



Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor

Koen Derks



Hoe voorspelt u een fout?



Stof tot nadenken

Door na te denken over deze vraag kunt u de statistische evaluatie bij de klant beter afstemmen op de praktijk.



Voorkennis

3 vormen van voorkennis belangrijk:

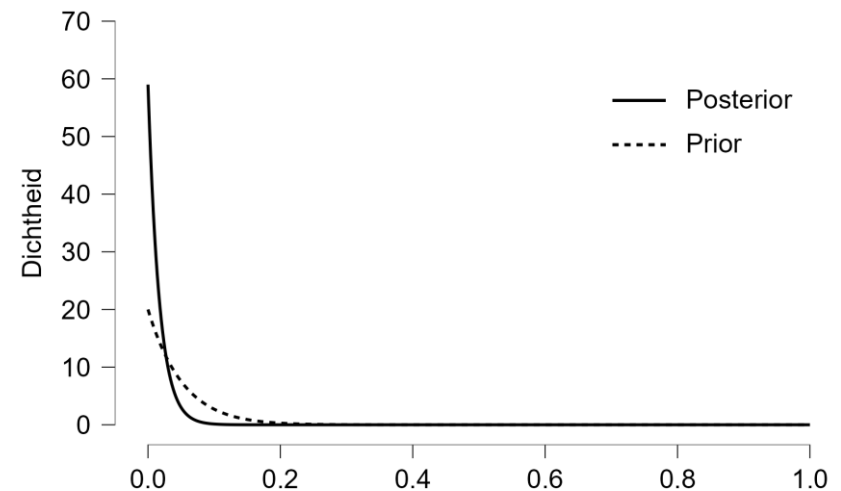
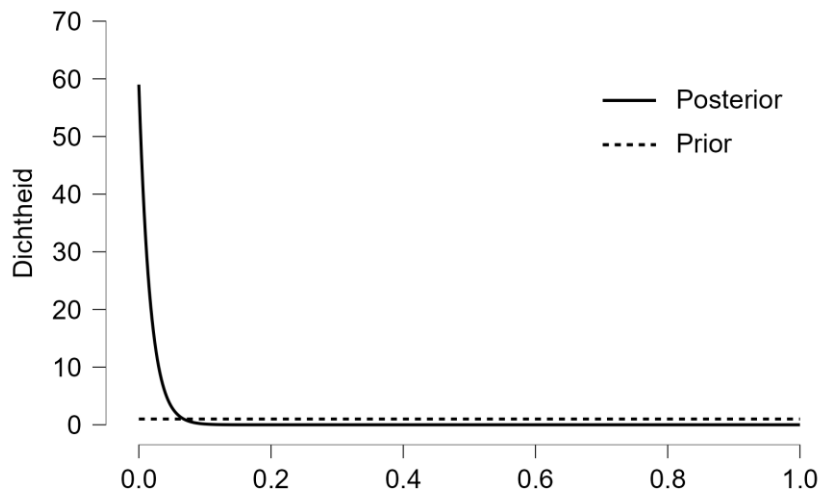
- Foutkans in de populatie
- Relatie tussen foutkansen in (eventuele) strata
- Relevante karakteristieken van de data



Foutkans populatie

Kennis over de foutkans in de populatie komt in de prior verdeling terecht:

- Kennis uit risico inschattingen
- Eerder genomen steekproef





Gelijkenis strata

Kennis over de gelijkenis tussen strata uit zich in het statistische model:

- *‘alles-op-een-hoop’ model:*
De strata zijn hetzelfde qua foutkans
- *‘ieder-apart’ model:*
De strata zijn compleet verschillend
- *Hiërarchisch model:*
De strata lijken deels op elkaar



Data toevoegen

Elk statistisch model (ongeacht de prior of de mate van informatie delen) kan uitgebreid worden met relevante karakteristieken van de data:

- Aantal *controls* voor een post
- Aantal werknemers in een filiaal (stratum)
- Dagen dat een betaling uitstond



Voordelen

- Relatie tussen data en foutkans wordt meegenomen
- Onderscheid maken tussen posten in de populatie
- Beter uitlegbaar hoe fouten tot stand komen.



Voorbeeld

- Populatie van 1000 posten
- 7 posten gestoken
- 1 fout gevonden
- Integrale data analyse geeft aantal *controls* per post



Voorbeeld

Post	Controls	Uitkomst
1	0	x
2	0	✓
3	0	
4	1	✓
5	1	✓
6	1	
7	2	✓
8	2	✓
9	2	✓
...
1000	2	

x = fout gevonden

✓ = geen fout gevonden



Voorbeeld

Methode 1: Evaluatie zonder extra data

- Model:

$$\theta_i = \theta$$

- θ is de foutkans in de populatie
- We kiezen een prior verdeling voor θ



Voorbeeld



Post	Controls	Uitkomst	Zonder extra data	Met extra data
1	0	x		
2	0	✓		
3	0		0.14 [0.031; 0.526]	
4	1	✓		
5	1	✓		
6	1		0.14 [0.031; 0.526]	
7	2	✓		
8	2	✓		
9	2	✓		
...
1000	2		0.14 [0.031; 0.526]	



Voorbeeld

Bij methode 1 wordt er **geen** onderscheid gemaakt tussen posten met verschillende aantallen *controls*.

Er kan dan ook **geen** statistische uitspraak gedaan worden over de relatie tussen het aantal *controls* en de foutkans.



Data toevoegen

Een *logistisch regressie model* bied de mogelijkheid om gemakkelijk extra data toe te voegen:

$$\text{logit}^{-1}(\theta_i) = \alpha + \beta^{\text{controls}} \cdot \text{controls}_i$$

- θ is de foutkans in een post
- We kiezen een prior verdeling voor α en β



Voorbeeld

Een voorspelling voor de foutkans gegeven

- $x = 0$ controls
- $x = 1$ controls
- $x = 2$ controls
- ...

+ een (gewogen) posterior verdeling voor de foutkans in de populatie



Voorbeeld

Post	Controls	Uitkomst	Zonder extra data	Met extra data
1	0	x		
2	0	✓		
3	0		0.14 [0.031; 0.526]	0.168 [0.01; 0.61] (↑)
4	1	✓		
5	1	✓		
6	1		0.14 [0.031; 0.526]	0.093 [0.013; 0.48] (↓)
7	2	✓		
8	2	✓		
9	2	✓		
...
1000	2		0.14 [0.031; 0.526]	0.086 [0.003; 0.46] (↓↓)



Voorbeeld

Bij methode 2 wordt er **wel** onderscheid gemaakt tussen posten met verschillende aantallen *controls*.

Er kan dan ook een statistische uitspraak gedaan worden over de relatie tussen het aantal *controls* en de foutkans.



Voordelen

- *Onderscheid*: Elke overgebleven post krijgt een andere voorspelde foutkans
- *Meer kennis*: Zowel een uitspraak op populatieniveau als per aantal *controls*
- *Beter uitlegbaar*: De relatie tussen het aantal *controls* en de foutkans wordt geschat



Stratificatie

- Extra data toevoegen op postniveau of op stratumniveau
- Onderscheid tussen strata beter in kaart brengen
- Relatie tussen extra data en foutkans wordt geschat



Voorbeeld

De auditor moet fouten consolideren over 20 filialen en besluit informatie te delen tussen de 20 strata.

Er is informatie beschikbaar over het aantal werknemers dat verantwoordelijk is voor de administratie in elk stratum.



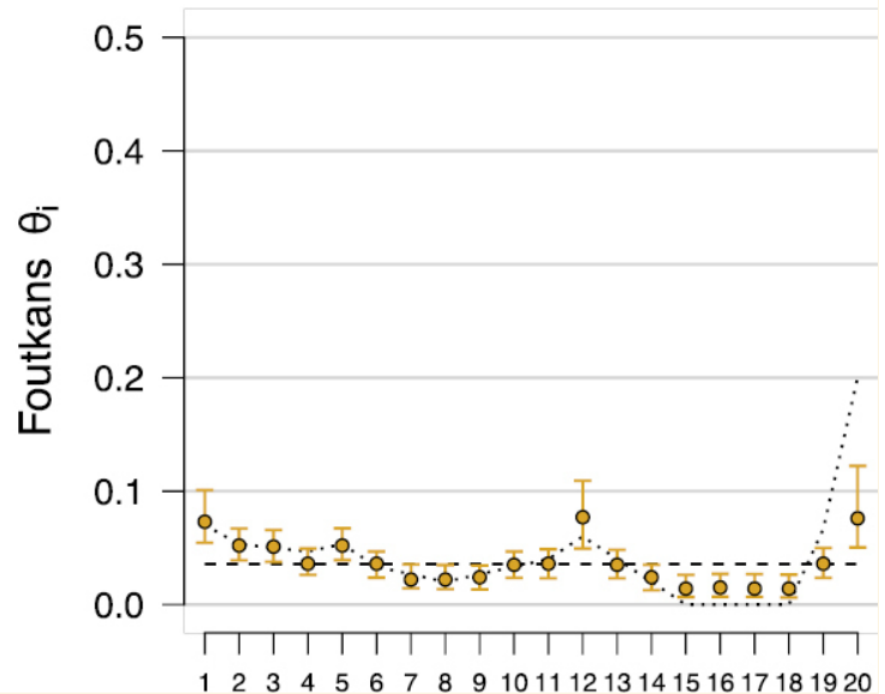
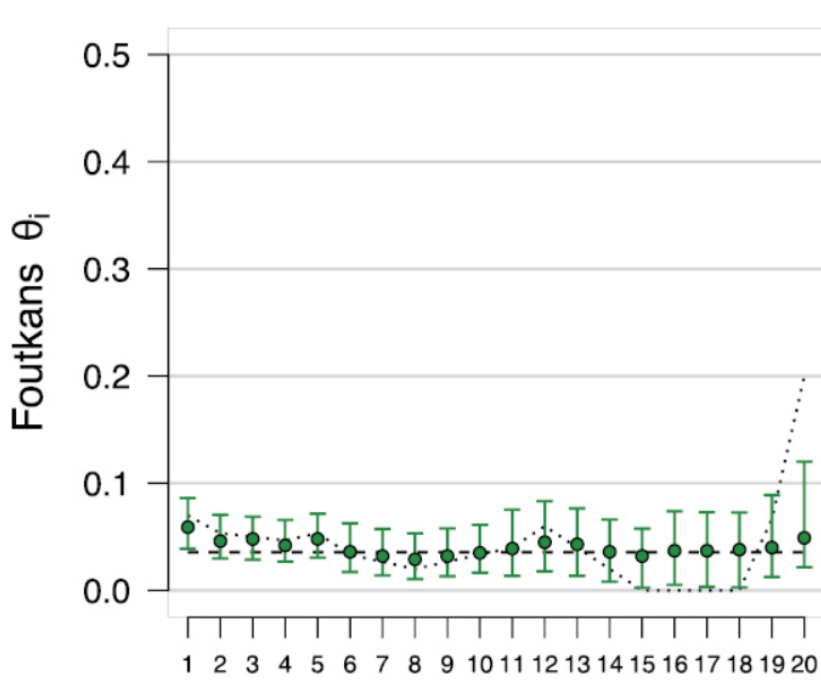
Gelijkenis strata

Ook hier weer belangrijk of er informatie gedeeld kan worden:

- *‘alles-op-een-hoop’ model:*
De strata zijn hetzelfde in de relatie werknemers - foutkans
- *‘ieder-apart’ model:*
De strata zijn compleet verschillend
- *Hiërarchisch model:*
De strata lijken deels op elkaar



Voorbeeld



Derks K, de Swart J, Wetzels R (2022) Een Bayesiaanse blik op gestratificeerde steekproeven heeft voordelen voor de auditor. Maandblad voor Accountancy en Bedrijfseconomie 96(1/2): 37-46. <https://doi.org/10.5117/mab.96.78836>



Conclusies

- Door relevante data mee te nemen in de evaluatie kan er (extra) onderscheid gemaakt worden tussen strata of posten
- Dit verrijkt de uitspraken die u kunt maken over de fouten in (delen van) de populatie
- Vormt een brug tussen de expertkennis van de auditor en het statistisch model.