

Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor

Koen Derks, Jacques de Swart, Ruud Wetzels

Received 3 December 2021 | Accepted 9 February 2022 | Published 14 March 2022

Samenvatting

Bayesiaanse statistiek is een manier om efficiëntie en transparantie bij gestratificeerde steekproeven te vergroten. De reden hiervoor is dat Bayesiaanse statistiek de auditor dwingt om expliciet te maken welke informatie en kennis gebruikt wordt bij de evaluatie van de steekproef. In dit artikel worden twee belangrijke vormen van voorkennis beschreven die de auditor moet valideren. Vervolgens zijn deze vormen van voorkennis vertaald naar statistische modellen, die worden gebruikt bij het doen van een gestratificeerde steekproef. Afsluitend wordt gedemonstreerd hoe sommige van die modellen kunnen leiden tot nauwkeurigere foutschattingen en een transparantere audit.

Relevantie voor de praktijk

Het stratificeren van steekproeven wordt in de auditpraktijk veelvuldig toegepast. In dit artikel is beschreven hoe twee vormen van voorkennis leiden tot statistische modellen voor de evaluatie van gestratificeerde steekproeven. Deze modellen kunnen resulteren in nauwkeurigere foutschattingen en meer transparantie. Auditors kunnen deze modellen zelf toepassen in de praktijk met behulp van de bijgeleverde R code in de online appendix (zie appendix bij dit artikel).

Trefwoorden

Audit, Bayesiaanse statistiek, stratificatie

1. Inleiding

Stratificatie is het opdelen van een populatie in deelpopulaties, strata genaamd. Belangrijk bij het definiëren van de strata is dat ze elkaar niet overlappen, dat de data binnen een stratum zoveel mogelijk op elkaar lijken (homogeen zijn), en dat de data tussen de strata zo min mogelijk op elkaar lijken (heterogeen). Immers, als een heterogene populatie data in homogeen(re) deelpopulaties opgedeeld kan worden, verhoogt stratificatie de nauwkeurigheid van de statistische foutschatting op populatieniveau (Touw and Hoogduin 2012). Het is daarom in veel gevallen efficiënter om de populatie op te delen in homogene strata en in elk stratum een steekproef te trekken. Merk hierbij op dat het niet toegestaan is om de strata te definiëren op basis van de steekproef zelf; dit dient te gebeuren voordat de steekproef

getrokken wordt. Zo kan bijvoorbeeld tegenwoordig op basis van kenmerken van de posten met behulp van (*un*) *supervised learning* een risicocategorie aan elke post worden toegekend, waarna de posten op basis van risicocategorie worden gestratificeerd. Een ander voorbeeld waarin stratificatie wordt toegepast, is wanneer een auditor uitkomsten van meerdere steekproeven, bijvoorbeeld over filialen, wil consolideren. In deze situaties is stratificatie de *go-to* tool die men gebruikt om zo efficiënt mogelijk met de beschikbare middelen (e.g., tijd, geld) om te gaan.

Hoeveel efficiënter de steekproefevaluatie is door stratificatie toe te passen, hangt af van de bepaling van de strata. De kunst is om de strata zo in te richten dat deze binnen zichzelf zo homogeen mogelijk zijn en tussen

de strata zo heterogeen mogelijk. Met andere woorden: het doel is de strata zo te definiëren dat de data binnen een stratum zoveel mogelijk op elkaar lijken, en zo min mogelijk lijken op de data in andere strata. Echter, het is onwaarschijnlijk dat de data tussen de strata op geen enkele manier op elkaar lijken; de strata komen immers uit dezelfde populatie. Desondanks houden de huidige methoden die binnen de audit worden gebruikt om gestratificeerde steekproeven te evalueren hier geen rekening mee. In de analyse van de data wordt aangenomen dat er geen enkele relatie bestaat tussen de verschillende strata. In veel situaties doet de auditor zichzelf hiermee tekort, want als de strata in enige mate hetzelfde zijn, kan er statistisch gezien informatie met betrekking tot de foutkans gedeeld worden tussen de strata.

In dit artikel wordt een hiërarchische aanpak beargumenteerd. Een onderliggende hiërarchische structuur in de data is bij uitstek relevant bij gestratificeerde steekproeven, waarbij een populatie bestaat uit een aantal strata en elk stratum uit een aantal posten. Er is dus een hiërarchie in de data, die we laten terugkomen in het statistische model. Bovendien kan op deze manier gelijktijdig een nauwkeurige schatting van de foutkans op populatieniveau worden gegeven, alsmede een nauwkeurige schatting van de foutkans op stratumniveau. In dit artikel wordt de totstandkoming van deze modellen besproken en de voorkennis die hieraan ten grondslag ligt.

De modellen in dit artikel worden toegepast in de context van de Bayesiaanse statistiek. Als alternatief voor de klassieke steekproef methodologie kan de Bayesiaanse statistiek voordelen hebben voor de auditpraktijk (Steele 1992; Van Batenburg et al. 1994; Stewart 2013). Bijvoorbeeld, zo kan in de Bayesiaanse statistiek voorkennis¹ over de foutkans in de populatie op een verantwoorde manier worden meegenomen in de evaluatie van de steekproef (Derks et al. 2021a) en is het toegestaan om continu bewijs te monitoren over de tijd (Wagenmakers et al. 2008). Dit is in flagrant contrast met de gebruikelijk frequentistische statistische analyses bij een gegeven steekproefomvang, die hun waarde verliezen bij uitbreiding van de steekproef (Touw and Hoogduin 2012).

De opzet van dit artikel is als volgt. In hoofdstuk 2 worden de kernconcepten uit de Bayesiaanse statistiek besproken. In hoofdstuk 3 wordt verder ingegaan op twee vormen van voorkennis die van toepassing zijn op de evaluatie van gestratificeerde steekproeven en de statistische modellen die hieruit volgen. Om te illustreren welk voordeel te behalen valt met deze statistische modellen, worden in hoofdstuk 4 de schattingen van de modellen besproken middels een eenvoudig voorbeeld. In het laatste hoofdstuk worden conclusies en aanbevelingen aangereikt voor het gebruik van deze modellen in de praktijk.

2. Bayesiaanse statistiek

Bij het evalueren van een steekproef is het doel van de auditor om een uitspraak te doen over een bepaald kenmerk

van een populatie posten, θ . Voor de leesbaarheid van dit artikel nemen we aan dat θ een foutkans is, maar – *mutatis mutandis* – geldt het hier beschrevene ook voor andere kenmerken, zoals het foutbedrag. Over het algemeen inspecteert de auditor niet de gehele populatie maar slechts een deelpopulatie, y , de data die afkomstig zijn van een steekproef uit deze populatie. Als gevolg hiervan moet de informatie uit de steekproef worden geëxtrapoleerd naar de populatie, waardoor onzekerheid en een kansuitspraak over θ worden geïntroduceerd.

De Bayesiaanse manier om een kansuitspraak te doen over het populatiekenmerk θ , gegeven de data y , is via de posterior-verdeling $p(\theta | y)$. De posterior-verdeling wordt gedefinieerd door de stelling van Bayes als het product van twee verdelingen, de prior-verdeling $p(\theta)$ en de *likelihood* functie $l(y | \theta)$, gegeven de data y :

$$p(\theta | y) \propto l(y | \theta) \times p(\theta). \quad (1)$$

Het symbool \propto geeft weer dat de expressie rechts van dit teken op schaling na gelijk is aan de kansverdeling links van dit teken. Zoals vergelijking 1 illustreert, houdt de stelling van Bayes in dat de informatie in de prior-verdeling wordt gecombineerd met de informatie uit de steekproefdata, om tot de posterior-verdeling te komen. In de volgende subsecties wordt kort ingegaan op deze drie fundamentele begrippen uit de Bayesiaanse statistiek: de prior-verdeling, de likelihood en de posterior-verdeling.

2.1. De prior-verdeling

De prior-verdeling $p(\theta)$ weerspiegelt de voorkennis over de foutkans θ voordat er informatie uit een steekproef is gezien door de auditor. Een adequate prior-verdeling kent een relatieve waarschijnlijkheid toe aan elke mogelijke waarde van θ , zodat de waarschijnlijkheid over alle mogelijke waarden van θ gelijk is aan één. Een eenvoudige prior-verdeling is een beta(1, 1)-verdeling voor θ (figuur 1). Dit is een uniforme verdeling, wat inhoudt dat elke waarde van de foutkans in de populatie θ *a priori* even waarschijnlijk is. Omdat de prior-verdeling een zo volledig mogelijke samenvatting is van de voorkennis van de auditor met betrekking tot de foutkans θ , dient deze als zodanig onderbouwd te worden. Voor meer achtergrond over de rol van de prior-verdeling en geavanceerdere prior-verdelingen dan een uniforme verdeling verwijzen we naar Derks et al. (2021a).

2.2. De likelihood

De *likelihood* representeert de informatie die de geobserveerde steekproefdata bevatten over de foutkans θ . De *likelihood* functie $l(y | \theta)$ kwantificeert de waarschijnlijkheid dat de steekproefdata optreden onder specifieke waarden van θ (Etz 2018). De kans op het waarnemen van k fouten in een steekproef van n posten, met een onderliggende foutkans θ , is binomiaal verdeeld en wordt

doorgaans genoteerd als binomiaal($k | n, \theta$). Stel dat de auditor een steekproef van $n = 20$ waarnemingen heeft getrokken waarin $k = 0$ fouten gevonden worden, dan zijn de data uit de steekproef binomiaal($k = 0 | n = 20, \theta$) verdeeld (zie figuur 1).

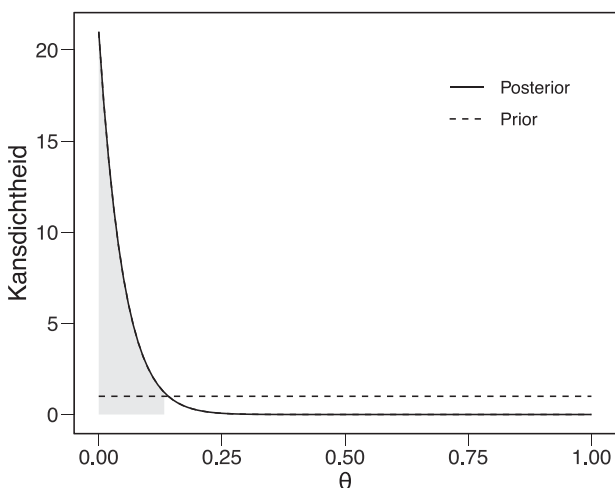
2.3. De posterior-verdeling

De posterior-verdeling $p(\theta | y)$ bevat de bijgewerkte kennis van de auditor over de foutkans θ na het evalueren van de data uit de steekproef y . De regel van Bayes bepaalt dat waarden van θ die de data uit de steekproef relatief goed voorspellen waarschijnlijker worden dan ze vooraf waren, terwijl waarden van θ die de steekproefuitkomsten relatief slecht voorspellen minder waarschijnlijk worden. De posterior-verdeling van θ is de bron van informatie voor de Bayesiaanse auditor, omdat de posterior-verdeling alle informatie bevat die de auditor vooraf had en alle verkregen informatie uit de steekproef.

Omdat de posterior-verdeling net als de prior-verdeling een kansverdeling is, kan de bijgewerkte kennis over de foutkans θ worden samengevat door middel van statistieken als het gemiddelde, de mediaan en de modus. De auditor kan bijvoorbeeld een uitspraak doen over de meest waarschijnlijke foutkans in de populatie door te kijken naar de modus van de posterior-verdeling. Evenzo kunnen percentielen van de posterior-verdeling worden geïnterpreteerd in termen van geloofwaardigheid (Kruschke en Liddell 2017).

In figuur 1 is de modus van de posterior-verdeling $\frac{k}{n} = 0$ (Myung 2003), wat inhoudt dat de meest waarschijnlijke foutkans in de populatie 0 procent is. Het 95e percentiel van de posterior-verdeling ligt op 0,133. Dit betekent dat, met een waarschijnlijkheid van 95 procent, de foutkans in de populatie lager is dan 13,30 procent. Het 95 procent geloofwaardigheidsinterval $[0, 0.133]$ is aangegeven met een grijs gebied in figuur 1.

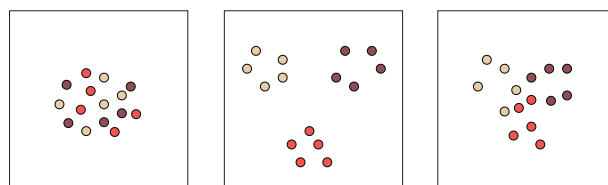
Figuur 1. De beta-prior-verdeling en posterior-verdeling (tevens de likelihood functie) na het zien van een steekproef van $n = 20$ posten en $k = 0$ fouten. Het grijze gebied geeft het 95 procent geloofwaardigheidsinterval aan.



3. Modellen voor gestratificeerde steekproeven

In dit hoofdstuk wordt ingegaan op twee vormen van voorkennis die de auditor moet valideren voorafgaand aan het gestratificeerde steekproefonderzoek. De eerste vorm van voorkennis heeft betrekking op in hoeverre de foutkans voor elk stratum hetzelfde is en – gerelateerd daaraan – in hoeverre er dus informatie kan worden gedeeld tussen de verschillende strata. Bijvoorbeeld: in sommige gevallen heeft de auditor voorkennis dat er geen relatie bestaat tussen de data afkomstig uit de verschillende strata. In dat geval dient er geen informatie gedeeld te worden tussen de strata. In een ander geval kan het zijn dat de foutkansen in de verschillende strata niet helemaal op dezelfde manier tot stand komen. Bijvoorbeeld doordat de interne beheersingsmaatregelen per stratum verschillen, maar er wel een bepaalde relatie is tussen de strata (bijvoorbeeld omdat deze alle dezelfde soort transacties bevatten). In dat geval dient er informatie tussen de strata gedeeld te worden, maar dienen verschillen tussen de strata wel te worden gerespecteerd. Als laatste kan het ook zo zijn dat de strata niet van elkaar verschillen qua opbouw en foutkansen. In dat geval kan de auditor alle informatie binnen alle strata in één keer analyseren. Figuur 2 illustreert deze drie verschillende smaken van voorkennis over in hoeverre informatie gedeeld kan worden tussen de strata.

Figuur 2. Grafische weergave van de drie verschillende smaken van voorkennis over in hoeverre informatie gedeeld kan worden tussen de strata. Afgebeeld zijn drie strata met daarin elk vijf posten (cirkels) waarbij de afstand tussen de posten de overeenkomstigheid tussen de foutkansen representeert. Van links naar rechts zijn afgebeeld: 1) een populatie waarin strata een overeenkomende foutkansen hebben en alle informatie met betrekking tot de foutkansen gedeeld kan worden tussen de strata, 2) een populatie waarin strata een andere foutkansen hebben en er geen informatie gedeeld kan worden tussen de strata, en 3) een populatie waarin strata deels dezelfde foutkansen hebben en er informatie gedeeltelijk gedeeld wordt tussen de strata.



Een tweede vorm van voorkennis die de auditor dient te evalueren is of het mogelijk is dat de auditor extra kennis toevoegt in de vorm van variabelen op stratumniveau, die gerelateerd zijn aan de foutkans in een stratum, zoals het aantal werknemers per filiaal, of de hoogte van de omzet per filiaal. Een modelmatige aanpak van stratificatie leent zich bij uitstek voor het toevoegen van dit soort extra kennis aan het statistische model en maakt de aannames van de auditor transparant. We onderscheiden hier voorkennis van extra kennis, omdat voorkennis in de prior-verdeling terechtkomt en extra kennis in het model terechtkomt.

Afhankelijk van de aanwezige voorkennis kan de auditor het statistisch model kiezen dat zo goed mogelijk overeenkomt met de beschikbare voorkennis. In dit hoofdstuk worden vier statistische modellen besproken en wordt aangegeven hoe ze zich verhouden tot de aanwezige voorkennis. Ook wordt behandeld welke voor- en nadelen deze modellen hebben.

3.1. Model 1: alle informatie gedeeld over strata, geen extra kennis

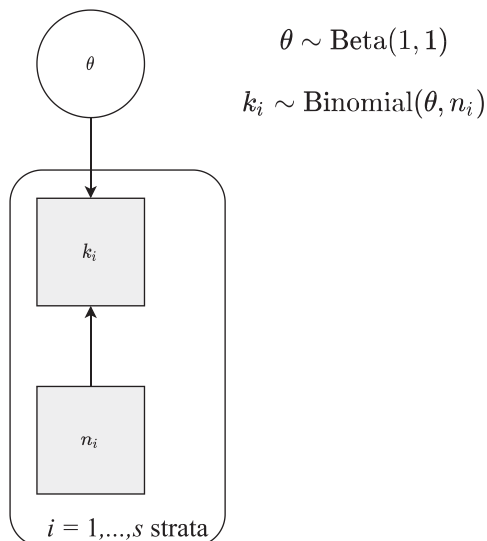
In model 1 neemt de auditor aan dat de data uit alle strata uitwisselbaar zijn. Met andere woorden: de strata zijn eigenlijk overbodig, alle informatie uit de strata wordt samengevoegd en als één geheel geanalyseerd. Statistisch wil dat zeggen dat er één parameter geschat wordt, $\theta \in [0, 1]$, die de foutkans in de populatie vertegenwoordigt. Omdat het een Bayesiaanse analyse betreft, krijgt deze parameter θ ook een prior-verdeling toegewezen, in dit geval een $\text{beta}(\alpha = 1, \beta = 1)$ prior-verdeling. De kansdichtheid f voor de $\text{beta}(\alpha, \beta)$ prior-verdeling wordt gegeven door vergelijking 2:

$$f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}. \quad (2)$$

Hierin stelt $B(\alpha, \beta)$ de beta-functie voor. Bij invulling van $\alpha = 1$ en $\beta = 1$ is de prior-verdeling gelijk aan een uniforme verdeling. Verder zijn de data uit de steekproef binomiaal verdeeld, met als kans op een fout de eerdergenoemde parameter θ . Figuur 3 is een grafische weergave (Lee en Wagenmakers 2013, p. 37) van het model waarin de informatie tussen de strata compleet wordt samengevoegd.

Als we uitgaan van de $\text{beta}(1, 1)$ prior-verdeling en de binomiale likelihood, kunnen de α en β parameters van de beta-posterior-verdeling voor de foutkans in de populatie vastgesteld worden. In principe kiest men in model 1 voor een gelijke trekkingskans in alle strata en verwacht

Figuur 3. Grafische weergave van een model om de foutkans θ in een populatie te schatten. Witte cirkels representeren niet geobserveerde continue parameters in het model en grijze vierkanten representeren geobserveerde discrete parameters.



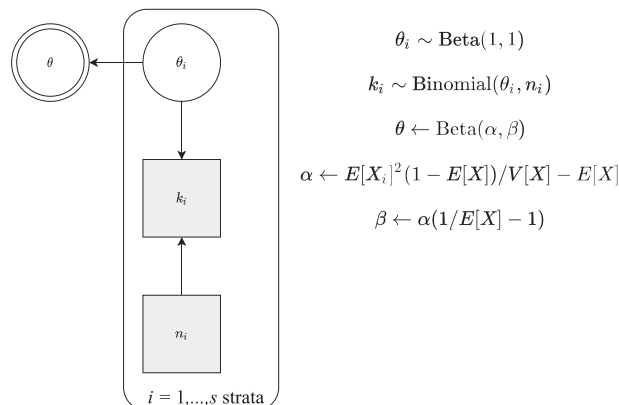
men in de steekproef ook geen grote verschillen in foutfracties per stratum. Als de trekkingskansen achteraf wel blijken te verschillen, én de voorkennis door een andere reden dan de steekproefuitslag toch niet blijkt te kloppen, én de foutfracties per stratum wel blijken te verschillen, is het mogelijk om met de binnen model 1 verzamelde data op basis van het gewogen aantal fouten toch de posterior-verdeling van θ te berekenen. Zie de appendix bij dit artikel voor hoe dit in zijn werk gaat.

Het eerste voordeel van dit model is dat het een erg eenvoudig model is en dus gemakkelijk te interpreteren. Het tweede voordeel is dat er in dit model maar één prior-verdeling gebruikt wordt om de foutkans in de populatie te schatten. Dit neemt niet weg dat er op basis van de beschikbare data alsnog de mogelijkheid is om de foutkans op stratumniveau te schatten (zie het hieropvolgende model 2). Een voorbeeld van een situatie waarin dit wenselijk kan zijn is in een group audit, waarin de strata componenten of werkmaatschappijen zijn met aparte auditors. In zo'n situatie kan deze aanpak passend zijn als de component auditors niet op de hoogte zijn van elkaars werk, maar de group auditor dit wel is. Het nadeel van dit model is dat de aanname dat alle foutkansen in alle strata gelijk worden geacht erg sterk is. Tot slot is een nadeel van deze methode dat het niet (eenvoudig) mogelijk is om extra kennis toe te voegen aan het model.

3.2. Model 2: geen informatie gedeeld over strata, geen extra kennis

In model 2 neemt de auditor aan dat er geen relatie bestaat tussen de data afkomstig uit de verschillende strata. Dit resulteert in een statistisch model waarin geen informatie tussen strata wordt gedeeld: voor elk stratum i wordt een aparte parameter voor de foutkans $\theta_i \in [0, 1]$ geschat. Elke parameter θ_i krijgt een $\text{beta}(1, 1)$ prior-verdeling toegewezen. De data uit de steekproef worden verondersteld voor elk stratum binomiaal verdeeld te zijn met voor elk stratum een foutkans θ_i . Figuur 4 is een grafische weergave van het model waarin geen informatie tussen de strata wordt samengevoegd.

Figuur 4. Grafische weergave van een model om de foutkansen θ_i in een aantal strata te schatten. De dubbele witte cirkel representeert θ . Deze is deterministisch bepaald (niet geobserveerd en continu) zodra de θ_i bekend zijn (zie appendix).



Gegeven de beta(1, 1) prior-verdeling en de binomiaal likelihood kunnen we de posterior-verdeling van θ_i vaststellen op een beta($1 + k_i$, $1 + n_i - k_i$) verdeling. Hoewel deze methode primair gericht lijkt te zijn op de schatting van foutkansen per stratum, is het dankzij de zogenoemde *Method of Moments* toch mogelijk om een verdeling van de foutkans op populatieniveau te bepalen. Deze zogenaamde momentenmethode wordt beschreven door Sellke (1983) en Stewart (2013) en veronderstelt dat de som van de individuele posterior-verdelingen per stratum te benaderen is met een nieuwe beta-verdeling, waarvan de parameters berekend worden uit de beta(α_i , β_i) posterior-verdelingen per stratum. Zie de appendix bij dit artikel voor deze berekeningen.

Zolang er geen voorkennis is over hoe de foutkans varieert per stratum, maar alleen dat deze varieert, kan de auditor naast de momentenmethode ook kiezen voor de eerdergenoemde gewogen methode. Bij gebrek aan stratumspecifieke informatie kan de prior immers op het niveau van de populatie worden gedefinieerd. De gevonden fouten per stratum, die als fractie van de steekproefomvang zullen verschillen als de voorkennis juist is, worden nu gewogen naar rato van de steekproefomvang alvorens ze verwerkt worden als *likelihood* in de posterior-verdeling.

Een voordeel van dit model is dat het leidt tot relatief eenvoudige statistische berekeningen: de parameters zijn gemakkelijk te schatten en te interpreteren. Het nadeel van dit model is dat de aanname dat alle foutkansen in de strata niet gerelateerd zijn aan elkaar, een erg restrictieve aanname is. Voor de momentenmethode geldt nog een ander nadeel: om te komen tot de kansverdeling voor de foutkansen in de populatie θ worden de onzekerheden uit de prior-verdelingen per stratum gestapeld bij het optellen. Het gebruik van de beta(1, 1) prior-verdeling per stratum, die grote foutkansen net zo waarschijnlijk acht als zeer kleine foutkansen, zal daarom in de momentenmethode bij veel strata een grote conservativiteit in de posterior-verdeling voor θ bewerkstelligen. De gewogen methode kent dit nadeel niet, maar wel een ander nadeel: de methode dwingt niet af hoe de steekproef over de strata verdeeld moet worden om op totaalniveau een representatief beeld van de gehele populatie te verkrijgen. Hoewel de gewogen methode vaak gebruikt wordt in onderzoeken op basis van surveys, is deze niet omstreden. Zie bijvoorbeeld Gelman (2007, p. 153), die zijn artikel opent met "*Survey weighting is a mess*". Overigens biedt Neyman allocatie (zie bijvoorbeeld Mathew et al. 2013) wel een handvat om de totale steekproefomvang over de strata te verdelen.

3.3. Model 3: deels informatie gedeeld over strata, geen extra kennis

In model 1 neemt de auditor aan dat de strata niet van elkaar verschillen en in model 2 neemt de auditor aan dat er geen relatie bestaat tussen de data afkomstig uit de verschillende strata. In model 3 zit de auditor conceptueel tussen model 1 en model 2 in. De data afkomstig uit de verschillende strata zijn niet totaal verschillend van elkaar; er zijn gelijkenissen tussen de data in

de verschillende strata. De sterke aanname dat de data binnen een stratum perfect homogeen zijn en dat de data tussen de strata perfect heterogeen zijn, is vaak niet realistisch. Een zekere mate van samenvoeging van deze aannames sluit dan beter aan op de aanwezige voorkennis.

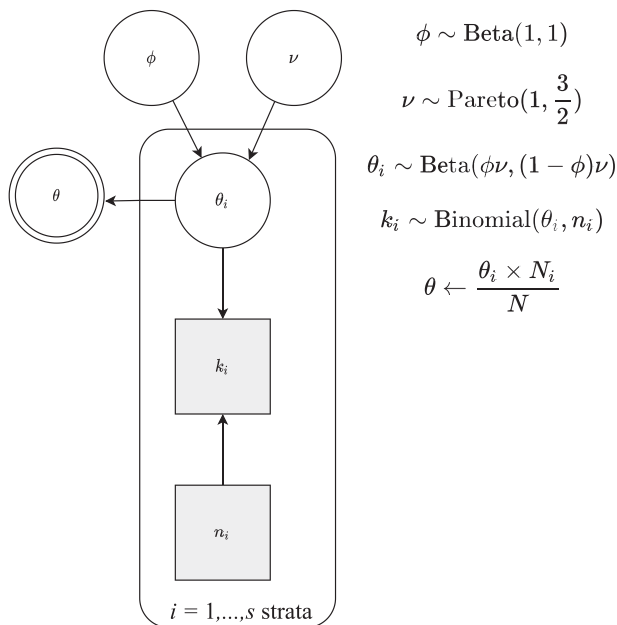
Hiervoor kan de auditor een hiërarchisch model specificeren. Hiërarchische modellen worden toegepast in allerlei wetenschappelijke vakgebieden om rekening te houden met onderliggende relaties tussen parameters in een model (Gelman et al. 2013, hoofdstuk 5). Een hiërarchisch model schat de foutkans van de populatie samen met de foutkansen per stratum, waarmee impliciet de hoeveelheid informatie wordt bepaald die wordt gedeeld tussen de strata. Dat heeft als gevolg dat dat schattingen van de foutkans op stratumniveau allemaal tot op zekere hoogte worden *gebiased* in de richting van de schatting van de foutkans op populatieniveau, en als gevolg winnen aan nauwkeurigheid (Efron and Morris 1977). Met andere woorden: dit model introduceert *bias* in de schattingen, met als resultaat kleinere variantie in de schattingen (Gelman et al. 2013, p. 101; Hastie et al. 2001, hoofdstuk 2).

Zoals in de eerdere twee modellen wordt er een prior-verdeling gedefinieerd voor de parameters die geschat worden. Omdat dit een hiërarchisch model is, zit er ook een hiërarchie in de constructie van de priors. De auditor kiest in dit geval een prior-verdeling voor de foutkans per stratum, waarbij de parameters van die prior zelf ook weer worden geschat (en dus ook een prior-verdeling nodig hebben, een zogenoemde hyperprior). In dit geval definiëren we een beta(α , β) verdeling als prior-verdeling voor de parameters θ_i . Een handige manier om de hyperpriors te specificeren, is door de α en β parameters van de beta-prior-verdeling per stratum te definiëren als $\alpha = \phi v$ en $\beta = (1 - \phi)v$ (Gelman et al. 2013, p. 111). Deze herparameterisering van de prior-verdeling op stratumniveau heeft als voordeel dat de hyperparameter $\phi \in [0, 1]$ is te interpreteren als de gemiddelde² foutkans in de populatie, en de hyperparameter $v > 1$ als de impliciete steekproefomvang in de prior-verdeling voor θ_i . Voor de hyperpriors in het model volgen we Gelman et al. (2013); een beta(1, 1) verdeling als prior-verdeling voor ϕ en een Pareto($1, 3/2$) verdeling als prior-verdeling voor v . Figuur 5 is een grafische weergave van het hiërarchische model waarin informatie tussen strata gedeeltelijk wordt samengevoegd.

Om tot een schatting voor de foutkans in de populatie θ te komen, wordt voor deze en de volgende methode *post-stratificatie* toegepast (Gelman and Little 1997). Hierbij wordt de steekproef gebruikt om de posterior-verdeling van de foutkans binnen elk stratum te schatten, waarna de posterior-verdelingen op stratumniveau geaggregeerd worden om tot een schatting op populatieniveau te komen. Dit doen we door de posterior-verdeling voor elk stratum te wegen naar het relatieve aandeel van dat stratum in de populatie.

Een voordeel van dit model is dat de aanname dat de data afkomstig uit de verschillende strata bepaalde overeenkomsten kunnen hebben, vaak realistisch is. Dit type model sluit dus in veel gevallen beter aan bij de

Figuur 5. Grafische weergave van een hiërarchisch model om de foutkansen θ in een aantal strata te schatten.



voorkennis van de auditor dan modellen zoals model 1 en model 2. Een nadeel van dit model is dat, door het introduceren van hiërarchie, de statistische modelstructuur complexer is dan bij de voorgaande modellen. Een ander nadeel is dat het definiëren van de hyperpriors niet triviaal is en betrokkenheid vergt van een statisticus die bekend is op dit vlak. In elk geval is het aan te bevelen om bij twijfel over de keuze van de hyperpriors verschillende prior-verdelingen te gebruiken en de daaruit volgende schatters ook te vergelijken. Door het effect van de hyperpriors in kaart te brengen, kan de auditor namelijk een uitspraak doen over de robuustheid van de schattingen. Als verschillende prior-verdelingen leiden tot dezelfde schattingen, kunnen de schattingen als robuust beschouwd worden en is de impact van de hyperprior dus verwaarloosbaar. Echter, als de hyperpriors cruciaal blijken te zijn voor de gevolgtrekking van de auditor, dan dient dit meegenomen te worden in de onderbouwing van de uiteindelijke prior-verdeling. In dit laatste geval is het aan te raden om verdere voorkennis te vergaren, die kan worden gebruikt om een meer informatieve hyperprior-verdeling te construeren.

Het opstellen van informatieve hyperpriors kan bijvoorbeeld door eerder vergaarde informatie (bijv. uit risico-inschattingen) mee te nemen in de prior-verdeling voor de gemiddelde fout in de populatie ϕ (Derks et al. 2021a). Daarnaast kan de prior-verdeling voor ν , die de hoeveelheid informatie bepaalt die in het model wordt samengevoegd, geïnformeerd worden door informatie met betrekking tot de gelijkens tussen de strata. Meer kansmassa bij grotere waarden van ν impliceert dat de auditor a priori meer belang toekent aan eigenschappen die een stratum erft uit de totale populatie (bijvoorbeeld een serie filialen met personeel dat rouleert over de filialen en één centrale AO/IC en één ERP-systeem); meer kansmassa bij kleinere waarden van ν impliceert dat de

auditor a priori minder belang toekent aan eigenschappen die een stratum erft uit de totale populatie (bijvoorbeeld filialen in een franchiseformule met alleen centrale inkoop). De auditor is hierbij niet gebonden aan de in dit artikel gebruikte Pareto- en beta-prior-verdelingen, maar kan de kansverdeling bepalen die het beste aansluit op de aanwezige voorkennis. We benadrukken dat zo'n specifieke prior voor ϕ en ν alleen nodig is indien de impact van de hyperprior groot is. Immers, als die impact klein is, zit de optimale mix tussen welk gedeelte een stratum erft uit de populatie en welk gedeelte binnen een stratum wordt bepaald, al verwerkt in de posterior-verdeling.

3.4. Model 4: deels informatie gedeeld over strata, met extra kennis

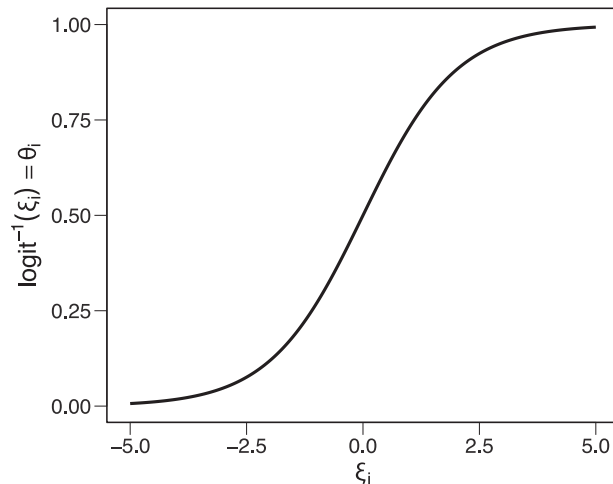
Model 4 is qua aannames vergelijkbaar met model 3, met één belangrijk verschil: De auditor neemt nog extra informatie mee in het model. Om dit te doen beschouwen we een alternatieve parametrisering in termen van de log-odds van de foutkans middels een parameter $\xi_i \in [-\infty, \infty]$. De log-odds zijn gelinkt aan de foutkans middels de logit-transformatie als:

$$\xi_i = \text{logit}(\theta_i) = \ln\left(\frac{\theta_i}{1-\theta_i}\right). \quad (3)$$

Zo komt bijvoorbeeld een foutkans van $\theta_i = 0,05$ overeen met odds van $\frac{0,05}{1-0,05} = \frac{1}{19}$ en daarmee log-odds van $\ln\left(\frac{1}{19}\right) = -2,94$. De steekproefuitkomsten in elk stratum zijn in dit model nog steeds binomiaal verdeeld, maar nu gekoppeld aan de parameter ξ_i middels de logit-transformatie. De inverse logit functie in vergelijking 4 wordt gebruikt om de log-odds ξ_i terug te transformeren naar een corresponderende foutkans θ_i , zie figuur 6. Wederom heeft elk stratum zijn 'eigen' log-odds, die gerepresenteerd worden door de parameter ξ_i .

$$\text{logit}^{-1}(\xi_i) = \frac{1}{1+e^{-\xi_i}} = \theta_i \quad (4)$$

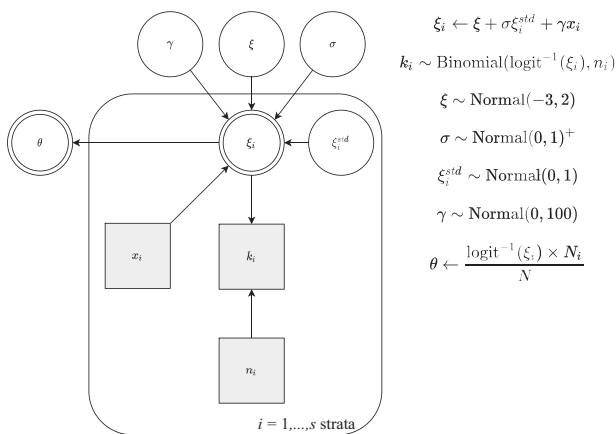
Figuur 6. De logistische (inverse logit) functie. De waarde voor de log-odds ξ_i kan variëren van $-\infty$ tot ∞ terwijl de waarde voor de foutkans θ_i altijd tussen 0 en 1 ligt.



Een groot voordeel van het formuleren van het probleem in termen van de log-odds is dat het gemakkelijk is om andere variabelen (op elk wenselijk niveau) toe te voegen aan het model (Gelman and Hill 2007). Hetzelfde principe van het vinden van een compromis tussen de twee uitersten van volledig samenvoegen van informatie en niet samenvoegen van informatie geldt namelijk ook voor uitgebreidere modellen. Het toevoegen van extra kennis kan nog een extra slag in efficiëntie bewerkstelligen, doordat er meer informatie tussen de strata gedeeld kan worden.

Figuur 7 is een grafische weergave van het logistische model waarin informatie tussen strata gedeeltelijk wordt samengevoegd en er extra kennis per stratum x_i wordt opgenomen. De getransformeerde foutkans heeft in dit model een relatie met de predictor x , zodat $\xi_i = \xi + \sigma \xi_i^{std} + \gamma x_i$. Net als in model 3 zullen ook hier keuzes gemaakt moeten worden voor de prior-verdelingen van σ , ξ_i^{std} , ξ en ook voor de extra kennis in de vorm van een parameter γ die de invloed van x_i modelleert. We volgen hierin Carpenter (2016): σ krijgt een standaard half-normaalverdeling als prior-verdeling en elke ξ_i^{std} krijgt een standaard normale verdeling als prior-verdeling. De parameter ξ krijgt als prior-verdeling een normale verdeling toegewezen met gemiddelde -3 en standaarddeviatie 2. Op de ongetransformeerde schaal ligt 95 procent van de kansmassa van deze prior-verdeling tussen de 0 en 71 procent, met als mediaan 4,7 procent. De parameter γ krijgt net als ξ een normale prior-verdeling toegewezen, maar nu met een verwachtingswaarde van 0 en een standaarddeviatie van 100, waarmee wordt aangegeven dat er vooraf geen idee bestaat over de omvang en de richting van de invloed van x .

Figuur 7. Grafische weergave van een hiërarchisch logistisch model om de log-odds ξ_i in een aantal strata te schatten middels een extra predictor.



Een nadeel van dit model is dat er, net als in het vorige model, steeds meer complexiteit wordt toegevoegd. Dit wil zeggen dat de auditor goed dient samen te werken met een statisticus, zodat alle aannames en statistische uitwerkingen daarvan goed kunnen worden uitgelegd en verantwoord. Een voordeel van dit model is dat het veel mogelijkheden biedt aan de auditor om extra kennis toe te voegen aan het model. Dat zal de transparantie

en efficiëntie ten goede komen, omdat de auditor dan op een transparante manier kan uitleggen welke impact extra kennis heeft op de uitkomsten. Het toevoegen van relevante extra kennis zal ook de schattingen van het model verbeteren, en dus de efficiëntie ten goede komen.

4. Voorbeeld: consolideren van fouten over filialen

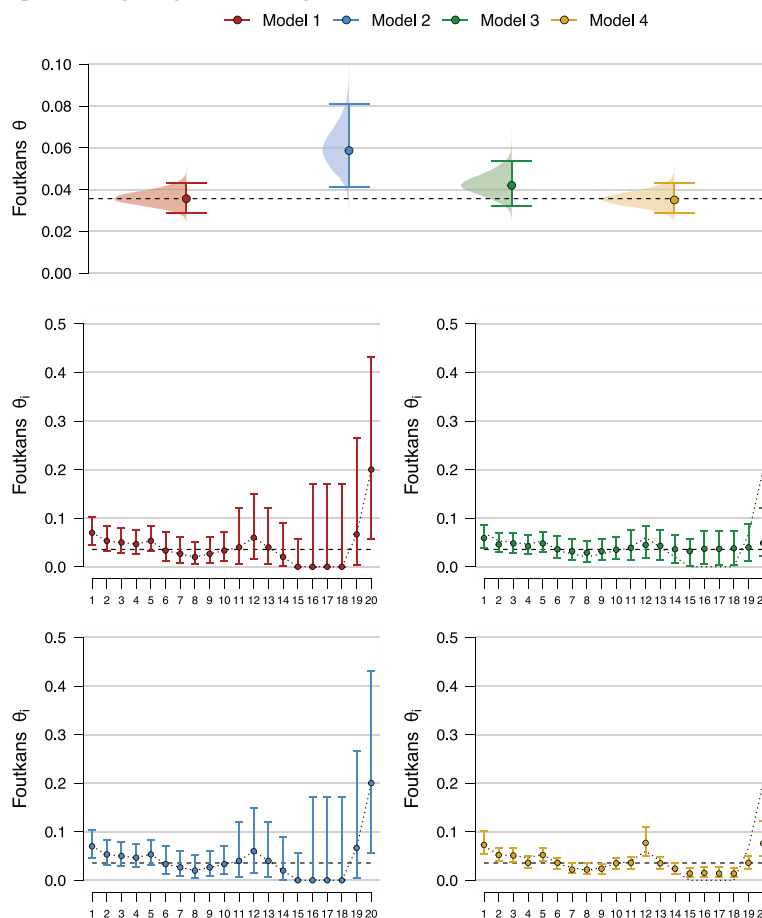
Om de voorgestelde modellen concreet te illustreren, wordt hier een voorbeeld behandeld van een audit van een organisatie, bestaande uit twintig filialen. De taak van de auditor is om 95 procent zekerheid te kunnen verschaffen over de foutkans in de populatie, bestaande uit twintig strata. In elk van de twintig filialen heeft een audit team een steekproef getrokken. Echter, in sommige filialen verwachtte de auditor meer fouten te vinden en heeft daarom bij deze filialen een grotere steekproef laten nemen dan in andere filialen. Het audit team rapporteert het aantal gevonden fouten per filiaal. Op het niveau van filiaal is ook extra kennis beschikbaar, in die zin dat de auditor verwacht dat het aantal werknemers binnen een filiaal dat verantwoordelijk is voor de administratie, de foutkans beïnvloedt. Tabel 1 geeft de beschikbare informatie en de steekproefuitkomsten per filiaal weer. Deze dataset wordt gebruikt om te illustreren hoe het samenvoegen van informatie kan helpen om nauwkeuriger te schatten.

Tabel 1. Overzicht van de beschikbare data en de steekproefuitkomsten per filiaal.

Filiaal	Aantal werknemers	Aantal posten	Aantal steken	Aantal fouten
1	5	5000	300	21
2	4	5000	300	16
3	4	5000	300	15
4	3	5000	300	14
5	4	5000	300	16
6	3	5000	150	5
7	2	5000	150	4
8	2	5000	150	3
9	2	5000	150	4
10	3	5000	150	5
11	3	10000	50	2
12	5	10000	50	3
13	3	10000	50	2
14	2	10000	50	1
15	1	10000	50	0
16	1	10000	15	0
17	1	10000	15	0
18	1	10000	15	0
19	3	10000	15	1
20	5	4000	15	3

Figuur 8 op de volgende pagina toont de schattingen van de vier modellen. Het bovenste paneel toont, per model, de 95 procent geloofwaardigheidsintervallen voor de schatting (i.e., de 95 procent meest waarschijnlijke waarden van de posterior-verdeling) van de foutkans in de populatie. De onderste vier panelen tonen de 95 procent geloofwaardigheidsintervallen om de geschatte foutkansen per stratum.

Figuur 8. Posterior-verdelingen en 95 procent-geloofwaardigheidsintervallen voor de foutkansen in de populatie (bovenste paneel) en in de twintig strata (onderste panelen, posterior-verdelingen weggelaten) volgens de vier modellen. De zwarte lijnen tonen respectievelijk de gevonden foutfractie in de populatie (gestreepte lijn) en de gevonden foutfracties per stratum (stippellijn). Voor model 1 is in het bovenste paneel de gewogen methode gebruikt; voor model 2 de momentenmethode.



Het bovenste paneel in figuur 8 laat zien dat het eerste model resulteert in een schatting van de foutkans van 3,56 procent en een bovengrens van 4,30 procent. Het geloofwaardigheidsinterval van deze schatting heeft een breedte van 1,40 procent. Het tweede model resulteert in een schatting voor de foutkans van 5,86 procent en een bovengrens van 8,10 procent. Het geloofwaardigheidsinterval heeft een breedte van 4,00 procent. Ten opzichte van het tweede model is het eerste model in dit geval $\frac{4,00 - 1,40}{4,00} = 65,00$ procent nauwkeuriger. Ter vergelijking: door gebruik van het derde model waarin informatie wordt samengevoegd, is de winst in nauwkeurigheid ten opzichte van het tweede model 47,25 procent. Als er extra informatie, in dit geval het aantal werknemers dat verantwoordelijk is voor de administratie, wordt meegenomen in het vierde model, is de nauwkeurigheid van de schatting 64,00 procent beter dan die van het tweede model. Naarmate er meer voorkennis aan model 4 toegevoegd wordt middels de prior-verdeling, of extra kennis middels een uitbreiding van het model, zal dit de nauwkeurigheid van de schattingen verder verbeteren.

De onderste vier panelen tonen dat de schattingen van de foutkansen per stratum van eerste twee modellen minder nauwkeurig zijn dan die van het derde en het vierde model. De modellen 1 en 2 hebben elk een gemiddelde nauwkeurigheid van 10,94 procent per stratum. Ter vergelijking: model 3 heeft een gemiddelde nauwkeurigheid

van 5,57 procent en model 4 heeft een gemiddelde nauwkeurigheid van 2,87 procent.

De panelen in de rechterkolom laten ook zien dat het samenvoegen van informatie als gevolg heeft dat de schattingen van de foutkansen in de strata naar het populatiegemiddelde getrokken worden, waardoor ze op elkaar gaan lijken. Dat gebeurt in grotere mate voor strata waar weinig is gestoken en in mindere mate voor strata waar veel stekken zijn genomen. Neem bijvoorbeeld stratum 20 waarin 15 posten zijn gestoken en de auditor de ‘pech’ had om 3 fouten te vinden. Bij het gebruik van model 1 en model 2 zal de auditor moeten concluderen dat dit stratum een uitzonderlijk hoge foutkans heeft ($\frac{3}{15} = 20$ procent), wat op basis van de steekproefdata uit de andere strata niet erg waarschijnlijk is. In model 3 en 4 wordt bij de schatting van de foutkans in dit stratum echter veel gewicht toegekend aan het populatiegemiddelde. Dat is ook logisch; bij weinig informatie is het immers beter om relatief meer te steunen op wat al bekend is. Model 3 en 4 geven daardoor een realistischer beeld van de foutkans in dit stratum.

5. Conclusie

In dit artikel zijn twee vormen van voorkennis besproken, die de auditor moet valideren bij het doen van

gestratificeerde steekproeven. We hebben daarbij laten zien dat, in vergelijking met huidige methoden, het gebruik van Bayesiaanse hiërarchische modellen kan bijdragen aan het verbeteren van de fout-schatting op populatieniveau alsmede aan de schattingen op stratumniveau. Het gebruik van de Bayesiaanse logistische aanpak (model 4) biedt bovendien kansen voor auditors om nog efficiënter met informatie om te gaan, omdat zowel voorkennis als extra data meegenomen kunnen worden in de evaluatie van de steekproef.

Merk op dat in de toepassing van deze modellen veelal prior-verdelingen zijn ingezet die, a priori, alle mogelijke parameterwaardes even waarschijnlijk achten. Het voordeel van deze opzet is dat deze prior-verdelingen relatief

gemakkelijk te begrijpen zijn, met als nadeel dat de schattingen voor de foutkansen hierdoor relatief onnauwkeurig en conservatief zijn. In de praktijk is het daarom aan te raden om voorkennis over de foutkans(en) te verzamelen en deze te verwerken in de prior-verdeling. In combinatie met de modelgebaseerde aanpak en de toevoeging van extra data kan zo de winst in efficiëntie volledig transparant gemaakt worden.

Onze hoop is dat, door deze methoden uiteindelijk beschikbaar te maken in de open-source software JASP for Audit (Derks et al. 2021b), auditors meer gebruik zullen maken van Bayesiaanse hiërarchische modellen om steekproeven te evalueren.

-
- **K. P. Derks MSc** is promovendus bij Nyenrode Business Universiteit en lid van de stuurgroep statistical auditing van het Limperg Instituut. Daarnaast is hij softwareontwikkelaar bij JASP, een gratis en open-source statistiekprogramma, waar hij de module voor statistische steekproeven onderhoudt.
 - **Prof. dr. J. J. B. de Swart MBA** is partner bij PwC Consulting en verantwoordelijk voor de data analytics groep. Daarnaast is hij hoogleraar Toegepaste Wiskunde aan Nyenrode Business Universiteit en voorzitter van de stuurgroep statistical auditing van het Limperg Instituut.
 - **Prof. dr. R. Wetzels** is director Data Analytics bij PwC Consulting. Daarnaast is hij hoogleraar Data Science bij Nyenrode Business Universiteit en bestuurslid van de faculty Business & Management van de NBA.
-

Noten

1. We spreken hier bewust van voorkennis, omdat de steekproef volgens het audit risk model als gegevensgerichte controle het sluitstuk van de audit is (De Swart et al. 2013).
2. Het gemiddelde van de beta(α, β) verdeling is $\alpha / (\alpha + \beta)$ en dus gelijk aan ϕ .
3. Immers, $v = \alpha + \beta = 1 + k + 1 + n - k = n + 2$.

Literatuur

- Van Batenburg PC, O'Hagan A, Veenstra RH (1994) Bayesian discovery sampling in financial auditing: A hierarchical prior model for substantive test sample sizes. *Journal of the Royal Statistical Society: Series D (The Statistician)* 43: 99–110. <https://doi.org/10.2307/2348936>
- Betancourt M, Girolami M (2015) Hamiltonian Monte Carlo for hierarchical models. *Current Trends in Bayesian Methodology with Applications* 79(30): 2–4. <https://doi.org/10.1201/b18502>
- Carpenter B (2016) Hierarchical Partial Pooling for Repeated Binary Trials. [online] mc-stan.org. <https://mc-stan.org/users/documentation/case-studies/pool-binary-trials.html> [Accessed 30 November 2021]
- Derks K, De Swart J, Van Batenburg P, Wagenmakers E-J, Wetzels R (2021a) Priors in a Bayesian audit: How integration of existing information into the prior distribution can improve audit transparency and efficiency. *International Journal of Auditing* 25(3): 621–636. <https://doi.org/10.1111/ijau.12240>
- Derks K, De Swart J, Wagenmakers E-J, Wille J, Wetzels R (2021b) JASP for Audit: Bayesian tools for the auditing practice. *Journal of Open Source Software* 6(68): 2733. <https://doi.org/10.31234/osf.io/9f6ub>
- Efron B, Morris C (1977) Stein's paradox in statistics. *Scientific American* 236(5): 119–127. <https://doi.org/10.1038/scientificamerican0577-119>
- Etz A (2018) Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science* 1(1): 60–69. <https://doi.org/10.1177/2515245917744314>
- Gelman A (2007) Struggles with survey weighting and regression modeling. *Statistical Science* 22(2): 153–164. <https://doi.org/10.1214/088342306000000691>
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis*, 3rd edn. Chapman & Hall/CRC Press, London. <https://doi.org/10.1201/b16018>
- Gelman A, Hill J (2007) *Data Analysis Using Regression and Multilevel-Hierarchical Models*. Cambridge University Press, Cambridge, United Kingdom. <https://doi.org/10.1017/CBO9780511790942>
- Gelman A, Little TC (1997) Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* 23(2): 127–135.
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning: Data mining, inference, and prediction*. Springer, New York. <https://doi.org/10.1007/978-0-387-21606-5>
- Kruschke JK, Liddell TM (2017) Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review* 25(1): 155–177. <https://doi.org/10.3758/s13423-017-1272-1>

- Lavallée P, Beaumont J-F (2015) Why we should put some weight on weights. *Survey Insights: Methods from the Field, Weighting: Practical Issues and ‘How to’ Approach*, Invited article, Retrieved from <https://surveyinsights.org/?p=6255>. <https://doi.org/10.13094/SMIF-2015-00001>
- Lee MD, Wagenmakers E-J (2013) *Bayesian cognitive modeling: A practical course*. Cambridge University Press.
- Mathew O, Sola AF, Oladiran BH, Amos AA (2013) Efficiency of Neyman allocation procedure over other allocation procedures in stratified random sampling. *American Journal of Theoretical and Applied Statistics* 2(5): 122–127. <https://doi.org/10.11648/j.aj-tas.20130205.12>
- Myung IJ (2003) Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* 47: 90–100. [https://doi.org/10.1016/S0022-2496\(02\)00028-7](https://doi.org/10.1016/S0022-2496(02)00028-7)
- Sellke T (1983) On the fitting of Pearson curves to sums of independent random variables. Technical Report SOL ONR 333, Department of Statistics, Stanford University. <https://doi.org/10.21236/ADA133153>
- Steele A (1992) *Audit risk and audit evidence: The Bayesian approach to statistical auditing*. Academic Press, London.
- De Swart JJB, Wille FJ, Majoor B (2013) Het ‘push left’-principe als motor van data analytics in de accountantscontrole. *Maandblad voor Accountancy & Bedrijfseconomie* 87: 425–433. <https://doi.org/10.5117/mab.87.11905>
- Stewart T (2013) *A Bayesian audit assurance model with application to the component materiality problem in group audits*. Ph.D. dissertation. VU University Amsterdam. Amsterdam: Limperg Instituut.
- Touw P, Hoogduin L (2012) *Statistiek voor audit en controlling*. Academic Service.
- Wagenmakers E-J, Lee M, Lodewyckx T, Iverson GJ (2008) Bayesian versus frequentist inference. In: *Bayesian evaluation of informative hypotheses*. Springer, New York, NY, 181–207. https://doi.org/10.1007/978-0-387-09612-4_9

Bijlage 1. Berekeningen

Model 1: Schatten van de foutkans op populatieniveau

Voor de schatting van de foutkans in de populatie bepaalt deze methode per stratum eerst een gewicht w_i :

$$w_i = \frac{N_i / n_i}{N / n}, \quad (5)$$

waarin

$$n = \sum_{i=1}^s n_i, N = \sum_{i=1}^s N_i. \quad (6)$$

Hierin is n_i het aantal gestoken posten in stratum i en N_i het totale aantal posten in stratum i . Dit soort weging komt ook vaak in *survey sampling* voor (Lavallée and Beaumont 2015). Merk op dat alle gewichten 1 zijn als de trekkingskansen per stratum gelijk zijn. Als we wederom uitgaan van de beta(1, 1) prior-verdeling en de binomiale likelihood, kunnen de α en β parameters van de beta-posterior-verdeling voor de foutkans in de populatie vastgesteld worden op basis van het gewogen aantal fouten:

$$\alpha = 1 + \sum_{i=1}^s (k_i \times w_i), \beta = 1 + n - \sum_{i=1}^s (k_i \times w_i). \quad (7)$$

Op basis van de posterior-verdeling met deze parameters kan bijvoorbeeld de meest waarschijnlijke fout in de populatie vastgesteld worden op $(\alpha - 1)/(\alpha + \beta - 2) = \sum_{i=1}^s w_i k_i / n$. Ook is het nog steeds mogelijk om posterior-verdelingen per stratum uit te rekenen, want deze zijn simpelweg gelijk aan beta($1 + k_i, 1 + n_i - k_i$). Immers, binnen een stratum hoeft er niet gewogen te worden.

Model 2: Schatten van de foutkans op populatieniveau

De momentenmethode werkt als volgt. Het eerste moment $E[X]$ (verwachtingswaarde) van de nieuwe (beta)verdeling van de foutkans in de populatie θ wordt berekend als

gewogen som van de eerste momenten per stratum $E[X_i]$, die gelijk zijn aan

$$E[X_i] = \frac{\alpha_i}{\alpha_i + \beta_i}, \quad (8)$$

waarin α_i en β_i de parameters van de beta-posterior-verdeling in stratum i voorstellen. In formulevorm ziet de berekening van deze gewogen som er als volgt uit:

$$E[X] = \sum_{i=1}^s N_i E[X_i] / N. \quad (9)$$

Wederom is hierin N_i het aantal posten per stratum en N het aantal posten in de populatie. Het tweede moment (variantie) wordt op vergelijkbare wijze berekend. Eerst wordt de variantie per stratum berekend als:

$$V[X_i] = \frac{\alpha_i \times \beta_i}{(\alpha_i + \beta_i)^2 \times (\alpha_i + \beta_i + 1)}, \quad (10)$$

en daarna de variantie voor de nieuwe verdeling als

$$V[X] = \sum_{i=1}^s N_i^2 V[X_i] / N^2. \quad (11)$$

Merk op dat de gewichten in vergelijking 11 gekwadeerd worden vanwege $V[cX] = c^2 V[X]$. Nu $E[X]$ en $V[X]$ bekend zijn, kunnen de α en β parameters van de beta-posterior-verdeling voor de foutkans in de populatie θ als volgt worden berekend:

$$\alpha = E[X]^2(1 - E[X])/V[X] - E[X], \beta = \alpha(1/E[X] - 1). \quad (12)$$

Bijlage 2. Online appendix

R code in de online appendix: <https://doi.org/10.17605/osf.io/zv6t3>