



Universiteit
Leiden
The Netherlands

E-values for anytime-valid inference with exponential families

Hao, Y.

Citation

Hao, Y. (2025, February 18). *E-values for anytime-valid inference with exponential families*. Retrieved from <https://hdl.handle.net/1887/4195433>

Version: 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/1887/4195433>

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

Samenvatting

Dit proefschrift richt zich voornamelijk op statistisch hypothesesetoetsen, een onderzoeksgebied met brede relevantie in verschillende academische disciplines en industrieën. Een voorbeeld dat laat zien hoe belangrijk dit is het ontwikkelen van nieuwe geneesmiddelen. Stel dat onderzoekers een nieuw medicijn aandragen dat bedoeld is om de bloeddruk te verlagen. De nulhypothese zou kunnen stellen dat het nieuwe medicijn geen effect heeft op de bloeddruk, terwijl de alternatieve hypothese suggereert dat het medicijn de bloeddruk wél verlaagt.

De onderzoekers voeren een klinische studie uit waarbij een groep het nieuwe medicijn krijgt en een andere groep een placebo. Stel dat na het verzamelen van de gegevens een verlaging van de bloeddruk in de eerste groep wordt waargenomen. Met een hypothesesetoets kan vervolgens (statistisch) uitgesloten worden dat de waargenomen verlaging op toeval berust. Als de toets aangeeft dat de bevinding statistisch significant is, dan concluderen de onderzoekers dat het medicijn effectief is. Zonder hypothesesetoetsing zouden ze niet in staat zijn om rigoureus te beoordelen of het medicijn echt werkt of dat de waarnemingen slechts het gevolg zijn van willekeurige variaties. Door een kader te bieden om fouten (zoals fout-positieve resultaten) te beheersen, zorgt hypothesesetoetsing ervoor dat de meeste medicijnen die goedgekeurd worden ook daadwerkelijk effectief zijn, wat cruciaal is voor de volksgezondheid en de vooruitgang van de medische wetenschap.

De meeste klassieke methoden voor hypothesesetoetsing vereisen echter dat onderzoekers – voordat de toets wordt uitgevoerd – een (vaste) steekproefgrootte bepalen. Nadat de steekproef van de vooraf vastgestelde grootte is verzameld, wordt de toets uitgevoerd en worden conclusies getrokken over het al dan niet verwerpen van de nulhypothese.

Deze vaste steekproefgroottebenadering kent verschillende beperkingen:

1. **Vastgestelde steekproefgrootte:** Onderzoekers moeten van tevoren beslissen

hoeveel gegevenspunten ze verzamelen, wat kan leiden tot studies waarbij het onderscheidend vermogen van de toets of te laag, of te hoog is.

2. **Geen tussentijdse analyse:** In klassieke hypothesetoetsing mogen onderzoekers doorgaans niet naar de gegevens kijken terwijl deze binnenkomen (om bevooroordeelde beslissingen te voorkomen) en moeten zij wachten tot de volledige steekproef beschikbaar is.

3. **Inflexibiliteit:** Als onverwachte resultaten optreden of als de steekproefgrootte na gegevensverzameling onvoldoende blijkt te zijn, kunnen onderzoekers de steekproefgrootte niet eenvoudig aanpassen zonder het risico te lopen de kans op een fout-positief resultaat substantieel te vergroten.

Ondanks deze beperkingen vormt de vaste steekproefgrootte-benadering al vele decennia de basis van statistische toetsing, en wordt deze nog steeds zeer breed toegepast. Moderne methoden, zoals zogenoemde altijd-valide toetsen (bijvoorbeeld op basis van e-waarden en e-processen), bieden echter meer flexibiliteit. Ze stellen onderzoekers in staat om het bewijs, gekwantificeerd in termen van de e-waarde, continu te evalueren terwijl gegevens worden verzameld, zonder een steekproefgrootte vooraf te specificeren. Voor e-processen hoeven zelfs de regels voor het stoppen van het experiment niet vooraf te worden bepaald.

Dit proefschrift richt zich voornamelijk op e-waarden en e-processen in de context van exponentiële families.

Hoofdstuk 2 behandelt het probleem van het bepalen of een steekproef verdeeld is volgens een distributie van een specifiek model binnen de exponentiële familie, in wezen het toetsen of een model correct is gespecificeerd. We willen bijvoorbeeld beoordelen of een steekproef een Gaussische verdeling volgt. In dit geval omvat de nulhypothese de volledige set Gaussische verdelingen, waardoor de nulhypothese composiet is. Het doel is de nulhypothese te verwerpen als er overtuigend bewijs is dat de steekproef niet verdeeld is volgens een Gaussische verdeling. We richten ons op de GRO (Growth-Rate Optimal) e-variabele, die doorgaans overeenkomt met een specifieke Bayes-factor en het hoogst onderscheidende e-vermogen heeft (d.w.z. het vermogen om alternatieven te detecteren). Het vinden van de a priori verdelingen voor de GRO e-variabele kan echter computationeel intensief zijn. Dit hoofdstuk toont aan dat in bepaalde situaties, de zogenoemde “eenvoudige gevallen”, de GRO e-variabele vereenvoudigt tot een aannemelijkheidsverhouding en biedt verschillende equivalenten voorwaarden waaronder een dergelijke aannemelijkheidsverhouding bestaat voor exponentiële familie-nulhypothesen. In dit hoofdstuk worden ook (GRO) e-processen afgeleid welke gebruikt kunnen worden om een altijd valide toets te construeren.

Hoofdstuk 3 breidt het werk van Hoofdstuk 2 uit door meer algemene theoretische resultaten te bieden voor verschillende e-variabelen in het kader van modelspecificatietoetsen, zowel voor eenvoudige als compositie hypothesen. Het toont aan dat in het “anti-eenvoudige geval” (het tegenovergestelde van het eenvoudige geval), de conditionele e-waarde asymptotisch het hoogst onderscheidende e-vermogen bereikt. Dit is bijzonder waardevol omdat de GRO e-variabele moeilijk te berekenen is in dergelijke gevallen, terwijl de conditionele e-variabele computationeel werkbaar is. Daarnaast introduceert en vergelijkt dit hoofdstuk verschillende soorten e-waarden, waaronder de GRO e-variabele, de conditionele e-variabele, de UI (Universal Inference) e-variabele en sequentiële e-variabelen, met een uitgebreide analyse van hun onderscheidende e-vermogen. Sommige van deze e-waarden vormen de basis van een e-proces, dat dan weer leidt tot een altijd valide toets.

Hoofdstuk 4 richt zich op het construeren van e-variabelen die gebruikt kunnen worden om te toetsen of k steekproeven allen hetzelfde verdeeld zijn volgens een verdeling van de exponentiële familie. Bijvoorbeeld, in het geval van twee-steekproeventoets ($k = 2$) kunnen deze methoden worden toegepast om te beoordelen of een nieuwe medische behandeling effectief is voor een bepaalde ziekte door de uitkomsten tussen de behandelings- en controlegroep te vergelijken. We introduceren vier soorten e-variabelen voor de k -steekproeventoets: de GRO e-variabele, een conditionele e-variabele, een gewogen middeling van e-variabelen en een pseudo-e-variabele. Deze e-variabelen worden vergeleken op basis van hun groeisnelheden onder alternatieve hypothesen, waarbij elke groep een andere, maar vaste, verdeling heeft uit dezelfde exponentiële familie. Het proefschrift biedt theoretische resultaten die aantonen dat bij kleine effectgroottes de e-variabelen vergelijkbaar gedrag vertonen. Ook worden gevallen geïdentificeerd waarin een e-variabele vereenvoudigt tot de GRO e-variabele, wat de computationele complexiteit vermindert. In complexere situaties worden algoritmen voorgesteld om de zogenoemde omgekeerde informatieprojectie te benaderen.

Hoofdstuk 5 richt zich op *growth-rate optimal in the worst-case* (GROW) e-variabelen. We analyseren de toepassing van GROW e-variabelen binnen een hypothesetoetsingskader voor multivariate verdelingen. In dit kader heeft de nulhypotheseverdeling P_0 een verwachting van nul, en worden verschillende alternatieve hypothesen $\mathcal{H}_1$ gedefinieerd door middel van verschillende gebieden voor de verwachting.

We laten zien dat de GROW e-variabele verbonden is met een nieuwe concentratieongelijkheid die we de *Csiszár-Sanov-Chernoff* (CSC) ongelijkheid noemen. Deze CSC-ongelijkheid breidt eerder werk uit naar multivariate gevallen met een convexe of begrensde alternatieve parameterregio rond 0. Een dergelijke ongelijkheid is

Samenvatting

waarschijnlijk waardevol in praktische toepassingen, zoals *online* (sequentieel) leren en vooral op het gebied van *bandit*-algoritmen.

Samenvattend: dit proefschrift breidt de theorie van e-variabelen binnen exponentiële families uit door verschillende e-variabelen te construeren en te analyseren, en biedt praktische inzichten voor situaties waarin traditionele p -waarde-gebaseerde toetsing tekortschiet.