# Sammenfatning

Udledningen af asymptotiske resultater inden for parametrisk statistisk inferens kan normalt henføres til en af de følgende tre kategorier:

(1) Der gives et stringent bevis for resultatet på basis af en liste af specificerede regularitetsbetingelser.

(2) Modellen begrænses til en klasse af 'pæne' modeller, fx klassen af (krumme) eksponentielle familier, og resultatet bevises stringent på basis af nogle få ekstra specificerede betingelser.

(3) Der gives et 'heuristisk bevis', som hævdes at holde under passende (ikke specificerede) regularitetsbetingelser.

Der er fordele og ulemper ved alle disse tre metoder. Fordelen ved den sidstnævnte er, at teorien hurtigere udvikles, når stringente beviser undlades. Ofte verificeres sådanne heuristisk opnåede resultater senere i form af stringente beviser. Ulempen er, at det kan være svært at vurdere hvilke af de heuristiske resultater, der således kan verificeres, og hvilke der ved en grundigere undersøgelse viser sig ikke at være korrekte.

Når det er muligt at benytte metode (2), bevares fordelen ved den heuristiske metode herved, idet de ekstra betingelser ofte er ganske simple. Samtidigt opnås præcise resultater. Den eneste ulempe synes at være indskrænkningen af resultatet til den betragtede klasse af modeller.

De åbenlyse fordele ved metode (1) modvirkes i nogen grad af omfanget af de beviser, der kræves, og især af, at de udledte betingelser ofte kan være særdeles vanskelige at verificere i konkrete anvendelser.

Formålet med nærværende afhandling er at udvide anvendeligheden af metode (2) ved at introducere en klasse af modeller (de analytiske modeller), som er tilstrækkelig 'pæn' til at opfylde regularitetsbetingelser af den type, der typisk kræves opfyldt i den asymptotiske teori, og samtidigt tilstrækkelig omfattende til at rumme et bredt spektrum af almindeligt anvendte statistiske modeller, herunder de (tilstrækkeligt glatte) krumme eksponentielle familier. Herved vil det være muligt ved udledningen af en række asymptotiske resultater at kombinere fordelene fra de tre ovennævnte metoder mod en, i mange tilfælde ubetydelig, indskrænkning af den betragtede klasse af modeller, forudsat at et tilstrækkeligt udvalg af tekniske resultater er bevist en gang for alle for denne klasse af modeller, således at udledningen af de asymptotiske resultater simplificeres. Det skal dog understreges, at teorien er begrænset til likelihood baseret parametrisk statistisk inferens, omend visse af metoderne er mere generelt anvendelige.

Indholdet af afhandlingen er derfor koncentreret om følgende fire hovedpunkter: at definere klassen af analytiske modeller, at udlede basale matematiske og sandsynlighedsteoretiske egenskaber for disse, at demonstrere bredden af denne klasse af modeller sammenholdt med almindeligt anvendte statistiske modeller, og at påvise dens anvendelighed ved udledningen af asymptotiske resultater. Derimod har det ikke været meningen her at videreudvikle den asymptotiske teori som sådan eller at give en mere eller mindre fyldestgørende oversigt over denne inden for rammen af den nye klasse af modeller.

I kapitel 1 indføres notationen, specielt i relation til multilineære afbildninger, og nogle matematiske og sandsynlighedsteoretiske resultater resumeres til brug for de senere kapitler. Afsnit 1 omhandler multilineære afbildninger mellem endeligt-dimensionale vektorrum, afsnit 2 omhandler differentiabilitet af afbildninger mellem sådanne rum, og i afsnit 3 defineres de analytiske afbildninger. Disse afsnit indeholder udelukkende velkendt matematisk teori, omend notationen afviger fra den, der normalt benyttes i statistisk litteratur. I afsnit 4 indføres momenter og kumulanter af flerdimensionale stokastiske variable som multilinearformer på det duale rum. I afsnit 5 bevises et par resultater vedrørende multilineære afbildninger, blandt andet det basale matematiske resultat, som ligger til grund for beviset for hovedsætningen i afhandlingen (sætning 2.4.2).

Kapitel 2 indeholder definitionen af, og den basale teori for klassen af analytiske modeller. Især er denne teori koncentreret i afsnittene 2–6, hvorimod afsnittene 7 og 8 er udløbere af teorien af potentiel interesse. Definitionen af de analytiske modeller er givet i afsnit 2. Den vedrører statistiske modeller parametriseret ved en parameter, der antager værdier i et endeligt-dimensionalt vektorrum. Den ene hovedbestanddel i definitionen er, at tætheden skal være positiv og *analytisk* som funktion af parameteren på et rum med sandsynlighed 1. Dette rum må ikke afhænge af parameterværdien, og modeller, hvori støtten for fordelingen afhænger af parameteren, er derfor ikke analytiske. Den anden hovedbestanddel af definitionen er en uniform momentbetingelse på de (uendeligt mange) afledede af log-likelihood funktionen (betingelse (iv) i definition 2.2.1). Afsnit 3 indeholder nogle basale tekniske resultater for de analytiske familier, fx bevises det, at den almindeligt benyttede ombytning af differentiation og integration af tæthedsfunktionen er legitim. Afsnit 4 indeholder hovedsætningen vedrørende klassen af analytiske modeller i form af nogle ækvivalente definitioner af denne klasse. Resultatet er, at den ovenfor nævnte uniforme momentbetingelse kan erstattes af en betingelse på kumulanterne til de afledede af log-likelihood funktionen. Denne betingelse fører frem til definitionen af en størrelse kaldt modellens *indeks*, eller mere udførligt modellens indeks for lineær normalitet, defineret i afsnit 5. Dette indeks er et ikke-negativt tal, som er nul, hvis og kun hvis modellen er en lineær normalfordelingsmodel med kendt varians. I overensstemmelse med at asymptotiske resultater normalt baserer sig på en approksimation af modellen med en sådan model, viser det sig at modellens indeks kan benyttes som et mål for fejlen ved denne form for approksimation. Denne påstand retfærdiggøres i kapitel 4; i første omgang vises det at modellens indeks er proportionalt med $1/\sqrt{n}$ for en model for $n$ uafhængige identisk fordelte observationer. I afsnit 6 vises, at analytiske omparametriseringer, sufficiente afbildninger af udfaldsrummet, og visse andre typer af operationer på analytiske modeller igen fører til analytiske modeller. Disse resultater er af central betydning ved udledningen af asymptotiske resultater, som det fremgår af kapitel 4. I afsnit 7 vises at en analytisk model kan approksimeres med en krum eksponentiel familie, hvorved fejlen begrænses af et udtryk, som først og fremmest afhænger af modellens indeks. I afsnit 8 vises at enhver analytisk model lokalt, omkring en fast parameterværdi, kan indlejres i en uendeligt-dimensional eksponentiel familie, og visse egenskaber ved den herved frembragte eksponentielle familie udledes.

Kapitel 3 indeholder eksempler på modeller og typer af modeller, som er analytiske, samt nogle eksempler på modeller, der ikke er analytiske. I dette kapitel godtgøres det, at eksponentielle familier er analytiske modeller og at analytisk parametriserede delfamilier af eksponentielle familier ligeledes er analytiske. Desuden fremgår det, at mange andre modeller tilhører klassen af analytiske modeller, og eksemplerne vedrørende modeller med ukendt positions- og skalaparameter antyder, at mange transformationsmodeller også er analytiske, omend dette emne ikke forfølges nærmere. Et eksempel baseret på Weibull fordelingen med ukendt potensparameter demonstrerer grænsen for de analytiske modeller ved at være et eksempel på en model, der uden at være analytisk besidder mange 'pæne' egenskaber.

I kapitel 4 demonstreres anvendeligheden af de analytiske familier ved udledningen af asymptotiske resultater i relation til statistisk inferens. Kun første ordens asymptotik betragtes i dette kapitel, hvori til gengæld vilkårlige følger af analytiske modeller behandles, således at resultaterne også er anvendelige for modeller for afhængige eller ikke identisk fordelte stokastiske variable. Hovedresultaterne, som vises i afsnittene 2–4, kan lidt upræcist udtrykkes ved, at de 'sædvanlige' asymptotiske resultater vedrørende maksimum likelihood estimation gælder for en vilkårlig følge af analytiske modeller, for hvilke modellernes indeks går mod nul. Specifikt vises blandt andet den asymptotiske normalitet af fordelingen af maksimum likelihood estimatoren. I resten af dette kapitel udledes betingelser for at modellernes indeks går mod nul i en række konkrete typer af modeller. Disse modeller tilhører enten klassen af generaliserede lineære modeller, som er mere præcist defineret i afsnit 5, eller klassen af generaliserede ikke-lineære modeller, som beskrives generelt i afsnit 12. Betydningen af disse eksempler er at vise, dels hvor let asymptotiske resultater udledes for disse typer af modeller, dels generaliteten af de i de første fire afsnit udledte resultater, idet der er tale om følger af ikke-identisk fordelte stokastiske variable.

I definitionen af de analytiske modeller indgår der betingelser vedrørende den uendelige følge af afledede af log-likelihood funktionen, og det er derved sikret, at også højere ordens asymptotiske resultater kan udledes for analytiske modeller. Dette er gjort i kapitel 5 for følger af uafhængige identisk fordelte observationer i afsnittene 2 og 3. Det primære formål med dette kapitel er at retfærdiggøre påstanden om, at analytiske modeller er 'regulære' i den forstand, at de opfylder almindelige regularitetsbetingelser fra asymptotisk statistisk teori. Derfor er resultaterne i afsnittene 3 og 4 vist ved, under henvisning til publicerede beviser for asymptotiske resultater, at godtgøre at betingelserne for disse resultater er opfyldt for analytiske modeller. Visse modifikationer af mindre betydning er dog nødvendige på grund af, at den valgte ramme for modellerne i de publicerede resultater ikke stemmer eksakt med den her benyttede. I afsnit 3 vises på denne måde, at fordelingen af maksimum likelihood estimatoren kan udvikles i en Edgeworth række, og i afsnit 4 vises konsistensen af maksimum likelihood estimatoren på et kompakt parameterrum.