CHALMERS | GÖTEBORGS UNIVERSITET

Nydanande metodik för parameteruppskattning inom ett empiriskt Bayesianskt ramverk

Pioneering methodology for parameter estimation within an empirical Bayesian framework

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Alexander Lehnberg
Emanuel Abada
Lukas Johansson
Malte Rundqvist
Sulaiman Carneil

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2021

Nydanande metodik för parameteruppskattning inom ett empiriskt Bayesianskt ramverk

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet
Sulaiman Carneil Malte Rundqvist Alexander Lehnberg
Kandidatarbete i matematik inom civilingenjörsprogrammet Maskinteknik vid Chalmers
Emanuel Abada
Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers
Lukas Johansson

Handledare: Krzysztof Podgórski

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2021

Förord

Det här arbetet handlar om huruvida empirisk Bayesiansk metodik kan användas som en metod för uppskattning av parametrar. Vi vill härmed tacka vår handledare Krzysztof Podgórski för hans hängivna engagemang i att vägleda oss genom arbetets gång.

Bidrag

En individuell tidslogg och en gemensam dagbok har förts över samtliga medförfattares bidrag. Tabellen nedan är en sammanfattning av ansvarfördelningen av arbetet mellan medlemmarna i gruppen. Vi vill dock betona att rapporten skrivits gemensamt och att alla har varit involverade i samtliga delar.

Sektion	**Huvudsaklig författare i ordning**	**Övrig information**
Populärvetenskaplig	Sulaiman	Reviderat av Malte, Emanuel och Alexander
Sammanfattning	Sulaiman	Reviderat av Alexander och Emanuel
Abstract	Sulaiman	Reviderat av Alexander och Emanuel
1 Inledning	Sulaiman, Malte, Emanuel	Alla författare har bidragit lika mychet. Alexander har bidragit till tidigare versioner samt revidering.
2 Teori		
2.1 Bayesiansk inferens	Malte, Sulaiman, Alexander	Alexander tre första, Malte mittersta och Sulaiman sista stycket. Emanuel har reviderat mellersta stycket och sista stycket.
2.2 Sannolikhetsfördelningar	Alexander	Malte, Lukas har bidragit till tidigare versioner.
2.3 MLE	Alexander	Reviderat av Sulaiman
2.4 Momentmetoden	Lukas	
2.5 Bootstrap	Malte, Emanuel	Reviderat av Alexander
3 WLE	Lukas	
3.1	Lukas	Sulaiman har bidragit till tidigare version.
3.2	Lukas, Alexander	
3.3	Lukas, Malte, Emanuel, Alexander	
4 Empiriska priori		
4.0	Malte, Lukas	Bidragit lika
4.1	Malte, Emanuel	Emanuel och Malte är båda huvudsakliga författare för stycket. Alexander har bidragit till tidigare version.
4.2	Alexander	Reviderat av Emanuel
5 WLE och flera okända parametrar	Emanuel, Malte	Emanuel har även bidragit till tidigare version. Reviderat av Sulaiman och Alexander.
5.1	Malte	Simulering av resultat har huvudsakligen gjort av Malte och Lukas. Emanuel har bidragit till tidigare version.
5.2	Lukas	Emanuel har bidragit till tidigare version och tabeller. Simulering av resultat har huvudsakligen gjort av Malte, Lukas och Emanuel.

6 Slutsatser	Lukas, Emanuel, Alexander	
Övrigt		
Referenser	Alla	Referensystem kontrollerat av Emanuel och Alexander
Kapten	Malte	Malte har utövat en ledande och organisatorisk roll under arbetsprocessen.

Populärvetenskaplig presentation

Matematiska tillämpningar genomsyrar stora delar av dagens samhälle. Vi försöker med hjälp av matematiska modeller beskriva naturliga fenomen för att kunna förutsäga framtiden som till exempel vädret eller räkna ut medelåldern i ett land utan att behöva ha ett register över alla invånares åldrar. Lösningen på problemen kretsar oftast kring uppskattning av någon parameter. En parameter kan beteckna medelåldern i ett land eller en egenskap i en matematisk modell med vars hjälp vi vill kunna beskriva ett naturligt fenomen.

Det finns idag standardmetoder för att uppskatta parametrar som exempelvis Maximum Likelihood Estimation(MLE) och Method of Moments(MoM). Men dessa metoder räcker inte till i vissa fall på grund av besvärliga beräkningsmässiga egenskaper. Det finns scenarion då det kan vara mer önskvärt att använda sig av en Bayesiansk behandling för uppskattning av parametrar, en metod artikulerad av den brittiske matematikern Thomas Bayes som grundar sig i Bayes formel. För att få en bättre förståelse för det Bayesianska tillvägagångssättet kan vi föreställa oss ett exempel där vi ska ta reda på om ett mynt är jämnt eller ojämnt. Låt oss anta att du hittar ett mynt vid rännstenen på gatan som du tycker har en underlig geometri. Baserat på din tidigare kunskap om mynt har du uppfattningen att singling av myntet X gånger borde leda till att klaven och kronan kommer upp ungefär $\frac{1}{2}$X gånger var. Det här sättet att använda och beskriva tidigare kunskap om ett fenomen kallas för den *priora fördelningen* eller bara *prior* och är alltså en modell som beskriver uppfattningen av ett fenomen innan observation av ny data. Men med tanke på myntets underliga geometri så tycker du att din priora modell inte kan stämma och bestämmer dig för att göra ett experiment där du singlar myntet 100 gånger. Experimentet resulterar i att du får klave 84 utav 100 försök. Denna typ av information kallas för *likelihood*, och är en modell som beskriver informationen i den observerade datan. Enligt det Bayesianska tankesättet borde vi ta hänsyn till både vår uppfattning innan observation av data, såväl som experimentets resultat för att få den mest sannolika uppskattningen. Sammanvägningen av priorifördelningen och likelihooden kallas i Bayesianska termer för den *posteriora fördelningen* som är vår uppdaterade uppfattning. Om vår priora uppfattning om myntet är att det är jämnt (50/50), och likelihooden säger att det är ojämnt (84/16), så borde vår posteriora uppfattning vara en kompromiss mellan dessa [12]. Vi antar således att myntet är ojämnt, men att det är jämnare än (84/16).

Det Bayesianska tillvägagångssättet har mött kritik när det kommer till hur man väljer priorifördelningen. Eftersom valet beror på typen av data, och personens subjektiva uppfattning menar kritikerna att det inte objektivt kan spegla verkligheten. Skepticismen grundar sig i osäkerheten kring huruvida datans beteendemönster kan beskrivas av den valda priorifördelningen, vare sig man har vetskap om typen av data i fråga. Likt exemplet ovan där ett test utförs för att undersöka om ett mynt är jämnt eller ej, baseras den priora fördelningen på en tidigare uppfattning om att alla mynt är jämna. Hur säkra kan vi egentligen vara på att vår tidigare uppfattning är korrekt?

Förutom problematiken kring valet av en priorifördelning är det som tidigare nämnt i vissa fall svårt eller till och med omöjligt att använda traditionella metoder för parameteruppskattning till följd av svårhanterliga likelihoodfunktioner för stora datamängder. Metoden Weighted Likelihood Estimation (WLE) som vi undersöker i detta arbete genererar en datadriven prior som är fördelad runt det sanna värdet på parametern vi vill uppskatta. Eftersom denna prior är till fullo bestämd av den observerade datan så är den empirisk och därav finns ingen subjektivitet och bias i den genererade uppskattningen. Ur priorifördelningen skapas ett antal uppskattningar med hjälp av diverse metoder som kallas estimatorer. De enskilda estimatorerna kan vara dåliga approximationer av parametern vi vill uppskatta, men de kommer att vara fördelade runt det sanna värdet av parametern. I WLE viktar vi uppskattningarna och summerar dessa till en viktad summa som blir den slutgiltiga uppskattningen av den okända parametern. Dessutom ger WLE oss ännu en fördel genom att vi får en fördelning över möjliga värden som parametern kan anta, vilket i sin tur möjliggör för oss att göra analyser över hur mycket den uppskattade parametern avviker från det sanna värdet [8].

Vi lyckas i projektet visa att WLE ofta ger likvärdiga, eller bättre estimat jämfört med tradi-

tionella metoder som MLE och MoM. Vi visar hur vi med WLE effektivt kan förbättra en rad dåliga estimat, nästan oavsett hur dålig estimatorn är. WLE är således en alternativ metod i situationer där det inte finns någon bra estimator att applicera direkt på ett stickprov. Resultaten är därför mycket positiva då det ger oss en anledning att använda oss av WLE för att kunna handskas med uppskattningar av parametrar i de fall där metoder som MLE blir för kostsamt att implementera.

Sammanfattning

Undersökningen behandlar en alternativ metod för parameteruppskattning kallad *Weighted Likelihood estimation* (WLE) och dess asymptotiska egenskaper i jämförelse med mer etablerade metoder. Kortfattat utnyttjar metoden parameteruppskattningar som empiriska priorifördelningar och summerar dessa med vikter proportionella mot likelihooden för att producera den posteriora fördelningen. För att genomföra denna studie undersöker vi inledningsvis fall där vi har en okänd parameter som skall uppskattas för att sedan utöka metoden till multipararameterfall. Vi tillämpar WLE på sannolikhetsfördelningar där metoder som Maximum Likelihood Estimation (MLE) kan visa sig vara omedgörliga till följd av exempelvis icke-explicita lösningar till likelihoodekvationen. Det visar sig att WLE i många fall lämpar sig till att förbättra estimat, både för en - och flera parametrar, även då estimaten som används som priorifördelning är av låg eller okänd kvalitet.

Abstract

This study investigates an alternative method for parameter estimation called *Weighted Likelihood Estimation* (WLE) and its asymptotic properties in comparison with other more established methods. In effect, WLE uses parameter estimates as empirical priors and summarise these with weights proportional to the likelihood to produce the posterior distribution. To explore the method, we initially investigate cases where we have one unknown parameter that is to be estimated. Further investigations were made regarding cases with more than one unknown parameter. We apply WLE on probability distributions where the implementation of methods such as the Maximum Likelihood Estimation (MLE) can be troublesome due to non-explicit solutions of the likelihood equation and other possibly complicating factors. After conducting some simulations it became evident that in many cases WLE is suitable for improving estimates in either single or multi parameter instances, even when the estimates used as prior distributions are of low or unknown quality.

Innehåll

1 Inledning

Vår intention med projektet är att granska och undersöka en alternativ metod för parameteruppskattning. Vanligt förekommande metoder för parameteruppskattning är till exempel Maximum Likelihood estimation och momentmetoden, men dessa går inte alltid att använda problemfritt, exempelvis eftersom det inte alltid är möjligt att explicit maximera likelihoodfunktionen. Ett typiskt exempel på problematiska sannolikhetsfördelningar är Cauchyfördelningen. Likelihoodfunktion till denna kan vara multimodal vilket ger dess motsvarande likelihoodekvation flera rötter och det blir därav svårt att hitta ett optimalt värde. En annan problematik vid uppskattning av en Cauchyfördelnings parametrar är att distributionen kan vara obegränsad vid ett eller flera typvärden.

Metoden som undersöks i detta projekt använder sig istället av Bayesiansk metodik och en empirisk priorifördelning. Metoden kan beskrivas använda Bayesiansk metodik snarare än att vara strikt Bayesiansk på grund av metodens användning av en empirisk prior. I traditionell Bayesiansk statistik kan denna priorifördelning väljas subjektivt med diverse gissningar baserat på annan information vilket i sin tur kan ligga till grund för bias.

Inspirationen till att undersöka metoden kommer från artikeln [8], där metoden undersöks i specialfallet då priorifördelningen ges av punktvis MLE. Metodiken kan i det fallet beskrivas på följande sätt. Antag att $\mathbf{x} = (x_1, \ldots, x_n)$ är ett stickprov ur täthetsfunktionen $f(x|\theta)$ där θ är en okänd distributionsparameter. MLE-estimatorn av θ baserat på datapunkt x_i är lika med $\hat{\theta}_i = v(x_i)$. Dessa estimatorer kombineras sedan till en viktad summa för att producera den slutgiltiga estimatorn enligt

$$\hat{\theta} = \sum_{i=1}^{n} \hat{\theta}_i w_i = \frac{\sum_{i=1}^{n} \hat{\theta}_i L(\hat{\theta}_i|\mathbf{x})}{\sum_{j=1}^{n} L(\hat{\theta}_i|\mathbf{x})} \tag{1}$$

där w_i är en vikt proportionell mot $L(\hat{\theta}_i|\mathbf{x})$ och normerad så att $\sum_{i=1}^{n} w_i = 1$. Vi kommer i vår uppsats att undersöka en generaliserad version av denna metod där den empiriska priorifördelningen $\hat{\theta}$ kan estimeras på fler sätt än punktvis MLE.

I den generaliserade metoden, vilken vi valt kalla Weighted Likelihood Estimation(WLE), använder vi en priorifördelning baserad på data genererad från stickprovet vars fördelningsparameter vi vill estimera. Estimaten till den empiriska priorifördelning kan genereras på flera olika sätt varav ett är att maximera likelihoodfunktionen för var och en av observationerna i stickprovet. En annan metod som undersöks i sektion 4 och 5 innefattar användning av bootstrap tillsammans med momentmetoden. Oavsett val av estimat låter vi priorifördelningen vara en likformig fördelning över dessa estimat. Den posteriora fördelningen kombineras sedan som en viktad summa och ger den slutgiltiga estimatorn. Vi kan på så vis ta en mängd dåliga estimat och generera ett bättre estimat från dem, vilket är metodens huvudsyfte.

En fördel med metoden är att posteriorfördelningen utgjord av vikterna w_i allokeras runt den slutliga estimatorn. Posteriorfördelningen ger på så sätt ett estimat av distributionen av $\hat{\theta}$, från vilken vi kan dra slutsatser om estimatorns säkerhet. De mer traditionella metoderna för parameteruppskattning ger endast ett uppskattningsvärde vilket inte ger någon information om variansen hos estimatorn.

I artikeln som föreslår metoden görs simuleringar av läges- och spridningsparametrarna med avseende på några vanliga distributioner. Vi har valt att återskapa några av dessa simuleringar för Cauchyfördelningen och Laplacefördelningen. Båda dessa fördelningar är bra exempel där den mer vanliga MLE metoden kan vara problematisk. Vid tillämpning av vanlig maximum likelihood kan Cauchydistributionen ge en likelihoodfunktion med flera rötter, vilket kan vara mycket tungt för datorer att optimera. Och likelihoodfunktionen för generaliserade Laplacefördelningar kan anta oändliga värden, vilket även komplicerar optimeringsproblemet [8].

Vi ser i Tabell 1 att resultaten från WLE-metoden och den traditionella MLE-metoden är relativt lika för Cauchyfördelningen. För lägre provstorlekar verkar MLE vara bättre, men allt eftersom provstorleken ökar ger WLE approximationer mer jämförbara med MLE.

Tabell 1: Medelvärde och Mean Square Error (MSE) för WLE och MLE efter 10 000 Monte Carlo simuleringar vid uppskattning av lägesparameter från en Cauchydistribution med lägesparameter $\theta = 2$ och spridningsparameter $\sigma = 3$. Dessa resultat är en återskapning av simuleringarna som gjordes i den föregående studien.

provstorlek	WLE(MSE)	MLE(MSE)
5	1.98(8.80)	2.11(2.576)
10	1.99(2.59)	2.04(1.68)
100	2.00(0.19)	2.00(0.18)

Att notera i Tabell 1 är att WLE saknar bias och presterar på en lika hög nivå som MLE för större provstorlekar. Däremot presterar WLE på en jämförelsevis lägre nivå för mindre provstorlekar. I Tabell 2 ser vi liknande resultat för Laplacefördelningen. Vår nya metod tycks dock här vara ungefär lika bra som MLE även för små värden på n, vilket kan ha att göra med att vi i Laplacedistributionen inte har lika stor risk för extrema utbölingar.

Tabell 2: Medelvärde och MSE för WLE och MLE efter 10 000 Monte Carlo simuleringar vid uppskattning av lägesparameter från en Laplacedistribution med lägesparameter $\theta = 2$ och spridningsparameter $\sigma = 1$. Dessa resultat är en återskapning av simuleringarna som gjordes i den föregående studien.

provstorlek	WLE(MSE)	MLE(MSE)
5	3.00(0.339)	2.99(0.349)
25	3.00(0.051)	3.00(0.053)
101	2.99(0.0110)	2.99(0.0113)

Dessa inledande exempel indikerar att WLE presterar förhållandevis bra i jämförelse med MLE med avseende på precision. Vi vill nu utvidga metoden genom bland annat bootstrapping och göra mer omfattande tester än vad som hittills publicerats, till exempel genom att inkludera fler scenarion och fördelningar i fallet av både en respektive flera variabler. I kapitel 2 av arbetet kommer vi att gå igenom en behandling av teorin som ligger till grund för vår undersökning, däribland distributioner som kommer att undersökas närmare med metoden samt tillämpning av bootstrap vid estimation med WLE. I kapitel 3 går vi djupare in på huvudämnet för detta arbete och behandlar teorin och tillämpningar av WLE. I kapitel 4 utvecklar vi tillvägagångssättet genom en bootstrapbehandling av metoden. I kapitel 5 utvidgar vi vår horisont genom att undersöka hur WLE lämpar sig vid estimation av flera parametrar. Avslutningsvis presenteras de slutsatser vi kommit fram till från resultaten som uppnåtts genom vår undersökning.

2 Bakomliggande teori och exempel

Nedan följer relevant bakomliggande teori som används i projektet. Syftet med avsnittet är att täcka det som direkt krävs för att förstå de beräkningar och resonemang som följer.

2.1 Bayesiansk inferens

Thomas Bayes var en brittisk matematiker och statistiker aktiv under 1700-talet. På grund av filosofiska motståndare och beräkningsbegränsningar innan IT-revolutionen var 'frekventisternas' synsätt dominerande fram till 2000-talet då Bayesiansk statistik såg ett lyft och tillämpningar hittades inom diverse områden som bland annat maskininlärning [13].

Den Bayesianska teorin bygger på Bayes sats

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{2}$$

och den fundamentala idén att sannolikheter påverkas av våra tidigare kännedomar. I Bayes sats kallas P(A) den priora sannolikheten (även priorifördelningen) för A och beskriver vår uppfattning om A innan bevisen, B, tagits i beräkning. Den posteriora sannolikheten, P(A|B), är således sannolikheten att vi ser bevisen givet vår priora uppfattning, kallat likelihood, multiplicerat med vår priora sannolikhet och delat med sannolikheten för bevisen. Den posteriora sannolikheten kan alltså ses som en reviderad version av den priora med hjälp av Bayes sats och mätdata. Denna uppdaterade sannolikhet är proportionell mot den priora, multiplicerat med likelihoodfunktionen enligt

$$\pi(\theta|x) \propto L(\theta|x) * \pi(\theta). \tag{3}$$

Den priora sannolikheten P(A), som har varit och är en stor källa till kritik av den Bayesianska teorin, är som nämnt vår tidigare kännedom eller tro om utsagan A. Skepticismen härstammar från subjektiviteten i valet av priorifördelning. Problemet kan emellertid kringgås genom användandet av en icke-subjektiv priorifördelning tagen ur datamängden, känd som en empirisk prior.

Om man är ny till Bayesiansk statistik kan det till en början vara svårförstått, då utnyttjandet av priorifördelningar i Bayes sats kan ses som väldigt partiskt. Ett viktigt motargument är att vi borde vara vaksamma när vi använder subjektiv data. Det kan ses som att vi gör vår undersökning partisk, men tvärt om så bekänner vi vår potentiella egna bias i undersökningen. Vi tillkännager att alla statistiska modeller är subjektiva. I följande exempel visar vi hur Bayesianskt tänkande effektivt kan användas i vardagen. Exemplet är inspirerat av en video av Julia Galef [5].

Exempel: Vi träffar en person som är blyg och ska gissa om denna person studerar IT eller ekonomi, givet att en hypotes måste vara sann. Vid första gissning tänker de flesta att studenten i fråga högst troligt studerar IT, då studenter från IT-fakulteten generellt sett är blygare. Vi antar att detta är sant och att andelen blyga studenter inom varje fakultet är

$$P(Blyg|IT) = 60\%, \qquad P(Blyg|Ekonomi) = 30\%. \tag{4}$$

Om detta är all information vi har, så är det helt korrekt att gissa att personen studerar IT. Men då glömmer vi en viktig del: Hur ser distributionen av elever på IT-fakulteten och Handelshögskolan ut? Enligt antagningstatisik från Göteborgs Universitet är det ca 2.8 gånger så många ekonomistudenter som IT-studenter [9],

$$P(IT) = \frac{2629}{10024} \approx 26\%, \qquad P(Ekonomi) = \frac{7395}{10024} \approx 74\%. \tag{5}$$

Detta är i Bayes sats tidigare känd information om distributionen, vilket vi självklart borde ta med i vår beräkning. Ekvation 2 blir i vårt fall

$$P(F_i|blyg) = \frac{P(Blyg|F_i) * P(F_i)}{P(Blyg)}. \tag{6}$$

Där F_i är fakulteten, IT eller ekonomi, som studenten tillhör. $P(Blyg)$ är den totala sannolikheten att någon är blyg. Denna sannolikhet ges av $P(Blyg) = 0.26 * 0.6 + 0.74 * 0.3 = 0.378$, vilket ger oss

$$P(IT|blyg) \approx 0.41\% \tag{7}$$

$$P(Ekonomi|blyg) \approx 0.59\%. \tag{8}$$

Efter att nu tagit med förhållandet mellan antalet IT- och ekonomistudenter ser vi hur svaret blir väldigt annorlunda. Enligt vår beräkning är det nu istället större chans att studenten studerar ekonomi. Att missa att ta med sig denna priora information, även känt som "Base rate fallacy", är ett vanligt misstag. I vårt exempel visste vi antagninstatistiken och kunde explicit beräkna sannolikheten. I vår vardag kan vi självklart sällan veta denna sannorlikhet, men vi kan ändå ha en uppfattning om hur distributionen ser ut. På så vis kan vi göra en subjektiv gissning om fördelningen. Vi tar hänsyn till att mer information finns om problemet än det som syns i datan vi betraktar.

Vi använder principen av ett Bayesianskt tillvägagångssätt i WLE när vi skapar en empirisk priorifördelning. I traditionell bayesiansk statistik antas denna fördelning subjektivt i förväg, likt exemplet ovan där vi sedan tidigare har vetskap om fördelningen av studenter och baserat på det ges en gissning av priorifördelningen. I det empiriska fallet har vi en datadriven priorifördelning. För att visualisera detta empiriska tillvägagångssätt tar vi ett nytt exempel.

Exempel: Vi går in i ett klassrum med 10 elever varav 6 är blyga. Klassen kommer antingen från IT- eller Handelsfackulteterna. Klassen har alltså en paramer F som betecknar dess fakultetstillhörighet. Givet att vi sedan tidigare vet att

$$P(Blyg|IT) = 60\%, \qquad P(Blyg|Ekonomi) = 30\%. \tag{9}$$

kan vi beräkna likelihood till

$$L(IT|x) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{10}{6} 0.6^6 0.4^4 = 0.25, \tag{10}$$

$$L(Ekonomi|x) = \binom{10}{6} 0.3^6 0.7^4 = 0.037. \tag{11}$$

Detta ger oss 87% chans att klassen kommer från IT-fakulteten. Dock har vi, som vi redan utforskat, troligen sedan tidigare en uppfattning om hur distributionen av elever ser ut, och vill ta med det i vår beräkning i form av en priorifördelning. Vi gissar, innan vi ens går in i klassrummet, att det är 30% chans att eleverna är från IT-fakulteten. Vi skulle då få med hjälp av Bayes sats 2 att

$$\pi(IT|x) = L(IT|x) * \pi(\theta = IT) = 0.25 * 0.3 = 0.075, \tag{12}$$

$$L(Ekonomi|x) = L(Ekonomi|x) * \pi(\theta = Ekonomi) = 0.037 * 0.7 = 0.0259. \tag{13}$$

Detta ger oss nu istället 74% chans att klassen är IT-studenter. I fallet då vi tänker att vi inte vet någonting om distributionen av elever innan vi går in i klassrummet skulle vi på klassiskt Bayesiankt vis anta en 50/50-fördelning. Ett annat alternativ är att generera en empirisk prior. En empirisk prior är när vi bildar oss en uppfattning om priorifördelningen genom att ta information från datan vi ska undersöka. När vi går in i klassrummet så ser vi att många elever är blyga, och eftersom vi sedan tidigare vet att $P(Blyg|IT) = 60\%$ kan vi tänka oss att vi gör en grov empirisk uppskattning av priorifördelningen till $\pi(IT) = 75\%$. Detta skulle ge

$$\pi(IT|x) = L(IT|x) * \pi(\theta = IT) = 0.25 * 0.75 = 0.1875, \tag{14}$$

$$\pi(Ekonomi|x) = L(Ekonomi|x) * \pi(\theta = Ekonomi) = 0.037 * 0.25 = 0.00925. \tag{15}$$

Vilket ger oss 95% sannolikhet att klassen kommer från IT-fakulteten. Denna empiriska uppskattningen av priorifördelningen kan göras på många sätt, ovan gissade vi på 75% då vi såg många blyga elever. Ett annat mer matematiskt sätt är punktvis MLE precis som metoden presenterad i inledningen. Detta tillvägagångssätt skulle i vårt exempel bli att varje blyg elev MLE-uppskattades till IT-student, medan varje oblyg uppskattades till ekonomistudent. Vår empiriska priorifördelning blir då, med 6 av 10 elever blyga, $\pi(IT) = 60\%$ och $\pi(Ekonomi) = 40\%$. Som vi ser kan denna priorifördelningen estimaras på olika sätt, även i det empiriska fallet. Gemensamt är att vi på något vis använder datan vi undersöker för att skapa priorifördelningen. I sektion 3 kommer vi utvidga metodiken presenterad i inledningen till att kunna använda fler empiriska priorifördelningar än punktvis MLE.

När den priora fördelningen $\pi(\theta)$ är tillgänglig kan vi härleda den posteriora fördelningen $\pi(\theta|x)$ från observation x med fördelning $f(x|\theta)$. Den här uppdaterade distributionen är då den sammanfattade informationen som är tillgänglig om θ, den integrerar samtidigt både den priora informationen och informationen som tillkommer från x. Detsamma gäller när ett stickprov $x_1, ..., x_n$ är tillgängligt. Således implicerar den Bayesianska versionen av likelihood-principen, en proposition som säger att: I en statistisk modell kan det utläsas alla, utifrån modellparametrarna relevanta, bevis ur ett stickprov i likelihoodfunktionen, att inferensen på θ borde enbart bero på den posteriora fördelningen $\pi(\theta|x)$. Även om θ nödvändigtvis inte är en slumpmässigt genererad variabel kan distributionen $\pi(\theta|x)$ användas som en vanlig distribution för att beskriva egenskaper hos θ. Summerande indicier för $\pi(\theta|x)$, som det posteriora medelvärdet, posteriora modalvärdet, posteriora variansen och den posteriora medianen kan användas. Till exempel, om parametern av vårt intresse är $h(\theta)$, är en möjlig estimator det posteriora medelvärdet $E^{\pi}(h(\theta)|x)$. Om ett val skall göras mellan indicierna ovan så kan det vara svårt att välja den bästa estimatorn [13]. För en sådan jämförelse bör således ett förlustkriterium användas, så som MSE, som vi hädan kommer att använda.

Medelvärdet och modalvärdet är båda estimatorer av var datan är centrerad. Medelvärdet är generellt sätt den vanligaste estimatorn, men är känslig för utbölingar. Då kan medianen eller modalvärdet vara en bättre estimator. Om den posteriora fördelningen av θ är symmetrisk kring dess modalvärde och dess väntevärde existerar är väntevärdet och modalvärdet lika, annars skiljer de sig åt.
Exempel: Anta att $(x_1, x_2..., x_n)$ är ett stickprov från en normalfördelning $N(\mu, \sigma_0^2)$ där $\mu \in R$ är okänd och σ_0^2 är känd. Vi antar att den priora fördelningen för μ är $N(\mu_0, \sigma_0^2)$ för några utvalda värden på μ_0 och σ_0^2. En uträkning ger oss följande uttryck för den posteriora fördelningen

$$N\left(\left(\frac{1}{\tau_0^2}\right)^{-1}\left(\frac{\mu_0}{\tau_0^2}+\frac{n}{\sigma_0^2}\bar{x}\right), \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma_0^2}\right)^{-1}\right)$$

Denna normalfördelning är symmetrisk kring modalvärdet och väntevärdet för den existerar, därav är modalvärdet och väntevärdet för den posteriora fördelningen lika.

2.2 Sannolikhetsfördelningar och deras parametrar

Vi skall i följande underrubriker diskutera de två primära sannolikhetsfördelningar som används i undersökningen, Cauchy- och Laplacefördelningen. Vi kommer kort prata om fördelningarnas parametrar, täthetsfunktioner och typiska egenskaper.

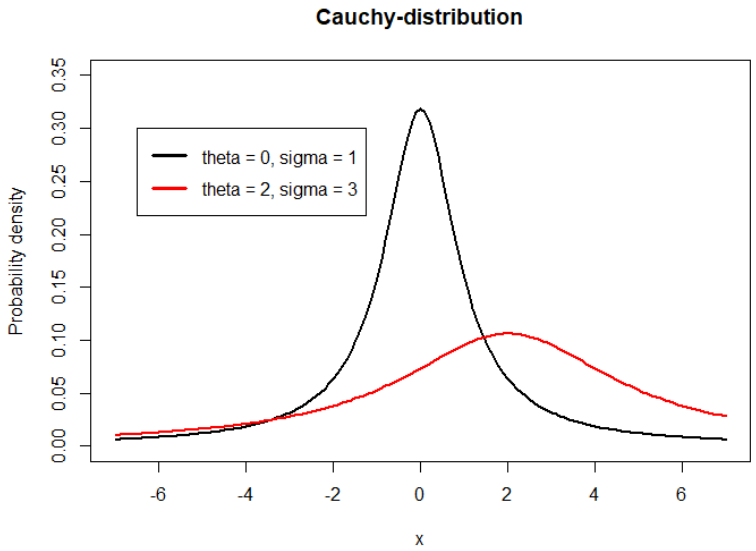
2.2.1 Cauchyfördelningen

Cauchyfördelningen beskriver kvoten mellan två oberoende Gaussfördelningar med väntevärde 0. Fördelningen är symmetrisk kring lägesparametern och har bara ett modalvärde. Då nämnaren i denna kvot kan väntas observeras mycket nära 0, ger detta Cauchyfördelningen säregna egenskaper som att varken medelvärde eller varians är definierat, vilket lett till att fördelningen ibland nämns som patologisk. Genom att studera grafen till täthetsfunktionen från en Cauchyfördelning ser vi att den har en typ av 'heavy tail' fördelning (se Figurer 1 & 4). Denna egenskap medför att ett stickprov från fördelningen ofta innehåller extrema värden som gör att de stora talens lag inte gäller för beräkning av distributionens medelvärde oavsett hur stort stickprov man har.

Cauchyfördelningens täthetsfunktion är definierad enligt

$$f(x|\theta,\sigma) = \frac{1}{\pi\sigma}\frac{1}{1+(\frac{x-\theta}{\sigma})^2}, \quad x \in \mathbb{R}, \tag{16}$$

där θ och σ är läges- respektive spridningsparametern.



Figur 1: Sannolikhetstäthet för Cauchyfördelningen med $\theta = 0, \sigma = 1$ (svart) och $\theta = 2, \sigma = 3$ (röd).

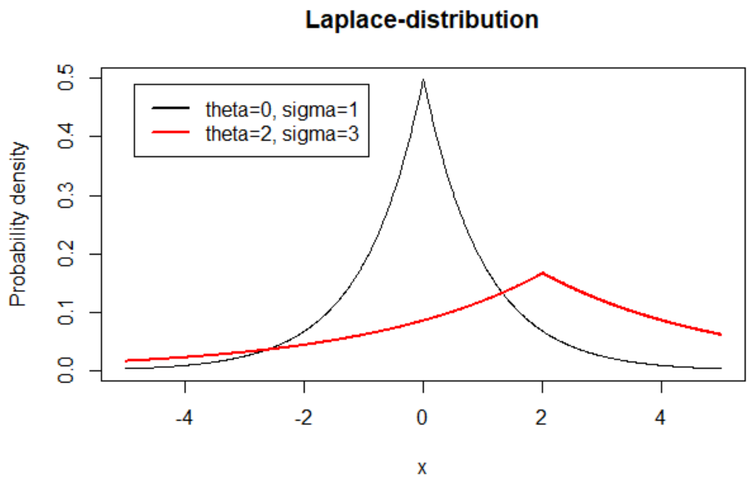
I och med att varken medelvärde eller varians är definierat för Cauchyfördelningen används sällan simpla estimatorer som stickprovets medelvärde. Det gäller emellertid att medianen är en opartisk estimator till θ och att mitten av IQR, Interquartile range, är en estimator till σ [6].

2.2.2 Laplacefördelningen

Laplacefördelningen tar formen av två exponentialfördelningar som möts på mitten, och är därav även känd som den dubbla exponentialfördelningen. Speciellt för Laplacefördelningen är att Maximum Likelihood estimatorn kan visas vara medianen. Dess täthetsfunktion ges av

$$f(x|\theta,\sigma) = \frac{1}{2\sigma}exp(-\frac{|x-\theta|}{\sigma}) \quad x \in \mathbb{R}, \tag{17}$$

där vi precis som i Cauchyfördelningen har parametrar θ och σ beskrivande modalvärdet och spridningen kring denna. I kontrast till Cauchyfördelningen är Laplaces väntevärde och varians väldefinierade, och fördelningen kan inte sägas ha ”heavy tails”, även om fördelningens spridning ökar för högre värden på σ, illustrerat i Figur 2. Faktum är att eftersom Laplacefördelningen utgörs av två exponentialfördelningar är den per definition gränsen för vad som avses med heavy tails-fördelning [7].



Figur 2: Sannolikhetstäthet för Laplacefördelningen med $\theta = 0, \sigma = 1$ (svart) och $\theta = 2, \sigma = 3$ (röd).

För parametriseringen av Laplacefördelningen detaljerad i ekvation 17 med θ som läges- och σ som spridningsparameter gäller det att medelvärdet och medianen är opartiska estimatorer till θ. MAD, mean absolute deviation, och $\frac{s}{\sqrt{2}}$, där s är stickprovets standardavvikelse, är estimatorer till σ [10].

2.3 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) är en av de mest använda metoderna för parameteruppskattning. Som benämningen antyder, är syftet med metoden att maximera likelihoodfunktionen

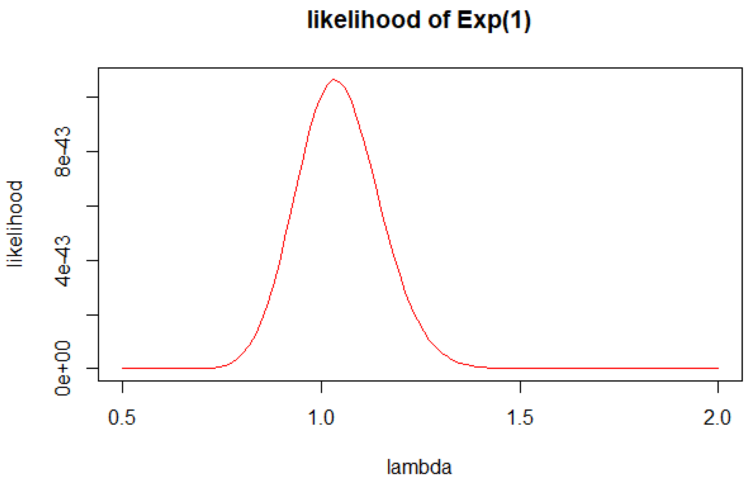
$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta). \tag{18}$$

givet stickprovet $\mathbf{x}$ vilket ger estimatorn

$$\hat{\theta}_{MLE} = \max_{\theta}(L(\theta|\mathbf{x})) \tag{19}$$

Eftersom likelihoodfunktionen är ett mätvärde på modellens anpassningsgrad med avseende på de okända parametrarna gäller det att maximivärdet kan tolkas som den bästa gissningen på dessa parametrar [2]. Som ses i ekvation 18 är likelihoodfunktionen definierad som en produkt av en sekvens av faktorer, vilka kan vara multivariata med flera lokala optimum och därför leda till mycket krävande optimeringsberäkningar.

I Figur 3 visas grafen för en likelihoodfunktion genererad av ett hundra exponentialfördelade punkter med väntevärde 1. Detta maximeringsproblem är mycket lätt att lösa i och med att grafen är kontinuerlig och endast har ett lokalt maximum. För exempelvis Cauchyfördelningen kan likelihoodfunktionen ha flera lokala optimum och för generaliserade Laplacedistributionen kan den anta oändliga värden, vilket försvårar maximeringsproblemet [8].



Figur 3: Likelihood som funktion av λ för ett stickprov av storlek hundra från en exponentialfördelning med $\lambda = 1$.

2.4 Momentmetoden

Momentmetoden är en metod för att uppskatta en eller flera distributionsparametrar $\theta_i, \ldots, \theta_k$ utifrån uppskattningar av distributionens moment. En distributions i:te moment μ_i ges av $\int_{-\infty}^{\infty}(x-c)^i f(x)$, där det första momentet, som är medelvärdet, beräknas med c=0. Andra momentet och över beräknas oftast centralt kring medelvärdet genom att låta c sättas till medelvärdet. Det andra centrala momentet blir då fördelningens varians.

Estimatorn för moment i givet ett stickprov $\mathbf{x}$ ges av $\hat{\mu}_i = \mathbf{E}(\mathbf{x}^i) = \frac{1}{n}\sum_{k=1}^{n} x_k^i$ vilket är en förhållandevis simpel beräkning. Om det går att uttrycka en distributionsparameter θ i distributionens moment genom någon funktion g s.a. $\theta = g(\mu_1, \ldots \mu_n)$, så kan vi estimera θ med $\hat{\theta} = g(\hat{\mu}_1, \ldots, \hat{\mu}_n)$. I exempelvis Laplacefördelningen kan momentmetoden användas för uppskattning av lägesparametern θ och spridningsparametern σ med $\hat{\theta} = \mu_1 = \mu$ och $\hat{\sigma} = \sqrt{2\hat{\mu}_2} = \sqrt{2}s$.

2.5 Bootstrap

Bootstrapmetoden är en metod som används för att få en bättre bild av ett stickprovs fördelning och säkerheten på olika estimatorer som kan användas på stickprovet. Metoden genererar ingen ny data och tillför heller ingen ny information, men erbjuder ett sätt att tydliggöra egenskaper hos informationen som redan finns. Bootstrapping introducerades av Bradly Efron under sent 1900-tal och är en väl beprövad metod. Ickeparametrisk Bootstrap, där vi drar ett antal stickprov med återläggning, $\mathbf{x}_1^*, \ldots, \mathbf{x}_B^*$ från ett originalstickprov $x_1, \ldots, x_n$, är en metod varifrån vi kan få information om säkerheten på estimatorer för distributionsparametrar beräknade från stickprovet. Från bootstrapstickprov estimeras θ_b^* för alla B bootstrapstickprov med en godtycklig estimator, till exempel MLE eller medelvärdet. Detta ger för tillräckligt stora värden på B en approximation till distributionen av θ, vilket kan ge oss en uppskattning av säkerheten i estimatorn, samt en bättre bild av parametern. Detta är särskilt användbart när det inte finns någon analytisk form eller normal teori för uppskattning av fördelningen av exempelvis medelvärde och varians.

Ickeparametrisk bootstrap kan felaktigt antas vara fri från antaganden, men metoden gör ett starkt antagande: att datan är diskret, och att värden som inte observeras i vårt originalstickprov är omöjliga. Detta kan ses som konstigt för kontinuerliga variabler, men om vi i vårt stickprov får exempelvis $(1.1, ...3.1, 3.5...5.2)$, så har ickeparametrisk bootstrap inget sätt att dra värden *mellan* 3.1 och 3.5, utan de ses som omöjliga. Ett exempel på hur bootstrap är en "naiv" estimator tas upp i ett blogginlägg [1], där en personifiering av bootstrap försöker uppskatta personers längd.

Efter två observationer vet bootstrap att personer kan vara 165 cm och 167 cm långa. Detta säger ingenting om sannolikheten för en längd på 166 cm, den sannolikheten förblir således 0 tills en sådan observation tillförs. Bootstrap är på så vis väldigt beroende på originaldata, och originaldatan behöver därav vara en bra representation av den faktiska fördelningen som undersöks för att få rimliga stabilitetsuppskattningar. Fördelar med bootstrap är att det är mycket enkelt att utföra och ger trots nämnda antaganden tydliga resultat. De punktvisa MLE estimationer vi använder i metoden presenterat i ekvation 1 kan ses som en priorifördelning för parametern θ som vi försöker uppskatta. I sektion 4 undersöker vi hur vi kan uppskatta denna priora distribution med hjälp av bootstrapping [4].

3 Viktat likelihood och empiriska priorifördelningar

Eftersom vi vill undersöka en generalisering av den föreslagna metoden i inledningen börjar vi således med att introducera metoden med en generell prior. Avsikten med metoden är att skapa en skälig estimering av en distributionsparameter θ_1 i en distribution med PDF $f(x|\boldsymbol{\theta})$, givet ett stickprov $\mathbf{x}$ från distributionen och en mängd lågkvalitativa estimatorer empiriskt framtagna från stickprovet. Vi gör ytterligare jämförelser med andra metoder samt utforskar olika sätt att välja och estimera den empiriska priorifördelningen.

3.1 Weighted likelihood med generell prior

Låt $\mathbf{x} = (x_1, \ldots, x_n)$ vara ett stickprov från en distribution med täthetsfunktion $f(\mathbf{x}|\theta)$, där θ är en godtycklig distributionsparameter. Låt vidare $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ vara en uppsättning godtyckliga, opartiska estimat av θ genererade från stickprovet av en funktion $g : \mathbb{R}^n \to \mathbb{R}^k$ s.a. $\hat{\theta} = g(\mathbf{x}, f)$, alltså där g är en godtycklig funktion som genererar estimat av θ.

Metoden genomförs enligt följande steg:

1. Välj funktionen $g(\cdot)$ och skapa en uppsättning estimatorer enligt $\hat{\theta} = g(\mathbf{x}, f)$
2. Ge varje estimator en vikt w_i proportionell mot dess likelihood givet stickprovet $\mathbf{x}$ enligt $w_i = \frac{L(\hat{\theta}_i : \mathbf{x})}{\sum_{j=1}^{k} L(\hat{\theta}_j : \mathbf{x})}$
3. Beräkna den slutliga estimatorn genom att summera varje estimat multiplicerat med dess vikt enligt $\hat{\theta}_{WLE} = \sum_{i=1}^{k} \omega_i \hat{\theta}_i$.

Dessa steg sammanfattas med följande ekvation:

$$\hat{\theta}_{WLE} = \sum_{i=1}^{k} \omega_i \hat{\theta}_i = \frac{\sum_{i=1}^{k} \hat{\theta}_i L(\hat{\theta}_i|\mathbf{x})}{\sum_{i=1}^{k} L(\hat{\theta}_i|\mathbf{x})} \tag{20}$$

Ett smidigt sätt att beskriva metoden är att använda Bayesiansk metodik. WLE använder i grunden inte ett Bayesianskt tankesätt eftersom ingen tidigare eller subjektiv information används, men den Bayesianska metodiken lånar sig väl för att beskriva metoden genom att låta priorifördelningen vara empiriskt framtagen från stickprovet. $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$ kan därför ses som en diskret prioridistribution med likformig sannolikhetsfördelning $\pi(\hat{\theta}_i) = \frac{1}{n}$. Genom att i ekvation 3 substituera A med $\hat{\theta}_i$ och B med $\mathbf{x}$, samt låta $w_i = \pi(\hat{\theta}_i|\mathbf{x}), i = 1, \ldots, n$ vara posteriorfördelningen, får vi

$$w_i = \pi(\hat{\theta}_i|\mathbf{x}) =\propto f(\mathbf{x}|\hat{\theta}_i)\pi(\hat{\theta}_i) = L(\hat{\theta}_i|\mathbf{x})\frac{1}{n} \propto L(\hat{\theta}_i|\mathbf{x})$$

Eftersom $(w_i, \ldots, w_n)$ är en sannolikhetsfördelning måste $\sum_{i=1}^{n} w_i = 1$ och därför dividerar vi med $\sum_{i=1}^{n} L(\hat{\theta}_i|\mathbf{x})$ och får posteriorfördelningen

$$w_i = \pi(\hat{\theta}_i|\mathbf{x}) = \frac{L(\hat{\theta}_i|\mathbf{x})}{\sum_{i=1}^{n} L(\hat{\theta}_i|\mathbf{x})}, i = 1, 2, \ldots, n. \tag{21}$$

Summan i ekvation 20, och därmed även den slutliga estimatorn, motsvarar således medelvärdet av den posteriora distributionen av θ. En av metodens viktiga egenskaper är att vi inte bara får estimatorn $\hat{\theta}_{WLE}$, utan även en approximation av estimatorns distribution $(w_1, \ldots, w_n)$. Denna approximation av distributionen på estimatorn är speciellt intressant eftersom mått på estimatorns säkerhet som till exempel varians kan estimeras. Metoden skiljer sig fundamentalt från andra optimeringsmetoder som MLE i att ett medelvärde beräknas utifrån en bestämd sekvens beräkningar utan att maximera eller optimera någon funktion. Detta kan ge metoden en beräkningsmässig fördel i de fall maximum likelihood är svårt att beräkna eller optimera, vilket kan vara fallet då likelihoodfunktionen är multimodal eller obegränsad i enstaka punkter.

Metodens huvudsakliga mål är att som tidigare nämnts skapa förhållandevis bra estimatorer från en mängd lågkvalitativa estimat. De kan vara få till antalet eller lågkvalitativa i sin natur. Vi ska studera metoden för olika typer av priorifördelningar, d.v.s. olika typer av funktionen g, där g kommer innefatta bland annat punktvis användning av MLE och bootstrapping.

3.2 Punktvis MLE som empirisk priorifördelning

Metoden som introducerats i tidigare litteratur använder WLE-metoden med en priorifördelning där varje estimator $\hat{\theta}_i$ är en MLE-skattning baserad på en datapunkt x_i i stickprovet. Detta ger att

$$g(\mathbf{x}, f) = (\max_\theta L(\theta|x_1), \dots, \max_\theta L(\theta|x_n)) = (\max_\theta f(x_1|\theta), \dots, \max_\theta f(x_n|\theta)) \tag{22}$$

Vi visar nu ett exempel för detta där vi antar $x \sim Exp(\lambda)$ och låter $\mathbf{x}$ innehålla n observationer från denna fördelning. Den punktvisa MLE-estimatorn är då definierad som $\hat{\theta}_i = x_i^{-1}$. Likelihood-funktionen för en observation x_i blir

$$L(\hat{\theta}_i|\mathbf{x}) = \left(\frac{1}{x_i}\right)^n \exp\left(\frac{-n\bar{x}}{x_i}\right).$$

och genom ekvation 20 når vi slutligen

$$\theta_{est} = \frac{\sum_{i=1}^n x_i^{-1} L(\theta_i : \mathbf{x})}{\sum_{i=1}^n L(\theta_i : \mathbf{x})} = \frac{\sum_{i=1}^n x_i^{-n-1} \exp\left(\frac{-n\bar{x}}{x_i}\right)}{\sum_{i=1}^n x_i^{-n} \exp\left(\frac{-n\bar{x}}{x_i}\right)}.$$

Vi skall i nästkommande stycke introducera en metod för att välja andra priorifördelningar än punktvis MLE med hjälp av bootstrap.

3.3 Parameteruppskattning för Cauchyfördelningen

Vi ska jämföra WLE-metoden med punkvis MLE som priorifördelning, med några andra metoder för att estimera lägesparametern för Cauchyfördelningen. Estimatorerna vi jämför är:

1. $\hat{\theta}_{initial} = x_1$, första värdet i stickprovet.
2. $\hat{\theta}_{mean} = \bar{\mathbf{x}}$, stickprovets medelvärde.
3. $\hat{\theta}_{median} = \tilde{\mathbf{x}}$, stickprovets median.
4. $\hat{\theta}_{MLE}$, maximum likelihood estimator.
5. $\hat{\theta}_{WLE}$, weighted likelihood estimator med punktvis MLE som prior.

Dessa metoder producerar alla estimat av θ av varierande kvalité. Den förstnämnda, $\hat{\theta}_{initial}$, är det första värdet ur ett stickprov vilket kan ge ett dåligt resultat. Samtidigt är det möjligt att ibland få ett förrädiskt bra resultat med denna metod trots att det inte är en pålitlig estimator. $\hat{\theta}_{mean}$ är medelvärdet av stickprovet vilket tillsammans med $\hat{\theta}_{initial}$ är erkänt dåliga estimatorer för Cauchy-fördelningen. Detta orsakas av de asymptotiska egenskaper vi diskuterat i avsnitt 2.3.1, nämnvärt heavy tails. Detta fenomen orsakar relativt höga sannolikheter för mycket extrema värden, vilket i sin tur påverkar estimatorer känsliga för utbölingar, så som medelvärdet. Estimatorn för medianen $\hat{\theta}_{median}$ är inte lika känslig för utbölingar och ger ett mer pålitligt resultat. Maximum likelihood estimatorn $\hat{\theta}_{MLE}$ är den vanligaste estimatorn och är relativt bra. $\hat{\theta}_{WLE}$ är vår nya metod för vilken vi ska jämföra prestanda.

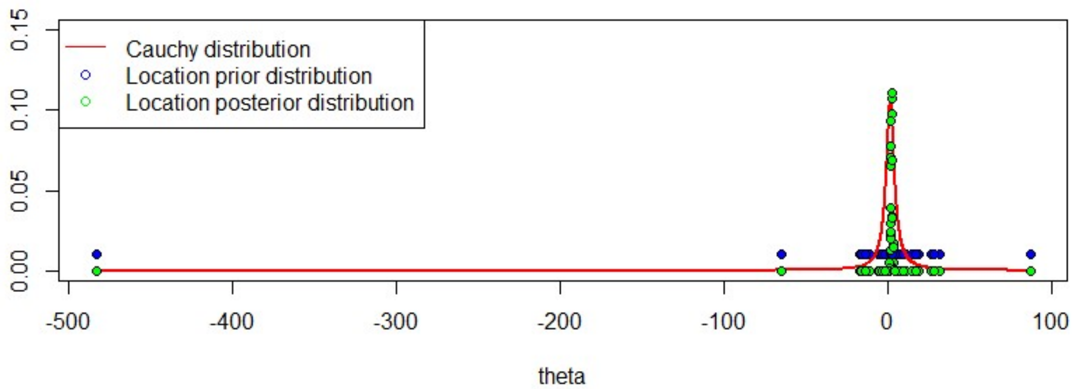
Metoderna undersöks på Cauchyfördelningen för $\theta = 2, \sigma = 3$ och för stickprovsstorlekarna $n = 5, 10, 100$. När vi mäter lägesparametern låter vi spridningsparametern vara känd och konstant. För att analysera en metods säkerhet använder vi Monte-Carlo metoden (MC) med 10 000 stickprov. Metodens estimator evalueras för varje stickprov och sedan beräknas medelvärde och

MSE på distributionen av estimat. Även om det MC-genererade medelvärdet för en metod är nära det faktiska värdet på parametern så beror den enskilda estimatorns säkerhet till hög grad på MSE hos MC-fördelningen. Således är MSE en bättre indikator för att jämföra estimatorers säkerhet då alla opartiska estimatorer asymptotiskt närmar sig det sanna värdet. Resultatet presenteras i tabell 3. Vi ser att $\hat{\theta}_{initial}$ och $\hat{\theta}_{mean}$ presterar väldigt dåligt, båda med MSE på flera tusen. Detta för att dessa estimatorer inte har någon mekanism för att minska de extrema utbölingarnas vikt vid beräkningen av estimatorn. WLE, MLE och $\hat{\theta}_{median}$ kompenserar alla för extrema utbölingar och presterar därför bättre. $\hat{\theta}_{median}$ ignorerar extrema utbölingar eftersom utbölingarna per definition hamnar på fördelningens kanter, långt från medianen. WLE kompenserar genom att vikten w_i för en utböling $\hat{\theta}_i = x_i$ blir mycket liten. MLE kompenserar genom att maxvärdet för likelihood-funktionen hamnar i området där flest datapunkter från stickprovet finns, vilket inte är vid utbölingar.

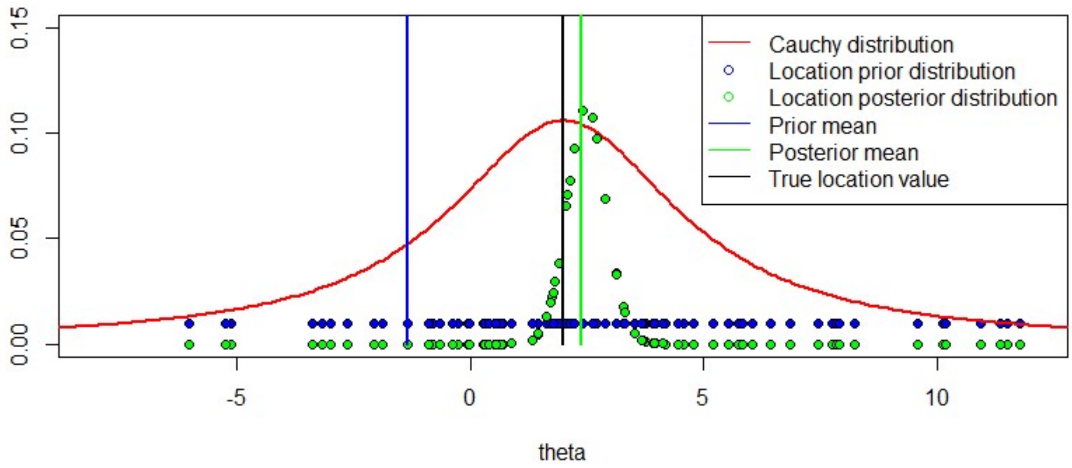
Tabell 3: Medelvärde och MSE för estimatorerna WLE, initialvärdet, Medelvärdet, Median och MLE, efter 10000 Monte-Carlo iterationer, för olika provstorlekar n.

provstorlek	WLE(MSE)	Initialvärde(MSE)	Medelvärde(MSE)	Median(MSE)	MLE(MSE)
5	1.981(8.79)	-5.884(387107)	-0.963(53044)	2.013(10.30)	2.112(2.576)
10	1.999(2.59)	1.962(55674)	-4.441(255329)	1.994(2.98908)	2.045(1.677)
100	1.996(0.190)	0.119(35876)	2.018(68480)	1.994(0.224)	1.99(0.184)

I figur 5 ser vi ett exempel på hur en extrem utböling ger ett stort inflytande på stickprovets medelvärde, som visualiseras med de blå vertikala strecket i figur 5. Genom viktningen i WLE metoden får den posteriora distributionen ett medelvärde mycket närmare den sanna lägesparametern visat av det gröna strecket. Det blir uppenbart hur WLE-metodens viktning av dåliga estimat minskar utbölingars inflytande och ökar inflytande från estimatrorer i ett mer precist område nära parameterns faktiska värde. Förmågan att ge utbölingar en låg inverkan på estimatorn är ytterligare en aspekt som gör WLE-metoden attraktiv då man har priora estimatorer av dålig eller okänd kvalitet. I fallet med Cauchyfördelningen presterade medianen på ungeför samma nivå som WLE och kan därför i detta fall vara att föredra på grund av dess enkelhet. Men för en generell distribution då man inte har en känt bra metod för att estimera en parameter, men flera dåliga estimatorer, och om MLE-metoden är osäker eller svårberäknad, kan WLE-metoden vara ett bra och simpelt alternativ för att generera ett estimat av god kvalitet.



Figur 4: Prior- och posteriordistribution för lägesparametern i Cauchydistributionen där den priora fördelningen är framtagen genom punktvis MLE och ett stickprov på storlek 100.



Figur 5: Prior- och posterior distribution för lägesparametern i Cauchydistributionen där den priora fördelningen är framtagen genom punktvis MLE och ett stickprov på storlek 100. Grafen visar samma data som i figur 4 men på ett mindre intervall centrerat kring det faktiska värdet på θ.

4 Bootstrapgenererade empiriska priorifördelningar

Vi såg i tabell 3 hur estimatorer som medelvärdet och initialvärdet var mycket dåliga även för stora datamängder. Vi ska nu undersöka huruvida det är möjligt att med bootstrap-metoden kombinera metoderna vi undersökt ovan, initialvärdet, medelvärde och median, med WLE för att förbättra estimeringarna. Detta tillvägagångssätt kan på Bayesianskt vis tolkas som att vi byter den nuvarande empiriska priorifördelning mot en annan empirisk fördelning. Tidigare har vi endast använt MLE punktvis för att skapa våra $\hat{\theta}$ till priorifördelningen, vilket kan beskrivas matematiskt med $\hat{\theta} = g(x, f)$, där $g(x, f) = (MLE(\theta|x_1), \ldots, MLE(\theta|x_n)) = (x_1, \ldots, x_n))$. Vi ska i detta avsnitt undersöka huruvida vi kan skapa den empiriska priorifördelning på andra vis. Denna undersökning genomförs genom att ta B bootstrapstickprov från stickprovet och därefter applicera estimatorerna vi undersökt i sektionen ovan på dessa, vilket kan beskrivas som $g(\mathbf{x}, f) = (\hat{\theta}_1, \ldots, \hat{\theta}_B) = (\bar{x}_1^*, \ldots, \bar{x}_B^*)$. Detta samband kan vara mycket användbart i fall då vi inte känner till det bästa sättet att skapa ett estimat och endast har tillgång till någon estimator, till exempel medelvärdet.

4.1 Bootstrap prior i Cauchyfallet

Vi genererade från samma stickprov som i undersökningarna i sektion 3 för $n = 10$ ett varierande antal bootstrap mängder, där alla hade samma storlek som det ursprungliga stickprovet, dvs $m = n = 10$. Resultaten presenteras i tabell 4 och visar på en betydlig förbättring i estimaten allt eftersom B ökar, speciellt för $\theta_{initial}$ och θ_{mean}. Notera hur alla mätvärden blivit signifikant förbättrade med möjligt undantag för median, då denna estimator redan från början är mycket bra och därav inte kan förbättras särskilt mycket. Nämnvärt är att MSE för medelvärdet och initialvärdet kan för låga antal bootstrapmängder vara instabilt och tycks minska asymptotiskt för stigande B.

Tidigare såg vi hur estimatorerna för initialvärdet och medelvärdet av stickprovet varit bristfälliga och genererade mycket dåliga estimat. Genom att nu i WLE använda initialvärdet, medelvärdet och medianen i vår funktion $g(\mathbf{x}, f)$ och applicera den på bootstrapmängderna genererade från x, ser vi hur vi kan vikta en rad dåliga estimat till ett bra om vi har tillräckligt många av dem. Intressant är att vi börjar se stora förbättringar redan för $B = 3$ och vid $B = 5$. Undantaget är för medelvärdet, där vi först vid $B = 100$ börjar få bra resultat.

Alltså går ett estimats kvalité och antalet bootstrapmängder hand i hand för att nå ett användbart resultat. Ett alltför dåligt estimat kommer att kräva allt större bootstrapmängder för att nå en punkt där MSE inte förändras särskilt mycket trots stora ökningar i antalet bootstrapmängder. Anledningen till varför medelvärdet inte blir optimalt när ett stickprov tas från en Cauchyfördelning med bootstrap har att göra med dess egenskaperna gällande medelvärdet, specifikt att det finns stor risk för extrema utbölingar. En extrem utböling kan förstöra ett medelvärdes precision, och med $m = 10$ riskerar vi att många av våra bootstrapmängder får med minst en utböling om den existerar i det originella stickprovet. Vi ser därför att medelvärdet blir vår klart sämsta estimator att applicera på bootstrapmängderna, och ger därav en mycket dålig empirisk priorifördelning. Men även i detta fall där vi har extremt dåliga estimat, kan vi alltså med hjälp av WLE effektivt kombinera dem till ett bra. Det blir tydligt att nästan oavsett hur dålig estimatorn är, så kan WLE förbättra estimaten från bootstrapmängderna. Den första datapunkten är en slumpmässig punkt ur det ursprungliga stickprovet vilket gör att metodiken för denna liknar vårt ursprungliga tillvägagångssätt där vi gjorde MLE punktvis. Resultaten blir därav mycket lika.

Tabell 4: Resultat från initialvärdet, medelvärdet och medianen applicerade på varierande antal bootstrapmängder av storlek $m = 10$ använt som priorifördelning i WLE på samma stickprov av storlek $n = 10$ som i tabell 3. I tabell 3 ser vi MSE-värden på 55674, 255329 och 2.98908 för initialvärdet, medelvärdet respektive medianen vilket kan jämföras med tabellen nedan. Vi ser en tydlig förbättring, särskilt för de 2 förstnämnda.

Bootstrapmängder	Initialvärde(MSE)	Medelvärde(MSE)	Median(MSE)
2	1.332(1487.31)	-0.154(21617.52)	1.96(4.754)
3	2.018(13.86)	-2.637(245665)	1.97(3.319)
5	2.009(4.67)	1.977(319.14)	1.991(2.82)
10	1.994(2.97)	2.154(79.88)	1.994(2.67)
100	2.002(2.60)	1.996(2.82)	1.995(2.54)

4.2 Metodjämförelser för Laplacefördelningen

Vi skall i detta avsnitt upprepa det vi gjort med Cauchyfördelningen ovanför med Laplacefördelningen. Vi kommer här fokusera exempelvis på att vidare undersöka hur WLE hanterar redan goda estimatorer.

Vi utförde liknande beräkningar och jämförelser till Cauchyfördelningen för Laplacefördelningen, presenterat i tabeller 5 och 6. Som tidigare är estimatorerna initialvärdet, medelvärdet och medianen. Resultaten syns i tabell 5. Metoderna användes sedan med WLE vilket visas i tabell 6. Vi ser här hur estimatorerna i relation till Cauchyfallet presterar bra redan innan vi applicerar WLE.

Tabell 5: Resultat för initialvärdet, medelvärdet och medianen, för Laplacefördelningen.

provstorlek	WLE(MSE)	Initialvärde (MSE)	MoM/Medelvärde(MSE)	Median(MSE)
5	3.001(0.324)	2.998(1.984)	3.002(0.395)	2.999(0.349)
10	3.000(0.141)	3.016(2.03)	2.996(0.199)	3.002(0.143)
100	3.000(0.0148)	2.973(1.96)	2.998(0.0199)	2.999(0.0116)

Nämnvärt, för Laplacefördelningen, är att MLE-estimatorn kan visas vara medianen. Mätvärdena i kolumn fyra är således mycket svåra att överträffa, vilket reflekteras i tabell 6.

Tabell 6: Resultat från modifierade initialvärdet, medelvärdet och medianen med konstant provstorlek 10, för Laplacefördelningen. Med undantag för initialvärdet, som är den klart sämsta estimatorn för Laplacefördelningen, ser vi små till inga förbättringar.

Bootstrapmängder	Initialvärde(MSE)	Medelvärde(MSE)	Median(MSE)
2	2.995(0.682)	2.998(0.264)	3.000(0.231)
5	2.995(0.285)	2.997(0.198)	2.998(0.174)
10	2.999(0.206)	2.998(0.181)	2.997(0.158)
100	2.996(0.150)	2.997(0.166)	2.998(0.143)

Som förväntat, ser vi försumbar förbättring för medelvärdet och inga för medianen vid applicering av WLE. I och med de mycket goda egenskaperna som dessa estimatorer besitter finns inte mycket att förbättra. Initialvärdet, å andra sidan, är en relativt lågpresterande estimator men som vi genom WLE kunde sänka variansen för markant, med så lite som två bootstrapmängder. Detta stärker således hypotesen att WLE effektivt kan användas i syfte att slå ihop anförskaffade mängder av ”dåliga” estimatorer.

5 WLE och flera okända parametrar

I detta avsnitt skall vi utvidga vår horisont genom att undersöka hur WLE-metoden presterar då flera parametrar är okända och skall uppskattas. Vi har valt att avgränsa oss till två parametrar och lämnar studier av 3 eller fler parametrar åt framtida studier.

För syftet har vi valt att utföra beräkningar med Laplacefördelningen då vi redan diskuterat och studerat fördelningen i fråga för en okänd parameter. Ett tidigare föreslaget tillvägagångssätt för att estimera parametrar för en Laplacefördelning är genom utnyttjande av Method of Moments. I sektion 4.2 använde vi MoMs första moment, medelvärdet, som estimator till lägesparametern θ. I artikeln presenteras hur man genom Method of Moments kan estimera fler parametrar av Laplacefördelningen än enbart lägesparametern. I vårt fall kan MoMs andra moment, variansen, användas som estimator till spridningsparametern σ. Dessa parametrar estimeras parvis ur bootstrapstickprov och ges en vikt w_i proportionell mot dess likelihood. Sedan summeras estimaten till en viktad summa genom WLE för att bilda slutgiltiga estimatorer för parametrarna. På detta sätt kan arbetssättet ses som en utveckling av tidigare metodik där parametrar estimeras parvis ur bootstrapstickprov [11].

Metoden, som beskriven i sektion 3.1, utvidgas därmed till följande. Låt $\mathbf{x} = (x_1, \ldots, x_n)$ vara ett stickprov från en distribution med täthetsfunktion $f(\mathbf{x}|\theta, \sigma)$, där θ och σ är distributionsparametrar. Låt vidare $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ och $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_k)$ vara en uppsättning godtyckliga, parvisa estimat av θ och σ genererade från stickprovet av en funktion $g : \mathbb{R}^n \to \mathbb{R}^{2\times k}$ s.a. $(\hat{\theta}, \hat{\sigma}) = g(\mathbf{x}, f)$, alltså där g är en godtycklig funktion som förväntas generera estimat av θ och σ.

1. Generera B bootstrapstickprov av storlek m från originalstickprovet på n datapunkter.

2. För varje stickprov B_i, applicera funktionen $g(\cdot)$ och skapa en uppsättning estimatorer enligt $(\hat{\theta}, \hat{\sigma}) = g(\mathbf{x_b}, f)$, där x_b är bootstrapstickprovet B_i.

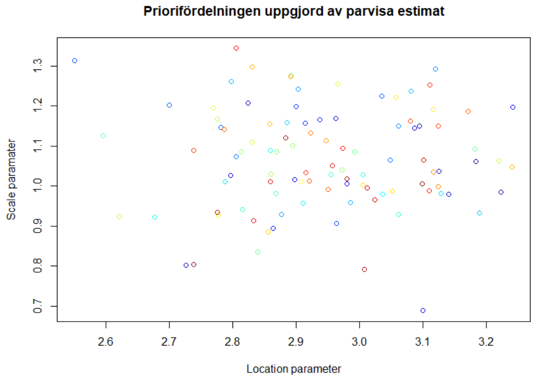
3. Ge varje par av estimat i $(\hat{\theta}, \hat{\sigma})$ en vikt w_i proportionell mot dess likelihood enligt $w_i = \frac{L(\hat{\theta}_i, \hat{\sigma}_i : \mathbf{x})}{\sum_{j=1}^{k} L(\hat{\theta}_j, \hat{\sigma}_j : \mathbf{x})}$, där x är orginalstickprovet.

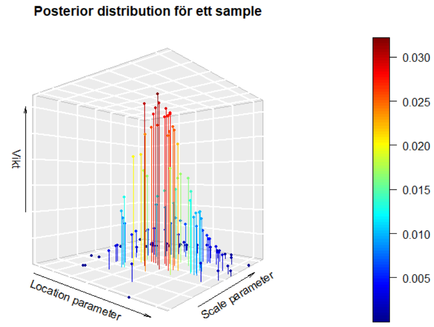
4. Beräkna den slutliga estimatorn genom att summera varje estimat multiplicerat med dess vikt enligt $\hat{\theta}_{WLE} = \sum_{i=1}^{k} \omega_i \hat{\theta}_i$ och $\hat{\sigma}_{WLE} = \sum_{i=1}^{k} \omega_i \hat{\sigma_i}$.

5.1 Estimatorernas fördelning

I avsnitt 3.1 diskuterades fördelen att användande av WLE resulterar i en approximation av parametrarnas fördelning, något vi sedan visualiserade i figur 5 i sektion 3.3. Vi ska nu presentera denna posteriordistribution för en Laplacedistribution med 2 okända parametrar.

I figur 6 ser vi alla parvisa estimat av lägesparametern θ och spridningsparametern σ givna av MoM på $B = 100$ bootstrap mängder av storlek $m = 100$ från ett originalstickprov av storlek $n = 100$ taget ur en Laplacedistribution med lägesparameter $\theta = 3$ och spridningsparameter $\sigma = 1$. De punktvisa estimaten har alla samma vikt i den priora distributionen och tolkas därför som lika viktiga. I figur 7 applicerar vi WLE på de 100 estimaten i figur 6. Estimaten får då olika stor vikt beroende på dess likelihood i förhållande till originalstickprovet. Vi viktar estimaten mot originaldatan för att se hur bra resultatet passar datan.





Figur 6: 100 parvisa estimat för läges- och spridningsparametern skapade genom Method of Moments på $B = 100$ bootstrapmängder av storlek $m = 100$ från ett originalstickprov av storlek $n = 100$. Priorifördelningen ges av en likformig fördelning över punkterna.

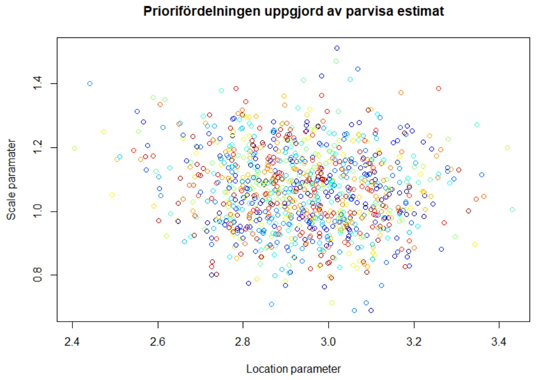
Figur 7: Posteriordistribution för estimaten i figur 6 med estimatens vikter givna i höjdled. Värden för läges- och distributionsparametrarna är samma som i figur 6 som därav uppgör botten av figuren.

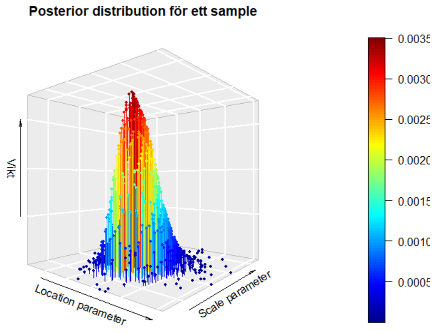
I tabell 7 ser vi resultatet av WLE-viktningen presenterad i figur 7 jämfört med att applicera MoM direkt på originalstickprovet. Viktningen tycks ge ett bättre resultat. Tidigare har vi använt MC-simulering och beräknat MSE för att uppskatta säkerheten i estimaten. Vid användning av ett enstaka stickprov kan säkerheten på estimatet vara svårberäknelig, men vid användning av WLE kan vi enkelt ta variansen av posteriorfördelningen. För MoM blir detta besvärligt då vi endast har en punkt. För tillräcklig stickprovsstorlek n blir variansen en grov uppskattning av säkerheten i estimatet och kan jämföras med MSE i tidigare tabeller. Anledning till att n måste vara tillräckligt stor ska vi undersöka i nästa sektion.

Tabell 7: MoM direkt på det originella stickprovet av storlek $n = 100$ jämfört med WLE efter MoM på $B = 100$ bootstrap mängder av storlek $m = 100$ från det originella stickprovet. Med WLE kan vi uppskatta variansen från posteriordistrubitonen.

θ_{mom}	σ_{mom}	θ_{WLE}(Variance)	σ_{WLE}(Variance)
2.938	1.075	2.964(0.0076)	1.02(0.0052)

Att öka antalet bootstrapmänger B är ofta ingen större svårighet då estimatorerna vi använder är enkla. Detta gör att vi enkelt kan erhålla en tydlig distribution av resultatet utan alltför krävande beräkningar. I Figur 9 ser vi att vi får en mycket tydligare posteriordistribution än i figur 7. Att vi får denna posteriordistrubtionen gratis är en stor fördel med vår metod jämfört med andra metoder för parameteruppskattning såsom MoM, särskilt i komplicerade flerdimensionella fall då denna kan vara svårberäknad.





Figur 8: 1000 parvisa estimat för läges- och spridningsparametern skapade genom Method of Moments på $B = 1000$ bootstrapmängder av storlek $m = 100$ från ett originalstickprov av storlek $n = 100$. Priorifördelningen ges av en likformig fördelning över punkterna.

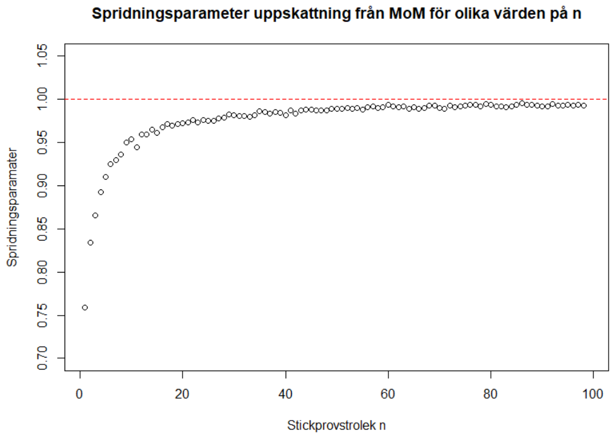
Figur 9: Posteriordistribution för estimaten i figur 8 med estimatens vikter givna i höjdled. Värden för läges- och distributionsparametrarna är samma som i figur 8 som därav uppgör botten av figuren.

5.2 Multi-parameteruppskattning för Laplacefördelningen

Som vi sett i tabell 6 är MoM-estimat på bootstrap tillsammans med WLE en nära opartisk estimator av Laplacedistributionens lägesparameter. I de vänstra kolumnerna av tabell 8 ser vi resultatet när samma metod används för att estimera både θ och σ samtidigt, igen med parametervärdena $\theta = 3$ och $\sigma = 1$. Även här ser vi att estimatorn av θ är nära opartisk, medan $\hat{\sigma}$ har ett negativt bias. En anledning till denna bias är att MoM estimatorn i sig är partisk vilket kan noteras från figur 10 där vi också ser att MoM-estimatorns bias ökar för lägre värden på n.

Tabell 8: Resultat för WLE med MoM som estimator på B bootstrapstickprov av storlek $m = 10$ från ett orginalstickprov av storlek $n = 10$ från Laplacefördelningen, jämfört med att applicera MoM direkt på orginalstickprovet.

Stickprovstorlek n	θ_{MoM}(MSE)	σ_{MoM}(MSE)
10	2.997(0.200)	0.951(0.108)
Bootstrapmängder B	**θ_{WLE}(MSE)**	**σ_{WLE}(MSE)**
2	2.999(0.260)	0.914(0.133)
5	2.997(0.191)	0.927(0.111)
10	2.997(0.176)	0.929(0.107)
100	2.998(0.165)	0.931(0.104)



Figur 10: MC-simulering av storlek 10000 för uppskattning av spridningsparametern med MoM för olika stickprovstorlekar n från en Laplacefördelning med $\theta = 3$ och $\sigma = 1$. Visar ett bias för låga värden på n.

Trots denna bias som kan uppstå vid användning av MoM för låga n verkar det som att WLE-metoden med bootstrap och MoM ger upphov till ett något större bias än att bara använda MoM på originalstickprovet, vilket kan ses i tabell 8 där $\hat{\sigma}_{MoM}$ ger ett bias på -0.5 medan $\hat{\sigma}_{WLE}$ ger ett bias på -0.7 även för stora värden på B. Vår teori om vad detta kan bero på är att en andel av bootstrapstickproven innehåller dubbletter av datapunkter, vilket är ett resultat av att bootstrapstickproven dras med återläggning. Varför detta skulle ge ett ökat bias kan förklaras med ett simpelt exempel: Antag att vi har ett stickprov av storlek 5 och tar av detta ett bootstrapstickprov av storlek 2. Sannolikheten att bootstrapstickprovet innehåller en dubblett av en och samma datapunkt är då 1/5 och ger i det fallet σ-estimatet 0, vilket är en orimligt låg estimator. Värdena i tabell 8 har visserligen ett högre värde på både n och m där båda är satta till 10, vilket minskar förekomsten av dubbletter, men det kan ändå tänkas att dubbletter även i detta fall get ett ökat bias i någon utsträckning.

Ett högre värde på n minskar alltså inverkan på bias från MoM och möjligtvis även inverkan från dubbletter i bootstrapstickproven. Av detta kan vi dra slutsatsen att ett relativt stort originalstickprov är önskvärt för att använda WLE med bootstrap och MoM för att estimera σ. Detta styrks också av tabell 9 där vi ser att samma typ av simuleringar, men med $n = 100$, minskar estimatorns bias avsevärt.

Tabell 9: Resultat för n=100, m=10 för Laplacefördelningen.

Bootstrapmängder	θ(MSE)	σ(MSE)
2	2.998(0.115)	0.977(0.0682)
5	3.003(0.044)	0.995(0.033)
10	3.000(0.025)	0.997(0.020)
100	3.000(0.012)	0.996(0.010)
1000	2.999(0.011)	0.996(0.010)

Estimaten då $n = 100$ är också mycket bra utöver ett minskat bias, trots storleken $m = 10$ på våra bootstrap stickprov. Om vi använder MoM direkt på stickprovet får vi här $\theta_{MoM}(MSE) = 2.999(0.012)$ och $\sigma_{MoM}(MSE) = 0.992(0.012)$. Genom WLE får vi för $B = 100$ något bättre MSE än MoM. Att öka B ytterligare efter detta ger ingen större förbättring.

6 Slutsatser

I sektion 3.3 såg vi hur WLE med en prior av lågkvalitativa estimat genererar en bra estimator som kompenserar för extrema utbölingar i priorifördelningen. Bootstrapping tillsammans med en simpel estimator från till exempel momentmetoden eller medelvärde gör det möjligt att generera prioriestimat som, trots en potentiellt låg kvalité, är tillräckliga för att tillsammans med WLE skapa estimatorer av önskvärd kvalité. Detta tillåter ett enkelt sätt att skaffa bra estimat av parametrar utan tillgång till någon känt bra estimator. I fallet med Cauchydistributionen är det välkänt att medianen är en bra estimator av lägesparametern, men om detta inte var känt hade WLE med bootstrap och medelvärde varit ett fullt gott alternativ. Metoden har även visats ge en marginell förbättring för prioriestimat av god kvalité, men det är med lågkvalitativa estimat metoden kan få ett betydande användningsområde. WLE tillsammans med bootstrapping och momentmetoden har också visat sig vara en lämplig metod för att estimera både läges- och spridningsparametern samtidigt, men att det då är önskvärt med större stickprov för att minska bias för spridningsparametern. Detta ger en potential för vidare studier att undersöka hur WLE tillsammans med bootstrap och MoM lämpar sig för uppskattning av ett godtyckligt antal distributionsparametrar av godtycklig typ. Vi har också sett att WLE-estimatorn saknar bias för lägesparametern i unimodala och symmetriska distributioner, även då lägesparametern estimeras tillsammans med spridningsparametern, och hur bias för spridningsparametern minskar för större stickprovsstorlekar. Det kan för vidare studier vara intressant att undersöka bias vid estimering av den logaritmerade spridningsparametern, eller användning av andra estimatorer med potentiellt lägre bias. Det visar sig att WLE inte bara är ett smidigt sätt att generera ett estimat utan också ger information om estimatets säkerhet genom posteriorifördelningen som produceras i metoden för att generera estimatet. Detta ger möjligheten att analysera estimatorns säkerhet även då det inte går att härleda analytiskt och det inte är möjligt att analysera statistiskt genom upprepade tester.

Trots att resultaten vi erhållit är lovande har studien endast undersökt WLE i några specialfall och fungerar därför bäst som konceptvalidering, och det krävs vidare studier för att kunna hävda våra slutsatser på en mer generell nivå utöver de exempel vi studerat. Sålunda finns det alla möjligheter för att göra vidare studier av WLE-metoden. Då studien har undersökt metoden på ett rent teoretiskt plan har vi inte heller undersökt möjliga praktiska applikationsområden, vilket också är ett intressant område för vidare studier, och det ska bli spännande att följa vilka applikationsområden som kan komma att upptäckas för WLE.

Referenser

[1] Rasmus Bååth. *The Non-parametric Bootstrap as a Bayesian Model*. 2015. URL: `http://www.sumsar.net/blog/2015/04/the-non-parametric-bootstrap-as-a-bayesian-model/`.

[2] Jonny Brooks. "Probability concepts explained: Maximum likelihood estimation". I: *Towards Data Science* (2018).

[3] Jason Brownlee. *A Gentle Introduction to the Bootstrap Method*. 2018. URL: `https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/`.

[4] Bradley Efron och R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman och Hall/CRC, 1994.

[5] Julia Galef. *A visual guide to bayesian thinking*. 2015. URL: `https://www.youtube.com/watch?v=BrK7X_XlGB8`.

[6] Stephanie Glen. "Cauchy Distribution: Simple Definition, PDF, Uses". I: *Statistics How To* (2015). URL: `https://www.statisticshowto.com/cauchy-distribution-2/`.

[7] Stephanie Glen. "Heavy Tailed Distribution Light Tailed Distribution: Definition Examples". I: *Statistics How To* (2016). URL: `https://www.statisticshowto.com/heavy-tailed-distribution/`.

[8] Md.Mobarak Hussein, Kozubowski.T.J och Podgórski.K. "A novel weighted likelihood estimation with empirical Bayes flavor". I: (2017).

[9] Karin Nordgren. *Mejl med statistik från Göteborgs Universitet för år 2020*. service@gu.se.

[10] Robert M. Norton. "The Double Exponential Distribution: Using Calculus to Find a Maximum Likelihood Estimator". I: *The American Statistician* (1984), s. 135–136.

[11] Krzysztof Podgórski och Jörg Wegener. "Estimation for Stochastic Models Driven by Laplace Motion,Communications in Statistics - Theory and Methods". I: (2011). URL: `http://dx.doi.org/10.1080/03610926.2010.499051`.

[12] John A. Rice. *Mathematical Statistics and Data Analysis, third edition.* 2007.

[13] Christian P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* Springer Science och Business Media, 2001.