korta fibrer, medan en positiv skevhet innebär att man har en längre svans uppåt mot långa fibrer. Framför allt vid årsring femton kan man se att skevheten verkar blir starkare (negativ) då årsringsbredden ökar. Det finns i dagsläget ingen skogligt förankrad förklaring på detta.



Figur 16. Väntevärde från generaliserad gammafördelning mot årsringsbredd för alla tre positioner.



Figur 17. Standardavvikelse från generaliserad gammafördelning mot årsringsbredd för alla tre positioner.



Figur 18. Skevhets från generaliserad gammafördelning mot årsringsbredd för alla tre positioner.

7 Diskussion och slutsatser

Segmenterad Weibull/Weibull anpassar sig bra till data från både gran och tall och får även bäst resultat i χ^2 -testet (se 4.2). Man kan dock ifrågasätta hur lämpligt det är att använda en fördelning med en sådan tydlig brytpunkt. Det är inte sannolikt att de sanna fiberlängderna i ett träd uppför sig så.

Av de sammanlagt åtta fördelningarna vi undersökt är generaliserad gamma den enda som anpassar sig bra till data och klarar av att hantera olika tecken på skevhet utan att för den skull överanpassas (som den segmenterade fördelningen gör med sin tvära brytning). I avsnitt 5.3 visar vi också att den klarar av att skatta sammanfattande mått mycket precist.

I avsnitt 5.1 fick vi resultatet att i genomsnitt 63 % av fibrerna bedömdes som kapade. Detta är en väldigt stor andel då det bara är fibrerna som hamnat i provstickans ändar som förväntas kapas vid provtagning. Orsaken kan vara att många fibrer slets itu vid behandlingen innan mätningen och därför uppfattades som kapade. Det mest troliga är dock att den manuella bedömningen av fiberbilderna varit hårdare än nödvändigt och därför bidragit (tillsammans med slumpen) till det oväntade resultatet. Positivt är dock att alla prover verkar ha bedömts lika hårt.

En ny undersökning med generaliserad gamma borde utföras där fördelningen testas på en större mängd träd för att säkerställa resultatet, men då vi använt både gran och tall och dessutom tallar med olika egenskaper (skyddade, utsatta, snabbt växande, långsamt växande, från södra och norra Sverige och från Norge) borde vi ändå ha fått ett resultat som kan appliceras på verkligheten. Metoderna som bl.a. Svensson (2007) utvecklat där man använt sig av borrhärnor på stående träd kan nu användas med generaliserad gammafördelning och intressanta mått för fiberlängderna kan beräknas för träd som fortfarande växer ute i skogen.

Det skulle vara intressant att köra två parallella studier där generaliserad gammafördelning skattades från både borrhärnor och stickor tagna från samma träd och bredvid varandra, för att se hur väl de överensstämmer. På så sätt skulle man få mer kunskap om hur den generaliserade gammafördelningen kan tillämpas på stående träd. Detta skulle i sin tur innebära att skogliga hypoteser som de i kapitel 6 skulle kunna testas med ett resultat att lita på.

Slutsatsen är att generaliserad gammafördelning är den hittills mest lovande fördelningen att använda för att skatta fiberlängdfördelningar i träd. Den kan med fördel användas redan nu men fler studier borde utföras på en större mängd träd för att få större kunskap om hur fördelningen beter sig.

8 Referenser

- Atmer, B. & Thörnqvist, T. (1982). *Fiberegenskaper i gran (Picea abies L. Karst) och tall (Pinus sylvestris L.)*. Rapport nr 134, Institutionen för virkeslära, SLU, Uppsala.
- Cao Q.V. & Wu Q. (2007). *Characterizing wood fiber and particle length with a mixture distribution and a segmented distribution*. *Holzforschung* 61(2):124-130.
- Ilvessalo-Pfäffli, M-S. (1995). *Fiber atlas. Identification of papermaking fibers*. Springer Verlag, Berlin.
- Johnson L.N., Kotz S. & Balakrishnan N. (1994). *Continuous univariate distribution, vol 1*. John Wiley & Sons, Inc.
- Liv, P (2005). *Skattning av fiberlängdsfördelning hos gran (Picea Abies)*. Diploma Work in Mathematical Statistics, Department of Mathematics and Mathematical Statistics, Umeå University, Sweden.
- Mörling T., Sjöstedt-de Luna S., Svensson I., Fries A. & Ericsson T. (2003). *A method to estimate fibre length distribution in conifers based on wood samples from increment cores*. *Holzforschung* 57:248-254.
- Nilsson, O. (2004). *Fiberlängdsvariation hos gran (Picea abies) och tall (Pinus sylvestris)*. Institutionen för skogsskötsel, SLU, Umeå. Examensarbeten 2004-13.
- Nyrén V. & Back E. (1959a). *The dimensions of tracheidal and parenchymatous ray cells of Pinus Silvestris pulpwood*. *Svensk papperstidning* 17:587-593.
- Nyrén V. & Back E. (1960). *The dimensions of tracheidal and parenchymatous ray cells of Picea Abies (Karst.) pulpwood*. *Svensk papperstidning* 18:619-624.
- Svensson I. (2007). *Estimation of wood fibre length distributions from censored mixture data*. PhD Dissertation, Department of Mathematics and Mathematical Statistics, Umeå University, Sweden.
- Svensson, I., Sjöstedt-de Luna, S. & Bondesson, L. (2006). *Estimation of wood fibre length distributions from censored data through an EM algorithm*. *Scandinavian Journal of Statistics*, 33:503-522.
- Zobel, B. & van Buijtenen, J.P. (1989). *Wood variation – its causes and control*. Springer Verlag, Berlin, New York, Tokyo.