



CHALMERS



GÖTEBORGS UNIVERSITET

Åldersbedömning av asylsökande

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers tekniska högskola

Viktor Erbro

Henrik Esmaili

Pontus Larsson

Jesper Olsson

James Pålsson

Erik Sörstadius

Åldersbedömning av asylsökande

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet

Jesper Olsson James Pålsson

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers

Viktor Erbro Henrik Esmaili
Pontus Larsson Erik Sörstadius

Handledare: Petter Mostad
Examinator: Marina Axelson-Fisk
 Maria Roginskaya

Institutionen för matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2018

Populärvetenskaplig presentation

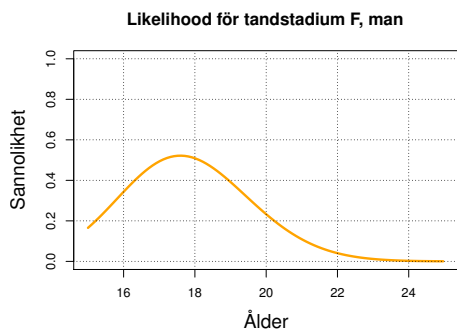
Under 2017 blossade en debatt upp kring Rättsmedicinalverkets (RMV) metod för att åldersbedöma ensamkommande asylsökande. I kritik från bland annat rättsläkare och statistiker hävdades att det fanns för litet underlag i form av vetenskapliga studier och bra statistisk grund för att kunna lita på resultaten från bedömningarna [10][18][20].

Den här rapporten presenterar en alternativ metod för åldersbedömning som använder sig av matematisk statistik och stokastisk optimering. Nedan ges ett exempel på ett fall där en fiktiv asylsökande åldersbedöms med metoden framtagen i rapporten.

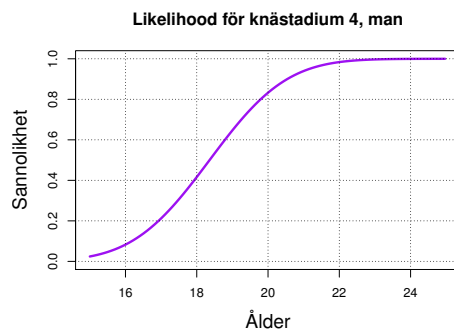
Pojken Kim vill söka asyl i Sverige och blir erbjuden ett åldersbedömningstest av Migrationsverket. Åldersbedömningstestet de erbjuder baseras på bedömningar på visdomständer och knäleder.

Visdomständer kategoriseras efter mognad enligt Demirjians metod, en skala från A till H , där H representerar fullmogen visdomstand [2]. Knäleder kategoriseras efter mognad enligt Krämers definition, en skala från 1 till 5, där stigande värden representerar en högre knäledsmognad [14]. Efter att ha gjort både magnetkameraundersökning av höger knäled samt röntgen av höger och vänster visdomstand i underkäken, bedöms Kim ha tandstadium F på vänster visdomstand samt mognadsstadium 4 på höger knäled. Höger visdomstand anses inte vara analyserbar.

För att ta fram sannolikheten att Kim är över 18 år används Bayes formel. Då behövs en *likelihood* samt en *prior* för att ta ut en *posteriorfördelning*. Idén bakom posteriorfördelningen är att använda en prior, som är gemensam för alla asylsökande, och uppdatera den med information om Kim beskrivet av en likelihood. En likelihood tolkas som sannolikheten att en individ har ett visst mognadsstadium givet en viss ålder. Man kommer sedan att få ut en posteriorfördelning som beskriver alla möjliga åldrar Kim kan ha. Priorerna ska beskriva åldersfördelningen på de som blir erbjudna ett åldersbedömningstest i Sverige. Den är baserad på data från de asylsökande som kom till Sverige 2015 och åldersbestämdes 2017.



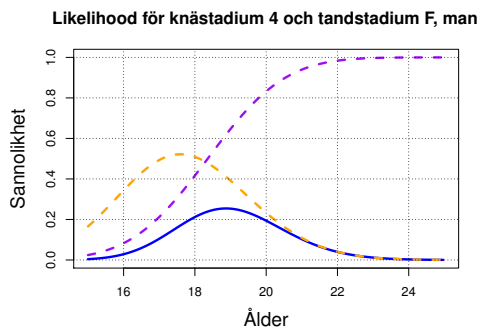
Figur 1: Sannolikheten att ha tandstadium F på vänster visdomstand i underkäken, för olika åldrar.



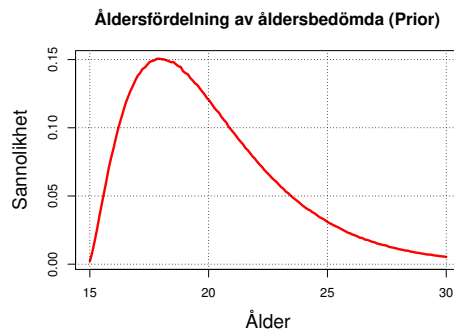
Figur 2: Sannolikheten att ha knästadium 4 på höger knäled, för olika åldrar.

För att kunna skapa en likelihood, för exempelvis mognadsstadier av tänder, krävs data i form av åldrar kopplade till mognadsstadierna. Då det här arbetet inte har haft tillgång till sådan data om åldrarna, har ett kompatibelt dataset framställts från att ha studerat publikationer om tandmognad från tandläkare. På motsvarande sätt skapas en likelihood för knäleder baserad på bedömningar av MR-bilder från radiologer. I figur 1 och 2 visas likelihoods för Kims mognadsstadier av knäled och tand.

Figur 3 presenterar produkten av likelihood från figur 1 och 2, vilket också är en likelihood. Den prior som vi har tagit fram visas i figur 4.

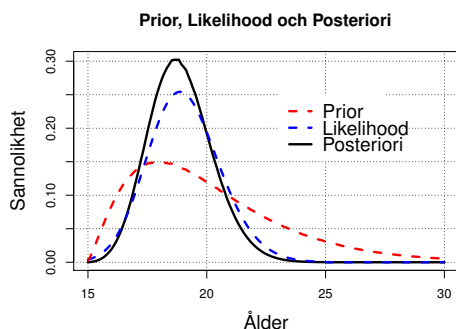


Figur 3: De streckade linjerna är samma likelihoods som i figur 1 och 2. Den blåa hel-dragna linjen är produkten av de två (sammanslagd likelihood).

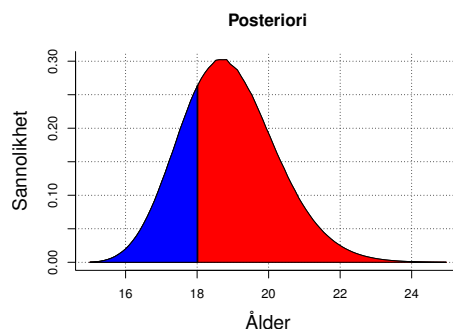


Figur 4: Prior för de som blir åldersbedömda av Rättsmedicinalverket.

Vidare fås posteriorifördelningen fram genom multiplikation av priorn sedd i figur 4 med produkten av likelihoods (figur 3). I figur 5 visas posteriorifördelningen normaliserad tillsammans med priorn och likelihood. Med posteriorifördelningen, vilken visas i figur 6, kan sannolikheten att Kim är över 18 år beräknas. Arean under kurvan är uppdelad i två. Den blåa arean motsvarar alla åldrar under 18 år och den röda arean motsvarar alla åldrar över 18 år. Man kan utifrån de här två areorna ta fram en sannolikhet för att Kim är över 18 år. I Kims fall är sannolikheten 78.8% att han är över 18 år.



Figur 5: Den svarta linjen är posteriorifördelningen. Den streckade röda linjen är priorn från figur 4 och den streckade blå linjen är den sammanslagda fördelningen i figur 3.



Figur 6: Posteriorifördelningen där arean under kurvan är uppdelad i två. Den blåa arean motsvarar alla åldrar under 18 år och den röda arean motsvarar alla åldrar över 18 år.

Hur bör ett beslut om Kim är över eller under 18 år fattas, givet Kims sannolikhet att vara över 18 år? Kostnader är alla möjliga fördelar och nackdelar som en individ, en grupp eller ett samhälle skulle kunna uppleva vid något beslut, översatt till någon gemensam skala. Det finns en kostnad, inte nödvändigtvis ekonomisk utan exempelvis politisk eller moralisk, associerad med att felbedöma Kim. Kostnaden för att bedöma ett barn som vuxen sätts som fyra gånger så stor som kostnaden att bedöma en vuxen som ett barn. Det här medför att det krävs att sannolikheten, att Kim är över 18 år, är över 80% för att bedöma honom som vuxen. Eftersom sannolikheten att Kim är över 18 år är 78.8%, vilket är mindre än 80%, bedöms Kim som under 18 år. RMV hade tagit beslutet att Kim är över 18 år endast eftersom han har ett moget knä. Observera att syftet med exemplet och projektet *inte* är att agera som en jämförelse mot RMV. Generellt sett ska det inte heller tolkas som att vår modell ger fördel åt beslut under 18 år, jämfört med RMV:s modell. Varför RMV:s beslut i frågan tas upp, är för att satisfiera den nyfiknes fundering. Nyttan med vår metod, jämfört med RMV:s metod, är att den har tydliga resultat och enkelt kan utökas om ny forskning publiceras.

Sammanfattning

Det här kandidatarbetet har tagit fram en åldersbedömningsmetod baserad på bayesiansk inferens. Som åldersindikatorer har mognadsstadier för vänster och höger visdomstand i underkäken och en knäled på godtycklig sida använts. I artiklar som kopplar ålder till mognadsstadium för knäled eller visdomstand finns rådata över åldrar *inte* presenterade. Information om åldrar är istället presenterade som beskrivande statistik.

Från den beskrivande statistiken har åldersdata framställts genom stokastisk optimering. Den stokastiska optimeringsalgoritmen *Basin-hopping* valdes för implementeringen. Därefter har en probitmodell ställts upp med användning av den framställda datan, varefter maximum likelihood estimation används för att hitta de modellparametrar som bäst beskriver probitmodellen. De framtagna modellparametrarna beskriver logistiska avbildningar. Från avbildningarna hämtas därefter sannolikheterna att tillhöra ett visst mognadsstadium givet den framställda åldersdatan. De här sannolikheterna beskriver *likelihood*.

Vidare har en apriorifördelning (*prior*) tagits fram för att motsvara åldersfördelningen på de som blir åldersbedömda av Migrationsverket. Den har utgått från antalet asylsökande, med tandstadie H på vänster eller höger visdomstand i underkäken, som åldersbedömdes år 2017. Därefter har antalet korrigerats med hänsyn till de fall då man har olika tandstadier på vänster och höger visdomstand. Framtagen prior beskrivs enligt $\text{Ålder} \sim \text{Gamma}(k = 2.4, \theta = 2.14) + 15$, trunkerad vid $\text{Ålder} = 30$ år.

Givet åldershypoteserna $H_0 = \{\text{ålder över 18 år}\}$ och $H_1 = \{\text{ålder under 18 år}\}$ ställs därefter likelihood ratio respektive prior odds upp varefter posteriori odds tas fram med Bayes teorem. Hypoteserna H_0 och H_1 har vardera en sannolikhet för att vara sann. Från posteriori oddset kan varje sannolikhet för att en hypotes stämmer tas fram.

Tröskelvärden, eller *cut-off*, för sannolikheterna har ansatts för att kunna testa metoden. Om en asylsökandes sannolikhet att vara över 18 år är större än dess tröskelvärde, bedöms den asylsökande till att vara över 18 år. Motsatt fall råder om sannolikheten att vara över 18 år är mindre än dess tröskelvärde.

Resultatet är ett flertal tabeller, baserade på olika publikationer, som kan användas för att åldersbedöma en asylsökande efter ett ansatt tröskelvärde.

Slutsatsen är att metoden kan tillämpas. Metoden ger ett explicit resultat i form av sannolikheter vilket kan utvidgas om ny forskning publiceras. Beslutstabellerna som har tagits fram kan användas som underlag för bedömning.

Abstract

This bachelor's thesis has developed a method to conduct age assesment based on Bayesian inference. The chosen age indicators are development stages of the lower left and right third molars as well as a knee joint on either side. There have been several publications dedicated to establish a connection between development stages of knee joints or wisdom teeth, with age. However the raw data used in such papers are only presented as descriptive statistics.

Using a stochastic optimization algorithm called *Basin-hopping*, a data set of ages fitting the descriptive statistics has been acquired. These ages are then used as input to a probit model, where maximum likelihood estimation has been utilized to find the model parameters which best describe the probit model itself. The obtained model parameters describe logistic mappings. From these mappings, the probabilities of having a certain development stage given the acquired age data are retrieved. These probabilities describe the *likelihood*.

Moreover, an apriori distribution (prior) has been produced to correspond to the age distribution of the individuals whose age are being assessed by the Swedish Migration Agency. It is based on the number of asylum seekers, having development stage H on any of the lower third molars, who were age assessed during the year 2017. Subsequently, the prior was corrected by taking into consideration the cases where the development stages on the right and left third molars are different. Thus yielding the prior described as the distribution $\text{Age} \sim \text{Gamma}(k = 2.4, \theta = 2.14) + 15$ truncated at 30 years of age.

Given age hypothesis $H_0 = \{\text{age above 18 years}\}$ and $H_1 = \{\text{age under 18 years}\}$ the likelihood ratio as well as the prior odds are computed. Consequently, the posteriori odds is found by an application of Bayes theorem. The hypothesis H_0 and H_1 each have a probability of being correct. The probabilities that each hypothesis is correct are obtained from the posterior odds.

Cut-offs have been set to test the method. If the probability of an asylum seeker being 18 years old is greater than the specified cut-off, the asylum seeker is determined to be over 18 years old. Similarly, should the probability be less than the cut-off, the asylum seeker is determined to be under 18 years old.

The result is a collection of tables, based on different papers, that can be used in age assessment of asylum seekers given a specified cut-off.

The conclusion is that the method is applicable. The method gives an explicit result in the shape of probabilities and can be expanded if new research is published. The produced decision tables can be used as a support to make decision of age assessment.

Innehåll

1	Inledning	1
1.1	Syfte	2
1.2	Problemformulering	2
1.3	Avgränsningar	2
1.4	Etik	3
2	Teori	4
2.1	Bayesiansk inferens	4
2.2	Statistisk beslutsteori	5
2.3	Likelihood	6
2.3.1	Markov chain Monte Carlo (MCMC)	6
2.3.2	Basin-hopping	7
2.3.3	Probitmodellen	8
3	Metod	9
3.1	Informationsextrahering och framställning av åldrar från artiklar	9
3.1.1	R:s definition av percentiler	9
3.1.2	Några exempel	10
3.1.3	Basin-hopping	10
3.2	Likelihoodfunktionen	11
3.3	Prior	11
3.3.1	Anpassning av data från RMV för att konstruera en prior	12
3.3.2	Kriterier för prior	12
3.4	Posteriori	13
3.5	Beslutsteori	13
4	Resultat	14
4.1	Basin-hopping och fördelningar från probit	14
4.2	Prior	16
4.2.1	Jämförelse av olika priors	16
4.3	Beslut från posteriorin	17
5	Diskussion	19
5.1	Likelihood	19
5.1.1	Basin-hopping	19
5.1.2	Fördelning från probitmodellen	20
5.2	Prior	20
5.3	Posterioriodds och beslut	21
5.3.1	Skillnad mellan sidor för visdomstånd	22
5.3.2	Skala för knäledsmognad	22
5.3.3	Jämförelse mellan tanddata	22
6	Slutsats	23

A	Teori	26
A.1	Fundamentala teorem inom statistik och sannolikhetslära	26
A.2	Markovkedjor	26
A.3	Monte Carlo-metoden	27
A.4	Några relevanta fördelningsfunktioner	27
A.4.1	Standardnormalfördelningen	27
A.4.2	Gammafördelningen	28
A.4.3	Bernouillifördelningen	28
A.4.4	Kontinuerlig likformig fördelning	28
B	Resultat	29
B.1	Beslut	29
B.2	Andra resultat från probitmodellen	32
C	Mognadsstadier	35
C.1	Demirjians metod	35
C.2	Krämers definition	36
C.3	Vieths skala	37

Förord

Den här rapporten är produkten av ett kandidatprojekt skriven under våren 2018 på Matematiska vetenskaper. Projektet har innefattat fyra elever från civilingenjörsprogrammet Teknisk matematik på Chalmers tekniska högskola, samt två elever från kandidatprogrammet Matematisk statistik på Göteborgs universitet.

Vår handledare Petter Mostad vill vi speciellt tacka för att ensam ha handlett oss i det här projektet.

Arbetet är en gruppinsats och varje medlem inom projektet har bidragit till hela rapporten genom skrivning, korrekturläsning och revidering. På grund av storleken på arbetet har alla medverkande bidragit på samtliga delar av rapporten beroende på behov som har uppstått.

Individuella tidsloggar har förts för varje medverkande i arbetet i form av ett excelark. Där presenteras vad varje enskild medlem har loggat sina timmar på.

Varje avsnitt har en eller flera huvudansvariga. De huvudansvariga kontrollerar innehåll, struktur samt språk för avsnittet de har blivit tilldelade. Sedan återges feedback till de som bidrar till det berörda avsnittet. Nedan presenteras de huvudansvariga för respektive avsnitt av rapporten:

Populärvetenskaplig presentation - Jesper Olsson & Erik Sörstadius

Sammanfattning - Pontus Larsson

Abstract - James Pålsson

Förord - James Pålsson

1. Inledning - Pontus Larsson & James Pålsson
2. Teori - Pontus Larsson & James Pålsson
3. Metod - Viktor Erbro & Jesper Olsson
4. Resultat - Viktor Erbro & Erik Sörstadius
5. Diskussion - Henrik Esmaili & Jesper Olsson
6. Slutsats - Henrik Esmaili & Jesper Olsson
7. Bilaga A - Viktor Erbro & Pontus Larsson
8. Bilaga B - Viktor Erbro & Erik Sörstadius
9. Bilaga C - Henrik Esmaili & James Pålsson

Kapitel 1

Inledning

I maj 2016 gav regeringen *Rättsmedicinalverket* (RMV) uppdraget att genomföra medicinska åldersbedömningar¹ på personer som söker uppehållstillstånd [30]. Ett halvår senare, i november 2016, redovisade RMV hur asylsökandes åldrar bör bedömas. Barn har annorlunda rättigheter än vad vuxna har, exempelvis skiljer sig straff för brott beroende på om den skyldige är över eller under 18 år. I fallet för asylsökande, ska barn behandlas olika enligt internationella konventioner [19]. Det är därför nödvändigt att göra åldersbedömning av asylsökande för att beskydda barnens rättigheter. Det leder till att specifikation av ålder vid asylsökningprocessen är viktig. Metoden RMV utformade har på senare tid blivit ifrågasatt av experter inom olika områden, däribland statistiker och läkare, på grund av bristande vetenskaplig korrekthet och exakthet [10][18][20].

Deras åldersbedömningstest består av analyser av MR-bilder² på knäleder och röntgenbilder av visdomständer. Tandläkare och radiologer gör sedan en avvägning kring huruvida tanden och/eller knäleden har uppnått ett mognadsstadium motsvarande en vuxen individ. Det mogna visdomstandstadiet är steg *H* enligt Demirjians metod³ [2]. För moget knästadium gäller steg 4 till 5 enligt Krämers definition⁴ [14]. Resultaten av bedömningarna från bilderna evalueras med beslutsmatrisen nedan (tabell 1.1), varefter ett beslut fattas [31]. Bedömningsmatrisen är framställd och måtten i den är fördefinierade av Rättsmedicinalverket.

Tabell 1.1: *Talar för* är ett mått som säger att utgående från populationsnivån är det betydligt mer sannolikt att personen är över/under 18 år. *Talar möjligen för* är ett mått som säger att utgående från populationsnivån är det sannolikt att personen är över/under 18 år. Fall där knäled eller visdomstand inte har nått fullmognad specificeras nedan som —. Endast de tre översta raderna leder till att den asylsökande registreras som över 18 år.

Resultat		Bedömning
Knäled	Visdomstand	
Fullmogen	Fullmogen	<i>Talar för att individen är 18 år eller äldre</i>
—	Fullmogen	<i>Talar för att pojken är 18 år eller äldre</i>
Fullmogen	—	<i>Talar för att pojken är 18 år eller äldre</i>
—	—	<i>Talar möjligen för att pojken är under 18 år</i>
—	Fullmogen	<i>Ingen bedömning avseende flickans ålder relativt 18-årsgränsen kan göras</i>
Fullmogen	—	<i>Ingen bedömning avseende flickans ålder relativt 18-årsgränsen kan göras</i>
—	—	<i>Talar för att flickan är under 18 år</i>

¹ Medicinsk åldersbedömning i den här rapporten kommer benämnas som åldersbedömning, men observera att de här två begreppen generellt sett är olika saker.

² MR, eller Magnetic Resonance imaging, är ett sätt att använda magnetvägor för att producera detaljerade bilder av kroppens interiör.

³ En skala från *A* till *H*, där *H* representerar fullmogen visdomstand. Se bilaga C.1.

⁴ En skala från 1 till 5, där stigande värden representerar en högre knäledsmognad. Se bilaga C.2.

1.1 Syfte

Syftet med arbetet är att ta fram en beslutsmetod för att åldersbedöma asylsökande i Sverige till över eller under 18 år. Tillvägagångssättet för det här projektet är bayesiansk statistik med underlag från medicinska åldersbedömningar.

1.2 Problemformulering

Målet är att bygga en ny statistisk beslutsmodell, utgående från bayesiansk inferens och beslutsteori, som ska användas för att åldersbedöma asylsökande till över eller under 18 år. Modellen ska fatta beslut givet bedömningar från MR-bild av knä, bedömningar från röntgen av visdomständer i underkäken, samt kön på individen. Motiveringen är att modellen skall utgå från liknande förutsättningar som RMV valt [31]. I mars 2018 togs en alternativ skala, Vieths skala⁵, fram för knäled vilken kommer jämföras med Krämers definition. RMV:s individuella bedömningar saknar faktisk ålder på den asylsökande; den är inte offentlig och är dessutom inte sparad i en form som kan användas av det här projektet. Det här medför att bedömningar som relaterar ålder till mognadsstadium inte finns [34].

Sådan information kommer istället hämtas från publicerade tidsskrifter som behandlat frågan om ålder kopplad till mognad av antingen knäled [23][36] eller visdomständer [5][6][24]. Datan till tidsskrifterna är inte heller offentliga och således finns inget dataset som explicit kopplar ålder till mognadsstadium tillgängligt för det här projektet. Däremot finns beskrivande statistik (descriptive statistics) som i det här projektet kommer hänvisa till antalet observationer n , medelvärde, standardavvikelse, minsta och största ålder samt övriga percentiler angivna i studien. Lösningen är att framställa ett dataset som följer den beskrivande statistiken ifråga. Den nyfikne läsaren kan se hur det här går till i kapitel 3.

Projektet delas in i fyra delproblem som formuleras nedan.

- Bayes
 - Bayesiansk inferens kommer utgöra grunden för arbetet. Därför kommer implementation av algoritmer och approximationer vara av bayesiansk karaktär.
- Likelihood och försök till återskapande av data
 - Dataset av åldrar kommer framställas från beskrivande statistik skildrad i vetenskapliga publikationer. Publikationerna har för avsikt att koppla ålder till mognadsstadium för någon av de tre skalorna. Därefter kommer logistiska avbildningar bestämmas från de framställda datasetten.
- Val av apriorifördelning (prior)
 - En prior ska tas fram för att motsvara åldersfördelningen på de som blir åldersbedömda av Migrationsverket.
- Beslut
 - När delproblemen ovan är lösta kan sannolikheten att en asylsökande är över 18 år tas fram. Givet ett ansatt tröskelvärde, som beskriver hur stor sannolikheten måste vara, kan ett beslut hurvida individen är över eller under 18 år fattas. När ett beslut kan fattas uppnås projektets syfte.

1.3 Avgränsningar

Följande kommer inte behandlas inom ramarna för projektet.

- I beslutssteget vid åldersbedömningen kommer två konkurrerande beslut, *bedömas som över 18 år* och *bedömas som under 18 år*, vägas mot varandra. Det innebär att en kostnad⁶ tillsätts respektive beslut. I projektet kommer exempelkostnader sättas på besluten. Det är enbart för att kunna köra numeriska implementationer.

⁵En skala från 1 till 6, där stigande värden representerar en högre knäledsmognad. Se bilaga C.3.

⁶För definition av vad kostnader är, se avsnitt 2.2.

- Vid åldersbedömning tittar RMV enbart på mognadsgrad i knäled och visdomständer. Samma avgränsning kommer användas för det här projektet.
- Tillgång till de individuella beslutsunderlagen RMV använder sig av saknas. Därför kommer bara vissa typer av jämförelser göras. Exempelvis hur vår metod skulle döma en fiktiv individ i relation till hur RMV:s metod skulle bedöma.
- Modellen som framställs kommer endast bygga på bayesiansk beslutsteori och kommer inte behandla det frekventistiska perspektivet.
- Endast individer med åldersindikatorer får bedömning. Minimumkravet för åldersbedömning med beslutsmodellen är ett fastställt mognadstadium för en visdomstand, alternativt knäled.
- Information kommer hämtas från publicerade tidsskrifter som behandlat frågan om ålder kopplad till mognad av knäled och visdomständer. Det förekommer att information som motsvarar flickor och pojkar med ålder långt ifrån 18 år (för unga eller gamla) inte finns tillgängligt. Underlag saknas för steg 1 och 5 i Krämers definition samt steg 1 i Vieths skala. Sannolikheten för en flicka eller en pojke att vara över 18 år när man har stadiet 1 med båda skalorna är så liten att den är försumbar. På samma sätt gäller det omvända, att om en flicka eller en pojke har det sista stadiet i någon knäledmognadsstadieskala, är sannolikheten att individen är under 18 år då återigen försumbar.

1.4 Etik

Etik är en viktig aspekt att beakta under projektets gång. Viktig information för projektet hämtas från publicerade tidsskrifter som behandlar frågan om ålder kopplad till mognad av antingen knäled eller visdomständer. Modellen som konstrueras i projektet använder mognadsstadier på knäleder och visdomständer för att åldersbedöma. Genom reverse-engineering är det möjligt att göra det omvända. Det vill säga att om en persons ålder är känd, så kan man uppskatta vad den personen har för knäled- och visdomstandsmognadsstadium. Därför är det viktigt att informationen om personerna i studierna är, och fortsätter vara, anonyma. Information om ålder och mognadsgrad ska exempelvis inte kunna kopplas till personnummer.

En annan viktig fråga är tillämpningen av resultatet. Under förutsättning att den i projektet framtagna metoden används vid fall av åldersbedömning kan individer komma att påverkas direkt såväl som indirekt. Direkta konsekvenser skulle kunna vara att premisserna för uppehållstillstånd förändras, antingen till det positiva eller negativa för den asylsökande. Det finns en rad indirekta konsekvenser som skulle kunna ske som en följd av någon direkt konsekvens. Exempelvis leder ökad asyl till en ekonomisk konsekvens för staten eftersom asylprocessen finansieras av den. Kostnaden för asylprocessen har varit ett ämne som debatterats frekvent och besluten som fattas av metoden kan påverka opinionen indirekt på det sättet.

För tillfället finns inte mycket forskning som rör kronologisk åldersbedömning av visdomständer i konjunktion med magnetrontgen av knäleder. En nytta med det här projektet är att det kan agera som en inspiration för vidare forskning av sådan karaktär. Projektresultatet kommer vara rent statistiskt och tar ingen ställning i etiska frågor. Modellen skulle dock mycket väl kunna användas i situationer där den har en inverkan i beslut av etisk karaktär.

Kapitel 2

Teori

Metoddelen i rapporten använder sig av tillämpad matematik, matematisk optimering och matematisk statistik. Beskrivet i problemformuleringen används det bayesiansk statistik samt algoritmer för framställning av dataset. I det här kapitlet presenteras den matematiska teorin bakom åldersbedömningen som kommer användas både i metoden och resultatet.

2.1 Bayesiansk inferens

Idén bakom bayesiansk inferens är att gå från en ursprunglig fördelning (prior) för en hypotes, H , till en uppdaterad fördelning för H där hänsyn har tagits till evidens. Den uppdaterade fördelningen kallas för posteriorifördelning, vars framtagande vilar på Bayes teorem [4].

Teorem 1 (Bayes teorem). *Låt A och B vara två händelser i utfallsrummet Ω . För $\mathbb{P}(B) > 0$ erhålls för den betingade sannolikheten $\mathbb{P}(A|B)$ att*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \quad (2.1)$$

Bevis. Betingad sannolikhet, som ett axiom inom sannolikhetsläran [4], är definierad enligt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B).$$

Vidare är snitt en kommuterande operator, det vill säga $A \cap B = B \cap A$, vilket medför

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(B|A)\mathbb{P}(A).$$

Enligt antagande är $\mathbb{P}(B) > 0$, varefter beviset följer efter division med $\mathbb{P}(B)$. □

För en godtycklig hypotes H och given Data, erhålls därför från Bayes teorem

$$\mathbb{P}(H|\text{Data}) = \frac{\mathbb{P}(\text{Data}|H)\mathbb{P}(H)}{\mathbb{P}(\text{Data})},$$

under förutsättning att $\mathbb{P}(\text{Data}) > 0$.

Givet två hypoteser, H_0 och H_1 , kan därefter ett oddsförhållande ställas upp enligt

$$\frac{\mathbb{P}(H_0|\text{Data})}{\mathbb{P}(H_1|\text{Data})} = \frac{\mathbb{P}(\text{Data}|H_0)}{\mathbb{P}(\text{Data}|H_1)} \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}, \quad (2.2)$$

där $\mathbb{P}(H_1|\text{Data})$, $\mathbb{P}(\text{Data}|H_1)$ och $\mathbb{P}(H_1)$ alla är skilda från noll. Fördelen med att ställa upp det på den här formen är att sannolikheten $\mathbb{P}(\text{Data})$ har eliminerats och behöver inte beräknas.

Vidare kan snittet av en familj av data, $\text{Data}_{\text{total}} = \cap_{i=1}^n \text{Data}_i$, beskrivas på den slutliga formen

$$\underbrace{\frac{\mathbb{P}(H_0|\text{Data}_{\text{total}})}{\mathbb{P}(H_1|\text{Data}_{\text{total}})}}_{\text{Posteriori odds}} = \underbrace{\prod_{i=1}^n \frac{\mathbb{P}(\text{Data}_i|H_0)}{\mathbb{P}(\text{Data}_i|H_1)}}_{\text{Likelihood ratio}} \cdot \underbrace{\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}}_{\text{Prior odds}}, \quad (2.3)$$

givet betingat oberoende observationer Data_i .

Prior odds är det ursprungliga oddset, det vill säga oddset utan evidens, och anger hur troligt det är att respektive hypotes dominerar över den andra. För en stor kvot gäller att sannolikheten att observera H_0 är större än sannolikheten att observera H_1 , och vice versa gäller för en liten kvot. Likelihood ratio är förhållandet mellan att datan stämmer givet hypotesen H_0 , och att datan stämmer givet den andra hypotesen H_1 . Produkten av likelihood ratio och prior odds bildar posteriori odds som är det uppdaterade oddset, det vill säga oddset med evidens. Det anger hur troligt det är att respektive hypotes dominerar över den andra givet datan.

2.2 Statistisk beslutsteori

När ett oddsförhållande mellan två hypoteser har tagits fram kan ett beslut över vilken hypotes som är värd att acceptera fattas. Frågan är; hur tar man ett sådant beslut? Svaret motiveras av en tvärvetenskaplig disciplin som kallas för (statistisk) beslutsteori [25].

Syftet med statistisk beslutsteori är att ta fram det optimala valet, givet ett ändligt urval att välja från. Betrakta scenariot att det finns två tillstånd i världen för en godtycklig individ enligt,

$$H_0 = \{\text{ålder över 18 år}\} \quad \text{och} \quad H_1 = \{\text{ålder under 18 år}\}.$$

Som beslutstagare i det här projektet är uppgiften att ta ett beslut för om individen i fråga är över eller under 18 år. För att underlätta beslutet så införs kostnadsvikter c_1 , c_2 , c_3 och c_4 , som svarar mot att en individ som är över 18 år bedöms som över 18 år, en individ som är över 18 år bedöms som under 18 år, en individ som är under 18 år bedöms som över 18 år samt att en individ som är under 18 år bedöms som under 18 år. Se tabell 2.1. Med kostnader menas alla möjliga fördelar och nackdelar som en individ, en grupp eller ett samhälle erfarar vid något beslut, översatt till någon gemensam monetär skala.

Tabell 2.1: Kostnader för att bedöma en individ som över eller under 18 år, givet tillstånden H_0 och H_1 .

Tillstånd	Bedöm som över 18 år	Bedöm som under 18 år
H_0	c_1	c_2
H_1	c_3	c_4

Från tabell 2.1 erhålls att följande kostnader är att förvänta om individen antingen bedöms att vara över eller under 18 år.

$$\mathbb{E}[\text{Bedöm som över 18 år}] = c_1\mathbb{P}(H_0) + c_3\mathbb{P}(H_1),$$

$$\mathbb{E}[\text{Bedöm som under 18 år}] = c_2\mathbb{P}(H_0) + c_4\mathbb{P}(H_1),$$

där $\mathbb{P}(H_0)$ är sannolikheten att individen är över 18 år och $\mathbb{P}(H_1)$ är sannolikheten att individen är under 18 år.

Kostnaden vid optimal bedömning är den som ger lägst kostnad, vilket medför

$$\begin{aligned} \mathbb{E}[\text{Optimal bedömning}] &= \min\left(\mathbb{E}[\text{Bedöm som över 18 år}], \mathbb{E}[\text{Bedöm som under 18 år}]\right) \\ &= \min\left(c_1\mathbb{P}(H_0) + c_3\mathbb{P}(H_1), c_2\mathbb{P}(H_0) + c_4\mathbb{P}(H_1)\right). \end{aligned}$$

Individen bedöms därför som över 18 år om

$$c_2\mathbb{P}(H_0) + c_4\mathbb{P}(H_1) > c_1\mathbb{P}(H_0) + c_3\mathbb{P}(H_1),$$

$$c_2 \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} + c_4 > c_1 \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} + c_3,$$

$$(c_2 - c_1) \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} > c_3 - c_4,$$

vilket medför att

$$\begin{aligned} \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} &> \frac{c_3 - c_4}{c_2 - c_1}, \quad c_2 - c_1 > 0, \\ \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} &< \frac{c_3 - c_4}{c_2 - c_1}, \quad c_2 - c_1 < 0, \end{aligned} \tag{2.4}$$

annars bedöms individen som under 18 år.

2.3 Likelihood

Från publikationer som har utforskat hur mognadsstadiernas åldersfördelningar ser ut, finns varken fördelningarna eller den bakomliggande åldersdatan att tillgå. Däremot finns beskrivande statistik för åldersdatan publicerad. I följande avsnitt beskrivs matematiken för att metoderna beskrivna i avsnitt 3.1 och 3.2 skall kunna implementeras.

Först presenteras teorin bakom de framställda datasetten. Ett dataset som svarar mot den beskrivande statistiken för åldersdatan tas fram med användning av den stokastiska optimeringsalgoritmen *Basin-hopping*, en metod som tillämpar minimering tillsammans med MCMC-metoden *Metropolis-Hastings*.

Därefter presenteras teorin bakom framtagning av åldersfördelningarna hos mognadsstadierna. De tas fram genom att använda de framställda datasetten som observerade parametrar till *probitmodellen*.

2.3.1 Markov chain Monte Carlo (MCMC)

Markov chain Monte Carlo-metoder, eller MCMC-metoder, är ett paraplybegrepp för alla simuleringsmetoder av en täthetsfunktion $f(y)$ som producerar ergodiska Markovkedjor $\{X^{(t)}\}_{t \geq 0}$ vars stationära fördelning är $f(y)$ [27]. Se avsnitt A.2 och A.3 i bilaga A för vidare läsning om Markovkedjor och Monte Carlo-metoden. Ett exempel på en MCMC-metod är Metropolis-Hastings-algoritmen.

Metropolis-Hastings

Målet med Metropolis-Hastings-algoritmen är att simulera från den (okända) täthetsfunktionen $f(y)$, med användning av en betingad täthet $q(y|x)$ som är känd så när som på en konstant.

$q(y|x)$ kallas för en *föreslagen täthetsfunktion* och kan väljas nästintill godtyckligt. Det enda kravet är att förhållandet $f(y)/q(y|x)$ måste vara känt upp till en konstant oberoende av x .

Implementationen av Metropolis-Hastings ser ut enligt följande [28].

Givet vektorn $x^{(t)}$ sådan att $X^{(t)} = x^{(t)}$

1. Generera vektorn $Y_t \sim q(y|x^{(t)})$.
2. Ta

$$X^{(t+1)} = \begin{cases} Y_t, & \rho(x^{(t)}, Y_t), \\ x^{(t)}, & 1 - \rho(x^{(t)}, Y_t), \end{cases}$$

där

$$\rho(x, y) = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\}, \quad (2.5)$$

är acceptanssannolikheten att acceptera värdet Y_t (Metropolis-Hastings-kriteriet).

Vad som återstår att visa är att Metropolis-Hastings-algoritmen producerar en kedja som har $f(y)$ som stationär fördelning.

Låt A beteckna händelsen att nästa värde i Metropolis-Hastings-algoritmen accepteras och \mathcal{X} definiera händelsen som beskriver alla möjliga vektorer $X^{(t+1)}$ kan anta. Vidare betecknar $\Omega = A \cup A^c$ händelsen att nästa värde accepteras eller inte accepteras. Då gäller för nästa tidssteg $X^{(t+1)}$ att

$$\begin{aligned} \mathbb{P}(X^{(t+1)} \in \mathcal{X} | X^{(t)} = x^{(t)}) &= \mathbb{P}(X^{(t+1)} \in \mathcal{X}, \Omega | X^{(t)} = x^{(t)}) = \mathbb{P}(X^{(t+1)} \in \mathcal{X}, (A \cup A^c) | X^{(t)} = x^{(t)}) \\ &= \mathbb{P}(X^{(t+1)} \in \mathcal{X}, A | X^{(t)} = x^{(t)}) + \mathbb{P}(X^{(t+1)} \in \mathcal{X}, A^c | X^{(t)} = x^{(t)}) \\ &= \int_{\mathcal{X}} \rho(x^{(t)}, y) q(y|x^{(t)}) dy + \mathbf{1}_{\mathcal{X}}(x^{(t)}) \mathbb{P}(A^c | X^{(t)} = x^{(t)}). \end{aligned}$$

Vidare gäller att

$$\mathbb{P}(A^c | X^{(t)} = x^{(t)}) = 1 - \underbrace{\int_{\mathcal{X}} \rho(x^{(t)}, y) q(y|x^{(t)}) dy}_{:= r(x^{(t)})}, \quad \mathbf{1}_{\mathcal{X}}(x^{(t)}) = \int_{\mathcal{X}} \delta_{x^{(t)}}(y) dy,$$

så att

$$\mathbb{P}(X^{(t+1)} \in \mathcal{X} | X^{(t)} = x^{(t)}) = \int_{\mathcal{X}} \underbrace{\rho(x^{(t)}, y)q(y|x^{(t)}) + r(x^{(t)})\delta_{x^{(t)}}(y)}_{=K(x^{(t)},y)} dy = 1.$$

Kärnan till Markovkedjan är således

$$K(x, y) = \rho(x, y)q(y|x) + r(x)\delta_x(y), \quad (2.6)$$

där $r(x) = 1 - \int_{\mathcal{X}} \rho(x, y)q(y|x) dy$ och δ_x är Diracmättet i x .

Från (2.5) erhålls efter multiplikation med $f(x)q(y|x)$

$$\begin{aligned} f(x)q(y|x)\rho(x, y) &= \min \{f(y)q(x|y), f(x)q(y|x)\} \\ &= f(y)q(x|y) \min \left\{ 1, \frac{f(x)q(y|x)}{f(y)q(x|y)} \right\} \\ &= f(y)q(x|y)\rho(y, x). \end{aligned}$$

Addition av $r(x)\delta_x(y)f(x)$ till båda led ger

$$f(x)K(x, y) = f(y)K(y, x),$$

eftersom $r(x)\delta_x(y)f(x) = r(y)\delta_y(x)f(y)$. Alltså uppfyller Metropolis-Hastings-algoritmen *the detailed balance condition* [28]. Det följer att

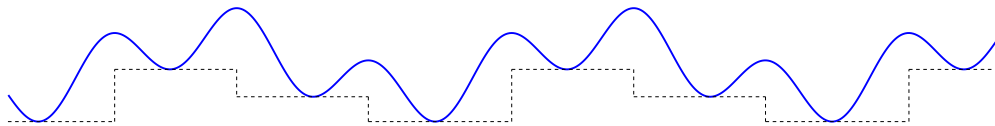
$$\int_{\mathcal{X}} K(x, y)f(x) dx = \int_{\mathcal{X}} K(y, x)f(y) dx = f(y) \underbrace{\int_{\mathcal{X}} K(y, x) dx}_{=1} = f(y),$$

vilket både innebär att $f(y)$ är den stationära fördelningen till kedjan $\{X^{(t)}\}_{t \geq 0}$, men också att så fort en vektor har simulerats från den stationära fördelningen så kommer även alla kommande vektorer simuleras från den [3].

2.3.2 Basin-hopping

Basin-hopping eller iterative local search är en stokastisk optimeringsalgoritm som tillämpar Metropolis-Hastings-algoritmen. Den uppkom i samband med att man ville hitta den lägsta energikonfigurationen i ett molekylärsystem [37].

Algoritmens syfte är att den skall finna ett globalt minimum genom att försöka dela in målfunktionen $E(x)$ så att lokala minimum delas in i områden, så kallade bassänger (basins). När ett lokalt minimum har hittats, vilket görs med en egen ansatt minimeringsmetod, så försöker algoritmen leta sig (hoppa) till en ny bassäng [16]. Se figur 2.1.



Figur 2.1: Exempel på hur Basin-hopping-algoritmen hade försökt dela in funktionen $\sin(x) + \cos(3x)$ i bassänger, för $x \in [-5\pi/2, 5\pi/2]$.

Hoppen är i själva verket en perturbation som tilläggs det befintliga minimat x_{old} . Perturbationerna görs stokastiskt med $\eta \sim q(y|x_{\text{old}}) = \text{Unif}(x_{\text{old}} - \Delta, x_{\text{old}} + \Delta)$, där Δ är den maximala steglängden ifrån vektorn x_{old} . Δ är en inparameter till algoritmen, varefter algoritmen självt anpassar maximala steglängden efter varje iteration [1]. Från den nya vektorn $x_{\text{old}} + \eta$ påbörjas därefter en minimering av målfunktionen.

Vidare accepteras vektorn associerad med det nya minimat, x_{new} , baserat på tillämpning av Metropolis-Hastings-kriteriet (2.5) på $E(x_{\text{new}})$ och $E(x_{\text{old}})$. Täthetsfunktionen $f(y)$ följer en simplifierad Boltzmann-fördelning enligt

$$f(y) = \exp\left(-\frac{y}{T}\right). \quad (2.7)$$

Att täthetsfunktionen $f(y)$ har det här utseendet är varför metoden har funnit sina tillämpningar inom kemifysik. Läsaren som kan sin statistiska mekanik känner igen Boltzmannfördelningen som en sannolikhetsfördelning att ett system ska befinna sig i ett visst tillstånd som funktion av systemets energi och temperatur, T . I vårt fall är T en godtycklig konstant.

Från (2.5) erhålls att acceptanssannolikheten för x_{new} blir

$$\rho(E(x_{\text{old}}), E(x_{\text{new}})) = \min \left\{ \exp \left(-\frac{(E(x_{\text{new}}) - E(x_{\text{old}}))}{T} \right), 1 \right\}.$$

Alltså accepteras den nya vektorn x_{new} med sannolikheter

$$\begin{cases} 1, & E(x_{\text{new}}) < E(x_{\text{old}}), \\ \exp \left(-\frac{(E(x_{\text{new}}) - E(x_{\text{old}}))}{T} \right), & E(x_{\text{new}}) \geq E(x_{\text{old}}), \end{cases} \quad (2.8)$$

Skulle den nya vektorn inte accepteras, så initieras istället ett nytt hopp. Vid färdig körning returneras vektorn som ger det minsta målfunktionsvärdet.

2.3.3 Probitmodellen

Probitmodellen är en regressionsmodell där responsvariabeln y_i är binär, det vill säga den kan endast anta värden $\{0, 1\}$. Den beskrivs

$$\begin{aligned} \mathbb{P}(y_i = 1|x_i) &= \Phi(x_i\beta), \\ \mathbb{P}(y_i = 0|x_i) &= 1 - \Phi(x_i\beta), \end{aligned} \quad (2.9)$$

där Φ är den kumulativa standardnormalfördelningen vars definition kan ses i avsnitt A.4, β är en vektor med modellparametrar och observationen $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ är en vektor med oberoende och likafördelade förklarande variabler [17].

För att hitta den mest lämpade probitmodellen används maximum likelihood estimation, en metod som baserat på likelihood optimerar parametrar till en fördelning givet observerad data som följer fördelningen [8]. De optimala parametrarna är de som returnerar maximal likelihood, vilket innebär att de är parametrarna som gör datan mest trolig att observera.

Från (2.9) erhålls att sannolikhetsfunktionen för probitmodellen är $f(y_i; p) = p^{y_i}(1-p)^{1-y_i}$ för $y_i \in \{0, 1\}$ och $p = \Phi(x_i\beta)$. Likelihoodfunktionen av β givet $y_i \in \{0, 1\}$ och observationen x_i beskrivs därför som

$$\mathcal{L}(\beta; y_i, x_i) = [\Phi(x_i\beta)]^{y_i} [1 - \Phi(x_i\beta)]^{1-y_i}.$$

Till följd, då de förklarande variablerna är oberoende och likafördelade, gäller för ett stickprov av observationer och en samling av responsvariabler att

$$\mathcal{L}(\beta; y, X) = \prod_{i=1}^N [\Phi(x_i\beta)]^{y_i} [1 - \Phi(x_i\beta)]^{1-y_i},$$

där X är stickprovet av storlek N och y är vektorn med responsvariabler för respektive observation. Genom att ta logaritmen över likelihoodfunktionen erhålls log-likelihoodfunktionen

$$l(\beta; y, X) = \sum_{i=1}^N y_i \ln [\Phi(x_i\beta)] + (1 - y_i) \ln [1 - \Phi(x_i\beta)]. \quad (2.10)$$

Eftersom att logaritmen är en strikt växande funktion gäller att det räcker att maximera log-likelihoodfunktionen för att maximera likelihoodfunktionen. Därför erhålls att de optimala modellparametrarna $\hat{\beta}$ är

$$\hat{\beta} = \arg \max_{\beta} l(\beta; y, X). \quad (2.11)$$

Kapitel 3

Metod

I metodkapitlet implementeras den matematiska teori bakom åldersbedömningen som presenterades i kapitel 2. Metoddelen är uppdelad i fyra delar. Den första delen leder fram till den stokastiska optimeringsalgoritmen Basin-hopping och bakgrund om percentiler som är nödvändig för att förstå optimeringsproblemet. Algoritmen används för att framställa åldrar på medverkande personer i undersökningar som kopplat ålder till mognadsstadium av antingen knäled eller visdomständer i underkäken. Andra delen beskriver hur de framställda åldrarna används för att skapa kumulativa fördelningar med hjälp av likelihoodfunktionen enligt probitmodellen. I den tredje delen beskrivs hur den slutgiltiga priorin genereras. Avslutningsvis presenteras hur posteriorin beräknas och hur beslut om en asylsökande är över eller under 18 år tas fram.

3.1 Informationsextrahering och framställning av åldrar från artiklar

I publicerade artiklar har åldersfördelningar för olika mognadsstadier presenterats i form av beskrivande statistik. Beskrivande statistik kommer i det här projektet hänvisa till antalet observationer n , medelvärde av stickprovet, standardavvikelse av stickprovet, minsta och största ålder samt övriga percentiler givna för åldersfördelningen i studien. I tabell 3.1 visas ett exempel på beskrivande statistik, i det här fallet för visdomständer på högersidan i överkäken, tandstadium B , flickor. Med parametrar menas populationsstatistikor i beskrivande statistik, förutom kolumnen för antalet, n . I tabell 3.1 är antalet parametrar 13: medelvärde, standardavvikelse och 11 percentiler (inklusive MIN och MAX).

Tabell 3.1: Beskrivande statistik för visdomständer på högersidan i överkäken, tandstadium B , flickor, Israel-datan.

n	MEAN	SD	MIN	0.5%ile	5%ile	10%ile	25%ile	50%ile	75%ile	90%ile	95%ile	99.5%ile	MAX
24	10.61	1.68	7.53	7.62	8.40	8.77	9.55	10.64	11.25	13.31	13.37	14.05	14.14

Målet är att framställa åldrarna på deltagarna i någon studie för något givet mognadsstadium och kön med hjälp av den beskrivande statistiken. Med åldrarna på deltagarna i studien menas åldrarna på deltagarna vid tidpunkten då studien gjordes.

3.1.1 R:s definition av percentiler

Låt x vara en sorterad vektor med åldrar av storlek n och låt p beteckna den p :te percentilen, det vill säga ett tal på intervallet $[0, 1]$. Åldern i den p :te percentilen definieras av kvantilfunktionen enligt

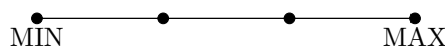
$$Q(x, p) = (1 - r)x_i + rx_{i+1}, \quad (3.1)$$

där i är indexeringen sådan att $i = \lfloor p(n - 1) + 1 \rfloor$ och $r = p(n - 1) + 1 - i$.

Notera att det här endast är en av definitionerna för kvantilfunktionen, och är ekvivalent med linjär interpolation mellan punkterna (p_j, x_j) , där $p_j = (j - 1)/(n - 1)$ för $j = 1, \dots, n$ [9]. Percentilerna i all beskrivande statistik från alla publikationer antas ha beräknats på det här sättet¹.

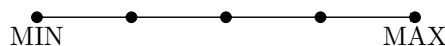
3.1.2 Några exempel

Nedan presenteras två exempel som visar att i många fall kommer det inte att gå att framställa en entydig uppsättning åldrar givet beskrivande statistik. Det första fallet som beskrivs kommer synliggöra när en entydig lösning existerar, medan det andra fallet illustrerar när fler än en lösning finns.



Figur 3.1: $n = 4$, MIN, MAX, medelvärde och standardavvikelse är givna.

Exempel 1. I figur 3.1 är $n = 4$. Låt den beskrivande statistiken i det här fallet enbart innehålla n , minsta ålder, största ålder, medelvärde och standardavvikelse. Den beskrivande statistiken används för att framställa åldrarna. Två av de fyra åldrarna är trivialt kända, den minsta respektive största åldern. Med endast medelvärdet kan inte de två resterande åldrarna bestämmas entydigt, eftersom systemet är överbestämt. Men eftersom även standardavvikelsen är given kan de två sista åldrarna bara placeras på ett sätt, och lösningen är entydig.



Figur 3.2: $n = 5$, MIN, MAX, medelvärde och standardavvikelse är givna.

Exempel 2. I figur 3.2 är $n = 5$. I övrigt skiljer sig inte den beskrivande statistiken från exemplet ovan. Precis som i det första exemplet är den minsta respektive största åldern trivialt kända. För att kunna bestämma de resterande åldrarna entydigt räcker det inte med att bara känna till medelvärdet och standardavvikelsen. Det betyder att de sista tre åldrarna kan bestämmas på många olika sätt och ändå passa in på den beskrivande statistiken. Det här systemet är överbestämt.

Generellt gäller att om stickprovsstorleken n är större än antalet parametrar, går det inte att fastställa en entydig lösning. När n är större än antalet parametrar går det fortfarande att få en lösning som uppfyller den beskrivande statistiken. Däremot går det inte att verifiera om de framställda åldrarna faktiskt är åldrarna på de som medverkade i studien. Värt att notera är att även om n är mindre än antalet parametrar är inte entydighet en nödvändighet, det beror på vilka referenspercentilerna är. I den beskrivande statistiken givet av Israel [5] och UK-Caucasian [6] är antalet parametrar 13. Därför, om $n > 13$, vilket är vanligt i de studierna, går det inte att framställa åldrarna entydigt.

3.1.3 Basin-hopping

Syftet med Basin-hopping i det här projektet är att bestämma en vektor med åldrar som uppfyller den beskrivande statistiken av mognadstadium för knäled och visdomständer.

Låt $\text{Ålder}_{\text{ref}}$ vara åldrarna för percentilerna i den beskrivande statistiken och x beteckna vektorn av framställda åldrar. Vidare kommer p beteckna en vektor med percentiler från den beskrivande statistiken. $Q(x, p)$ definierades i avsnitt 3.1.1 och är således åldrarna för percentilerna i x beräknade med hjälp av linjär interpolation enligt (3.1). Målfunktionen definieras som $E(x) = \text{Ålder}_{\text{ref}} - Q(x, p)$. Framställning av åldrar kan därför ställas upp som följande minimeringsproblem (MP).

¹Det här är ett antagande, eftersom det i praktiken är omöjligt att testa, i alla studier, om åldrarna på percentilerna verkligen har beräknats fram med ovan definition av percentiler. I alla de fall där det har undersökts huruvida det är ovan definition av linjär interpolation som använts har det bekräftats.

$$\begin{aligned} & \text{minimize} && E(x) \\ & \text{subject to} && \text{mean}(x) = \text{MEAN} \\ & && \text{sd}(x) = \text{SD} \\ & && \text{min}(x) = \text{MIN} \\ & && \text{max}(x) = \text{MAX} \end{aligned} \tag{MP}$$

(MP) använder Basin-hopping, som beskrevs i avsnitt 2.3.2, för att lösa optimeringsproblemet. Den kräver en initial gissning vilken väljs att vara en vektor med MIN som minsta ålder, MAX som största ålder, ekvidistanta avstånd mellan de mellanliggande åldrarna och totalt n åldrar. Maximala steglängden Δ sätts till $\Delta = 0.5$, temperaturen T sätts till $T = 1$ och antalet iterationer, eller hopp, sätts till 100. Som minimeringsmetod har vi valt att använda Sequential Least Squares Programming (SLSQP).

SLSQP är en icke-linjär optimeringsmetod som relaxerar minimeringsproblemet Lagrangsianskt, var-efter sökriktning tas ut genom att kvadratisk approximerar Lagrangefunktionen och linjärt approximerar bivillkoren. Man kan visa att det här är ekvivalent med ett linjärt minstakvadraten-problem, vilket har gett upphov till metodens namn [11][12][13]. I och med att fokus för det här arbetet inte ligger på optimering så presenteras inte det här.

3.2 Likelihoodfunktionen

Med hjälp av Basin-hoppingalgoritmen framställs vektorer av åldrar för alla kategorier av mognadsstadi-er i en studie från den beskrivande statistiken i artiklar. Med probitmodellen fördelas datan från de framställda vektorerna till två fall, ett 0-fall och ett 1-fall som i (2.9) är då den stokastiska variabeln y_i antar värde 1 eller 0. Exempelvis kan åldrar från tanddata för ett moget stadium H tillhöra 1-fallet och då tillhör resten av åldrarna för de andra stadierna 0-fallet. I R fördelas datan för de två fallen till två vektorer.

Första steget i probitmodellen är att genom MLE maximera likelihooden att parametrarna för en kumulativ fördelning passar datan. Med utgång från (2.10) definieras den kumulativa fördelningen som MLE utförs på som sannolikheten att en ålder från datan tillhör 1-fallet. Den kumulativa fördelning är i R definierad som en funktion. För att skilja åldersdata som tillhör 1-fallet mot de som tillhör 0-fallet är två funktioner definierade. Den ena funktionen tar sannolikheten att åldern tillhör 0-fallet och returnerar den naturliga logaritmen av den sannolikheten. Den andra funktionen gör samma sak med sannolikheten att den tillhör 1-fallet. Ytterligare en funktion är definierad som utifrån de definierade vektorerna summerar åldrarnas logaritmerade sannolikheter. Funktionen tar de två parametrarna för fördelningen som ska optimeras och summerar. För att hitta vilka parametrar som ger högst likelihood används R-funktionen *optim*. Med de optimerade parametrarna visar den logistiska avbildningen av den kumulativa fördelningen en approximativ sannolikhet att en ålder tillhör 1- eller 0-fallet. Fördelningskurvor från R presenteras i resultatet. Vidare genom att låta till exempel mognadsstadierna G och H tillhöra 1-fallet kan en approximativ sannolikhet för en ålder att tillhöra endast stadium G lösas ut. Subtraktion av sannolikheten att tillhöra endast stadium H från att tillhöra antingen stadium G eller H ger sannolikheten för att endast tillhöra G . På samma sätt går det att lösa ut den approximativa sannolikheten för en ålder att tillhöra alla mognadsstadium. Det som kan hända är att kurvorna korsar varandra och därför bildas en negativ sannolikhet. Det kan hända om många höga åldrar hamnar i ett lägre mognadsstadium relativt högre mognadsstadi-er. I sådana fall försummas sannolikheten här som noll.

3.3 Prior

Priorn kommer utgöra en täthetsfunktion för den asylsökandes ålder innan någon mätning har gjorts. Det antas att framtida asylsökande som åldersbedöms kommer från samma åldersfördelning som de som åldersbedömdes 2017. Data som RMV sparar från tidigare ärenden kan då användas för att bestämma priorn för framtida asylsökande. RMV har sparar två dataset vilka kommer användas för att ta fram en prior. Det första datasettet innehåller totala antalet asylsökande som fått ett åldersbedömningstest under 2017 i Sverige vilka var 9617 stycken [34]. Det andra datasettet innehåller antalet asylsökande som haft tandstadium H på någon av de visdomständerna, vilket är 4908 stycken [34] [35].

Enligt Migrationsverket erbjuds endast en åldersbedömning till de asylsökande där det finns en osäkerhet kring åldern. De asylsökande med ett utseende tydligt pekande på att de är under 18 år kommer alltså inte att bli åldersbedömda. Asylsökande med ett utseende som tydligt indikerar att de är över 18 år enligt handläggaren kommer inte att bli erbjuden ett gratis åldersbedömningstest. Den asylsökande i fråga kan fortfarande göra ett åldersbedömningstest på egen bekostnad. Det är handläggare på migrationsverket som tar beslut ifall den asylsökande ska få ett åldersbedömningstest[7]. Det antas att om man är yngre än en viss ålder har man ett utseende som tydligt pekar på att man är under 18 år och kommer därför aldrig bli åldersbedömd. På samma sätt antas att om man är över en viss ålder kommer man aldrig bli erbjuden ett gratis åldersbedömningstest. Med anledning av det här motiveras att åldersspannet för priorn är begränsat. Vi väljer att sätta åldersspannet för priorn till (15, 30) år. Priorn representerar inte åldersfördelningen av alla som söker asyl i Sverige utan endast gruppen av de asylsökande som blev åldersbedömda av RMV under 2017.

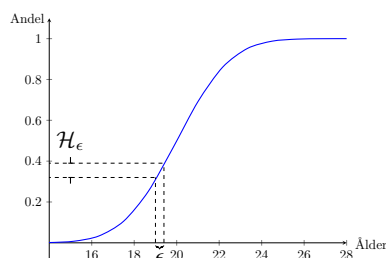
3.3.1 Anpassning av data från RMV för att konstruera en prior

Datan från RMV, där 4908 av 9617 asylsökande har bedömts ha tandstadium H , kommer ligga till grund för att ta fram priorn. Märk att priorn är gemensam för alla asylsökande och att den inte beror på vilka undersökningar som görs på individen som åldersbedöms. Dock behöver den här andelen, 4908/9617, korrigeras för att kunna användas eftersom Migrationsverkets bedömning för när en asylsökande har fullmoget tandstadium skiljer sig från vår bedömning. Migrationsverket bedömer efter det maximala av två mognadsstadier medan vår bedömning fuserar endast på vänster visdomstand.

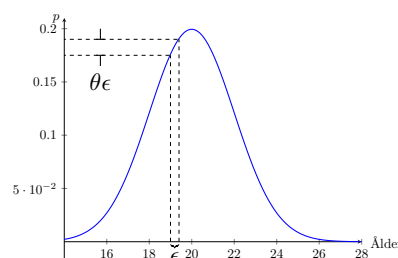
Låt $X|Y$ beteckna en förkortning för att ha tandstadium X på vänster visdomstand och Y på höger visdomstand. Låt $\{A - G\}$ vara något tandstadium A, B, C, D, E, F eller G . Totala antalet individer som har $H|\{A - G\}$, $\{A - G\}|H$ eller $H|H$ är 4908 [34]. $\{A - G\}|H$ motsvarar ett scenario där den asylsökande inte har tandstadium H på vänster visdomstand och ska därför inte inkluderas. Nedan beskrivs hur $\{A - G\}|H$ kvantifieras.

I en undersökning konstaterades att 12.4% av alla undersökta hade skilda tandstadier på de nedre visdomständerna[26]. Det antas att om man har tandstadium H på en visdomstand så är sannolikheten 12.4% att man har tandstadium $\{A - G\}$ på den andra visdomstanden. Det här är ett antagande då det inte nödvändigtvis är så i fallet då man har tandstadium H . 12.4% motsvaras av unionen av $H|\{A - G\}$ och $\{A - G\}|H$. Andelen som har $\{A - G\}|H$ antas vara hälften, det vill säga 6.2%. På grund av bilateral symmetri antas $H|\{A - G\}$ och $\{A - G\}|H$ vara lika sannolika. Givet att en asylsökande har tandstadium H på höger tand antas att 6.2% kommer ha $\{A - G\}$ på andra tanden. $\{A - G\}|H$ kvantifieras därför till $0.062 \cdot 4908 \approx 304$. Det här betyder att totala antalet som har tandstadium H på vänster visdomstand kommer vara $4908 - 304 = 4604$. Andelen med tandstadium H på vänster visdomstand antas därför vara $4604/9617$ för priorn. I nästa avsnitt beskrivs hur den här andelen beräknas för en fördelning och generella kriterier för priorn bestäms.

3.3.2 Kriterier för prior



Figur 3.3: Kumulativ fördelning från probit-modellen för tandstadium H för vänster visdomstand för pojkar.



Figur 3.4: Stickprov från godtyckligt förslag till prior, θ . Här är p tätheten.

Låt \mathcal{H} vara fördelningen i figur 3.3 för tandstadium H för vänster visdomstand för pojkar från Israel [5], se figur 3.3. \mathcal{H} är endast baserad på pojkars mognadsstadie då pojkar utgjorde 96.5% av alla åldersbedömningar 2017 [29]. Θ är en *gissning* för fördelningen av priorn och θ är ett stort stickprov

från Θ med stickprovsstorlek S , se figur 3.4. ϵ_i är ett litet åldersintervall som kommer användas för att diskretisera \mathcal{H} och θ . ϵ_1 är det första åldersintervallet och ϵ_N är det sista åldersintervallet. Låt θ_{ϵ_i} vara antalet observationer i intervallet ϵ_i i θ och låt \mathcal{H}_{ϵ_i} vara det genomsnittliga värdet av \mathcal{H} på intervallet ϵ_i . H_θ är antalet med tandstadium H från ett stickprov θ . H_θ räknas ut enligt följande.

$$H_\theta = \sum_{i=1}^N \mathcal{H}_{\epsilon_i} \cdot \theta_{\epsilon_i} \quad (3.2)$$

Följande kriterier måste vara uppfyllda för att θ ska kunna bedömas som rimlig: $H_\theta/S \approx 4604/9617$. Θ ska ha sin maximala täthet vid 18 års ålder. Θ ska vara nollskiljd på åldersintervallet $(15, 30)$ [7]. Sannolikhetsdensiteten ska gå mot noll då Θ närmar sig $\{15, 30\}$, alternativt så har sannolikhetsdensiteten en liten diskontinuitet då åldern passerar $\{15, 30\}$. Det vill säga då $\epsilon \ll 1$ kommer för $t \in \{15, 30\}$ $|\Theta(t - \epsilon) - \Theta(t + \epsilon)|$ vara litet i förhållande till den maximala tätheten.

3.4 Posteriori

För två hypoteser $H_0 = \{\text{ålder över 18 år}\}$ och $H_1 = \{\text{ålder under 18 år}\}$ kan posteriorioddsen tas fram med hjälp av Bayes formel (2.2). Från den framtagna prior (3.3) hämtas sannolikheterna att hypoteserna stämmer. Med probitmodellen hämtas den approximerade sannolikheten att en ålder tillhör det uppmätta mognadsstadiet för knäleder och höger respektive vänster visdomstånd. För prior odds definieras $\mathbb{P}(H_0)$ och $\mathbb{P}(H_1)$ som i (2.3). Från samma ekvation hittas Likelihood ratio. Den kommer nu beskrivas för $\mathcal{L}_r(H_0)$ (Likelihood ratio H_0) som

$$\mathcal{L}_r(H_0) = \frac{\mathbb{P}(\text{Data}_1|H_0)\mathbb{P}(\text{Data}_2|H_0)}{\mathbb{P}(\text{Data}_1|H_1)\mathbb{P}(\text{Data}_2|H_1)} \quad (3.3)$$

där Data_1 är data från knäleder och Data_2 är data från antingen höger eller vänster visdomstånd. Det antas att indikatorerna från knäled och visdomstånd är betingat oberoende. Posteriorin ser nu ut som

$$\frac{\mathbb{P}(H_0|\text{Data}_{\text{total}})}{\mathbb{P}(H_1|\text{Data}_{\text{total}})} = \mathcal{L}_r(H_0) \cdot \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}. \quad (3.4)$$

3.5 Beslutsteori

För att styra beslutet vid åldersbedömning viktas en kostnad av ett beslut. Förhållandet mellan kostnaderna är det som styr vikten av beslutet. För att kunna föra ett beslut i resultatet skulle kostnaderna i beslutsteorin *exempelvis* kunna sättas till $c_1 = c_4 = 0$, $c_3 = 10$ samt $c_2 = 1$. Den här vägningen innebär att oddsförhållandet mot att ett barn bedöms som över 18 och en vuxen som under 18 år är tio mot ett. Med de kostnaderna kommer beslut tas enligt följande genom ekvation (2.4)

$$\begin{aligned} H_0 &\rightarrow \frac{\mathbb{P}(H_0|\text{Data}_{\text{total}})}{\mathbb{P}(H_1|\text{Data}_{\text{total}})} > 10, \\ H_1 &\rightarrow \frac{\mathbb{P}(H_0|\text{Data}_{\text{total}})}{\mathbb{P}(H_1|\text{Data}_{\text{total}})} < 10. \end{aligned} \quad (3.5)$$

Vidare, då hypoteserna i det här fallet är disjunkta och utgör hela utfallsrummet

$$\frac{\mathbb{P}(H_0|\text{Data}_{\text{total}})}{\mathbb{P}(H_1|\text{Data}_{\text{total}})} = \text{Odds} \implies \mathbb{P}(H_0|\text{Data}_{\text{total}}) = \text{Odds} \cdot \mathbb{P}(H_1|\text{Data}_{\text{total}}), \quad (3.6)$$

så att från $\mathbb{P}(H_0|\text{Data}_{\text{total}}) + \mathbb{P}(H_1|\text{Data}_{\text{total}}) = 1$ och (3.6) erhålls

$$\mathbb{P}(H_0|\text{Data}_{\text{total}}) = \frac{\text{Odds}}{\text{Odds} + 1}, \quad \mathbb{P}(H_1|\text{Data}_{\text{total}}) = \frac{1}{\text{Odds} + 1}.$$

Kostnaderna ovan, vilka leder till tröskelvärden, kommer användas i rapporten för att ta beslut. Med tröskelvärde menas den sannolikhet att vara över 18 år en asylsökande minst behöver ha för att bedömas som över 18 år. Även andra tröskelvärden kommer att användas. Ett annat tröskelvärde kommer att användas vilket motsvarar att oddsförhållandet mot att ett barn bedöms som över 18 och en vuxen som under 18 år är fyra mot ett.

Kapitel 4

Resultat

I det här kapitlet presenteras resultatet av att använda metoderna beskrivna i föregående kapitel. Först presenteras åldrar som framställts med hjälp av Basin-hoppingalgoritmen, sedan olika logistiska avbildningar från probitmodellen som bygger på framställda åldrar. Den valda priorn introduceras och jämförs med andra möjliga priors. Besluten som tas givet kombination av mognadsstadium på visdomstand och knäled för olika priors jämförs.

Det huvudsakliga resultatet som presenteras bygger på Israel-datan för visdomständer och Ottow-datan för knäleder [5][23]. Båda källorna använder sig av skalorna presenterade i inledningen och har en stor undersökningsgrupp. Ottow-datan använder sig av Krämers definition där statistik finns för stadier 2 – 4. För knädatan ges fem percentiler, medelvärde och standardavvikelse. Antalet deltagande i studien var 325 pojkar och 333 flickor från Tyskland. Från Israel-datan finns alla mognadsstadier i Demirjians metod beskrivna på samma sätt som i tabell 3.1. För vänster visdomstand ingår 529 flickor och 487 pojkar i studien. För höger visdomstand ingår 492 flickor och 566 pojkar i studien. Tanddata framställs även från en ytterligare studie, UK-Caucasian-datan, där alla stadier från Demirjians metod [6] undersöks. För vänster visdomstand ingår 1012 flickor och 757 pojkar i studien. För höger visdomstand ingår 1013 flickor och 947 pojkar i studien. En studie som framöver benämns som Spanien-datan har bedömningar från stadium C till H enligt Demirjians metod [24]. 590 flickor och 464 pojkar ingår i studien. En studie som framöver benämns som Vieth-datan har bedömningar från stadie 2 till 6 enligt Vieths skala [36]. 350 flickor och 344 pojkar ingår i studien.

Den beskrivande statistik som finns i tandstudierna skiljer sig mellan vänster och höger sida. Konsekvensen blir olika resultat beroende på vilken tand som bedöms.

4.1 Basin-hopping och fördelningar från probit

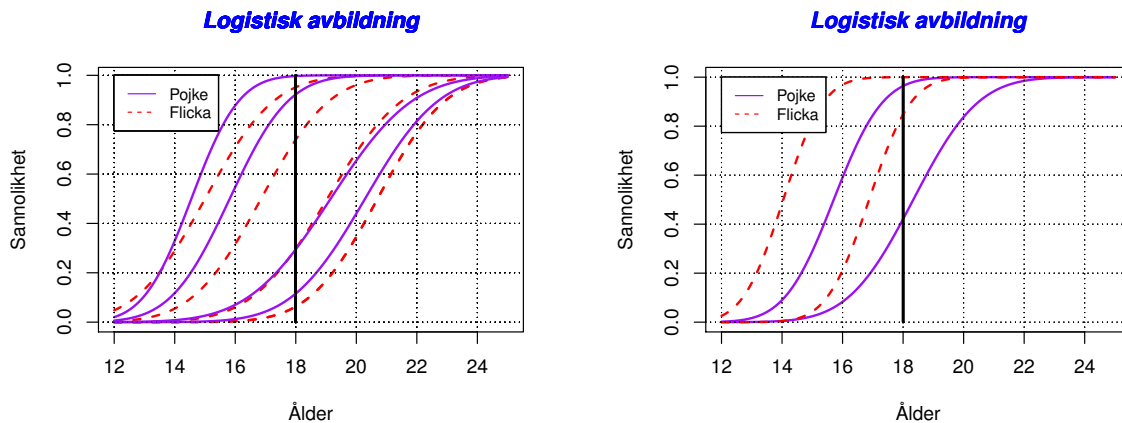
I tabell 3.1 visades beskrivande statistik för visdomständer på högersidan i överkäken, tandstadium B , flickor. Med hjälp av Basin-hopping har en vektor av åldrar, se listing 4.1 som uppfyller den beskrivande statistiken skapats. Den initiala gissningen i det här är fallet är en vektor där $n = 24$, vars lägsta ålder, $MIN = 7.53$ och vars högsta ålder, $MAX = 14.14$, ekvidistanta avstånd mellan de mellanliggande åldrarna och totalt 24 åldrar. Åldrarna för percentilerna i den beskrivande statistiken är avrundade till två decimaler i alla artiklar.

Listing 4.1: Optimering med $n = 24$ för någon kategori med Basin-hopping. Felet ligger i fun och är 0.004 avrundat. Felet kommer ifrån percentilerna i den skapade vektorn av åldrar. Medelvärdet och standardavvikelsen från den beskrivande statistiken är 10.61 respektive 1.68, samma som för den skapade vektorn av åldrar. Jämför med tabell 3.1 i avsnitt 3.1.

```
1 fun: 0.003993776618722578
2   x: array([ 7.53
3           9.23005552,  9.65664759,  9.72315514,  9.74003138,  9.74015265,
4           9.74358018, 10.63300065, 10.6469998 , 11.22649014, 11.22733612,
5           11.22733612, 11.22733612, 11.22733612, 11.31799026, 11.65068618,
6           13.17000004, 13.36999995, 13.36999995, 14.14      ])
7 MEAN: 10.61
8 STD: 1.68000001492
9 Percentiles: [ 7.62254179  8.40000058  8.76999968  9.54999957 10.64000023 11.24999965
10 13.30999997 13.36999995 14.05144999]
```

4.1. BASIN-HOPPING OCH FÖRDELNINGAR FRÅN PROBIT

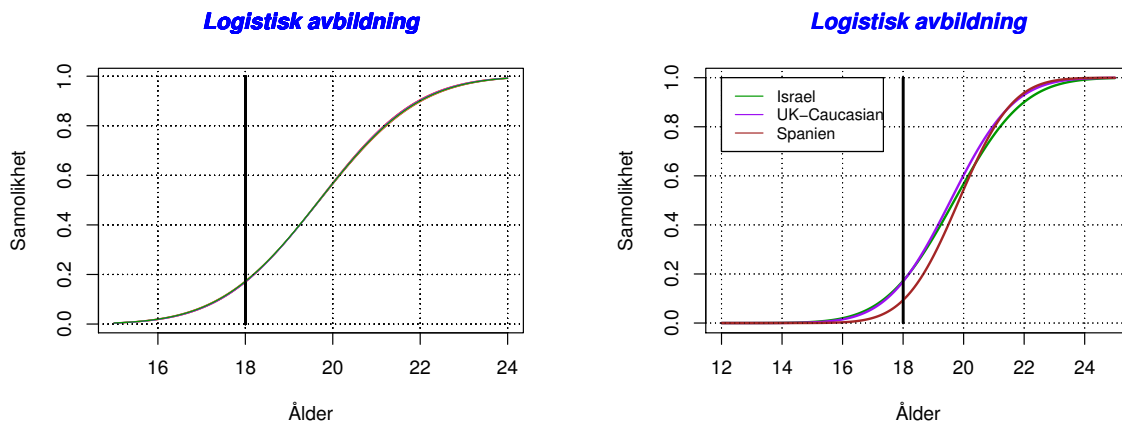
De logistiska avbildningar från probitmodellen använder dataset framställda med Basin-hopping. Det kommer vara Israel-datan och Ottow-datan som används om inte annat nämns.



(a) I figuren visas probits avbildning för höger visdomstanddatans de olika mognadsstadier från D eller mindre till H .

(b) I figuren visas probits avbildning för knådatans olika mognadsstadier från 2 eller mindre till 4 eller större.

Figur 4.1: I figurerna 4.1a och 4.1b jämförs olika mognadsstadier och kön. Flickors knäleder mognar snabbare än pojkars vilket förklarar skillnaderna i figuren [14]. Det svarta strecket representerar 18 års gränsen.



(a) I figuren visas probits avbildning för pojkars vänster visdomstanddata för tillhörighet till mognadsstadium H eller inte. Här är 15 olika dataset framställda med Basin-hopping för att urskilja om fördelningen saknar entydighet då dataseten gör det.

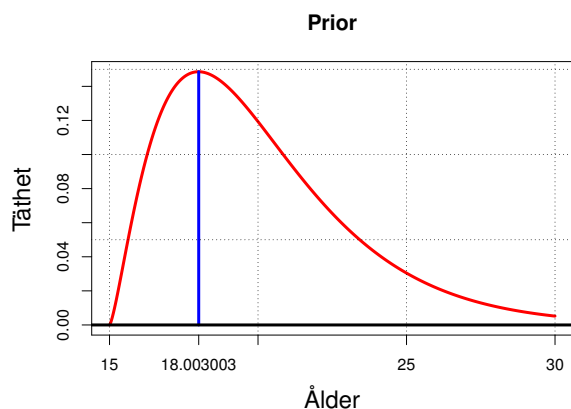
(b) I figuren visas probits avbildning för pojkars vänster visdomstanddata för tillhörighet till mognadsstadium H eller inte. En kurva med Israel-datan, en med UK-Caucasian-datan och en med Spanien-datan.

Figur 4.2: I figur 4.2a har 15 olika dataset från Basin-hopping genererats från samma beskrivande statistik. Eftersom entydighet vid framställandet har saknats har åldrarna i dataseten skiljts sig från varandra 3.1.2. Skillnaderna har däremot inte haft någon stor påverkan på fördelning vilket syns i figuren.

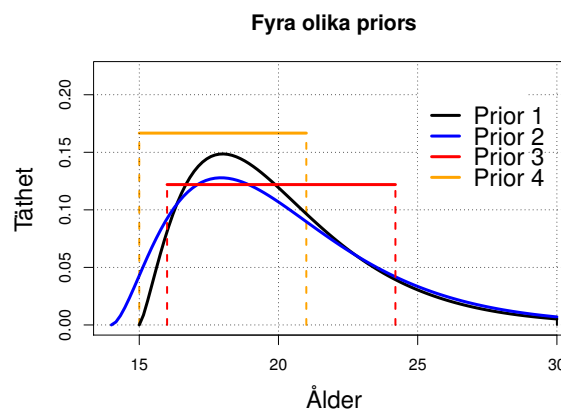
Figur 4.2b är tre kurvor med data från tre olika källor där åldersdatan är framställd genom Basin-hopping. Det finns tydliga skillnader mellan de olika kurvorna vilket innebär att resultatet kan variera beroende på vilken källa som används.

4.2 Prior

I figur 4.3 visas den valda priorn. Analys av priorn, enligt kriterierna i avsnitt 3.3.2, ger att ett stort stickprov, där stickprovsstorleken är 9617000, ger att andelen med tandstadie H blir $4614.4/9617 \approx 4604/9617$. Den maximala tätheten är vid $18.003 \approx 18$. Vidare observeras att tätheten är nollskild och kontinuerlig på intervallet $(15, 30)$, och att sannolikheten går mot 0 då fördelningen närmar sig 15. Det finns en diskontinuitet vid $\text{Ålder}=30$ där storleken på diskontinuiteten är liten. Tätheten av åldersfördelningen kan accepteras enligt kriterierna och antas därför beskriva priorn.



Figur 4.3: Den valda priorn följer en förskjuten åldersfördelning beskriven av $\text{ålder} \sim \text{Gamma}(2.4, 2.14) + 15$ vilken är trunkerad vid 30 års ålder. Priorn antas vara samma för båda könen.



Figur 4.4: Prior 1 är $\text{Gamma}(2.4, 2.14) + 15$ fördelad. Prior 2 är $\text{Gamma}(2.75, 2.25) + 14$ fördelad. Prior 3 är uniformt fördelad på $(16, 24.2)$. Prior 4 är uniformt fördelad på $(15, 21)$.

Olika priors kan användas för att ta fram en posteriorifördelning. Prior 1 från figur 4.3 är den slutgiltigt valda priorn. Prior 2, prior 3 och prior 4 kommer att undersökas för att se vilka beslut de kommer leda till och de besluten kommer jämföras med de besluten prior 1 leder till. Prior 1, 2 och 3 uppfyller alla att endelen med tandstadie H är ungefär $4604/9617$ medan prior 4 inte gör det. För ett stort stickprov är antalet som har tandstadium H i prior 4 mindre än $4604/9617$. I avsnitt 4.2.1 visas ett urval av vilka beslut från vad de fyra priors leder till.

4.2.1 Jämförelse av olika priors

Tabell 4.1 anger sannolikheten för att den asylsökande är över 18 år givet de olika priors och en given kombination av mognadstadium. Olika tröskelvärden kan användas för att läsa ut beslut.

4.3. BESLUT FRÅN POSTERIORIN

Tabell 4.1: Sannolikheten att vara över 18 givet uppmätt data för de fyra olika priors framtagna i avsnitt 4.2. Uppmätt data förkortas till exempel *Flicka Vänster Tand H Knä 4* vilket betyder att det är en flicka som har tandstadium *H* på vänster nedre visdomstand samt mognadsstadie 4 på det undersökta knäet. Tabellen är färgkodad där **röd** indikerar att sannolikheten att vara äldre än 18 år är större än 0.9091. **Blå** indikerar att sannolikheten att vara äldre än 18 år är mindre än 0.8. Resterande sannolikheter är svarta. Alla andra kombinationer av mognadsstadier har sannolikhet mindre än 0.8

Data uppmätt	Prior 1	Prior 2	Prior 3	Prior 4
Flicka Vänster Tand H Knä 4	0.9874	0.9891	0.9870	0.9606
Flicka Vänster Tand G Knä 4	0.9032	0.9054	0.9203	0.8864
Flicka Höger Tand H Knä 4	0.9887	0.9903	0.9886	0.9654
Flicka Höger Tand G Knä 4	0.8803	0.8820	0.8985	0.8648
Pojke Vänster Tand H Knä 3	0.8054	0.8035	0.8187	0.8761
Pojke Vänster Tand H Knä 4	0.9865	0.9882	0.9870	0.9644
Pojke Vänster Tand G Knä 4	0.8870	0.8880	0.9034	0.8761
Pojke Vänster Tand F Knä 4	0.7881	0.7854	0.8025	0.7745
Pojke Höger Tand H Knä 3	0.8214	0.8201	0.8359	0.8128
Pojke Höger Tand H Knä 4	0.9900	0.9913	0.9900	0.9695
Pojke Höger Tand G Knä 4	0.9085	0.9101	0.9247	0.8938
Pojke Höger Tand F Knä 4	0.8066	0.8045	0.8246	0.7869

Om tröskelvärde är $10/11 \approx 0.9091$ kan man se att prior 1 och 2, som är snarlika, inte ger samma sannolikheter, men samma beslut fattas i alla fall förutom *Pojke Höger Tand G Knä 4*. Val av olika priors kommer alltså inte att ha stor påverkan på beslut om vilka som kommer bedömas till över eller under 18 år om tröskelvärde är ≈ 0.9091 . Om tröskelvärde är 0.8 överensstämmer alla beslut för prior 1 och prior 2.

Prior 3 är den prior som generellt sett ger störst sannolikhet att den åldersbedömda är över 18. Prior 4 är den prior som inte uppfyller det översta kriteriet från Avsnitt 3.3.2 och är därför troligtvis inte en bra uppskattning av den riktiga åldersfördelningen. Trots det här avviker posteriorisannolikheterna från prior 4 och från prior 3 typiskt sett med mindre än 3 procentenheter.

4.3 Beslut från posteriorin

De beslut vi tycker bör användas presenteras i tabell B.1 och tabell B.2 i appendix. De baseras på Israel-datan (tand) i konjunktion med Ottow-datan (knä). Nedan visas ytterligare beslutstabeller baserad på annan data, och några jämförelser görs mellan beslut för olika data. I tabellerna 4.2 presenteras beslut baserade på Vieth-data (knä) och Isreal-datan. Tröskelvärde är satt till 0.9091. Tabellen är färgkodad där **röd** indikerar att sannolikheten att vara äldre än 18 år är större än 0.9091. **Blå** indikerar att sannolikheten att vara äldre än 18 år är mindre än 0.8. Resterande sannolikheter är svarta. Alla andra kombinationer av mognadsstadier, de som inte är med i tabellen, har sannolikhet mindre än 0.8.

Tabell 4.2: Beslut med Vieth-datan och Israel-datan

Data från undersökning	Beslut	Sannolikhet att vara över 18	Data från undersökning	Beslut	Sannolikhet att vara över 18
Flicka Vänster Tand H Knä 5	Över 18	0.9851	Pojke Vänster Tand H Knä 5	Över 18	0.9839
Flicka Vänster Tand H Knä 6	Över 18	0.9987	Pojke Vänster Tand H Knä 6	Över 18	0.9978
Flicka Vänster Tand G Knä 5	Under 18	0.8919	Pojke Vänster Tand G Knä 5	Under 18	0.8915
Flicka Vänster Tand G Knä 6	Över 18	0.9609	Pojke Vänster Tand G Knä 6	Över 18	0.9222
Flicka Vänster Tand F Knä 5	Under 18	0.7558	Pojke Vänster Tand F Knä 5	Under 18	0.8074
Flicka Vänster Tand F Knä 6	Under 18	0.8530	Pojke Vänster Tand F Knä 6	Under 18	0.8225
Flicka Höger Tand H Knä 5	Över 18	0.9866	Pojke Höger Tand H Knä 5	Över 18	0.9878
Flicka Höger Tand H Knä 6	Över 18	0.9989	Pojke Höger Tand H Knä 6	Över 18	0.9985
Flicka Höger Tand G Knä 5	Under 18	0.8671	Pojke Höger Tand G Knä 5	Över 18	0.9104
Flicka Höger Tand G Knä 6	Över 18	0.9459	Pojke Höger Tand G Knä 6	Över 18	0.9436
Flicka Höger Tand F Knä 5	Under 18	0.7394	Pojke Höger Tand F Knä 5	Under 18	0.8234
Flicka Höger Tand F Knä 6	Under 18	0.8425	Pojke Höger Tand F Knä 6	Under 18	0.8494

Tabell 4.3: I tabellen jämförs skillnaden mellan två olika artiklar om tandmognad, Israel (tand)-Ottow mot Uk Caucasian (tand)-Ottow. Märk att data från undersökning endast är på vänster tand.

Data från undersökning	Israel-Ottow	UK Caucasian-Ottow
Flicka Vänster Tand H Knä 4	0.9874	0.9835
Flicka Vänster Tand G Knä 4	0.9032	0.853
Pojke Vänster Tand H Knä 3	0.8054	0.822
Pojke Vänster Tand H Knä 4	0.9865	0.9877
Pojke Vänster Tand G Knä 4	0.887	0.8795

Beroende på om besluten är baserade på Israel-datan eller UK caucasian-datan kommer sannolikheterna, som visas i tabell 4.3, att skilja sig. Man kan dock observera att oavsett om tröskelvärdet är satt till 0.9091 eller 0.8 kommer alla beslut att vara samma.

Tabell 4.4: Sannolikheten att vara över 18 år för två olika skalor för knämognad, Vieths skala och Krämers definition. Beslut har tagits med ett tröskelvärde på 0.9091.

Data från undersökning (Vieths skala)	Beslut	Sannolikhet att vara över 18	Data från undersökning (Krämers definition)	Beslut	Sannolikhet att vara över 18
Flicka Knä 2	Under 18	0.0008	Flicka Knä 2	Under 18	0.0284
Flicka Knä 3	Under 18	0.0444	Flicka Knä 3	Under 18	0.2851
Flicka Knä 4	Under 18	0.2411	Flicka Knä 4	Under 18	0.8412
Flicka Knä 5	Under 18	0.8198	Pojke Knä 2	Under 18	0.1986
Flicka Knä 6	Över 18	0.9658	Pojke Knä 3	Under 18	0.5455
Pojke Knä 2	Under 18	0.1219	Pojke Knä 4	Över 18	0.9092
Pojke Knä 3	Under 18	0.1959			
Pojke Knä 4	Under 18	0.4388			
Pojke Knä 5	Över 18	0.9094			
Pojke Knä 6	Över 18	0.9779			

För komplett tabell baserad på Israel-Vieth, se tabell B.4.

Kapitel 5

Diskussion

5.1 Likelihood

5.1.1 Basin-hopping

I det här avsnittet diskuteras fördelar respektive nackdelar med att ha valt Basin-hopping som optimeringsalgoritm, samt förutsättningarna för optimeringen överlag och förslag på andra sätt att optimera.

Basin-hoppingalgoritmen löser problemet med att framställa åldrar från den beskrivande statistiken. Minimeringen fungerar till synes väl eftersom felet ofta blir litet. Skillnaden på felet som fås från olika körningar, med samma inparametrar, är mycket litet. Det trots att vektorerna av åldrar som framställts kan se olika ut, vilket motiverades i avsnitt 3.1.2. Det finns därför anledning att tro att målfunktionen inte är glatt. Men SLSQP, som används för minimeringen, är en BFGS-metod som är en erkänt bra metod för icke-glatta funktioner [15].

Att $f(y)$, täthetsfunktionen, följer Boltzmandistributionen (2.7) har konsekvenser för hur ofta en ny vektor accepteras. Eftersom felet som varje vektor i regel associeras med är väldigt små är sannolikheten att acceptera den nya vektorn, och hoppa, stor. Därför, givet $f(y)$, är området av målfunktionen som undersöks stor.

Det finns två generella saker om förutsättningarna för optimeringen som är värda att ta upp. För det första, att vektorn av åldrar som Basin-hopping producerar alltid har ett nollskiljt fel beror inte, nödvändigtvis, på Basin-hoppingalgoritmen *per se*. Att felet är nollskiljt beror bland annat på att åldrarna för percentilerna i den beskrivande statistiken är avrundade. Problemet som försöker minimeras har alltså ett fel redan i indatan till optimeringsalgoritmen. Antagligen gäller att felet blir större ju större avrundningarna är. Det är svårt att se hur man på ett enkelt sätt kan komma runt det här trångmålet. En möjlig anledning till att målfunktionen eventuellt inte är glatt är just på grund av avrundningarna.

För det andra, som beskrevs i avsnitt 3.1.2, är den vektor av åldrar som framställs, nödvändigtvis, inte entydig. Det här är ett generellt konstaterande och har inte heller något att göra med Basin-hopping som algoritm. I vårt fall är det här egentligen inte något problem eftersom vi har visat att de logistiska avbildningarna från probitmodellen i avsnitt 4.1 i stort sett ser identiska ut även om de är byggda från olika vektorer av åldrar, som kommer från samma beskrivande statistik. Se figur 4.2a.

Ett sätt att förbättra implementeringen av Basin-hoppingalgoritmen är att sätta steglängden Δ och konstanten T , så att hoppen mellan bassänger görs på ett sådant sätt att det globala minimumet hittas med större sannolikhet och/eller snabbare. Exempelvis, om $E(x^{(t)} + \eta) \geq E(x^{(t)})$ och om T är fixt, går sannolikheten att $x^{(t)} + \eta$ accepteras mot 1 enligt (2.8) för tillräckligt små fel. Om $T \rightarrow 0$ går sannolikheten att acceptera $x^{(t)} + \eta$ mot 0. $T = 1$ leder till ökad beräkningskomplexitet. Det är också möjligt att skriva en egen täthetsfunktion $f(y)$ så att nya vektorer accepteras eller förkastas med någon sannolikhet som passar vår situation bättre. Men som sagt, målfunktionen är okänd vilket gör det svårt att veta vad $f(y)$ borde vara eller vad Δ och T borde sättas till.

Att använda någon annan algoritm för att framställa åldrar är också möjligt. För att minska beräkningskomplexiteten skulle exempelvis Simulated annealing kunna användas. Simulated annealing är en teknik som, likt Basin-hopping, används för att finna ett globalt minimum. Den hoppar också mellan bassänger, men minimerar inte målfunktionen. Genom att minska T efterhand minskar sannolikheten att acceptera nya vektorer, och tillslut är sannolikheten 0 att acceptera en ny vektor. Då antas att ett

globalt minimum har hittats. På så sätt tappar möjligen en Simulated annealing-teknik i precision, men har å andra sidan mindre beräkningskomplexitet än Basin-hopping.

5.1.2 Fördelning från probitmodellen

Ett problem är att åldersfördelningen på de undersökta i publikationerna är okänd. Det är möjligt genom de beskrivna metoderna att framställa dataset, däremot går det inte utifrån det att sätta en exakt fördelning. Det finns både för lite information från tidskrifterna och från de framställda dataseten för att säkert sätta en fördelning. Det går däremot att analysera och komma väldigt nära en bra fördelning. Att ansätta probitmodellen innebär att en specifik fördelning, till exempel en normalfördelning, antas vara den sanna fördelning hos datasettet. För att kunna använda MLE måste en fördelning väljas. Ifall en normalfördelning ansätts följer även egenskapen att fördelningen är symmetrisk. Symmetri verkar inte vara fallet, vilket diskuteras i 5.2. Istället verkar fördelningen vara assymmetrisk. Att anta en gammafördelning är ett assymmetriskt alternativ. För den beskrivande statistiken som används i projektet betar sig en gammafördelning dock väldigt likt en normalfördelning, se figur B.1. Att anta en fördelning, oavsett om det är en normalfördelning eller gammafördelning, påverkar resultatet i stor grad. Sannolikheterna från den kumulativa fördelningen är endast approximativ. Det innebär att de uppmätta resultaten endast kan ta ett beslut som approximativt är korrekt. Bästa lösningen på att hitta den sanna fördelningen är storskalig insamling av människors åldrar och mognadsstadium.

När parametrarna optimeras med MLE i (2.11) hittas de parametrar som passar flest åldrar (data) bäst. MLE hittar alltså de parametrar som passar så många åldrar som möjligt för en vald fördelning. Enbart med hjälp av fler stickprov kan en mer exakt parameter bestämmas.

En alternativ metod är att försöka att anpassa datan genom att minska felet för alla stickprovspunkter med hjälp av en minsta-kvadratmetod. Fördelen med det här tillvägagångssättet är att det är en relativt lätt metod att implementera och kräver ingen förhandsgissad kurva. Nackdelen är att det enbart minskar felet till stickproven och inte i stor utsträckning tar hänsyn till avstickande värden vilket kan göra att de negligeras av metoden.

5.2 Prior

I det här avsnittet diskuteras val av prior. Olika fördelningar är möjliga att ha som grund för priorn. Normalfördelning av priorn utesluts som alternativ eftersom priorn antas vara assymmetrisk. Antagandet görs på grund av att det finns större personliga incitament för fler individer över 18 år att bli åldersbedömda av RMV än individer under 18 år. Om en 25-åring vill bli beviljad asyl har den asylsökande inget att förlora på att bli åldersbedömd. Dessutom kommer en asylsökande med prepubertalt utseende inte behöva göra ett åldersbedömningstest för att bli bedömd som under 18 år [7]. Däremot antas sannolikheten att en 25-åring blir åldersbedömd vara nollskiljd. Det här påvisar att priorn är assymmetrisk runt 18 års ålder.

Den enklaste priorn är uniformt fördelad på ett intervall, vilket motsvarar att sannolikheten är lika stor att man får ett åldersbedömningstest oavsett vilken ålder man är på det givna intervallet. Det finns dock en brist med en uniform prior, nämligen att den är diskontinuerlig i gränserna, vilket inte är realistiskt vid åldersbedömning. Exempelvis, för en uniformt fördelad prior på intervallet (15, 21) är tätheten 0 om den asylsökande är (14,99) år. Är den asylsökande 15,01 år gammal är det lika sannolikt att bli åldersbedömd som om den asylsökande hade varit 18,00 år gammal.

Gammafördelningen har valts som prior då den kan uppfylla kriterierna i avsnitt 3.3.2. Den maximala tätheten för priorn antas vara vid 18 års ålder. Det här är rimligt på grund av två anledningar. Dels är det runt 18 års ålder Migrationsverket har i uppgift att bedöma om den asylsökande är över eller under. Dels, om den asylsökandes ålder är långt över 18 år antas det vara mindre sannolikt att den asylsökande överväger ett åldersbedömningstest överhuvudtaget. Ett scenario som skulle tala emot att maximal täthet är vid 18 års ålder kan illustreras med ett exempel. Låt säga att 100 % av alla asylsökande, med verklig ålder 18, åldersbedöms medan 80% av alla asylsökande med ålder 20 åldersbedöms. Om då 100 stycken 18åriga asylsökande och 200 stycken 20åriga asylsökande söker asyl kommer alltså 100 18åringar att bli åldersbedömda med 160 20åriga asylsökande åldersbedöms. Vid ett sådant här scenario skulle den maximala tätheten för priorn inte vara vid 18 år. Det finns dock inget som tyder på att fler 20åringar söker asyl jämfört med 18åringar. Om det däremot är lika många 20åringar som 18åringar som söker asyl är det desto rimligare att anta att maximala tätheten är vid 18 år. På grund av diskussionen ovan samt

enligt tolkning av information från Migrationsverket[7] reduceras åldersspannet för priorn till mellan 15 och 30 år. Vi har valt det här åldersspannet då det verkar vara en bra uppskattning. Vill man ytterligare undersöka lämpliga kriterier för priorn uppmuntrar vi till vidare studie i ämnet.

Nedan följer en sammanställning av brister med metoden som använts för att ta fram priorn.

- Hur fördelningen, i det här fallet från probitmodellen, framställs kommer att påverka sannolikheten att tillhöra mognadsstadium H och därför också val av prior. Den matematiska påverkan syns i (3.2).
- Det finns inget som säger att flickor och pojkar kommer från samma åldersfördelning, dock tas inte det här hänsyn till i priorn. Det här är på grund av dels att antalet flickor som blivit åldersbedömda av RMV under 2017 endast är 337 och dels för att ingen statistik hittades om antalet av de åldersbestämda med fullmoget tandstadium uppdelat per kön [29][34]. Priorn för båda könen antas därför vara samma.
- Sannolikheten för $H|\{A - G\} \cup \{A - G\}|H$ uppskattades till 0.124, givet att man har minst ett tandstadie H . För att få en bättre uppskattning behövs en studie som preciserar antalet individer som har en given kombination av visdomständer där individernas ålder sparas. Alla kombinationer av tandstadium och åldersfördelningar för de undersökta behövs. En sådan undersökning hittades inte och kunde därför inte användas.
- Givet kriterierna från avsnitt 3.3.2 är den framtagna priorn inte entydig. Observera att liknande priors inte påverkar besluten stort varpå alla priors som uppfyller kriterierna från avsnitt 3.3.2 hade kunnat användas.

För att konstruera priorn valdes antalet med tandstadium H på vänster visdomstand som ensam grund för ekvation (3.2). Det hade varit fördelaktigt om andelen med moget knä och mogen visdomstand stämt överens med datan från åldersbedömningarna av RMV under 2017 för den valda priorn. Anledningen till att informationen om knä inte användes för att ta fram en prior är för att fördelningarna var, inom rimliga gränser, oförenliga. Det vill säga att det fanns ingen åldersfördelning vilken stämde överens med andelen med moget knästadie samt andelen med moget tandstadie. Det här beror möjligtvis på att bedömningarna i de använda studierna inte utförts identiskt med hur RMV utfört bedömningar. På grund av att det finns fler och större studier om ålder korrelerat till visdomständer än studier om ålder korrelerat till knästadie kommer priorn endast ta hänsyn till andelen med moget tandstadium på nedre vänster visdomstand.

Om data i framtiden sparas om vilka tandstadium och knästadium den åldersbedömda har kan man använda den här informationen för att uppdatera priorn. En brist i framtagandet av använd prior har varit att vi endast vetat hur många som haft tandstadie H . Hade $X|Y$ -data sparats för varje asylsökande hade man kunnat precisera priorn ytterligare. Man kan även konstruera olika priors beroende på vilket land den asylsökande kommer ifrån om data om mognadsgrad och land från en asylsökande sparas i framtiden.

5.3 Posteriorodds och beslut

I teori- och metodavsnitten för beslutsteorin definierades kostnader för beslut vid åldersbedömning. Kostnaderna som används bildar en nedre gräns (tröskelvärde) enligt (3.5) i avsnitt 3.5. Vid ett lägre värde bedöms färre individer som under 18 år och risken att ett barn bedöms som vuxen ökar. Den risken minskar för högre tröskelvärden. Ett för högt värde ökar istället risken att en vuxen blir bedömd som ett barn. Det går inte att säga säkert hur stora felmarginalerna är. Efter de beskrivna antaganden och approximationer som görs för att få fram posteriorin kan felet ha stor påverkan på resultatet. Vägning av kostnader i beslutmodellen bör därför ta hänsyn till den osäkerheten.

Populationerna från artiklarna som datan har hämtats från kommer i huvudsak från västerländska länder. Forskning inom åldersbedömning har presenterat underlag för att åldersfördelningar skiljer sig utifrån vilka socioekonomiska förhållanden som råder[22][32][33]. Därför bör även ursprung tas hänsyn till när bedömning görs eftersom att testet kan ge olika utslag beroende på individens ursprung. Det mest fördelaktiga hade varit ifall RMV hade gjort en egen undersökning av hur tänder och knäleders mognad kopplar till ålder. Man hade alternativt kunnat göra en djupare metaanalys där man väger ihop ett flertal artiklar för att på bästa sätt representera RMVs bedömning. Man hade även kunnat göra

bedömningar baserat på fler betingat oberoende indikatorer än endast knän och tänder. Exempelvis hade man kunnat undersöka nyckelben eller fotled för att få en bättre helhetsbedömning. Fler betingat oberoende åldersindikatorer kan vara fördelaktigt för att precisera åldern ytterligare.

5.3.1 Skillnad mellan sidor för visdomständer

I tabell 4.2 och 4.1 kan man vid enkel observation se att skillnaderna på sannolikheter beroende på om vänster eller höger visdomstand testas är väldigt små. Generellt ligger skillnaderna mellan noll till tre procentenheter både för män och kvinnor. I de tabeller som presenterades i resultat var det endast ett fall i tabell 4.1 och tre fall i 4.2 som gav olika beslut beroende på sida. Skillnaden mellan de olika sidorna kommer från att den beskrivande statistiken från studierna skiljer sig beroende på vilken sida som undersökts. Det finns därför grund börja tvivla på antagandet om bilateral symmetri i avsnitt 3.3.1.

Vid en implementering av metoden bör man välja en av tänderna, höger eller vänster, och alltid bedöma efter enbart den. Först om tanden på den valda sidan inte finns eller inte är analyserbar, bedöms visdomstanden på andra sidan.

5.3.2 Skala för knäledsmognad

Från tabell 4.4 kan det utläsas att enligt Krämers definition bedöms flicka alltid som under 18 år givet om tröskelvärde är 10/11. Däremot kan beslut fattas för om pojkar är över eller under 18 år.

Från tabell 4.4 kan det även utläsas att Vieth skala och Krämers definition bedömer likvärdigt för pojkar, för samma tröskelvärde, samt att Vieth kan åldersbedöma flickor till även över 18 år. Alltså bedömer vi att Vieths skala är mer passande för att åldersbedöma flickor och pojkar jämfört med Krämers definition. Enligt vår metod är sannolikheten att en flicka är över 18 år 0.8412, om hon endast bedöms med knäledsmognadsstadium 4. Om tröskelvärde är satt till under 0.8412 kan då flickor bedömas till över 18 år med endast knäled.

5.3.3 Jämförelse mellan tanddata

I figur 4.2b jämförs dataset från olika publicerade tidsskrifter. Det finns små skillnader mellan källorna men även likheter. I tabell 4.3 jämfördes sannolikheterna från Israel-studien och UK-studien mot varandra för vänster visdomstand. De sannolikheter som utvanns visade sig inte påverka besluten för tröskelvärdena 0.8 och 0.9091. De skillnader som finns ger däremot olika sannolikheter och tyder på att bättre undersökningar kan göras för att få en bättre åldersfördelningen. I appendix B.5 finns en tydligare figur som visar hur stor skillnaderna är mellan fler mognadsstadier. En optimal undersökning är en som bedömer mognadsstadier likadant som RMV samt att de som bedömts i undersökningen kommer från liknande socioekonomiska förhållanden som de bedömda asylsökande.

Kapitel 6

Slutsats

I den här rapporten har en ny metod presenterats för att ta beslut huruvida en asylsökande är över eller under 18 år. Författarna tycker att metoden kan användas för åldersbedömning i Sverige.

För att metoden ska fungera behöver RMV göra ett antal ändringar. En ändring är att tandläkare och radiologer anger mognadsstadier för visdomständer och knäled. Det är alltså inte tillräckligt att ange om knäled eller visdomständer uppnått slutstadium, vilket görs i dagsläget.

Man behöver även definiera vilken visdomstand bedömningen ska baseras på. Eftersom rapporten inte innefattar hur man på bästa sätt använder två visdomständer för att ta fram posteriorodds bör man välja en av tänderna, höger eller vänster, och bedöma efter enbart den valda. Först om visdomstanden på den valda sidan inte finns eller inte är analyserbar, bedöms visdomstanden på andra sidan. En vidareutveckling av projektet skulle kunna vara att ta båda visdomständerna i beaktning.

Utöver de föreslagna förändringarna bör även ett beslut om tröskelvärdets storlek tas. Författarna föreslår att någon som har underlag inom relevanta områden skall bestämma ett rimligt tröskelvärde. Personer som sedan skall fatta beslut om en asylsökande är över eller under 18 år kan använda tabeller B.1 och B.2 med det valda tröskelvärdet.

Ett eventuellt framtida projekt är att testa hur metoden presterar gentemot RMVs. Det hade varit intressant att undersöka kombinationerna av mognadsstadier som motsvarar de asylsökande som åldersbedömdes 2017. Anledningen är för att se hur många som hade bedömts annorlunda jämfört med vår metod. Ett annat förslag är att utföra en metaanalys av fler studier och mognadsdefinitioner då det har inte funnits tillräckliga resurser eller tid för att kunna utföra det här som delmål i vårt kandidatarbete.

Litteraturförteckning

- [1] The SciPy community. Basing-hopping in Python using the SciPy package, May 2018.
- [2] A. Demirjian, H. Goldstein, and J. M. Tanner. A New System of Dental Age Assessment. *Human Biology*, 45(2):211–227, 1973.
- [3] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995.
- [4] Donald Gillies. *Philosophical Theories of Probability*. Routledge, 2000.
- [5] Dental Age Research London Information Group. Dental Age Assessment - Reference data set, Israel, 2013.
- [6] Dental Age Research London Information Group. Dental Age Assessment - Reference data set, UK - Kent, 2017.
- [7] Migrationsverket Helene Hedebris, stabsjurist. Uttalande om åldersbedömning, April 2018. Mailkorrespondens.
- [8] Christophe Hurlin. Maximum likelihood estimation, 9 December 2013.
- [9] Rob J. Hyndman and Yanan Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.
- [10] Linnea Johansson. Nya siffror: 80 procent av undersökta ensamkommande över 18 år men metoden får kritik, 02 augusti 2017.
- [11] Robert Johansson. *Optimization*, pages 147–168. Apress, Berkeley, CA, 2015.
- [12] Johannes Hugo Kitschke. Investigation of the convergence behaviour of different optimization algorithms using shape optimization in flow problems, feb 2014.
- [13] Dieter Kraft. *A software package for sequential quadratic programming*. DFVLR-FB, 1988.
- [14] Jan Alexander Krämer, Sven Schmidt, Kai-Uwe Jürgens, Markus Lentschig, Andreas Schmeling, and Volker Vieth. Forensic age estimation in living individuals using 3.0t mri of the distal femur. *International Journal of Legal Medicine*, 128(3):509–514, May 2014.
- [15] Adrian S. Lewis and Michael L. Overton. Nonsmooth optimization via BFGS, 2008.
- [16] Marco Locatelli and Fabio Schoen. *Global Optimization: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2013.
- [17] J. S. Long. *Regression models for categorical and limited dependent variables*. SAGE Publications, 2 edition, 1997.
- [18] Petter Mostad and Fredrik Tamsen. Error Rates for Unvalidated Medical Age Assessment Procedures, May 2018.
- [19] Förenta Nationerna. Förenta nationernas konvention om barnets rättigheter, November 1989.

- [20] Tommie Olofsson, Håkan Sandler, Anna Ybo, and Ann-Sofie Ceciliaason. “RMV:s modell ett hafsverk av självutnämnda experter”, 20 november 2017.
- [21] Andreas Olze, Dominique Bilang, Sven Schmidt, Klaus-Dieter Wernecke, Gunther Geserick, and Andreas Schmeling. Validation of common classification systems for assessing the mineralization of third molars. *International Journal of Legal Medicine*, 119:22–6, 11 2005.
- [22] Andreas Olze, Andreas Schmeling, Mari Taniguchi, Hitoshi Maeda, Piet van Niekerk, Klaus-Dieter Wernecke, and Gunther Geserick. Forensic age estimation in living subjects: the ethnic factor in wisdom tooth mineralization. *International Journal of Legal Medicine*, 118:170–173, 2004.
- [23] Christian Ottow, Ronald Schulz, Heidi Pfeiffer, Walter Heindel, Andreas Schmeling, and Volker Vieth. Forensic age estimation by magnetic resonance imaging of the knee: the definite relevance in bony fusion of the distal femoral- and the proximal tibial epiphyses using closest-to-bone t1 tse sequence. *European Radiology*, Dec 2017.
- [24] José Luis Prieto, Elena Barbería, Ricardo Ortega, and Concepción Magaña. Evaluation of chronological age based on third molar development in the spanish population. *International Journal of Legal Medicine*, 119:349–354, 2005.
- [25] John A. Rice. *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press., third edition, 2006.
- [26] Olaf H. Drummer Richard B. BassedEmail, Christopher Briggs. The incidence of asymmetrical left/right skeletal and dental development in an australian population and the effect of this on forensic age estimations. *International Journal of Legal Medicine*, 126, 3 2012.
- [27] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [28] Christian P. Robert and George Casella. *Introducing Monte Carlo Methods with R (Use R)*. Springer-Verlag, Berlin, Heidelberg, 1st edition, 2009.
- [29] Rättsmedicinalverket. Aktuell statistik, 19 april 2018.
- [30] Rättsmedicinalverket. Medicinska åldersbedömningar, 16 januari 2018.
- [31] Rättsmedicinalverket. Metoder för medicinska åldersbedömningar, 16 januari 2018.
- [32] A. Schmeling, W. Reisinger, D. Loreck, K. Vendura, W. Markus, and G. Geserick. Effects of ethnicity on skeletal maturation: consequences for forensic age estimations. *International Journal of Legal Medicine*, 113(5):253–258, Aug 2000.
- [33] Andreas Schmeling, Ronald Schulz, Bettina Danner, and Friedrich W. Rösing. The impact of economic progress and modernization in medicine on the ossification of hand and wrist. *International Journal of Legal Medicine*, 120(2):121–126, Mar 2006.
- [34] Rättsmedicinalverket Stina Björklund. Statistik om tidigare åldersbedömningar av RMV, Mars 2018. Mailkorrespondens.
- [35] Rättsmedicinalverket Stina Björklund. Uttalande om röntgen av visdomständer, Mars 2018. Mailkorrespondens.
- [36] Volker Vieth, Ronald Schulz, Walter Heindel, Heidi Pfeiffer, Boris Buerke, Andreas Schmeling, and Christian Ottow. Forensic age assessment by 3.0t mri of the knee: proposal of a new mri classification of ossification stages. *European Radiology*, Mar 2018.
- [37] D. Wales and J. Doye. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *eprint arXiv:cond-mat/9803344*, March 1998.

Bilaga A

Teori

A.1 Fundamentala teorem inom statistik och sannolikhetslära

Teorem 2 (Centrala gränsvärdesatsen). Låt X_1, X_2, \dots vara en oändlig följd av oberoende och likafördelade stokastiska variabler med väntevärde μ och med standardavvikelse $\sigma > 0$. Låt vidare den stokastiska variabeln

$$S_n = X_1 + X_2 + \dots + X_n$$

beteckna summan av de första n stokastiska variablerna i följd.

Då gäller att

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a),$$

där $\Phi(x)$ betecknar fördelningsfunktionen för en standardiserad normalfördelning.

Teorem 3 (Stora talens lag på stark form). Låt X_1, X_2, \dots vara en oändlig följd av oberoende och likafördelade stokastiska variabler med väntevärde μ . Låt vidare den stokastiska variabeln

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

beteckna medelvärdet av de första n stokastiska variablerna i följd.

Om väntevärdet μ är ändligt, så gäller det att

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} M_n = \mu \right) = 1.$$

A.2 Markovkedjor

En tidshomogen Markovkedja $\{X(t)\}_{t \geq 0}$ är en sekvens stokastiska variabler sådan att täthetsfunktionen av $X(t)$ betingad av alla tidigare stokastiska variabler, endast beror på den senaste av dem, nämligen $X(t-1)$. Den här betingade sannolikheten betecknas K och kallas för kärnan av Markovkedjan,

$$\pi(X(t) | X(0), X(1), \dots, X(t-1)) = \pi(X(t) | X(t-1)) = K(X(t-1), X(t)).$$

En stationär fördelning $f(x)$ till Markovkedjan $\{X(t) : t \geq 0\}$ uppfyller

$$f(y) = \int_{\Omega} K(x, y) f(x) dx, \quad \int_{\Omega} K(y, x) dx = 1.$$

Definition A.2.1. Ett stadie i en Markovkedja sägs vara återkommande om

$$\mathbb{P}(X(t) = i, t \in [1, \infty) | X(0) = i) = 1,$$

alltså att sannolikheten för att starta i i och sedan nå i igen är 1 om stadie kan bytas oändligt antal gånger.

Definition A.2.2. En Markovkedja anses *kommunicera* om $i \leftrightarrow j \forall i, j \in \omega$, det vill säga att från att man startar i stadie i så kan man nå j och vice versa.

$$\mathbb{P}(X(t) = j | X(0) = i) > 0.$$

Definition A.2.3. En Markovkedja sägs vara aperiodisk om för alla stadier i är $d_i = 1$ där

$$d_i = \text{sgd}\{n \geq 1 : \mathbb{P}(X_n = i | X_0 = i) > 0\}. \quad (\text{A.1})$$

Där sgd betecknar *största gemensamma delare*.

Definition A.2.4. Ett tillstånd sägs vara ergodiskt om den är aperiodisk och om den förväntade återkommande tiden är ändlig.

En Markovkedja $\{X(t)\}_{t \geq 0}$ som uppfyller samtliga definitioner konvergerar mot en unik stationär fördelning $f(y)$ då $t \rightarrow \infty$.

A.3 Monte Carlo-metoden

Monte Carlo-metoden bygger på att det finns en täthetsfunktion π som är okänd men kan simuleras. För att lösa det här problemet låtes en dator simulera en sekvens X_1, X_2, \dots av iid¹ stokastiska variabler som följer den okända täthetsfunktionen. Väntevärdet bestäms därefter genom

$$\mathbb{E}_\pi[g(X)] = \int_\Omega g(x)\pi(x)dx$$

för en godtycklig funktion g , exempelvis $g(x) = x$, och utfallsrum Ω . Om X är en diskret stokastisk variabel fås

$$\mathbb{E}_\pi[g(X)] = \int_\Omega g(x)\pi(x)dx = \sum_{i \in \Omega} g(x_i)p(x_i).$$

Här är $p(x)$ massfunktionen av fördelningen π . I många fall är det svårt att utvärdera summan eller integralen och då beräkna väntevärdet. Vi inför $\mu = \mathbb{E}_\pi[g(X)]$ för värdet vi vill beräkna. Vi inför också $Y = g(X)$. Om vi kan simulera oberoende värden (X_1, X_2, \dots) som är realiseringar av en stokastisk process med fördelning π . Approximationen $\hat{\mu}_n \approx \mu$ kan då skrivas som

$$\hat{\mu}_n = \hat{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Det här konvergerar mot μ då $n \rightarrow \infty$ enligt *stora talens lag i stark form* (teorem 3). Variansen av Y är definierad som

$$\sigma^2 = \int_\Omega [g(x) - \mu]^2 \pi(x) dx.$$

Här kan en approximation också göras för fördelningens standardavvikelse $s \approx \sigma$ enligt kvadratroten av

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n [Y_i - \hat{\mu}_n]^2.$$

Monte Carlo-metodens uppskattade fel erhålls från *centrala gränsvärdessatsen* (teorem 2) som säger att standardavvikelsens konvergensthastighet mot 0 är $1/\sqrt{n}$ och felet är definierat som s/\sqrt{n} .

A.4 Några relevanta fördelningsfunktioner

A.4.1 Standardnormalfördelningen

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (\text{A.2})$$

¹Oberoende likafördelade stokastiska variabler eller independent and identically distributed.

A.4.2 Gammafördelningen

$$\text{Gamma}(x; \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \gamma\left(\alpha, \frac{x}{\lambda}\right), \quad (\text{A.3})$$

där $\Gamma(a)$ är gammafunktionen och

$$\gamma\left(\alpha, \frac{x}{\lambda}\right) = \int_0^{x/\lambda} t^{\alpha-1} e^{-t} dt. \quad (\text{A.4})$$

I rapporten används *shape* och *scale* för att beskriva Gammafördelningar.

A.4.3 Bernoullifördelningen

$$\text{Bernoulli}(n, p) = \begin{cases} 1 - p, & n = 0, \\ 1, & n = 1. \end{cases}$$

A.4.4 Kontinuerlig likformig fördelning

$$\text{Unif}(a, b) = \begin{cases} 0, & x < 0, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases}$$

Bilaga B

Resultat

B.1 Beslut

Tabell B.1: Tröskelvärde vald till 0.9091. Undersökningar Israel och Ottow. För att läsa av vilket beslut RMV hade tagit, se tabell 1.1

Data från undersökning	Beslut	Sannolikhet att vara över 18	Data från undersökning	Beslut	Sannolikhet att vara över 18
Flicka Vänster Tand H Knä 2	Under 18	0.1425	Pojke Vänster Tand H Knä 2	Under 18	0.4311
Flicka Vänster Tand H Knä 3	Under 18	0.5604	Pojke Vänster Tand H Knä 3	Under 18	0.8054
Flicka Vänster Tand H Knä 4	Över 18	0.9874	Pojke Vänster Tand H Knä 4	Över 18	0.9865
Flicka Vänster Tand G Knä 2	Under 18	0.0926	Pojke Vänster Tand G Knä 2	Under 18	0.3119
Flicka Vänster Tand G Knä 3	Under 18	0.4576	Pojke Vänster Tand G Knä 3	Under 18	0.6393
Flicka Vänster Tand G Knä 4	Under 18	0.9032	Pojke Vänster Tand G Knä 4	Under 18	0.887
Flicka Vänster Tand F Knä 2	Under 18	0.059	Pojke Vänster Tand F Knä 2	Under 18	0.243
Flicka Vänster Tand F Knä 3	Under 18	0.3667	Pojke Vänster Tand F Knä 3	Under 18	0.535
Flicka Vänster Tand F Knä 4	Under 18	0.7742	Pojke Vänster Tand F Knä 4	Under 18	0.7881
Flicka Vänster Tand E Knä 2	Under 18	0.022	Pojke Vänster Tand E Knä 2	Under 18	0.1162
Flicka Vänster Tand E Knä 3	Under 18	0.2219	Pojke Vänster Tand E Knä 3	Under 18	0.3256
Flicka Vänster Tand E Knä 4	Under 18	0.5608	Pojke Vänster Tand E Knä 4	Under 18	0.5121
Flicka Vänster Tand D Knä 2	Under 18	0.0094	Pojke Vänster Tand D Knä 2	Under 18	0.0542
Flicka Vänster Tand D Knä 3	Under 18	0.131	Pojke Vänster Tand D Knä 3	Under 18	0.1998
Flicka Vänster Tand D Knä 4	Under 18	0.4203	Pojke Vänster Tand D Knä 4	Under 18	0.3577
Flicka Vänster Tand C Knä 2	Under 18	0.0029	Pojke Vänster Tand C Knä 2	Under 18	0.0116
Flicka Vänster Tand C Knä 3	Under 18	0.0536	Pojke Vänster Tand C Knä 3	Under 18	0.0607
Flicka Vänster Tand C Knä 4	Under 18	0.2434	Pojke Vänster Tand C Knä 4	Under 18	0.1403
Flicka Vänster Tand B Knä 2	Under 18	0	Pojke Vänster Tand B Knä 2	Under 18	0.001
Flicka Vänster Tand B Knä 3	Under 18	0	Pojke Vänster Tand B Knä 3	Under 18	0.0056
Flicka Vänster Tand B Knä 4	Under 18	0	Pojke Vänster Tand B Knä 4	Under 18	0.0142
Flicka Vänster Tand A Knä 2	Under 18	0.0073	Pojke Vänster Tand A Knä 2	Under 18	0.0088
Flicka Vänster Tand A Knä 3	Under 18	0.1162	Pojke Vänster Tand A Knä 3	Under 18	0.0496
Flicka Vänster Tand A Knä 4	Under 18	0.4246	Pojke Vänster Tand A Knä 4	Under 18	0.1221
Flicka Höger Tand H Knä 2	Under 18	0.16	Pojke Höger Tand H Knä 2	Under 18	0.4395
Flicka Höger Tand H Knä 3	Under 18	0.5815	Pojke Höger Tand H Knä 3	Under 18	0.8214
Flicka Höger Tand H Knä 4	Över 18	0.9887	Pojke Höger Tand H Knä 4	Över 18	0.99
Flicka Höger Tand G Knä 2	Under 18	0.0805	Pojke Höger Tand G Knä 2	Under 18	0.3166
Flicka Höger Tand G Knä 3	Under 18	0.4332	Pojke Höger Tand G Knä 3	Under 18	0.6576
Flicka Höger Tand G Knä 4	Under 18	0.8803	Pojke Höger Tand G Knä 4	Under 18	0.9085
Flicka Höger Tand F Knä 2	Under 18	0.0483	Pojke Höger Tand F Knä 2	Under 18	0.2311
Flicka Höger Tand F Knä 3	Under 18	0.3402	Pojke Höger Tand F Knä 3	Under 18	0.5342
Flicka Höger Tand F Knä 4	Under 18	0.7588	Pojke Höger Tand F Knä 4	Under 18	0.8066
Flicka Höger Tand E Knä 2	Under 18	0.0229	Pojke Höger Tand E Knä 2	Under 18	0.1052
Flicka Höger Tand E Knä 3	Under 18	0.2273	Pojke Höger Tand E Knä 3	Under 18	0.3053
Flicka Höger Tand E Knä 4	Under 18	0.5741	Pojke Höger Tand E Knä 4	Under 18	0.4883
Flicka Höger Tand D Knä 2	Under 18	0.0099	Pojke Höger Tand D Knä 2	Under 18	0.0281
Flicka Höger Tand D Knä 3	Under 18	0.1356	Pojke Höger Tand D Knä 3	Under 18	0.119
Flicka Höger Tand D Knä 4	Under 18	0.4266	Pojke Höger Tand D Knä 4	Under 18	0.2342
Flicka Höger Tand C Knä 2	Under 18	0.0008	Pojke Höger Tand C Knä 2	Under 18	0.0051
Flicka Höger Tand C Knä 3	Under 18	0.014	Pojke Höger Tand C Knä 3	Under 18	0.0291
Flicka Höger Tand C Knä 4	Under 18	0.0742	Pojke Höger Tand C Knä 4	Under 18	0.0734
Flicka Höger Tand B Knä 2	Under 18	0	Pojke Höger Tand B Knä 2	Under 18	0
Flicka Höger Tand B Knä 3	Under 18	0	Pojke Höger Tand B Knä 3	Under 18	0
Flicka Höger Tand B Knä 4	Under 18	0	Pojke Höger Tand B Knä 4	Under 18	0
Flicka Höger Tand A Knä 2	Under 18	0.0092	Pojke Höger Tand A Knä 2	Under 18	0.0357
Flicka Höger Tand A Knä 3	Under 18	0.1373	Pojke Höger Tand A Knä 3	Under 18	0.1582
Flicka Höger Tand A Knä 4	Under 18	0.4755	Pojke Höger Tand A Knä 4	Under 18	0.3198

Tabell B.2: Beslut om endast tanddata eller knädata finns. Undersökningar Israel och Ottow. För att läsa av vilket beslut RMV hade tagit, se tabell 1.1

Data från undersökning	Beslut	Sannolikhet att vara över 18	Data från undersökning	Beslut	Sannolikhet att vara över 18
Flicka Vänster Tand H	Över 18	0.9812	Pojke Vänster Tand H	Över 18	0.9581
Flicka Vänster Tand G	Under 18	0.8584	Pojke Vänster Tand G	Under 18	0.7546
Flicka Vänster Tand F	Under 18	0.6865	Pojke Vänster Tand F	Under 18	0.612
Flicka Vänster Tand E	Under 18	0.4239	Pojke Vänster Tand E	Under 18	0.3267
Flicka Vänster Tand D	Under 18	0.2639	Pojke Vänster Tand D	Under 18	0.1825
Flicka Vänster Tand C	Under 18	0.1111	Pojke Vänster Tand C	Under 18	0.0466
Flicka Vänster Tand B	Under 18	0	Pojke Vänster Tand B	Under 18	0.0037
Flicka Vänster Tand A	Under 18	0.2526	Pojke Vänster Tand A	Under 18	0.037
Flicka Höger Tand H	Över 18	0.9833	Pojke Höger Tand H	Över 18	0.9682
Flicka Höger Tand G	Under 18	0.8254	Pojke Höger Tand G	Under 18	0.7869
Flicka Höger Tand F	Under 18	0.6612	Pojke Höger Tand F	Under 18	0.6232
Flicka Höger Tand E	Under 18	0.4375	Pojke Höger Tand E	Under 18	0.3028
Flicka Höger Tand D	Under 18	0.2713	Pojke Höger Tand D	Under 18	0.1001
Flicka Höger Tand C	Under 18	0.0258	Pojke Höger Tand C	Under 18	0.0209
Flicka Höger Tand B	Under 18	0	Pojke Höger Tand B	Under 18	0
Flicka Höger Tand A	Under 18	0.3015	Pojke Höger Tand A	Under 18	0.1389
Flicka Vänster H Knä 2	Under 18	0.0284	Pojke Vänster H Knä 2	Under 18	0.1986
Flicka Vänster H Knä 3	Under 18	0.2851	Pojke Vänster H Knä 3	Under 18	0.5455
Flicka Vänster H Knä 4	Under 18	0.8412	Pojke Vänster H Knä 4	Över 18	0.9092

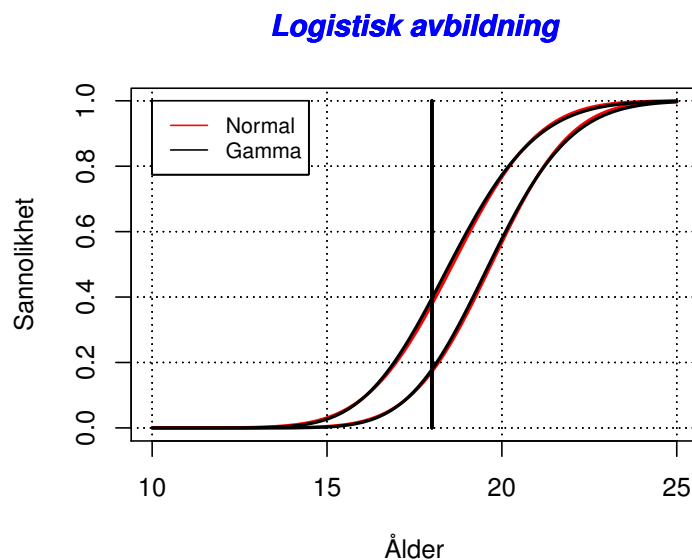
Tabell B.3: Tröskelvärde vald till 0.9091. Undersökningar UK Caucasian och Ottow. För att läsa av vilket beslut RMV hade tagit, se tabell 1.1

Data från undersökning	Beslut	Sannolikhet att vara över 18	Data från undersökning	Beslut	Sannolikhet att vara över 18
Flicka Vänster Tand H Knä 2	Under 18	0.1609	Pojke Vänster Tand H Knä 2	Under 18	0.4566
Flicka Vänster Tand H Knä 3	Under 18	0.5732	Pojke Vänster Tand H Knä 3	Under 18	0.822
Flicka Vänster Tand H Knä 4	Över 18	0.9835	Pojke Vänster Tand H Knä 4	Över 18	0.9877
Flicka Vänster Tand G Knä 2	Under 18	0.0951	Pojke Vänster Tand G Knä 2	Under 18	0.3386
Flicka Vänster Tand G Knä 3	Under 18	0.4448	Pojke Vänster Tand G Knä 3	Under 18	0.6493
Flicka Vänster Tand G Knä 4	Under 18	0.853	Pojke Vänster Tand G Knä 4	Under 18	0.8795
Flicka Vänster Tand F Knä 2	Under 18	0.0508	Pojke Vänster Tand F Knä 2	Under 18	0.2096
Flicka Vänster Tand F Knä 3	Under 18	0.3241	Pojke Vänster Tand F Knä 3	Under 18	0.4688
Flicka Vänster Tand F Knä 4	Under 18	0.6645	Pojke Vänster Tand F Knä 4	Under 18	0.686
Flicka Vänster Tand E Knä 2	Under 18	0.0116	Pojke Vänster Tand E Knä 2	Under 18	0.0946
Flicka Vänster Tand E Knä 3	Under 18	0.1342	Pojke Vänster Tand E Knä 3	Under 18	0.2838
Flicka Vänster Tand E Knä 4	Under 18	0.3554	Pojke Vänster Tand E Knä 4	Under 18	0.4567
Flicka Vänster Tand D Knä 2	Under 18	0.0119	Pojke Vänster Tand D Knä 2	Under 18	0.0635
Flicka Vänster Tand D Knä 3	Under 18	0.156	Pojke Vänster Tand D Knä 3	Under 18	0.2253
Flicka Vänster Tand D Knä 4	Under 18	0.4742	Pojke Vänster Tand D Knä 4	Under 18	0.3935
Flicka Vänster Tand C Knä 2	Under 18	0.0115	Pojke Vänster Tand C Knä 2	Under 18	0.0702
Flicka Vänster Tand C Knä 3	Under 18	0.1575	Pojke Vänster Tand C Knä 3	Under 18	0.2542
Flicka Vänster Tand C Knä 4	Under 18	0.5057	Pojke Vänster Tand C Knä 4	Under 18	0.4585
Flicka Vänster Tand B Knä 2	Under 18	0.0074	Pojke Vänster Tand B Knä 2	Under 18	0.0247
Flicka Vänster Tand B Knä 3	Under 18	0.1157	Pojke Vänster Tand B Knä 3	Under 18	0.1161
Flicka Vänster Tand B Knä 4	Under 18	0.4171	Pojke Vänster Tand B Knä 4	Under 18	0.2451
Flicka Vänster Tand A Knä 2	Under 18	0.0044	Pojke Vänster Tand A Knä 2	Under 18	0.002
Flicka Vänster Tand A Knä 3	Under 18	0.0786	Pojke Vänster Tand A Knä 3	Under 18	0.0128
Flicka Vänster Tand A Knä 4	Under 18	0.3292	Pojke Vänster Tand A Knä 4	Under 18	0.0358

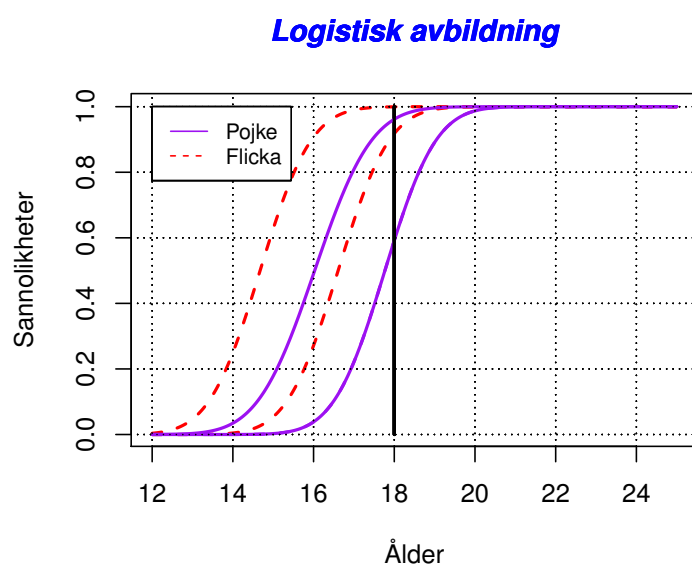
Tabell B.4: Tröskelvärde vald till 0.9091. Undersökningar Israel och Vieths

Data från undersökning	Beslut	Sannolikhet att vara över 18	Data från undersökning	Beslut	Sannolikhet att vara över 18
Flicka Vänster Tand H Knä 2	Under 18	0.0097	Pojke Vänster Tand H Knä 2	Under 18	0.3368
Flicka Vänster Tand H Knä 3	Under 18	0.1878	Pojke Vänster Tand H Knä 3	Under 18	0.4104
Flicka Vänster Tand H Knä 4	Under 18	0.4866	Pojke Vänster Tand H Knä 4	Under 18	0.6463
Flicka Vänster Tand H Knä 5	Över 18	0.9851	Pojke Vänster Tand H Knä 5	Över 18	0.9839
Flicka Vänster Tand H Knä 6	Över 18	0.9987	Pojke Vänster Tand H Knä 6	Över 18	0.9978
Flicka Vänster Tand G Knä 2	Under 18	0.004	Pojke Vänster Tand G Knä 2	Under 18	0.2234
Flicka Vänster Tand G Knä 3	Under 18	0.1285	Pojke Vänster Tand G Knä 3	Under 18	0.3026
Flicka Vänster Tand G Knä 4	Under 18	0.3986	Pojke Vänster Tand G Knä 4	Under 18	0.5292
Flicka Vänster Tand G Knä 5	Under 18	0.8919	Pojke Vänster Tand G Knä 5	Under 18	0.8915
Flicka Vänster Tand G Knä 6	Över 18	0.9609	Pojke Vänster Tand G Knä 6	Över 18	0.9222
Flicka Vänster Tand F Knä 2	Under 18	0.0022	Pojke Vänster Tand F Knä 2	Under 18	0.1642
Flicka Vänster Tand F Knä 3	Under 18	0.0854	Pojke Vänster Tand F Knä 3	Under 18	0.2388
Flicka Vänster Tand F Knä 4	Under 18	0.3193	Pojke Vänster Tand F Knä 4	Under 18	0.4552
Flicka Vänster Tand F Knä 5	Under 18	0.7558	Pojke Vänster Tand F Knä 5	Under 18	0.8074
Flicka Vänster Tand F Knä 6	Under 18	0.853	Pojke Vänster Tand F Knä 6	Under 18	0.8225
Flicka Vänster Tand E Knä 2	Under 18	0.0006	Pojke Vänster Tand E Knä 2	Under 18	0.068
Flicka Vänster Tand E Knä 3	Under 18	0.0344	Pojke Vänster Tand E Knä 3	Under 18	0.118
Flicka Vänster Tand E Knä 4	Under 18	0.1907	Pojke Vänster Tand E Knä 4	Under 18	0.2962
Flicka Vänster Tand E Knä 5	Under 18	0.5409	Pojke Vänster Tand E Knä 5	Under 18	0.5785
Flicka Vänster Tand E Knä 6	Under 18	0.5942	Pojke Vänster Tand E Knä 6	Under 18	0.4936
Flicka Vänster Tand D Knä 2	Under 18	0.0002	Pojke Vänster Tand D Knä 2	Under 18	0.0285
Flicka Vänster Tand D Knä 3	Under 18	0.0154	Pojke Vänster Tand D Knä 3	Under 18	0.0567
Flicka Vänster Tand D Knä 4	Under 18	0.1116	Pojke Vänster Tand D Knä 4	Under 18	0.1882
Flicka Vänster Tand D Knä 5	Under 18	0.4023	Pojke Vänster Tand D Knä 5	Under 18	0.4527
Flicka Vänster Tand D Knä 6	Under 18	0.4118	Pojke Vänster Tand D Knä 6	Under 18	0.3193
Flicka Vänster Tand C Knä 2	Under 18	0.0001	Pojke Vänster Tand C Knä 2	Under 18	0.0055
Flicka Vänster Tand C Knä 3	Under 18	0.0049	Pojke Vänster Tand C Knä 3	Under 18	0.0127
Flicka Vänster Tand C Knä 4	Under 18	0.0459	Pojke Vänster Tand C Knä 4	Under 18	0.0615
Flicka Vänster Tand C Knä 5	Under 18	0.2309	Pojke Vänster Tand C Knä 5	Under 18	0.2457
Flicka Vänster Tand C Knä 6	Under 18	0.2005	Pojke Vänster Tand C Knä 6	Under 18	0.105
Flicka Vänster Tand B Knä 2	Under 18	0	Pojke Vänster Tand B Knä 2	Under 18	0.0004
Flicka Vänster Tand B Knä 3	Under 18	0	Pojke Vänster Tand B Knä 3	Under 18	0.0011
Flicka Vänster Tand B Knä 4	Under 18	0	Pojke Vänster Tand B Knä 4	Under 18	0.0063
Flicka Vänster Tand B Knä 5	Under 18	0	Pojke Vänster Tand B Knä 5	Under 18	0.0378
Flicka Vänster Tand B Knä 6	Under 18	0	Pojke Vänster Tand B Knä 6	Under 18	0.0091
Flicka Vänster Tand A Knä 2	Under 18	0.0002	Pojke Vänster Tand A Knä 2	Under 18	0.004
Flicka Vänster Tand A Knä 3	Under 18	0.0122	Pojke Vänster Tand A Knä 3	Under 18	0.0097
Flicka Vänster Tand A Knä 4	Under 18	0.0983	Pojke Vänster Tand A Knä 4	Under 18	0.0508
Flicka Vänster Tand A Knä 5	Under 18	0.4045	Pojke Vänster Tand A Knä 5	Under 18	0.2289
Flicka Vänster Tand A Knä 6	Under 18	0.423	Pojke Vänster Tand A Knä 6	Under 18	0.0885
Flicka Höger Tand H Knä 2	Under 18	0.0124	Pojke Höger Tand H Knä 2	Under 18	0.3439
Flicka Höger Tand H Knä 3	Under 18	0.2072	Pojke Höger Tand H Knä 3	Under 18	0.4175
Flicka Höger Tand H Knä 4	Under 18	0.5063	Pojke Höger Tand H Knä 4	Under 18	0.6571
Flicka Höger Tand H Knä 5	Över 18	0.9866	Pojke Höger Tand H Knä 5	Över 18	0.9878
Flicka Höger Tand H Knä 6	Över 18	0.9989	Pojke Höger Tand H Knä 6	Över 18	0.9985
Flicka Höger Tand G Knä 2	Under 18	0.0032	Pojke Höger Tand G Knä 2	Under 18	0.2262
Flicka Höger Tand G Knä 3	Under 18	0.1135	Pojke Höger Tand G Knä 3	Under 18	0.3065
Flicka Höger Tand G Knä 4	Under 18	0.3766	Pojke Höger Tand G Knä 4	Under 18	0.5386
Flicka Höger Tand G Knä 5	Under 18	0.8671	Pojke Höger Tand G Knä 5	Över 18	0.9104
Flicka Höger Tand G Knä 6	Över 18	0.9459	Pojke Höger Tand G Knä 6	Över 18	0.9436
Flicka Höger Tand F Knä 2	Under 18	0.0016	Pojke Höger Tand F Knä 2	Under 18	0.1526
Flicka Höger Tand F Knä 3	Under 18	0.0714	Pojke Höger Tand F Knä 3	Under 18	0.2273
Flicka Höger Tand F Knä 4	Under 18	0.2945	Pojke Höger Tand F Knä 4	Under 18	0.4497
Flicka Höger Tand F Knä 5	Under 18	0.7394	Pojke Höger Tand F Knä 5	Under 18	0.8234
Flicka Höger Tand F Knä 6	Under 18	0.8425	Pojke Höger Tand F Knä 6	Under 18	0.8494
Flicka Höger Tand E Knä 2	Under 18	0.0006	Pojke Höger Tand E Knä 2	Under 18	0.0607
Flicka Höger Tand E Knä 3	Under 18	0.0356	Pojke Höger Tand E Knä 3	Under 18	0.1072
Flicka Höger Tand E Knä 4	Under 18	0.1953	Pojke Höger Tand E Knä 4	Under 18	0.279
Flicka Höger Tand E Knä 5	Under 18	0.5537	Pojke Höger Tand E Knä 5	Under 18	0.559
Flicka Höger Tand E Knä 6	Under 18	0.6137	Pojke Höger Tand E Knä 6	Under 18	0.4666
Flicka Höger Tand D Knä 2	Under 18	0.0002	Pojke Höger Tand D Knä 2	Under 18	0.0141
Flicka Höger Tand D Knä 3	Under 18	0.0162	Pojke Höger Tand D Knä 3	Under 18	0.0299
Flicka Höger Tand D Knä 4	Under 18	0.1156	Pojke Höger Tand D Knä 4	Under 18	0.1164
Flicka Höger Tand D Knä 5	Under 18	0.4085	Pojke Höger Tand D Knä 5	Under 18	0.338
Flicka Höger Tand D Knä 6	Under 18	0.4196	Pojke Höger Tand D Knä 6	Under 18	0.1933
Flicka Höger Tand C Knä 2	Under 18	0	Pojke Höger Tand C Knä 2	Under 18	0.0023
Flicka Höger Tand C Knä 3	Under 18	0.0013	Pojke Höger Tand C Knä 3	Under 18	0.0057
Flicka Höger Tand C Knä 4	Under 18	0.0124	Pojke Höger Tand C Knä 4	Under 18	0.0306
Flicka Höger Tand C Knä 5	Under 18	0.0707	Pojke Höger Tand C Knä 5	Under 18	0.1552
Flicka Höger Tand C Knä 6	Under 18	0.0491	Pojke Höger Tand C Knä 6	Under 18	0.0508
Flicka Höger Tand B Knä 2	Under 18	0	Pojke Höger Tand B Knä 2	Under 18	0
Flicka Höger Tand B Knä 3	Under 18	0	Pojke Höger Tand B Knä 3	Under 18	0
Flicka Höger Tand B Knä 4	Under 18	0	Pojke Höger Tand B Knä 4	Under 18	0
Flicka Höger Tand B Knä 5	Under 18	0	Pojke Höger Tand B Knä 5	Under 18	0
Flicka Höger Tand B Knä 6	Under 18	0	Pojke Höger Tand B Knä 6	Under 18	0
Flicka Höger Tand A Knä 2	Under 18	0.0002	Pojke Höger Tand A Knä 2	Under 18	0.0176
Flicka Höger Tand A Knä 3	Under 18	0.0152	Pojke Höger Tand A Knä 3	Under 18	0.0379
Flicka Höger Tand A Knä 4	Under 18	0.116	Pojke Höger Tand A Knä 4	Under 18	0.1497
Flicka Höger Tand A Knä 5	Under 18	0.4538	Pojke Höger Tand A Knä 5	Under 18	0.4318
Flicka Höger Tand A Knä 6	Under 18	0.4957	Pojke Höger Tand A Knä 6	Under 18	0.276

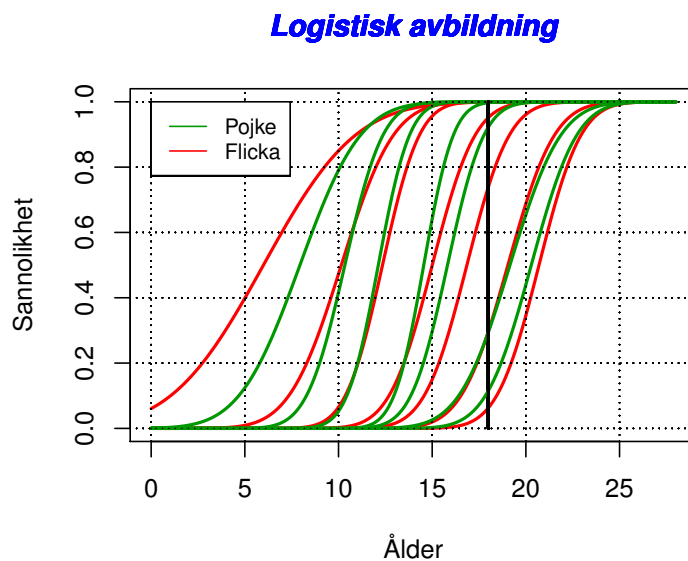
B.2 Andra resultat från probitmodellen



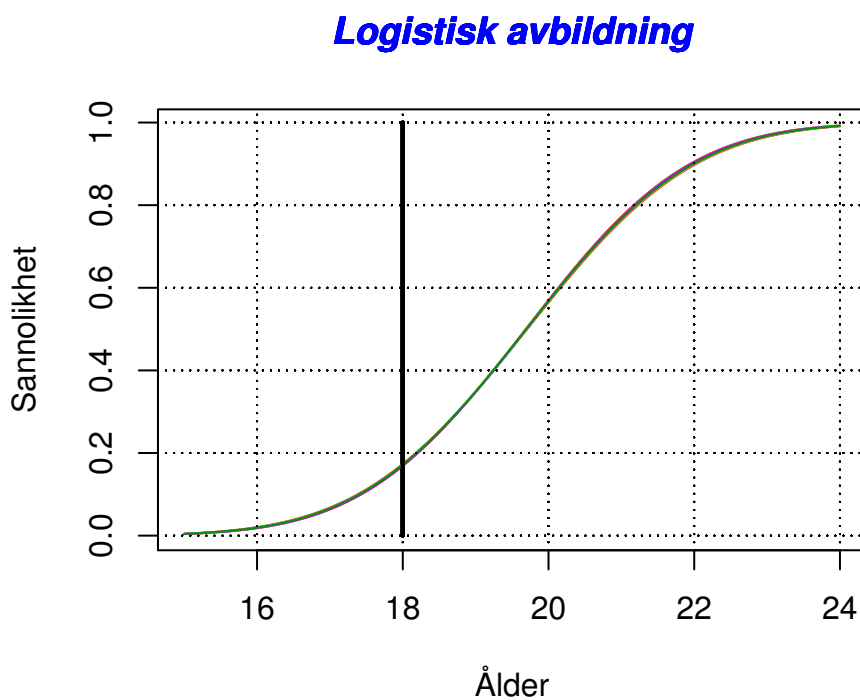
Figur B.1: I figuren visas logistisk avbildning för mäns vänster visdomstanddata för tillhörighet till mognadsstadium H eller inte. De två kurvorna är avbildningar där den ena är med standardnormalfördelningen från och den andra med gammafördelning från samma model som probit. Båda kurvorna bygger på samma dataset framställt från Basin-hopping. Från figuren syns det att de kumulativa fördelningar betar sig väldigt lika. Egenskaperna mellan fördelningarna skiljer sig generellt. Normalfördelning är symmetrisk vilket inte är fallet för gammafördelningen. Gammalfördelningen har däremot egenskapen att bete sig som en normalfördelning för små standardavvikelser. Resultatet säger därför mer om fördelningarna en om datasettet.



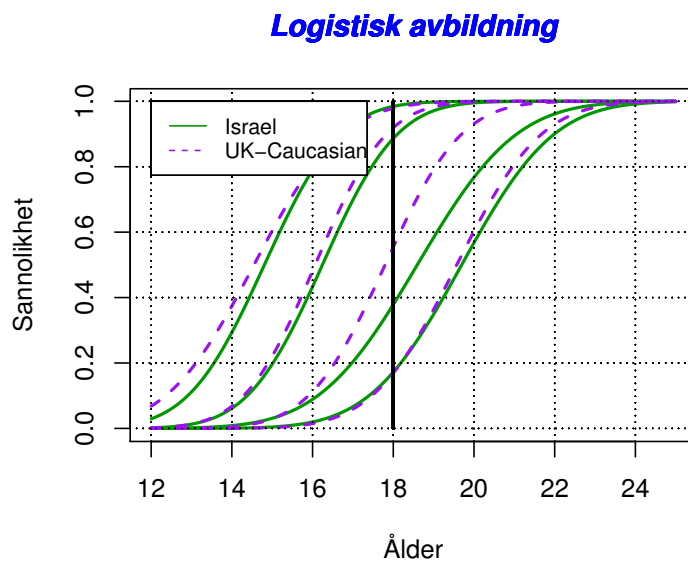
Figur B.2: I figuren visas probits avbildning för Vieth-datans olika mognadsstadier från 3 eller mindre till 5 eller större.



Figur B.3: I figuren visas probits avbildning för Israel-datans höger visdomstands olika mognadsstadier från $A - H$.



Figur B.4: I figuren visas probits avbildning för pojkars vänster visdomstanddata för tillhörighet till mognadsstadium H eller inte. Här är 15 olika dataset framställda med Basin-hopping för att urskilja om fördelningen saknar entydighet då dataseten gör det.



Figur B.5: I figuren visas probits avbildning för Israel-datans och UK-Caucasian olika mognadsstadier från D eller mindre till H .

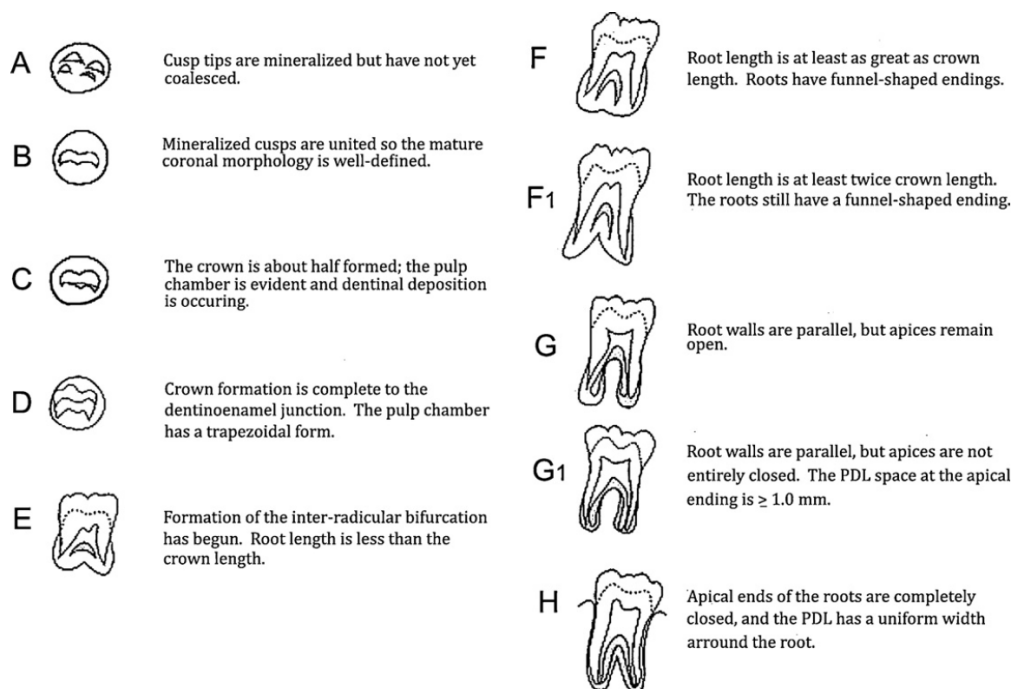
Bilaga C

Mognadsstadier

I rapporten presenteras siffror och bokstäver som motsvarar ett mognadsstadium enligt antingen Demirjians metod, Krämers definition och Vieths skala. Följande bilaga handlar om att ge bakgrund till vad varje mognadsstadium definieras som, istället för att endast betraktas som en bokstav eller siffra. Det skall vara klart att arbetet *inte* bedömer vad framställd röntgenbild på visdomständer eller MR-bild på knäled, motsvarar för mognadsgrad. Det ansvaret ligger hos expert i form av exempelvis radiolog, som nämndes i introduktion.

C.1 Demirjians metod

En metastudie publicerades av Olze et al. (2004)[21] i syftet att avgöra giltigheten av fem vanliga åldersklassificeringssystem för visdomständer: Gleiser och Hunt (1955); Demirjian et al. (1973); Gustafson och Koch (1974); Harris och Nortje (1984); samt Kullman et al. (1992). Olze et al. fastställde att Demirjians metod fick högst betyg enligt något mått som baserades hur väl observerat stadium stämmer med det faktiska stadium, samt kunde hitta ett samband mellan stadium definierat av metod och faktiska ålder.

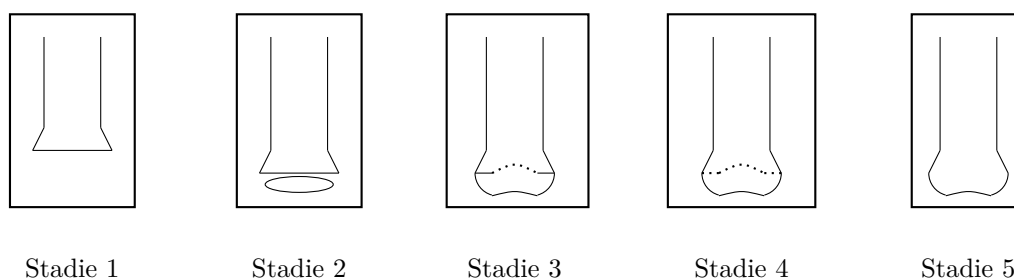


Figur C.1: Mognadsstadier enligt Demirjian metod, bilden är typisk i tidsskrifter som handlar om tandmognad och ålder.

C.2 Krämers definition

Användandet av visdomständer som mätobjekt för att åldersbedöma är välbeprövat under lång tid och har därför stark forsknings underlag, som påpekades i föregående avsnitt, C.1. Eftersom åldersbedömning har blivit mer aktuellt så är det rimligt att använda ytterligare mätobjekt som underlag för beslut. RMV har valt att använda MR-bilder av knäled som en metod för att förstärka bedömningen för pojkar. Metoden har än så länge visat sig vara för osäker för flickor, eftersom att flickors skelett är färdigmoget tidigare än pojkars. Metoden publicerades av Krämer et al.(2014)[14] och går bra att åldersbedöma pojkar med. Stickprovet, $n_{man} = 166$, i tidsskiften är väldigt obalanserat fördelat mellan knästadierna, 70% av observationerna är på stadie 4, alltså redan vuxna män.

Skalan inte rekommenderad att användas enligt the Study Group on Forensic Age Diagnostics (AG-FAD) [36].



Figur C.2: Olika tillväxtstadier för knäleden

Steg 1 Epifysen¹ har inte börjat förbena sig.

Steg 2a Längden av epifysens förbening är en tredjedel eller mindre jämfört med bredden på metafysens ände.

Steg 2b Längden av den förbenade epifysen är mellan en till två tredje delar jämfört med bredden av de benändarnas metafysdelar.

Steg 2c Längden av den förbenade epifysen är över två tredje delar jämfört med bredden av metafysen.

Steg 3a Epifys-metafys fusionen har fullbordat mellan en tredjedel eller mindre av det tidigare gapet mellan epifysen och metafyserna.

Steg 3b Epifys-metafys fusionen har fullbordat mellan en och två tredjedelar av det tidigare gapet mellan epifysen och metafyserna.

Steg 3c Epifys-metafys fusionen har fullbordat över två tredjedelar av det tidigare gapet mellan epifysen och metafyserna.

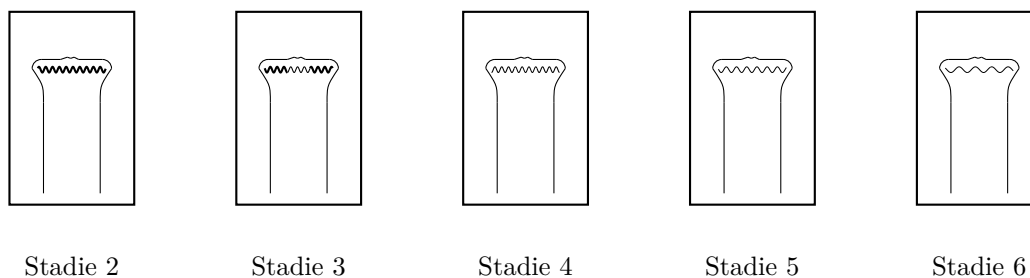
Steg 4 Det epifyserade brosket är nu fullt förbenat och ärret från epifysen är synligt.

Steg 5 Det epifyserade brosket har nu sammanväxt helt och ärret från epifysen är inte längre synligt.

¹Epifysis är den rundade änden av ett långt ben.

C.3 Vieths skala

Det skall observeras att till skillnad från föregående två skalor så är Vieths väldigt ny, den hade ännu inte publicerats när det här projektet påbörjades. Eftersom den mäter knäled och definierar liknande men fortfarande olika steg från Krämers så har den valts att undersökas. Som påpekas i 5 ger Vieths skala tydligare steg än Krämers definition gör för flickors knäledsmognadstadium, skillnaden är likvärdig för pojkar. [36]. De olika stadierna i figur C.3 är beskrivna i [36].



Figur C.3: Mognadsstadietutveckling enligt Vieths skala