



Universiteit
Leiden
The Netherlands

Modelling long term survival with non-proportional hazards

Perperoglou, A.

Citation

Perperoglou, A. (2006, October 18). *Modelling long term survival with non-proportional hazards*. Retrieved from <https://hdl.handle.net/1887/4918>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4918>

Note: To cite this publication please use the final published version (if applicable).

Sammenvatting

De toepassing van statistische modellen in medisch onderzoek is een belangrijk onderdeel van de statistiek. Dit proefschrift behandelt modellen voor de analyse van medische overlevingsduurgegevens. Het bevat een inleiding en vijf wetenschappelijke artikelen, waarvan drie gepubliceerd zijn en twee aangeboden aan wetenschappelijke tijdschriften, en een conclusie en een appendix. De inleiding geeft de benodigde voorkennis over de kern van de overlevingsduuranalyse, namelijk het Cox model. Daarbij worden ook de belangrijkste ideeën geïntroduceerd die ten grondslag liggen aan de uitbreidingen van dit model die gepresenteerd worden in het vervolg van dit proefschrift.

Hoofdstuk 2 behandelt gereduceerde rang-methoden voor het modelleren van niet-proportioneel risico. Het principe van gereduceerde rang-regressie wordt hier toegepast op overlevingsduuranalyse en tot in detail uitgewerkt. Het hoofdstuk presenteert methoden om de parameters van het model te schatten en behandelt de vraag hoe de rang en de gebruikte tijdsfuncties gekozen moeten worden. Deze methoden worden vervolgens toegepast op een kleine dataset van 358 ovariumtumorpatiënten. De resultaten worden vergeleken met uitkomsten van het Cox model en het *gamma frailty* model.

In hoofdstuk 3 komt een efficiënt algoritme aan de orde dat zowel gebruikt kan worden voor het schatten van gereduceerde rang-modellen als voor het schatten van Cox modellen met tijdafhankelijke effecten van de covariaten. De rekenkundige uitdagingen die deze modellen met zich meebrengen worden uitgelegd in de inleiding, waarna het schattingsalgoritme wordt geïntroduceerd. Op basis van dit algoritme is een softwarepakket voor R geschreven, dat beschikbaar is vanaf de website van de auteur. In het hoofdstuk wordt dit programma vergeleken met reeds beschikbare routines voor **S-plus** en **SAS**, op basis van een gesimuleerde dataset. Bovendien worden gereduceerde rang-methoden geïllustreerd aan de hand van dezelfde dataset die in hoofdstuk 1 gebruikt werd en vervolgens met behulp van een grotere dataset van 2433 borstkankerpatiënten.

Het volgende hoofdstuk gaat in op het *gamma frailty* (Burr) model, en de beperkingen die de aanname van constante *frailty* oplegt. De aannemelijkheid van deze aanname wordt kort ter discussie gesteld. Vervolgens worden in dit hoofdstuk alternatieve modellen besproken waarbij de *frailty* tijdsafhankelijk

is. De meeste van deze modellen gebruiken echter complexe mathematische uitdrukkingen die het schatten moeilijk maken. Hierdoor zijn deze modellen beperkt bruikbaar in toegepast statistisch onderzoek. Een generalisatie van het Burr model kan daarom een bruikbaar alternatief zijn, omdat het op een flexibele manier convergerend en niet-convergerend risico kan modelleren. Deze generalisatie wordt gedetailleerd beschreven en geëvalueerd met behulp van gesimuleerde data.

Het gegeneraliseerde Burr model wordt vervolgens toegepast op verschillende datasets. Bij een onderzoek naar patiënten die aan longkanker lijden vertoonde het gefitte model convergerend risico; in dit geval waren de resultaten vrijwel gelijk aan het eenvoudige Burr model. In een tweede toepassing van het model op data van een grootschalige registratie van patiënten met acute myeloïde leukemie toonde het gegeneraliseerde model aan dat het risico een complexer patroon liet zien. Het risico convergeerde gedurende de eerste vier jaar, maar divergeerde daarna tot proportionaliteit.

Een belangrijk aspect van statistisch onderzoek is de vraag welk model het meest passend is voor een bepaald type data en hoe de kwaliteit van aangepaste modellen te evalueren. In hoofdstuk 5 worden verschillende modelleringstrategieën vergeleken: modellen met tijdsafhankelijke effecten van de covariaten, *frailty*-modellen, en herstelmodellen (*cure rate models*) passeren ieder de revue als mogelijke aanpak voor het modelleren van de dataset van 2433 borstkankerpatiënten die al in hoofdstuk 3 aan de orde kwam. Deze data worden hier opnieuw gedetailleerd beschreven, voordat zeven verschillende modellen erop worden toegepast. De bruikbaarheid van elk van deze modellen wordt bekeken en de resultaten vergeleken. Om de voorspellende kracht van de benaderingen in kaart te brengen worden pseudo-observaties en Brier scores gebruikt. Deze twee maten worden bovendien samengenomen, waarmee een nieuwe manier ontstaat om het voorspellend vermogen van modellen te bestuderen.

Het zesde hoofdstuk van dit proefschrift behandelt manieren om met overdispersie in gegeneraliseerde lineaire modellen om te gaan. Een nieuwe klasse van modellen wordt geïntroduceerd, waarin een extra vector van parameters wordt toegevoegd met één nieuwe parameter voor elke individuele observatie-eenheid. Om deze parameters te schatten wordt er een penalty toegevoegd aan de likelihood van het model. Deze penalty is nodig om identificeerbaarheid te garanderen, om collineariteit te verminderen in de schattingvergelijkingen en om de effectieve dimensie van het model te reduceren. De benodigde theorie wordt uitgewerkt voor Poisson- en binomiaal verdeelde data, maar ook voor sterftetabellen. Verschillende toepassingen worden gegeven. Een

efficiënt algoritme voor het uitrekenen van de deviantie-effecten wordt in detail uitgewerkt.

De conclusie stippelt lijnen uit voor toekomstig onderzoek op het gebied van dit proefschrift. Tenslotte is er een appendix die toelicht hoe het pakket `coxvc` gebruikt kan worden. Dit pakket voor R werd geschreven om Cox modellen met tijdsafhankelijke effecten van covariaten en gereduceerde rang-modellen toe te kunnen passen.

