



Universiteit  
Leiden  
The Netherlands

## Bayes and Networks

Gao, F.

### Citation

Gao, F. (2017, May 23). *Bayes and Networks*. Retrieved from <https://hdl.handle.net/1887/49012>

Version: Not Applicable (or Unknown)

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

Downloaded from: <https://hdl.handle.net/1887/49012>

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

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/49012> holds various files of this Leiden University dissertation.

**Author:** Gao, F.

**Title:** Bayes and networks

**Issue Date:** 2017-05-23

Dit proefschrift bestaat uit vijf artikelen, met twee thema's: Bayes en netwerken; en drie verschillende onderwerpen.

Deel I van het proefschrift behandelt een inverse probleem uit de niet-parametrische Bayesiaanse statistiek. Niet-parametrische Bayes is de statistische methodologie om, gebruikmakend van a-priori kennis, niet-parametrische objecten te identificeren, zoals dichtheidsfuncties, verdelingen en andere objecten die niet met eindig veel parameters kunnen worden beschreven. Inverse problemen refereren naar de situatie dat geen directe waarnemingen van het doel-object beschikbaar zijn, en het object moet worden gereconstrueerd uit indirecte data die gecontamineerd zijn met ruis. In Hoofdstuk 1 bestuderen we de schatting van de mengverdeling uit waarnemingen met Laplace ruis verdeeld door de mengverdeling te voorzien van een Dirichlet proces a-priori verdeling. We leiden een contractiesnelheid af voor de corresponderende a-posteriori verdeling, zowel voor de mengverdeling relatief ten opzichte van de Wasserstein metriek als voor de mengseldichtheid relatief ten opzichte van de Hellinger en  $L_q$  metriekeken. Opmerkelijk is dat het resultaat de gebruikelijke interpretatie tegensprekt van een Laplace mengsel als een 1-Hölder functie, dat een langzamere snelheid dan hier bewezen suggereert.

In Deel II van het proefschrift bestuderen we de statistische schatting van voorkeurs-aanhechting netwerken. Het voorkeurs-aanhechting netwerk is een populair model voor dynamische netwerken die groeien door het toevoegen van nieuwe knooppunten, welke volgens een voorkeurs-mechanisme worden verbonden met bestaande knooppunten. Het model biedt een mogelijke verklaring voor schaal-vrije netwerken, die overal voorkomen. De voorkeurs-aanhechting functie, welke de graad van een knooppunt naar de voorkeurskans afbeeldt, bepaalt het gedrag van het model. We bestuderen het inverse probleem van het vinden van deze functie uit data aangaande een waargenomen netwerk. We nemen bijvoorbeeld een sociaal netwerk waar, en willen het mechanisme bepalen dat verantwoordelijk is voor de groei van het netwerk.

Hoofdstuk 2 is een inleiding in Deel II. Allereerst schetsen we het belang van de relatief nieuwe discipline van netwerk wetenschap en beschrijven in het kort hoe deze wetenschap tot bloei is gekomen. Daarna zetten we de geschiedenis van het voorkeurs-aanhechting netwerk uiteen, en geven een precieze wiskundige formulering van het model.

In Hoofdstuk 3 introduceren we een empirische schatter van het

algemeen sublineaire voorkeurs-aanhechting model. We beschrijven een superkritisch CMJ vertakkingsproces raamwerk, dat in zekere zin op dezelfde manier evolueert als het voorkeurs-aanhechting netwerk. Binnen het vertakkingsproces raamwerk geven we een bewijs van de bijna-zekere consistentie van de empirische schatter, gebruikmakend van klassieke resultaten over vertakkingsprocessen. In een numerieke studie illustreren we de eigenschappen van de schatter.

In Hoofdstuk 4 beschouwen we de schatting van de affiene parameter en de machtswet exponent in het voorkeurs-aanhechting model met stochastische initiële graden. Dit model combineert de effecten van “de rijken worden rijker” (door het voorkeurs-aanhechting mechanisme) en “rijk door geboorte” (door de stochastische initiële graden) en is voldoende flexibel als model voor netwerken in de wereld om ons heen. We leiden de aannemelijkheidsfunctie af en laten zien dat de meest aannemelijke schatter asymptotisch normaal en efficiënt is. We introduceren ook een quasi-meest aannemelijke schatter die niet afhankelijk is van het waarnemen van de rij initiële graden. We illustreren de kracht van de schatter in een simulatie.

In Hoofdstuk 5 keren we terug naar het algemene sublineaire voorkeurs-aanhechting model en nemen aan dat de voorkeurs-aanhechting functie een parametrische vorm bezit, d.w.z. de voorkeurs-aanhechting functie is bepaald door eindig veel parameters. Gebruikmakend van het superkritische vertakkingsproces raamwerk van Hoofdstuk 3 en relevante klassieke resultaten, bestuderen we de asymptotische limiet van de aannemelijkheidsfunctie, en laten zien dat de meest aannemelijke schatter asymptotisch normaal is. We bediscussiëren ook een remedie voor de afhankelijkheid van de aannemelijkheidsfunctie van de geschiedenis van het netwerk, leidend tot een schatter die nog steeds asymptotisch normaal is.

Het laatste deel, Deel III, van het proefschrift handelt over het modelleren van het film-acteur netwerk door het voorkeurs-aanhechting netwerk model. Op basis van de vrij beschikbare data van de Internet Movie Database (IMDb) ontwerpen we een nieuw model met twee lagen—één laag van acteurs en een tweede laag van films, waarin de knooppunten tussen de lagen verbonden worden als een acteur speelt in een film en binnen de laag van acteurs als twee acteurs in dezelfde film spelen. In het model heeft iedere film een stochastisch aantal acteurs en moet een stochastische fractie acteurs kiezen uit het bestaande acteurs netwerk met voorkeur voor acteurs die in veel films hebben gespeeld. We ontdekken dat de graadverdeling van het netwerk een machtswet volgt en passen het model aan de IMDb data aan. We simuleren de evolutie van netwerken volgens het aangepaste model en vinden dat het nieuwe model een realistische fit geeft.