A.R. Harinandansingh

Meervoudig toetsen met de horseshoe prior

Bachelorscriptie

Scriptiebegeleiders: B. Szabó, PhD
S.L. van der Pas, MSc MA

Datum Bachelorexamen: 17 augustus 2016

Mathematisch Instituut, Universiteit Leiden

Inhoudsopgave

1 Inleiding

Het meervoudig toetsen van een verzameling hypothesen houdt in dat deze simultaan worden getoetst. Dit zou gedaan kunnen worden door iedere hypothese afzonderlijk te toetsen. Het probleem dat hierbij optreedt is dat er een grote kans is op het maken van een Type 1 fout. Een Type 1 fout wilt zeggen dat we de nulhypothese ten onrechte verwerpen. Voor een verzameling van $n = 10$ onafhankelijke hypothesen en een significantieniveau van 5% is de kans op een Type 1 fout gelijk aan 0.40. Een significantieniveau van 5% betekent dat we van de 100 toetsen, gemiddeld vijf keer de nulhypothese ten onrechte verwerpen. Voor $n = 40$ onafhankelijke hypothesen bij hetzelfde significantieniveau is de kans echter een flink stuk groter, namelijk 0.87. In de praktijk treden er problemen op waarbij geldt dat $n = 1000$ of een veelvoud hiervan. De genoemde kans nadert dan de waarde 1. Bekende toetstechnieken om dit probleem tegen te gaan zijn bijvoorbeeld de Bonferroni correctie en de Benjamini-Hochberg methode (zie [6]), waarbij met name de laatstgenoemde methode gebruikelijk is voor het tegengaan van het meervoudig toetsingsproblemen.

Toepassingen van het meervoudig toetsen vinden we bijvoorbeeld in de astronomie bij het detecteren van supernova's en in de geneeskunde. Meer details hierover worden gegeven in het volgende hoofdstuk.

In 2010 kwam Carvalho, Polson en Scott in hun artikel 'The horseshoe estimator for sparse signals' (zie [9]) met een nieuwe benadering voor *sparse* problemen. Zij introduceerden de horseshoe prior. In deze scriptie onderzoeken we of de horseshoe prior, waarbij er gebruik wordt maakt van Bayesiaanse statistiek, ons een goede toetsingsprocedure oplevert bij het meervoudig toetsen. We gebruiken een eenvoudig sparse probleem bij het meervoudig toetsen. Om te bepalen of de horseshoe prior geschikt is, voeren we simulaties uit in verschillende situaties. Alle simulaties zijn uitgevoerd met het sofwarepakket *R*.

Wat we willen bij het bepalen van de geschiktheid van de horseshoe prior is een goede *trade-off* tussen Type 1 en Type 2 fouten. We maken een Type 2 fout als we de nulhypothese niet verwerpen, terwijl deze niet waar is. Natuurlijk willen we dat beide type fouten zo klein mogelijk zijn, maar in de praktijk blijkt dit lastig. Als de ene type fout toeneemt dan neemt de andere type fout af en andersom. Als we een goede balans hebben tussen Type 1 en Type 2 fouten, kunnen we zeggen dat de horseshoe prior geschikt is. Het is echter relatief wat we kunnen aanmerken als een goede balans. Hiervoor zijn veel mogelijkheden en in de praktijk hangt dit af van de toepassing van het probleem.

2 Sparsity

Sparsity is een verschijnsel dat bij verschillende problemen optreedt binnen de statistiek. Dit houdt in dat we in het algemeen een groot aantal parameters hebben waarvan er slechts een paar relevant zijn. Sparse problemen vormen een nogal grote klasse. Het probleem dat in deze scriptie wordt beschouwd, is één van de simpelste gevallen.

Wiskundig kunnen we ons probleem als volgt beschrijven. We observeren een vector

$$Y = (Y_1, Y_2, \ldots, Y_n) \tag{1}$$

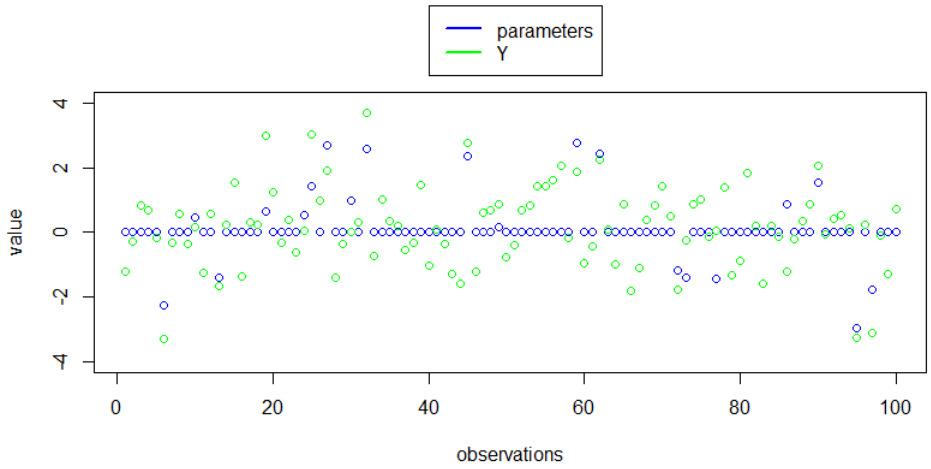
Er geldt dat voor $i = 1, \ldots, n$

$$Y_i = \theta_{0,i} + \epsilon_i \tag{$*$}$$

Hierbij geldt dat $\theta_{0,i}$ onbekende parameters zijn, waarbij de meeste gelijk zijn aan nul. ϵ_i kunnen we opvatten als ruis, met $\epsilon_i \sim \mathcal{N}(0, 1)$. Deze zijn onafhankelijk en identiek verdeeld.

Ons doel is om de parameters $(\theta_{0,i})$ ongelijk aan nul te vinden. Omdat we te maken hebben met een geringe aantal parameters ongelijk aan nul, spreken we over het schatten van een sparse vector.

Ter illustratie van het probleem zijn in Figuur 1 honderd parameters weergegeven. Dit zijn de blauwe cirkels. Zoals we kunnen zien zijn de meeste gelijk aan nul. Wanneer we de waarde van de standaard normaal verdeelde ruis bij de waarde van de parameters optellen krijgen we de observaties Y_i, voor $i = 1, \ldots, n$. Dit zijn de groene cirkels.



Figuur 1: Een sparse model met de parameters en de observaties

Echter zijn alleen de observaties bekend. Aan de hand hiervan moeten we bepalen welke parameters ongelijk zijn aan nul.

In het vervolg zullen we zo nu en dan de vergelijking trekken met signalen. Parameters ongelijk aan nul corresponderen dan met signalen. Deze zijn niet altijd detecteerbaar.

2.1 Toepassingen

Sparse modellen worden toegepast in verschillende vakgebieden. Hieronder volgen voorbeelden uit de astronomie en geneeskunde.

- Beeldverwerking: Het verwijderen van ruis op de achtergrond op afbeeldingen van hemellichamen.

  Een *Chandra X-Ray observatory* is een satelliet die astronomische waarnemingen doet op basis van röntgenstraling die wordt uitgezonden door hemellichamen. Afbeeldingen van astronomische objecten die worden verkregen met een Chandra X-Ray observatory zijn veelal in iedere pixel aan ruis onderhevig. Ruis ontstaat door hemelse bronnen en kosmische straling.

  Om ruis op de achtergrond te verwijderen, kan gebruik worden gebruikt van meervoudig toetsen. Voor n pixels krijgt iedere pixel i, met $i = 1, \ldots, n$, een intensiteitswaarde λ_i toegekend. De intensiteitswaarde van de achtergrondruis wordt bepaald op een waarde η. Voor de meeste pixels geldt dat $\lambda_i = \eta$. Bovendien is η klein in vergelijking met $\max\{\lambda_i : i = 1, \ldots, n\}$. Om te bepalen of we in iedere pixel te maken hebben met ruis of een signaal wordt voor iedere i een tweezijdige hypothese toets uitgevoerd. De nulhypothese $H_{0,i}$ wordt getoetst tegenover de alternatieve hypothese $H_{1,i}$, waarbij

  - $H_{0,i} : \lambda_i = \eta$
  - $H_{1,i} : \lambda_i > \eta$

  voor $i = 1, \ldots, n$.

  Dit is een voorbeeld uit de praktijk waar er gebruik wordt gemaakt van meervoudig toetsen. Bovendien hebben we te maken met een sparse structuur, omdat voor de meeste pixels geldt dat $\lambda_i = \eta$.

  Afbeeldingen van bijvoorbeeld planeten die zijn verkregen met bovengenoemde satelliet zijn nogal wazig. Met bekende toetstechnieken voor meervoudig toetsen kan de afbeelding dan minder wazig worden gemaakt door de ruis te verwijderen, zodat meer detailles naar voren komen.

- Het vinden van supernova's.

  Het detecteren en classificeren van voorbijgaande objecten in het heelal vormt een onderzoeksgebied binnen de astronomie. Dit wordt doorgaans gedaan op basis van astronomische afbeeldingen, waarbij iedere pixel een intensiteit met betrekking tot de hoeveelheid licht met zich meedraagt. Wanneer de intensiteit zich boven een bepaalde drempelwaarde bevindt, dan maakt deze pixel deel uit van een voorbijgaand object. Zo'n pixel wordt een *source* pixel genoemd (zie [3]). Een voorbijgaand object is dan een verzameling van source pixels. Men maakt onderscheid tussen enerzijds een source pixel en anderzijds een achtergrond pixel, die geen deel uitmaakt van het object. Dit laatste kunnen we vergelijken met de ruis zoals in het vorige voorbeeld.

  Een voorbeeld van zo'n voorbijgaand object is een supernova. Een supernova is het verschijnsel waarbij een ster op een spectaculaire wijze explodeert. Dit gaat veelal gepaard met grote hoeveelheden uitgestraalde licht. Een supernova kan enkelen weken duren, voordat deze vervaagd. Om een supernova te detecteren kan data worden verzameld gedurende verschillende nachten. Daglicht kan namelijk ongewenste ruis opleveren. Data kan per nacht verschillen vanwege bijvoorbeeld bewolking en maanlicht.

  Aangenomen wordt dat de intensiteit van de achtergrond pixel A_i gedurende nacht i normaal verdeeld is, met $A_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ en de intensiteit van de source pixel S_i gedurende nacht i normaal verdeeld is met $S_i \sim \mathcal{N}(\mu_i + \psi_i, \sigma_i^2)$. Om objecten te detecteren kan ook hier gebruik worden gemaakt van meervoudig toetsen, waarbij voor $i = 1, \ldots, n$

  - $H_{0,i} : \psi_i = 0$
  - $H_{1,i} : \psi_i > 0$.

  Op astronomisch foto's is een supernova bijvoorbeeld te herkennen door een gele vlek. Dit correspondeert met het uitgestraalde licht. In de bijbehorende gele pixels geldt dan dat $\psi_i > 0$. Dit is slechts in een klein aantal pixels van de afbeelding het geval, waardoor het vinden van supernova's beschouwd kan worden als een sparse probleem.

- Geneeskunde: Het detecteren van genen die verantwoordelijk zijn voor ziekte.

  Voordat men op eventuele genezing overgaat, wordt er in de medische wereld bij ziekte van een patiënt eerst nagegaan welke genen verantwoordelijk zijn voor ziekte. Dit zijn de genen die een afwijking hebben ten opzichte van de rest. Het aantal genen dat een afwijking heeft is ten opzichte van het geheel klein. Daarom is dit een sparse probleem. Ook kan gebruik worden gemaakt van meervoudig toetsen om de genen te detecteren die verantwoordelijk zijn voor ziekte.

3 Meervoudig toetsen: deel 1

Bij ons probleem $(*)$ hebben we een verzameling observaties die in een vector Y zitten. Ons doel is om de parameters $\theta_{0,i}$, voor $i = 1, \ldots, n$, te vinden die ongelijk zijn aan nul. Om hiertoe een poging te ondernemen kunnen we gebruik maken van meervoudig toetsen.

Bij meervoudig toetsen hebben we een verzameling hypothesen die we simultaan willen gaan testen. Een manier om dit te doen is door iedere hypothese afzonderlijk te gaan testen.

Voor $i = 1, \ldots, n$, testen we de volgende nulhypothese.

- $H_{0,i} : \theta_{0,i} = 0$
- $H_{1,i} : \theta_{0,i} \neq 0$

In dit hoofdstuk bekijken we verschillende methoden die we kunnen gebruiken om de hypothesen te testen.

3.1 Type 1 en 2 fouten

Bij de Neyman-Pearson benadering (zie [4]) voor het testen van hypothesen staan twee fouten centraal. We onderscheiden de volgende twee fouten.

- Type 1 fout: H_0 wordt verworpen, terwijl deze waar is.
- Type 2 fout: H_0 wordt niet verworpen, terwijl deze niet waar is.

Schematisch kunnen we dit als volgt weergeven.

	H_0 waar	H_0 niet waar
H_0 verwerpen	Type 1 fout	✓
H_0 niet verwerpen	✓	Type 2 fout

Het symbool ✓ betekent in dit geval dat de juiste beslissing wordt genomen.

Daarnaast is de volgende terminologie gebruikelijk:

- Het significantieniveau $\boldsymbol{\alpha}$ is de kans op het maken van een Type 1 fout.
- De kans op het maken van een Type 2 fout wordt genoteerd met $\boldsymbol{\beta}$.

Voorbeelden

Om voorbeelden van Type 1 en 2 fouten te geven die kunnen optreden in de praktijk kunnen we de genoemde toepassingen van sparsity uit sectie 2.1 gebruiken. Het hangt echter maar net van de toepassing af welke fout vervelender kan zijn.

Het onderzoeken van nieuwe regio's en objecten in de ruimte vergt veel tijd en is bovendien kostbaar. Bij het vinden van supernova's willen astronomen daarom

bijvoorbeeld zo weinig mogelijk valse detecties doen. Voor astronomen is het daarom vooral belangrijk om het aantal Type 1 fouten minimaal te houden.

Bij de bestrijding van ziekten daarentegen kunnen Type 2 fouten nadelige gevolgen hebben. Stel dat de nulhypothese luidt dat een gen niet verantwoordelijk is voor ziekte en dat dit volgens de alternatieve hypothese wel zo is. Dan kan het niet verwerpen van de nulhypothese, terwijl deze niet waar is, als gevolg hebben dat de genezing van een bepaalde ziekte niet succesvol zal verlopen.

3.2 Simple multiple testing

Als statistische toets gebruiken we de likelihood ratio. De likelihood ratio defniëren we als

$$L(Y) = \frac{f_0(Y)}{f_1(Y)},$$

met $Y = (Y_1, Y_2, \ldots, Y_n)$ zoals in (1).

$f_0(Y)$ en $f_1(Y)$ zijn de likelihood functies voor respectievelijk de nulhypothese en alternatieve hypothese.

Omdat $\epsilon_i \sim \mathcal{N}(0,1)$ en $Y_i = \theta_{0,i} + \epsilon_i$, geldt dat $Y_i \sim \mathcal{N}(\theta_{0,i}, 1)$ voor $i = 1, \ldots, n$.

Daarnaast maken we onderscheid tussen een enkelvoudige hypothese en een samengestelde hypothesen. Eerstgenoemde specificeert de parameters(s) voor een kansverdeling. Bij een samengestelde hypothese is dat niet het geval. Onze $H_{0,i}$ is een voorbeeld van een enkelvoudige hypothese, terwijl $H_{1,i}$ een samengestelde hypothese is.

We willen nu $L(Y_i)$ bepalen voor $i = 1, \ldots, n$, maar voordat we dit doen introduceren we eerst het Neyman-Pearson lemma.

Lemma 3.1 (Neyman-Pearson (uit [4]))**.** *Veronderstel dat H_0 en H_1 enkelvoudige hypothesen zijn en dat de toets die H_0 verwerpt als de likelihood ratio kleiner is dan een kritieke waarde c bij een significantieiniveau α de grootste kans heeft op het terecht verwerpen van de nulhypothese.*
Dan heeft iedere andere toets met een significantieniveau kleiner dan of gelijk aan α een kleinere of even grote kans op het terecht verwerpen van de nulhypothese als bij de likelihood ratio toets.

De kans op het terecht verwerpen van de nulhypothese noteren we met $1 - \beta$, met β zoals in de vorige paragraaf.

Het lemma vereist dat we twee enkelvoudige hypothesen hebben, terwijl bij ons probleem de alternatieve hypothese samengesteld is. In dit geval kunnen we het lemma uitbreiden, door gebruik te maken van een *uniformly most powerful (UMP)* toets. Een toets is UMP als de kans op het terecht verwerpen van de nulhypothese het grootst is bij iedere enkelvoudige alternatieve hypothese voor gegeven significantieniveau α. Volgens het lemma is de likelihood ratio een UMP toets.

We kunnen $L(Y_i)$ schrijven als

$$
\begin{aligned}
L(Y_i) &= \frac{f_0(Y_i)}{f_1(Y_i)} \\
&= \frac{\frac{1}{\sqrt{2\pi}} \cdot \exp[-\frac{1}{2}(Y_i - \theta_{0,i})^2]}{\frac{1}{\sqrt{2\pi}} \cdot \exp[-\frac{1}{2}(Y_i - \theta_{1,i})^2]} \\
&= \frac{\exp[-\frac{1}{2}(Y_i - \theta_{0,i})^2]}{\exp[-\frac{1}{2}(Y_i - \theta_{1,i})^2]}
\end{aligned}
$$

$\theta_{1,i}$ staat hier voor de waarde van $\theta_{0,i}$ in het geval de alternatieve hypothese geldt. De nulhypothese wordt verworpen als de ratio zich onder een kritieke waarde c bevindt. In dat geval is de kans op het terecht verwerpen van de nulhypothese het grootst volgens bovenstaand lemma. Daarom willen we dat de ratio relatief klein is.

Kleine waarden voor $L(Y_i)$ komen overeen met kleine waarden voor

$$
-(Y_i - \theta_{0,i})^2 + (Y_i - \theta_{1,i})^2.
$$

Dit kunnen we schrijven als

$$
\begin{aligned}
&= -Y_i^2 + 2 \cdot Y_i \cdot \theta_{0,i} - \theta_{0,i}^2 + Y_i^2 - 2 \cdot Y_i \cdot \theta_{1,i} + \theta_{1,i}^2 \\
&= 2 \cdot Y_i \cdot (\theta_{0,i} - \theta_{1,i}) + \theta_{1,i}^2 - \theta_{0,i}^2.
\end{aligned}
$$

Merk op dat we in het begin van dit hoofdstuk hebben vermeld dat voor ons probleem geldt dat $H_{0,i} : \theta_{0,i} = 0$. Invullen van $\theta_{0,i} = 0$ geeft

$$
-2 \cdot Y_i \cdot \theta_{1,i} + \theta_{1,i}^2. \tag{2}
$$

Kleine waarden voor $L_(Y_i)$ komen overeen met kleine waarden voor (2).

Als $\theta_{1,i} > 0$, dan is (2) klein als Y_i positief en groot is.
Als $\theta_{1,i} < 0$, dan is (2) klein als $-Y_i$ negatief en klein is.

Als $|Y_i|$ dus groot is, is de likelihood ratio klein en zullen we de nulhypothese verwerpen.

Uit lemma 3.1 volgt dat de toets met het grootste onderscheidend vermogen, d.w.z. waarbij de kans op het terecht verwerpen van de nulhypothese het grootst is, de nulhypothese verwerpt als $|Y_i| > c_1$ voor zekere $c_1 \in \mathbb{R}$. De waarde c_1 wordt gekozen zodanig dat $P(|Y_i| > c_1) = \alpha$ in het geval $H_{0,i}$ waar is, waarbij α het significantieniveau is. Merk op dat Y_i onder $H_{0,i}$ een verwachting heeft van 0 en een standaardafwijking van 1. Met R kunnen we nu c_1 bepalen. Merk op dat we te maken hebben met een tweezijdige toets. Bovendien kiezen we $\alpha = 0.05$.

```
> c_1 = qnorm(1-0.05/2)
> c_1
[1] 1.959964
```

In de simulaties zullen we de gevonden waarde voor c_1 afronden op twee decimalen.

Het experiment waarbij we iedere hypothese afzonderlijk testen noemen we in het vervolg 'simple multiple testing'.

3.2.1 Simulaties

Nu bekijken we wat resultaten van simulaties. We nemen een vector θ_0 ter lengte $n = 1000$, waarbij de laatste 100 elementen een waarde $A \neq 0$ toegekend krijgen. Bij het testen van de verzameling hypothesen gebruiken we een significantieniveau van $\alpha = 0.05$. Vervolgens bepalen we de proportie van Type 1 en Type 2 fouten. De proporties worden als volgt bepaald.

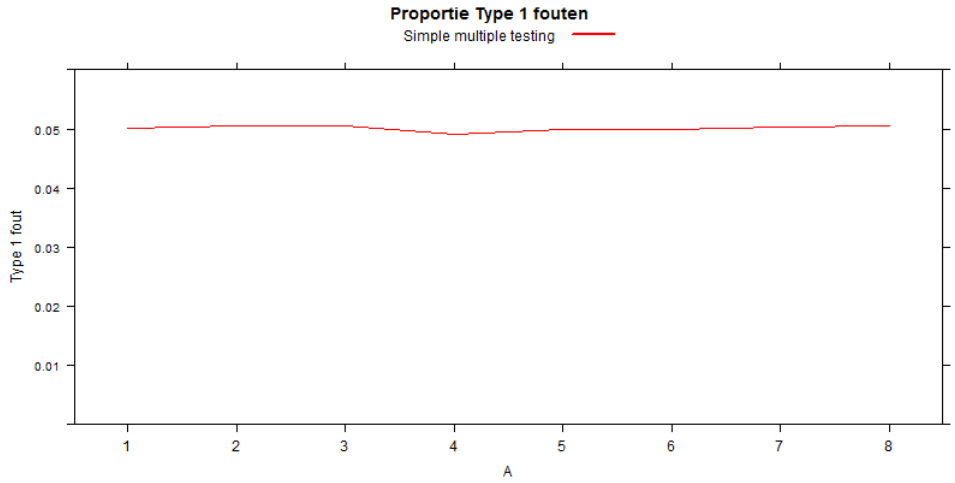
Voor $i = 1, \ldots, 900$ geldt

$$\text{Proportie Type 1 fouten} = \frac{\#\{|Y_i| > 1.96\}}{\#\{\theta_{0,i} = 0\}} = \frac{\#\{|Y_i| > 1.96\}}{900}$$

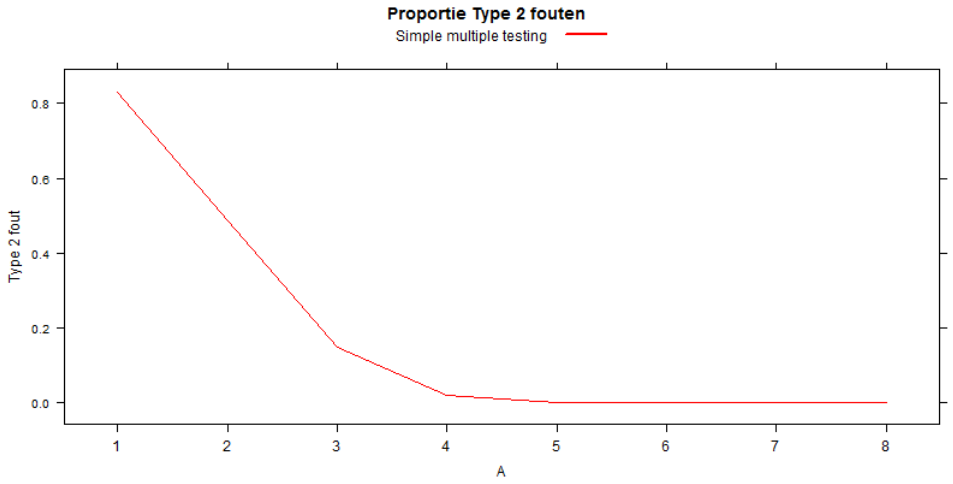
En voor $i = 901, \ldots, 1000$ geldt

$$\text{Proportie Type 2 fouten} = \frac{\#\{|Y_i| \leq 1.96\}}{\#\{\theta_{0,i} \neq 0\}} = \frac{\#\{|Y_i| \leq 1.96\}}{100}$$

Voor acht verschillende waarden van A zetten we de Type 1 en 2 fouten uit in een plot.



Figuur 2: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$ bij Simple multiple testing



Figuur 3: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$, bij Simple multiple testing

We zien dat de proportie van Type 2 fouten groter wordt naarmate we voor A kleinere waarden kiezen. Een kleine waarde voor $|A|$ betekent immers een waarschijnlijk kleine waarde voor $|Y|$, waardoor de kans groter is dat deze onder de grens van 1,96 ligt voor $\alpha = 0.05$. Het significantieniveau $\alpha = 0.05$ betekent dat de kans op het onterecht verwerpen van de nulhypothese per definitie gelijk is aan 0.05. Dit is de reden dat de proportie van Type 1 fouten begrensd is door 0.05.

Daar waar de Type 1 fouten min of meer gelijk zijn voor verschillende waarden van A, gaan de Type 2 fouten naar nul voor $A \geq 5$. We willen dat de proporties van fouten zo klein mogelijk zijn. Bovendien willen we een redelijke balans tussen de fouten hebben. Zoals we kunnen zien gaat dit niet lukken met simple multiple testing voor voornamelijk $A \leq 3$. Daarom kijken we naar alternatieve methoden.

3.3 Bonferroni Correctie

De family-wise error rate (FWER) is de kans dat we minstens één keer H_0 ten onrechte verwerpen (zie [6]). Er geldt voor n onafhankelijke tests en een significantieniveau α:

$$\begin{aligned} FWER &= P(\text{minstens één } H_0 \text{ ten onrechte verwerpen}) \\ &= 1 - P(\text{geen enkele } H_0 \text{ ten onrechte verwerpen}) \\ &= 1 - (1-\alpha)^n \end{aligned}$$

Wanneer we voor iedere i de nulhypothese afzonderlijk gaan testen, is er een grote kans dat we een Type 1 fout maken. Het volgende voorbeeld laat dit zien.

Voorbeeld 3.2. Voor $n = 100$ en $\alpha = 0.05$ krijgen we:

$$\begin{aligned} FWER &= 1 - (1 - \alpha)^n \\ &= 1 - (1 - 0.05)^{100} \\ &\approx 0.99. \end{aligned}$$

Uit het voorbeeld volgt dat het vrijwel zeker is dat we minstens één keer de nulhypothese ten onrechte verwerpen. We willen dat $FWER \leq \alpha$, voor n tests. In dat geval spreken we over het hebben van controle over de family-wise error rate. De Bonferroni Correctie biedt ons hiervoor mogelijk uitkomst.

Bonferroni Correctie (zie [5]): Kies als significantieniveau $\frac{\alpha}{n}$ voor iedere test i, met $1 \leq i \leq n$.

Voorbeeld 3.3. Als we de Bonferroni Correctie toepassen voor dezelfde waarden van n en α als in het vorige voorbeeld, krijgen we

$$\begin{aligned} FWER &= 1 - (1 - \frac{\alpha}{n})^n \\ &= 1 - (1 - \frac{0.05}{100})^{100} \\ &\approx 0.049. \end{aligned}$$

We zien dat de family-wise error rate een flink stuk kleiner is geworden. De Bonferroni Correctie is echter een conservatieve methode. Dit betekent dat we H_0 niet snel zullen verwerpen. Voor $n = 1000$ vinden we met behulp van R voor $\alpha = 0.05$ een hogere grenswaarde voor de Bonferroni correctie.

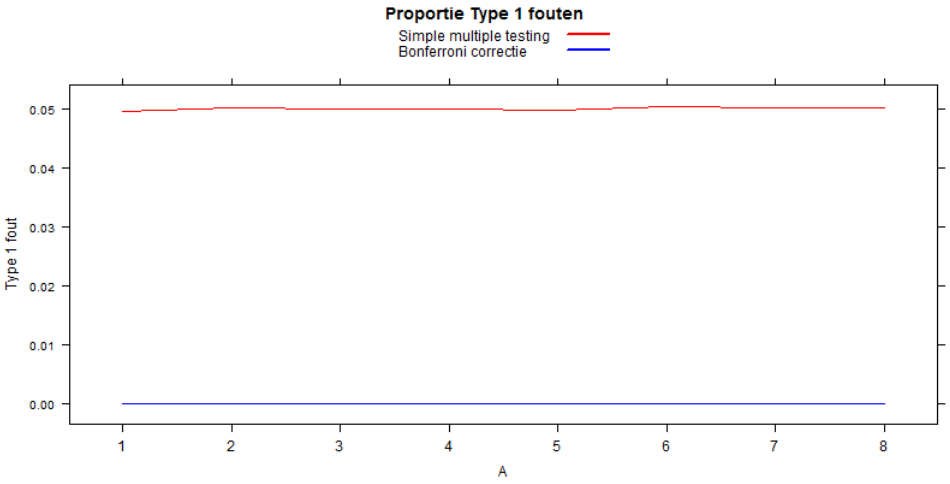
```
> qnorm(1-((0.05/2)/1000))
[1] 4.055627
```

De grenswaarde van 4.06 is nogal hoog. Bij een vector θ_0 met 1000 elementen, waarvan er 100 ongelijk zijn aan nul, zullen er weinig observaties boven de grenswaarde uitkomen. Dit verklaart waarom de Bonferroni correctie conservatief is. Het gevolg is dat de kans waarschijnlijk is dat we een groot aantal Type 2 fouten maken. Dit wordt duidelijk in de simulaties.

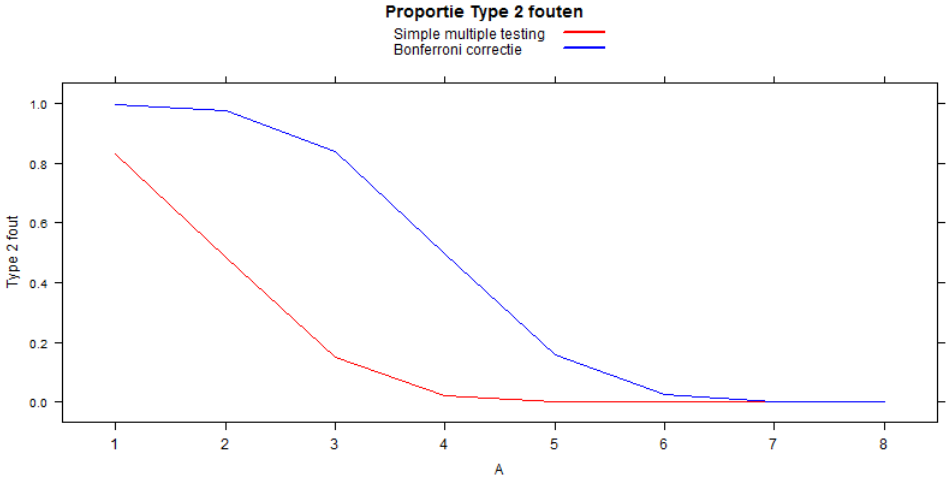
3.3.1 Simulaties

We kiezen voor de simulaties opnieuw voor $\alpha = 0.05$ en $n = 1000$, net als in de vorige paragraaf. Voor acht verschillende waarden van A voegen we de Type 1 en 2 fouten die optreden bij de Bonferroni correctie bij de plots uit de vorige paragraaf.

We krijgen de volgende resultaten.



Figuur 4: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$



Figuur 5: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$

Na het toepassen van de Bonferroni correctie zien we dat de proportie van Type 1 fouten nul nadert. De proportie van Type 2 fouten is in de meeste gevallen echter een flink stuk groter dan bij simple multiple testing. Hoe lager de gekozen waarde voor A is, des te groter wordt de proportie van Type 2 fouten. Dit is ook wat we verwachten. Een lage waarde voor A geeft in de meeste gevallen lage waarden voor de observaties, waardoor deze zich onder de grens van 4.06 bevinden. Hierdoor zullen we een echt signaal niet snel detecteren. De Bonferroni richt zich op het verkleinen van de Type 1 fouten, maar houdt geen rekening met de Type 2 fouten.

3.4 Benjamini-Hochberg methode

Daar waar de family-wise error rate zich richt op de kans op minstens een Type 1 fout, kwamen Benjamini en Hochberg in 1995 in hun artikel 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing' (zie [6]) met een alternatieve benadering voor het probleem dat optreedt bij meervoudig toetsen. Zij introduceerden de False Discovery Rate.

De volgende uiteenzetting is vooral gebaseerd op de lecture notes die onder [7] en [8] staan vermeld bij de referenties.

Definitie 3.4. De False Discovery Proportie (FDP) definiëren we als

$$FDP = \frac{\#\text{ ten onrechte verworpen nulhypothesen}}{\#\text{ verworpen nulhypothesen}}$$

Definitie 3.5. De False Discovery Rate (FDR) definiëren we als

$$FDR = \mathbb{E}[FDP]$$

Om controle te houden over de FDR willen we dat $FDR \leq \alpha$, waarbij α opnieuw het significantieniveau is. De eis dat $FDR \leq \alpha$ is minder streng dan $FWER \leq \alpha$. Om ervoor te zorgen dat $FDR \leq \alpha$ kunnen we de Benjamini-Hochberg methode gebruiken. Voordat we dit doen, leggen we eerst uit wat een p-waarde is.

De p-waarde is de kans dat onder de nulhypothese een extremere resultaat wordt behaald, dan dat er daadwerkelijk wordt geobserveerd. Hoe kleiner de p-waarde is, des te waarschijnlijker het is om de nulhypothese te verwerpen. In het algemeen wordt de nulhypothese verworpen als de p-waarde kleiner dan of gelijk aan het significantieniveau α is.

De volgende stappen worden doorlopen bij de Benjamini-Hochberg methode.

Benjamini-Hochberg methode

1. Bepaal de p-waarden p_i voor de corresponderende nulhypothesen $H_{0,i}$, voor $1 \leq i \leq n$. Merk op dat we bij ons probleem te maken hebben met een tweezijdige overschrijdingskans.

2. Rangschik de p-waarden van klein naar groot. We krijgen

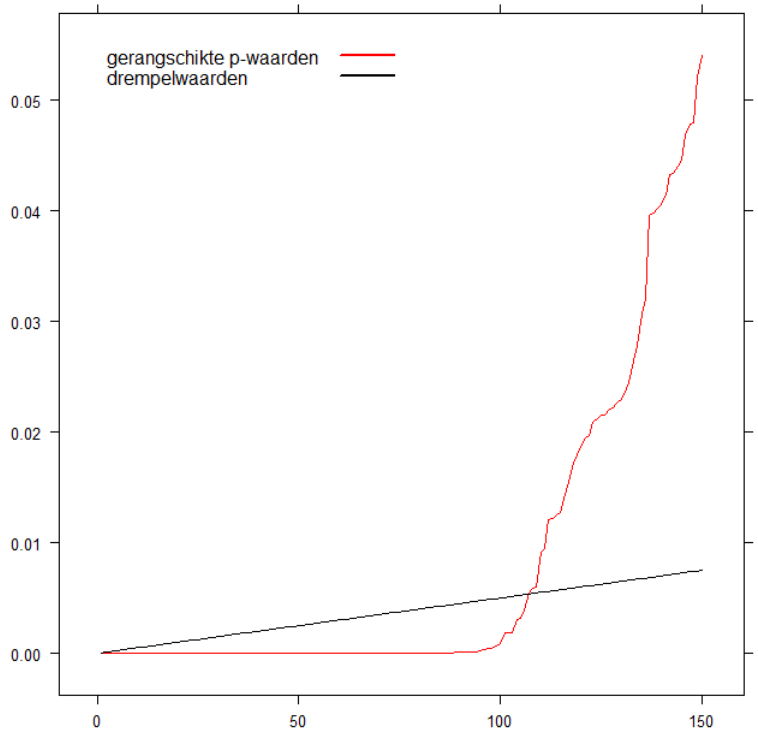
$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$$

3. Vergelijk iedere p-waarde $p_{(i)}$ met de corresponderende drempelwaarde $\frac{i \cdot \alpha}{n}$, voor $1 \leq i \leq n$.

4. Neem $j = max\{i : p_{(i)} < \frac{i \cdot \alpha}{n}\}$. Verwerp $H_{0,i}$ voor $i = 1, \ldots, j$.

Ter verduidelijking staat een simulatie van de procedure afgebeeld in het volgende figuur.

We hebben gekozen voor een vector θ_0 ter lengte 1000, waarbij 100 elementen de waarde $A = 5$ hebben gekregen. Hier staan de eerste 150 gerangschikte p-waarden uitgezet tegenover hun bijbehorende drempelwaarden.

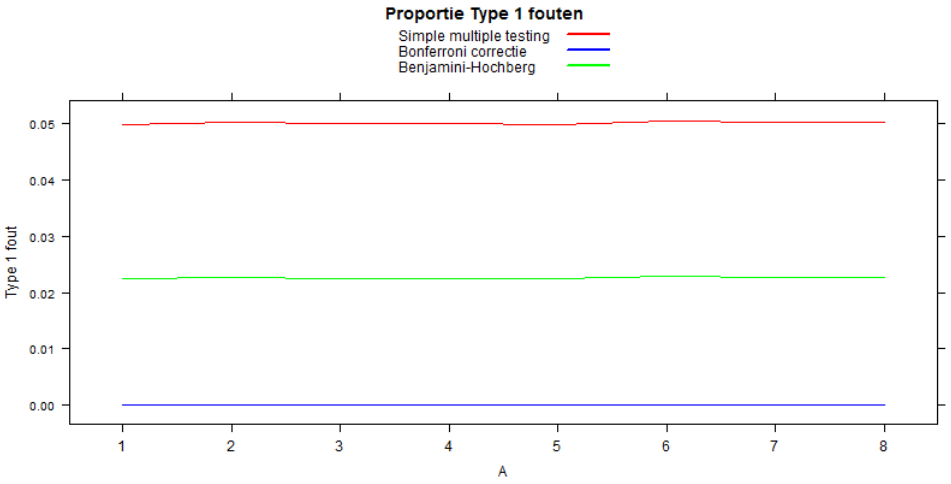


Figuur 6: De eerste 150 gerangschikte p-waarden uitgezet tegenover hun corresponderende drempelwaarden

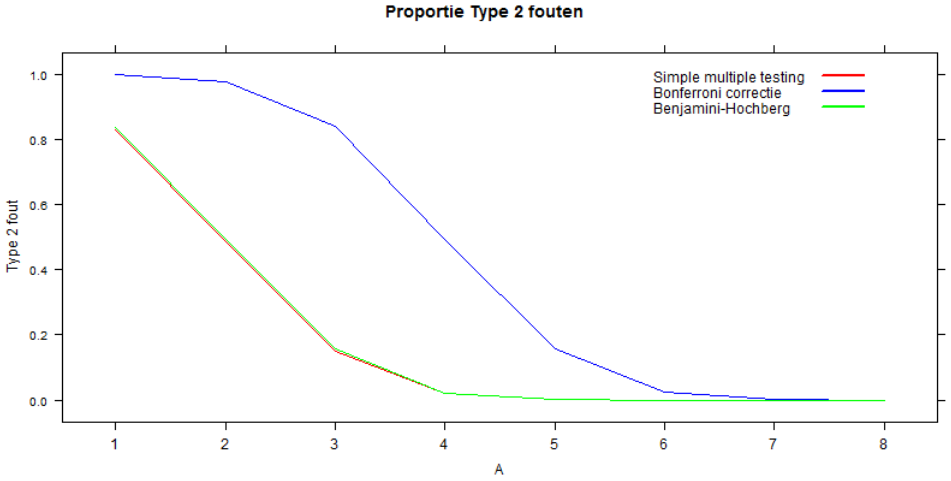
Uit de plot volgt dat we ongeveer de eerste 110 nulhypothesen zullen verwerpen.

3.4.1 Simulaties

We bekijken de resultaten van de simulaties nadat we de Benjamini-Hochberg methode hebben gebruikt. We nemen opnieuw $n = 1000$ en $\alpha = 0.05$. θ_0 bevat 100 elementen ongelijk aan nul. De resultaten voegen we bij de plots uit de vorige paragraaf.



Figuur 7: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$



Figuur 8: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$

Wat onmiddelijk opvalt is dat de Type 2 fouten bij de Benjamini-Hochberg methode een flink stuk kleiner zijn dan bij de Bonferroni correctie. Deze proporties van Type 2 fouten bij de Benjamini-Hochberg methode verschillen echter niet veel van de resultaten bij simple multiple testing.

Als we kijken naar de Type 1 fouten, dan zijn de proporties bij de Benjamini-Hochberg methode ongeveer twee keer zo klein als bij simple multiple testing. De proporties waren echter al niet groot, dus bij de Benjamini-Hochberg methode ook niet. De Bonferroni correctie doet het hier nog wel beter bij de Type 1 fouten, al zijn alle waarden bij de drie methoden klein.

3.5 Vervolg

Als we de resultaten uit de grafieken bekijken, dan levert de Bonferroni correctie niet het gewenste resultaat op vanwege het enorme verschil tussen de proporties van Type 1 en 2 fouten. Voor $A \geq 4$ zien we dat de proporties bij simple multiple testing en de Benjamini-Hochberg methode redelijk in balans zijn. Bij $A \leq 3$ zien we dat er een omslag plaatsvindt en wordt het verschil groter en groter.

In een poging om toch een redelijke balans tussen de proporties van Type 1 en 2 fouten te vinden waarbij deze minimaal zijn, gebruiken we een sparse Bayesiaanse toetstechniek, namelijk de horseshoe prior. Voordat we dit doen, behandelen we eerst het concept van Bayesiaanse statistiek.

4 Bayesiaanse statistiek

In de Bayesiaanse statistiek beschouwen we de gezochte waarde als een stochastische variable, waarvan we a priori een inschatting kunnen maken. Stel we hebben:

$$Y \sim P_\theta, \quad \theta \sim f(\theta)$$

Dat wil zeggen een stochast Y met een zekere kansverdeling P die afhangt van θ. Deze θ heeft een eigen kansverdeling $f(\theta)$.

$f(\theta)$ noemen we de prior kansverdeling. Dit is onze initiële inschatting voor θ. Met behulp van de regel van Bayes krijgen we een posterior verdeling voor θ, namelijk $f(\theta|Y)$. We updaten onze initiële inschatting voor θ.

4.1 De regel van Bayes

De regel van Bayes kunnen we als volgt definiëren.

Definitie 4.1. Zij A en B twee gebeurtenissen met $P(B) > 0$. Dan geldt

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Dit leiden we als volgt af. De formule voor voorwaardelijke kansen wordt gegeven door:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Op dezelfde manier kunnen we schrijven:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Hieruit volgt:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In het geval we twee continue stochasten X en Y hebben met respectievelijk de kansdichtheden f_X en f_Y, kunnen we de regel van Bayes schrijven als:

$$f(x|y) = \frac{f_Y(y|x) \cdot f_X(x)}{\int f_Y(y|x) \cdot f_X(x) dx}$$

Voorbeeld 4.2. (Twee Dobbelstenen)

Stel we hebben

- 1 zuivere dobbelsteen

- Een dobbelsteen met kans $\frac{1}{2}$ op het gooien van zowel een 1 als een 4

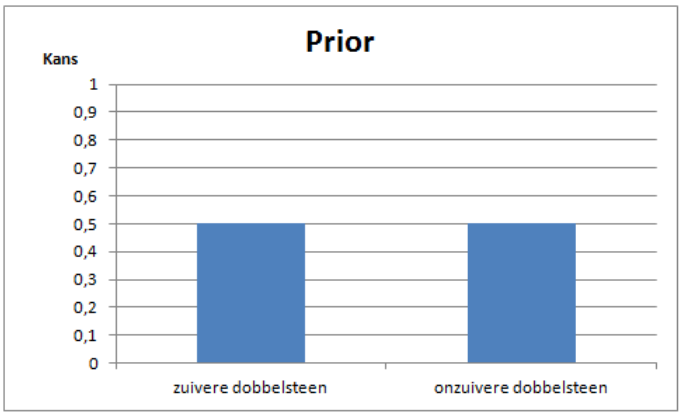
We gooien met een willekeurig gekozen dobbelsteen een 4.

Wat is de kans dat we met de zuivere dobbelsteen hebben gegooid?

Oplossing: Zij Y het aantal ogen dat we gooien en θ de gebeurtenis dat we gooien met de zuivere dobbelsteen. Gevraagd wordt $P(\theta|Y = 4)$. Hiervoor kunnen we de Regel van Bayes gebruiken:

$$P(\theta|Y = 4) = \frac{P(Y = 4|\theta) \cdot P(\theta)}{P(Y = 4)}$$

Het ligt voor de hand om als prior verdeling voor θ te kiezen voor gelijke kansen om te gooien met zowel de zuivere als de onzuivere dobbelsteen, namelijk een half.



Figuur 9: De prior verdeling voor θ

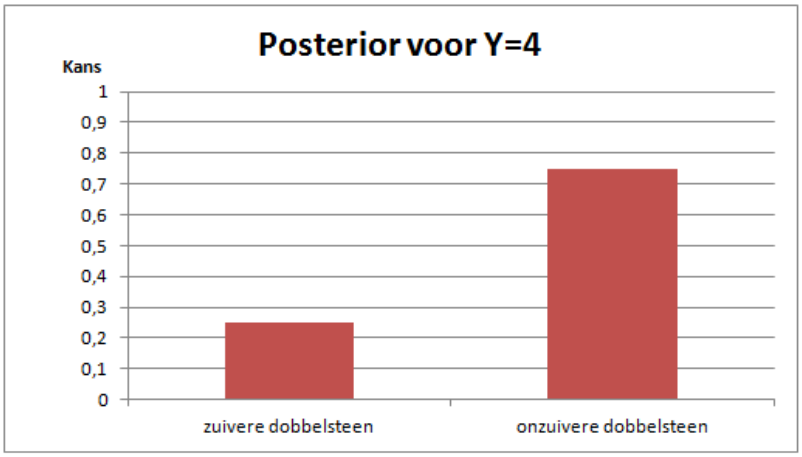
Er geldt:

- $P(Y = 4|\theta) = \frac{1}{6}$
- $P(\theta) = \frac{1}{2}$
- $P(Y = 4) = \frac{1}{6} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{12} + \frac{1}{4} = \frac{4}{12}$

Dit geeft

$$P(\theta|Y = 4) = \frac{P(Y = 4|\theta) \cdot P(\theta)}{P(Y = 4)} = \frac{\frac{1}{6} \cdot \frac{1}{2}}{\frac{4}{12}} = \frac{1}{4}$$

De posterior verdeling voor θ gegeven $Y = 4$ staat weergegeven in het volgende figuur.



Figuur 10: De posterior verdeling van θ voor $Y = 4$

Door gebruik te maken van de beschikbare gegevens en informatie hebben we de prior kansverdeling met behulp van de regel van Bayes bijgewerkt tot een posterior kansverdeling.

Na het bepalen van de posterior verdeling in het voorbeeld zullen we schatten dat we met de onzuivere dobbelsteen hebben gegooid. Dit is de MAP (maximum a posteriori) schatter. In het algemeen geeft de posterior verdeling ons informatie over de waarschijnlijkheid van de te schatten parameters gegeven de geobserveerde data.

Bij de horseshoe prior in het volgend hoofdstuk zullen we ook aspecten van de bijbehorende posterior gaan gebruiken, namelijk de verwachting en variantie.

5 Horseshoe prior

In het artikel 'The horseshoe estimator for sparse signals' (zie [9]) kwamen Carvalho, Polson en Scott met een nieuwe benadering voor sparse problemen. Zij noemden dit destijds de 'horseshoe estimator.' Voor een sparse vector θ stelden zij een model op voor het schatten en voorspellen van deze vector. Dit model werd de horseshoe prior genoemd. In een eerdere versie van [9] uit 2008, gebruikten zij de horseshoe voor meervoudig toetsen. Met behulp van Bayesiaanse statistiek kunnen we een posterior dichtheid afleiden voor de sparse vector θ. De bijbehorende posterior mean geeft ons dan nieuwe inzichten over θ.

5.1 Posterior mean

De horseshoe prior luidt als volgt:

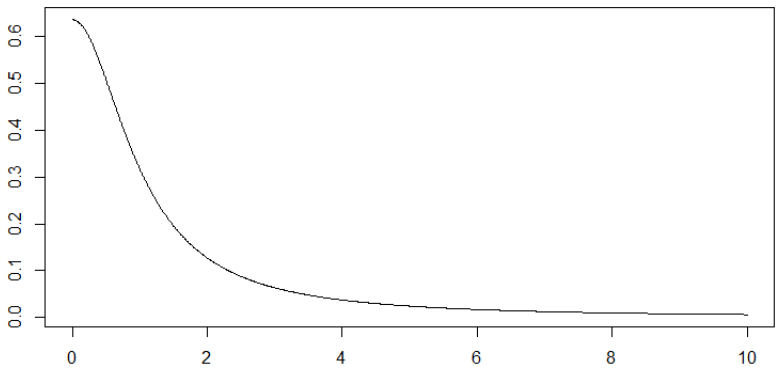
Prior verdeling: Voor $i = 1, \ldots, n$:

- $\theta_i | \lambda_i, \tau \sim \mathcal{N}(0, (\lambda_i \tau)^2)$

  Dus θ_i is normaal verdeeld met verwachting 0 en variantie $(\lambda_i \tau)^2$. De parameters λ_i en τ van de prior verdeling noemen we hyperparameters.

- $\lambda_i \sim C^+(0, 1)$.

  Dit wil zeggen dat λ_i een 'half-Cauchy' verdeling heeft met kansdichtheid $p(\lambda) = \dfrac{2}{\pi(1 + \lambda^2)}$.



Figuur 11: Een weergave van de $C^+(0, 1)$ verdeling

  In dit geval is λ_i een lokale hyperparameter, omdat deze wordt gespecificeerd door een kansverdeling. De exacte waarde is niet bekend.

- $\tau = \dfrac{\#\{\theta_i \neq 0\}}{n}$.

  τ is gelijk aan het aantal elementen ongelijk aan nul gedeeld door het totaal aantal elementen in θ_0. De keuze hiervoor is gebasseerd op resultaten uit

[11], net als de keuze voor τ in de simulaties in het volgende hoofdstuk. Daar nemen we

$$\tau = \frac{\#\{\theta_i \neq 0\}}{n} \cdot \sqrt{2\frac{n}{\#\{\theta_i \neq 0\}}}.$$

Hier geldt dat τ een globale hyperparameter is, omdat τ een vooraf gespecificeerde waarde heeft.

Met Bayesiaanse statistiek willen we nu de posterior mean $\mathrm{E}[\theta_i|y]$ afleiden. Dit is de verwachtingswaarde van θ_i gegeven de vector y met de observaties. De posterior mean geeft ons een schatting voor θ_i.

De posterior mean (zie [10]) kunnen we bepalen met de formule

$$\mathrm{E}[\theta_i|y] = y + \frac{m'_\tau(y)}{m_\tau(y)}.$$

De regel van Bayes geeft ons:

$$\pi(\theta_i|y_i) = \frac{f_{\theta_i}(y_i) \cdot \pi(\theta_i)}{\int_{-\infty}^{\infty} f_{\theta_i}(y_i) \cdot \pi(\theta_i) d\theta_i}. \tag{3}$$

We hebben de volgende kansdichtheden:

- $p(y_i - \theta_i) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(y-\theta_i)^2}{2}}$
- $\pi(\theta_i) = \displaystyle\int_0^\infty \pi(\theta_i|\lambda_i) \cdot \pi(\lambda_i) d\lambda_i$
- $\pi(\theta_i|\lambda_i) = \dfrac{1}{\sqrt{2\pi}\lambda_i\tau} \cdot e^{-\frac{\theta_i^2}{2(\lambda_i\tau)^2}}$
- $\pi(\lambda_i) = \dfrac{2}{\pi(1+\lambda^2)}$

Dan kunnen we (3) schrijven als

$$\pi(\theta_i|y_i) = \frac{p(y_i - \theta_i) \cdot \pi(\theta_i)}{\int_{-\infty}^{\infty} p(y_i - \theta_i) \cdot \pi(\theta_i) d\theta_i}. \tag{4}$$

We definiëren

$$m_\tau(y) = \int_{-\infty}^{\infty} p(y_i - \theta_i) \cdot \pi(\theta_i) d\theta_i. \tag{5}$$

Er geldt dat $m_\tau(y)$ de marginale verdeling van y is.

Substitueren we de bovenstaande kansdichtheden in (5), dan krijgen we

$$\begin{aligned}
m_\tau(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(y-\theta_i)^2}{2}} \cdot \left[\int_0^{\infty} \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{\theta_i^2}{2(\lambda_i\tau)^2}} \cdot \frac{2}{\pi(1+\lambda_i^2)} d\lambda_i \right] d\theta_i \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(y-\theta_i)^2}{2}} \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{\theta_i^2}{2(\lambda_i\tau)^2}} \cdot \frac{2}{\pi(1+\lambda_i^2)}\ d\theta_i\ d\lambda_i \\
&= \int_0^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(y-\theta_i)^2}{2}} \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{\theta_i^2}{2(\lambda_i\tau)^2}} d\theta_i \right] \frac{2}{\pi(1+\lambda_i^2)}\ d\lambda_i \\
&= \int_0^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{(\lambda_i\tau)^2\cdot(y^2-2y\cdot\theta_i+\theta_i^2)+\theta_i^2}{2(\lambda_i\tau)^2}} d\theta_i \right] \frac{2}{\pi(1+\lambda_i^2)}\ d\lambda_i\ .
\end{aligned}$$

De termen in de integraal tussen de vierkante haken willen we schrijven in dezelfde vorm als de kansdichtheid $g(\theta_i)$ van de normale verdeling met verwachting μ_* en variantie σ_*^2:

$$\begin{aligned}
g(\theta_i) &= \frac{1}{\sqrt{2\pi}\sigma_*} \cdot e^{-\frac{(\theta_i-\mu_*)^2}{2\sigma_*^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_*} \cdot e^{-\left[\frac{\theta_i^2}{2\sigma_*^2} - \frac{\theta_i\mu_*}{\sigma_*^2} + \frac{\mu_*^2}{2\sigma_*^2}\right]}.
\end{aligned}$$

We kunnen schrijven:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{(\lambda_i\tau)^2\cdot(y^2-2y\cdot\theta_i+\theta_i^2)+\theta_i^2}{2(\lambda_i\tau)^2}} d\theta_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau} \cdot e^{-\frac{(1+(\lambda_i\tau)^2)\cdot\theta_i^2-(\lambda_i\tau)^2\cdot 2y\cdot\theta_i+(\lambda_i\tau)^2\cdot y^2}{2(\lambda_i\tau)^2}} d\theta_i\ .
\end{aligned}$$

Hieruit volgt:

$$\frac{1}{2\sigma_*^2} = \frac{1+(\lambda_i\tau)^2}{2(\lambda_i\tau)^2} \Longrightarrow \sigma_*^2 = \frac{(\lambda_i\tau)^2}{1+(\lambda_i\tau)^2}$$

$$\frac{\mu_*}{\sigma_*^2} = y \Longrightarrow \mu_* = \sigma_*^2 \cdot y = \frac{(\lambda_i\tau)^2}{1+(\lambda_i\tau)^2} \cdot y$$

$$\frac{\mu_*^2}{2\sigma_*^2} = \frac{\frac{(\lambda_i\tau)^4}{(1+(\lambda_i\tau)^2)^2}\cdot y^2}{2\cdot\frac{(\lambda_i\tau)^2}{1+(\lambda_i\tau)^2}} = \frac{(\lambda_i\tau)^2}{2\cdot(1+(\lambda_i\tau)^2)} \cdot y^2\ .$$

Zodat

$$
\begin{aligned}
& e^{-\frac{(1+(\lambda_i\tau)^2)\cdot\theta_i^2-(\lambda_i\tau)^2\cdot 2y\cdot\theta_i+(\lambda_i\tau)^2\cdot y^2}{2(\lambda_i\tau)^2}} \\
&= e^{\frac{(1+(\lambda_i\tau)^2)\cdot\theta_i^2}{2(\lambda_i\tau)^2}-y\cdot\theta_i+\frac{(\lambda_i\tau)^2\cdot y^2}{2\cdot(1+(\lambda_i\tau)^2)}-\frac{(\lambda_i\tau)^2\cdot y^2}{2\cdot(1+(\lambda_i\tau)^2)}+\frac{y^2}{2}} .
\end{aligned}
$$

Er geldt

$$
-\frac{(\lambda_i\tau)^2\cdot y^2}{2\cdot(1+(\lambda_i\tau)^2)}+\frac{y^2}{2}=\frac{y^2}{2\cdot(1+(\lambda_i\tau)^2)}.
$$

Dit geeft uiteindelijk

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau}\cdot e^{-\frac{(\lambda_i\tau)^2\cdot(y^2-2y+\theta_i^2)+\theta_i^2}{2(\lambda_i\tau)^2}}d\theta_i \\
&= e^{-\frac{y^2}{2\cdot(1+(\lambda_i\tau)^2)}}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{2\pi}\cdot\lambda_i\tau}\cdot\frac{\sqrt{2\pi}\cdot\lambda_i\tau}{\sqrt{1+(\lambda_i\tau)^2}} \\
&\cdot\int_{-\infty}^{\infty} e^{-\frac{(\theta_i-\frac{(\lambda_i\tau)^2}{1+(\lambda_i\tau)^2}\cdot y)^2}{2\cdot\frac{(\lambda_i\tau)^2}{1+(\lambda_i\tau)^2}}}\cdot\frac{1}{\sqrt{2\pi}\cdot\frac{\lambda_i\tau}{\sqrt{1+(\lambda_i\tau)^2}}}\,d\theta_i \\
&= \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1+(\lambda_i\tau)^2}}\cdot e^{-\frac{y^2}{2(1+(\lambda_i\tau)^2)}}\cdot 1 .
\end{aligned}
$$

Zodat

$$
m_\tau(y)=\int_0^{\infty}\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1+(\lambda_i\tau)^2}}\cdot e^{-\frac{y^2}{2(1+(\lambda_i\tau)^2)}}\cdot\frac{2}{\pi(1+\lambda_i^2)}d\lambda_i . \tag{6}
$$

Neem $u=\dfrac{1}{1+(\lambda_i\tau)^2}$. Dit geeft

$$
\frac{du}{d\lambda_i}=-\frac{1}{(1+(\lambda_i\tau)^2)^2}\cdot 2\tau^2\lambda_i \tag{7}
$$

$$
\implies du=-\frac{2\tau^2\lambda_i}{(1+(\lambda_i\tau)^2)^2}d\lambda_i . \tag{8}
$$

En

$$\frac{1}{u} = 1 + (\lambda_i \tau)^2 \tag{9}$$
$$\implies \frac{1}{u} - 1 = (\lambda_i \tau)^2 \tag{10}$$
$$\implies \frac{\frac{1}{u} - 1}{\tau^2} = \lambda_i^2 \tag{11}$$
$$\implies 1 + \frac{\frac{1}{u} - 1}{\tau^2} = 1 + \lambda_i^2 \ . \tag{12}$$

Vullen we (8) en (12) in in (6), dan krijgen we

$$\begin{aligned}
m_\tau(y) &= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \sqrt{u} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \frac{2}{\pi \cdot (1 + \frac{\frac{1}{u}-1}{\tau^2})} \cdot \frac{(\frac{1}{u})^2}{2\tau^2 \cdot \sqrt{\frac{\frac{1}{u}-1}{\tau^2}}} \, du \\
&= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \sqrt{u} \cdot \frac{1}{1 + \frac{\frac{1}{u}-1}{\tau^2}} \cdot \frac{(\frac{1}{u})^2}{\tau \cdot \sqrt{\frac{1}{u} - 1}} \, du \\
&= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \frac{(u)^{-1\frac{1}{2}}}{\tau \cdot \sqrt{\frac{1}{u} - 1} + \frac{1}{\tau} \cdot (\frac{1}{u} - 1)^{1\frac{1}{2}}} \, du \\
&= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \frac{(u)^{-1\frac{1}{2}}}{\sqrt{\frac{1}{u} - 1} \cdot (\tau + \frac{1}{\tau} \cdot (\frac{1}{u} - 1))} \, du \\
&= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \frac{(u)^{-1\frac{1}{2}}}{(\frac{1}{u} - 1) \cdot (\tau \cdot (\frac{1}{u} - 1)^{-\frac{1}{2}} + \frac{1}{\tau} \cdot (\frac{1}{u} - 1)^{\frac{1}{2}})} \, du \\
&= -\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot u} \cdot \frac{(u)^{-\frac{1}{2}}}{\tau \cdot (1-u) \cdot ((\frac{1}{u} - 1)^{-\frac{1}{2}} + \frac{1}{\tau^2} \cdot (\frac{1}{u} - 1)^{\frac{1}{2}})} \, du \ .
\end{aligned}$$

Substitutie van $z = 1 - u$ geeft

$$\begin{aligned}
m_\tau(y) &= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{(1-z)^{-\frac{1}{2}}}{\tau \cdot z \cdot ((\frac{1}{1-z} - 1)^{-\frac{1}{2}} + \frac{1}{\tau^2} \cdot (\frac{1}{1-z} - 1)^{\frac{1}{2}})} \, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{1}{\tau \cdot z \cdot (1-z)^{\frac{1}{2}} \cdot ((\frac{z}{1-z})^{-\frac{1}{2}} + \frac{1}{\tau^2} \cdot (\frac{z}{1-z})^{\frac{1}{2}})} \, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{1}{\tau \cdot z \cdot (1-z)^{\frac{1}{2}} \cdot (z^{-\frac{1}{2}} \cdot (1-z)^{\frac{1}{2}} + \frac{1}{\tau^2} \cdot z^{\frac{1}{2}} \cdot (1-z)^{-\frac{1}{2}})} \, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{1}{\tau \cdot (1-z) \cdot z^{\frac{1}{2}} + \frac{1}{\tau} \cdot z^{1\frac{1}{2}}} \, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{1}{z^{\frac{1}{2}} \cdot (\tau \cdot (1-z) + \frac{1}{\tau} \cdot z)} \, dz
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}}}{\tau \cdot (1-z) + \frac{1}{\tau} \cdot z}\, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 \cdot (1-z) + z}\, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 - \tau^2 \cdot z + z}\, dz \\
&= \int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz\ .
\end{aligned}
$$

Hieruit volgt

$$
m'_\tau(y) = -\int_0^1 y \cdot (1-z) \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz\ .
$$

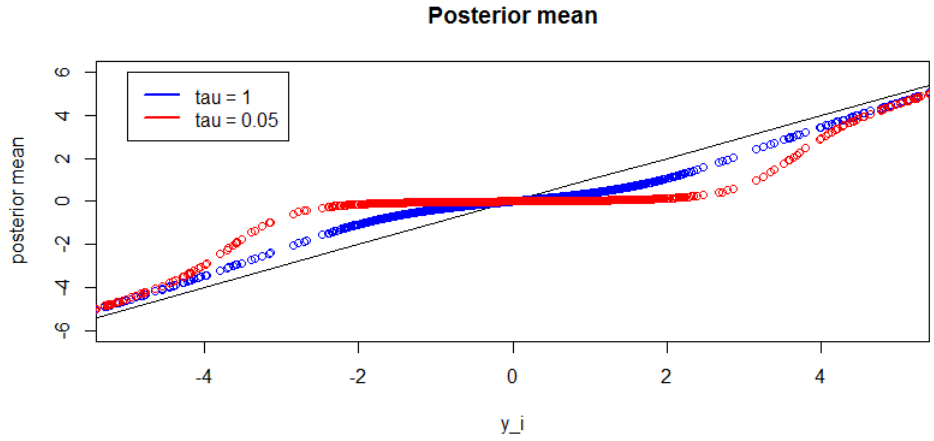
Deze uitdrukkingen voor $m_\tau(y)$ en $m'_\tau(y)$ kunnen we nu substitueren in de eerder genoemde formule voor de posterior mean:

$$
\mathrm{E}[\theta_i|y] = y + \frac{m'_\tau(y)}{m_\tau(y)}.
$$

We krijgen voor observatie i, met $1 \leq i \leq n$

$$
\begin{aligned}
\mathrm{E}[\theta_i|y_i] &= y_i + \frac{m'_\tau(y_i)}{m_\tau(y_i)} \\
&= y_i + \left(\frac{-\int_0^1 y_i \cdot (1-z) \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y_i^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz}{\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y_i^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz} \right) \\
&= y_i \cdot \left(1 - \frac{\int_0^1 (1-z) \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y_i^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz}{\int_0^1 \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y_i^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz} \right) \\
&= y_i \cdot \left(1 - \frac{\int_0^1 z^{-\frac{1}{2}} \cdot (1-z) \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2}z} dz}{\int_0^1 z^{-\frac{1}{2}} \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2}z} dz} \right).
\end{aligned}
$$

Voor twee verschillende waarden voor τ is in het volgende figuur de posterior mean weergegeven samen met de identiteit. We zien dat de elementen van de posterior mean zich dichter bij de identiteit bevinden naarmate τ groter wordt.



Figuur 12: De posterior mean voor twee verschillende waarden voor τ

De globale hyperparameter τ vertoont een bepaalde *shrinkage* gedrag. Zoals we eerder hebben vermeld, geldt $\theta_i|\lambda_i, \tau \sim \mathcal{N}(0, (\lambda_i\tau)^2)$. Voor kleine waarden van τ zit er bij een plot van θ_i meer concentratie rond nul. Dit betekent dan weer dat de plot van deze zelfde kleine waarden van τ, voor hogere waarden van $|y_i|$, dichter bij nul blijft zitten in vergelijking met hogere waarden voor τ. Dit kunnen we goed zien in Figuur 12. De plot voor $\tau = 0.05$ blijft voor meer waarden van $|y_i|$ dicht bij nul zitten. Hier is er dan meer concentratie rond nul.

We kunnen de plot in Figuur 12 opdelen in drie soorten gebieden: een gebied waar we duidelijk te maken hebben met signalen, een gebied waar we vrijwel alleen ruis hebben en een gebied waar er twijfel is over of we te maken hebben met een signaal. Voor $0 \leq |y_i| \leq 2$ hebben we vooral te maken met ruis en voor grofweg $|y_i| \geq 3$ hebben we te maken met signalen, al hangt de keuze af van de gekozen waarde voor τ. In het tussenliggende gebied is het onduidelijk of we te maken hebben met signalen. De posterior variance, die we introduceren in de volgende paragraaf, kan ons hierbij mogelijk helpen. De grootte van de drie genoemde gebieden is overigens wel afhankelijk van het aantal elementen in θ_0.

5.2 Posterior Variance

Naast de posterior mean kunnen we ook de posterior variance bepalen. Deze is vooral geschikt om de onzekerheid van de Bayesiaanse procedure die we in de vorige paragraaf hebben doorlopen te karakteriseren.. Het helpt ons om relatief kleine signalen uit θ_0 te detecteren.

Met behulp van [10] vinden we

$$Var[\theta_i|y, \tau] = 1 + \frac{\partial^2}{\partial y^2} \log m_\tau(y)\Big|_{y=y_i} \tag{13}$$

$$= 1 + \frac{\partial}{\partial y} \frac{m'_\tau(y)}{m_\tau(y)}\Big|_{y=y_i} \tag{14}$$

$$= 1 + \frac{m_\tau(y) \cdot m''_\tau(y) - [m'_\tau(y)]^2}{[m_\tau(y)]^2}\Big|_{y=y_i} \tag{15}$$

$$= 1 + \frac{m''_\tau(y)}{m_\tau(y)} - \left[\frac{m'_\tau(y)}{m_\tau(y)}\right]^2\Big|_{y=y_i} \tag{16}$$

De posterior variance is dus eigenlijk gewoon de afgeleide van de posterior mean.

In de vorige paragraaf hebben we $m_\tau(y)$ en $m'_\tau(y)$ bepaald. Dan kunnen we ook $m''_\tau(y)$ bepalen. Hiervoor gebruiken we de productregel.

$$m''_\tau(y) = -\int_0^1 (1 - y^2 \cdot (1-z)) \cdot (1-z) \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\pi} \cdot e^{-\frac{1}{2}y^2 \cdot (1-z)} \cdot \frac{z^{-\frac{1}{2}} \cdot \tau}{\tau^2 + (1-\tau^2)z}\, dz$$

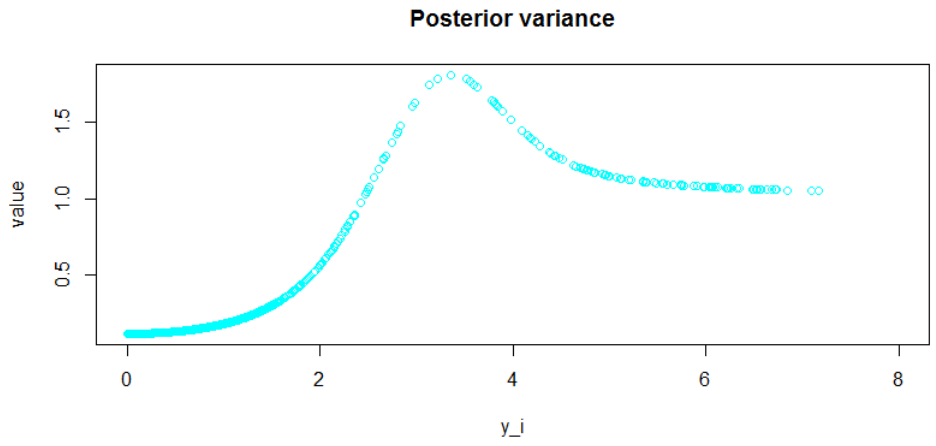
Substitueren we $m_\tau(y)$, $m'_\tau(y)$ en $m''_\tau(y)$ in (16) dan vinden we voor de posterior variance (zie [11]):

$$Var[\theta_i | y_i, \tau] = \frac{I_{\frac{1}{2}}(y_i)}{I_{-\frac{1}{2}}(y_i)} + y_i^2 \cdot \left[\frac{I_{\frac{3}{2}}(y_i)}{I_{-\frac{1}{2}}(y_i)} - \left(\frac{I_{\frac{1}{2}}(y_i)}{I_{-\frac{1}{2}}(y_i)}\right)^2\right], \tag{17}$$

waarbij

$$I_k(y_i) = \int_0^1 z^k \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{1}{2}y_i^2 \cdot z}\, dz\ .$$

In het volgende figuur zien we een plot van de posterior variance.



Figuur 13: De posterior variance

We zien hier een bepaalde curve, waarbij voor (relatief) hoge waarden voor de observaties de posterior variance de waarde 1 nadert. Dit gedeelte rechts van de

curve, correspondeert met het gebied waar de signalen makkelijk te detecteren zijn. In het gedeelte links van de curve hebben we te maken met ruis. In het tussenliggende gebied is het niet zo makkelijk te zeggen of we te maken hebben met signalen. Hier zijn de varianties namelijk het hoogst. Hoge varianties corresponderen met meer spreiding ten opzichte van de verwachtingswaarde in een plot van de normale verdeling. Dit betekent dat het lastiger is om te zeggen of we te maken hebben met een signaal.

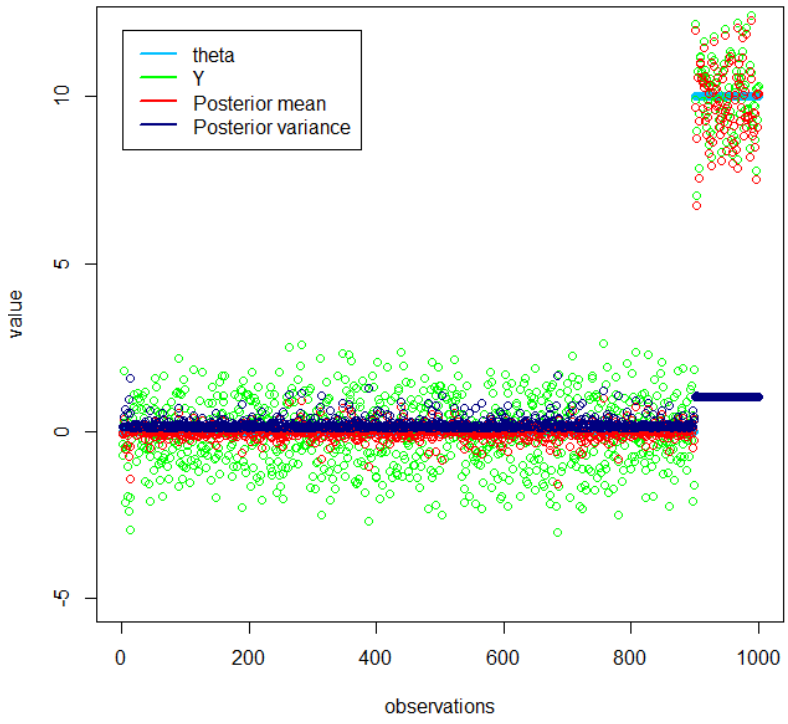
Overigens is de plot van de posterior variance symmetrisch ten opzichte van de y-as.

5.3 Visuele weergave

Wat komt er met behulp van de posterior mean en posterior variance uit voor de parameters als slechts de observaties uit de vector Y zijn gegeven? We hebben dit al enigzins kunnen zien in de vorige twee figuren, maar om hier een beter beeld van te krijgen voeren we simulaties uit in verschillende situaties. We nemen verschillende lengtes voor θ_0 en veranderen het aantal elementen ongelijk aan nul.

De parameters ongelijk aan nul in θ_0 krijgen een waarde A. Voor grote waarden van A, dat wil zeggen voor $A > 5$, zijn de signalen makkelijk te detecteren. Dit verandert naarmate we A verlagen.

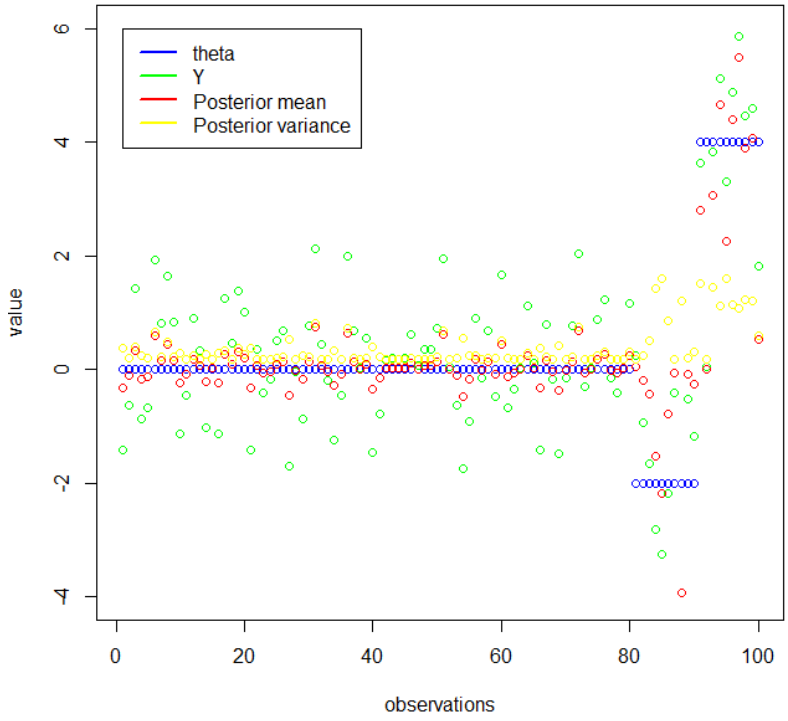
Allereerst kiezen we een vector θ_0 met lengte 1000. De laatste 100 elementen nemen we gelijk aan A.



Figuur 14: A = 10

In bovenstaande figuur is het voor $A = 10$ duidelijk dat de elementen ongelijk aan nul uit θ_0 na het gebruiken van de posterior mean ook ongelijk zijn aan nul. De rode cirkels in de plot blijven immers ver van de nul weg.
Omdat de observaties hier relatief groot zijn, zijn de donkerblauwe cirkels bijna allemaal gelijk aan 1 voor de elementen ongelijk aan nul. Hieruit kunnen we opmaken dat we te maken hebben met signalen.

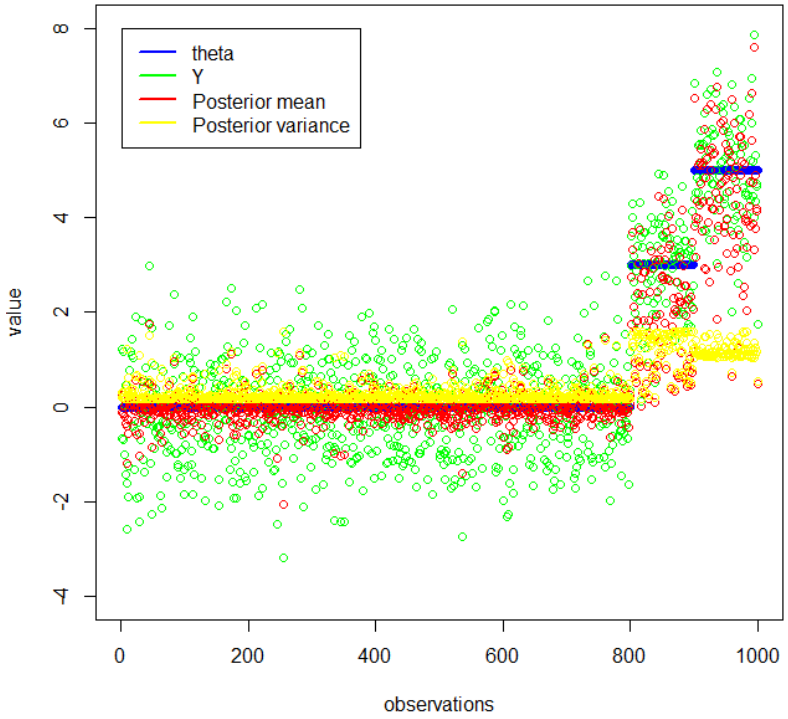
In de volgende twee figuren kiezen we voor twee verschillende niet-nul elementen: A en B. Bovendien is de lengte van θ_0 in het volgende figuur gelijk aan 100. De elementen 81 t/m 90 zijn gelijk aan B en de laatste tien elementen zijn gelijk aan A.



Figuur 15: A = 4, B = -2

We kiezen eerst voor $A = 4$ en $B = -2$. We zien dat dat voor $B = -2$ de meeste rode cirkels de neiging hebben om dicht bij nul te komen. Dit verandert bij de laatste 10 elementen. Daar zijn er minder rode cirkels die dicht bij nul zitten.
Als we kijken naar de resultaten die de posterior variance ons geeft dan zien we dat voor $B = -2$ ongeveer de helft van de gele cirkels zich rond de waarde 1 begeeft. We kunnen hier niet met zekerheid zeggen of we te maken hebben met signalen. Bij de laatste 10 elementen zitten bijna alle gele cirkels tussen de waarden 1 en 2. Hier is het waarschijnlijk dat we te maken hebben met signalen, aangezien ook bijna alle rode cirkels die horen bij de posterior mean relatief ver van de nul zitten.

Tenslotte kiezen we nu weer voor een vector θ_0 met lengte 1000. Nu krijgen de elementen 801 t/m 900 de waarde $B = 3$ en de laatste 100 elementen de waarde $A = 5$.



Figuur 16: A = 5, B = 3

Bovenstaand figuur laat vooral zien wat we verwachten. Hoe dichter de observaties bij nul zitten, des te vaker zitten de corresponderende posterior mean cirkels dichter bij nul.

Bij de posterior variance zit ruim de helft van de gele cirkels voor de observaties 801 t/m 900 rond de 1.5. Bij de laatste 100 observaties zijn slechts twee elementen kleiner dan 1. De overigen zitten voornamelijk tussen 1 en 1.2. Van deze laatste 100 observaties kunnen we zeggen dat het signalen zijn.

6 Meervoudig toetsen: deel 2

Zoals eerder gezegd testen we voor $i = 1, \ldots, n$, de volgende nulhypothese.

- $H_{0,i} : \theta_{0,i} = 0$
- $H_{1,i} : \theta_{0,i} \neq 0$

De methoden die we tot zover hebben gebruikt zijn:

- Simple multiple testing
- Bonferroni correctie
- Benjamini-Hochberg methode

Nu we de horseshoe posterior mean en posterior variance hebben besproken, kunnen we gaan kijken wat voor resultaten dit geeft bij het meervoudig toetsen. Voordat we hiermee beginnen bespreken we eerst de keuze voor een *threshold*.

6.1 Threshold Posterior mean

In de plots in het vorige hoofdstuk hebben we kunnen zien dat de posterior mean θ_i voor $i = 1, \ldots, n$ vaak schat op een waarde dicht bij nul als $|A| \leq 3$. Dit hangt echter wel af van de waarde van n, wat we in dezelfde plots hebben kunnen zien. Hoe kunnen we dan bepalen of we de betreffende θ_i wel of niet kunnen aanmerken als signaal? Om dit te doen kiezen we een drempelwaarde, een *threshold*.

Zoals we al eerder hebben gezien geldt voor de posterior mean

$$\mathrm{E}[\theta_i|y_i] = y_i \cdot \left(1 - \frac{\displaystyle\int_0^1 z^{-\frac{1}{2}} \cdot (1-z) \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z} dz}{\displaystyle\int_0^1 z^{-\frac{1}{2}} \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z} dz} \right). \tag{18}$$

We introduceren de volgende notatie

$$\hat{\theta}_{0,i} = \mathrm{E}[\theta_i|y_i].$$

Dan kunnen we (18) schrijven als

$$\hat{\theta}_{0,i} = y_i \cdot c(y_i)$$

met

$$c(y_i) = \left(1 - \frac{\displaystyle\int_0^1 z^{-\frac{1}{2}} \cdot (1-z) \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z} dz}{\displaystyle\int_0^1 z^{-\frac{1}{2}} \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z} dz} \right).$$

Merk op dat voor $0 \leq z \leq 1$

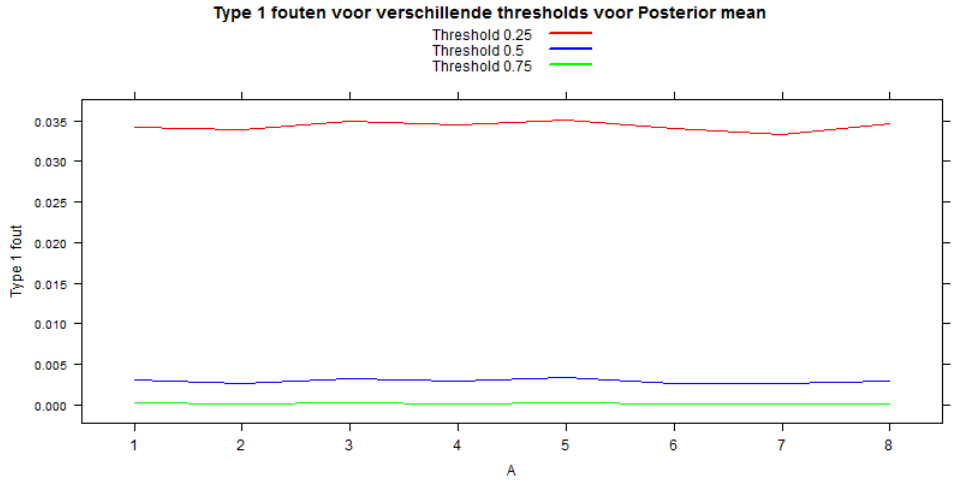
$$\begin{aligned} & z^{-\frac{1}{2}} \cdot (1-z) \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z} \\ \leq\ & z^{-\frac{1}{2}} \cdot \frac{1}{\tau^2 + (1-\tau^2)z} \cdot e^{\frac{y_i^2}{2} z}. \end{aligned}$$

Hieruit volgt $0 \leq c(y_i) \leq 1$.

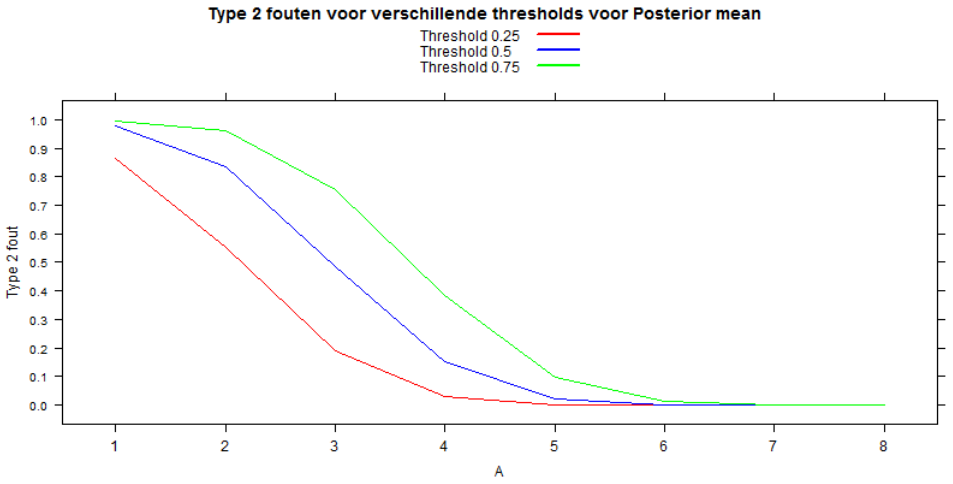
Als onze observatie y_i klein is, dan geldt $\hat{\theta}_{0,i} \approx 0$. Als y_i niet dicht bij nul zit, dan kan het zo zijn dat $\hat{\theta}_{0,i}$ wel of niet dicht bij nul zit. Dit hebben we kunnen zien in de plots in sectie 5.3. Als $\hat{\theta}_{0,i}$ dicht bij nul zit terwijl dat voor y_i niet geldt, dan is het mogelijk dat $\hat{\theta}_{0,i}$ zich onder de gekozen threshold bevindt, waardoor we mogelijk een signaal zullen missen. Om dit soort situaties te voorkomen, vergelijken we niet $\hat{\theta}_{0,i}$ met de threshold, maar $c(y_i)$.

Als $c(y_i)$ groter is dan de gekozen threshold, dan hebben we te maken met een (sparse) signaal. Als $c(y_i)$ kleiner is dan of gelijk aan de gekozen threshold, dan verwerpen we de nulhypothese niet.

In de volgende twee figuren staan de Type 1 en 2 fouten uitgezet voor drie verschillende thresholds, namelijk $\frac{1}{4}$, $\frac{1}{2}$ en $\frac{3}{4}$, voor $A = 1, 2, \ldots, 8$. Er zijn 100 simulaties uitgevoerd. θ_0 bevat opnieuw 1000 elementen, waarvan er 900 gelijk zijn aan nul.



Figuur 17: Proportie Type 1 fouten voor thresholds $\frac{1}{4}$, $\frac{1}{2}$ en $\frac{3}{4}$ voor de posterior mean



Figuur 18: Proportie Type 2 fouten voor thresholds $\frac{1}{4}$, $\frac{1}{2}$ en $\frac{3}{4}$ voor de posterior mean

We zien wat fluctuaties rond 0.04 bij de Type 1 fouten bij een threshold gelijk aan $\frac{1}{4}$. Voor hogere thresholds liggen de Type 1 fouten dicht bij nul.

Het interessante deel bevindt zich hier bij de Type 2 fouten. Hoe hoger de threshold, des te hoger de Type 2 fout voor de meeste waarden van A. Dit is niet verrassend. Een hoge threshold betekent dat het waarschijnlijker is dat $c(y_i)$ zich onder de threshold bevind. Het tegenovergestelde is het geval bij een lage threshold.

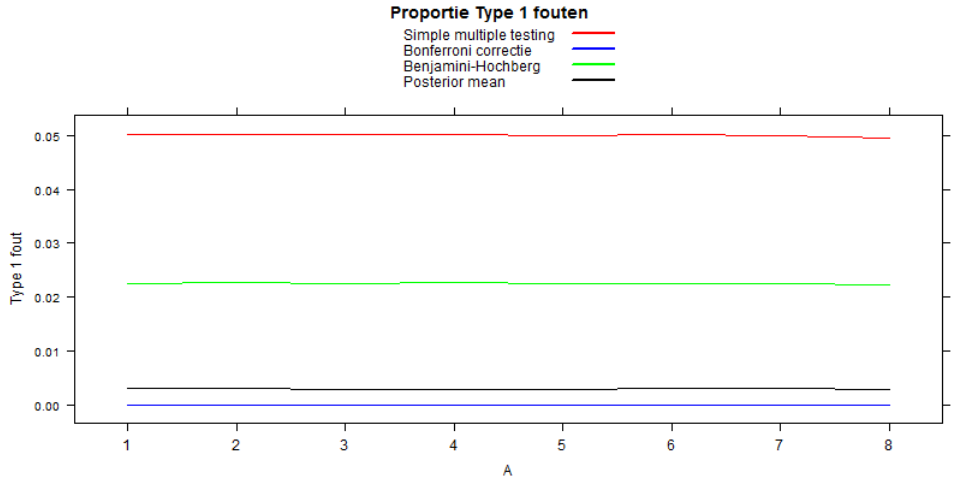
Om een goed beeld te krijgen van de parameters in de vector θ_0, kiezen we voor een threshold die niet te hoog en ook niet te laag is. Het is echter relatief wat we kunnen aanmerken als hoog of laag. In de praktijk hangt de keuze voor een threshold af van de toepassing van het probleem. Bij een te lage threshold zal het aantal elementen dat als signaal kan worden beschouwd te groot zijn, omdat de nullen worden meegerekend. Bij een te hoge threshold zullen we signalen missen, wat resulteert in een hogere Type 2 fout.

We zagen eerder al dat $0 \leq c(y_i) \leq 1$. Om te voldoen aan de eis dat de gekozen threshold niet te hoog of te laag is, kiezen we in onze verdere berekeningen voor een threshold gelijk aan een half.

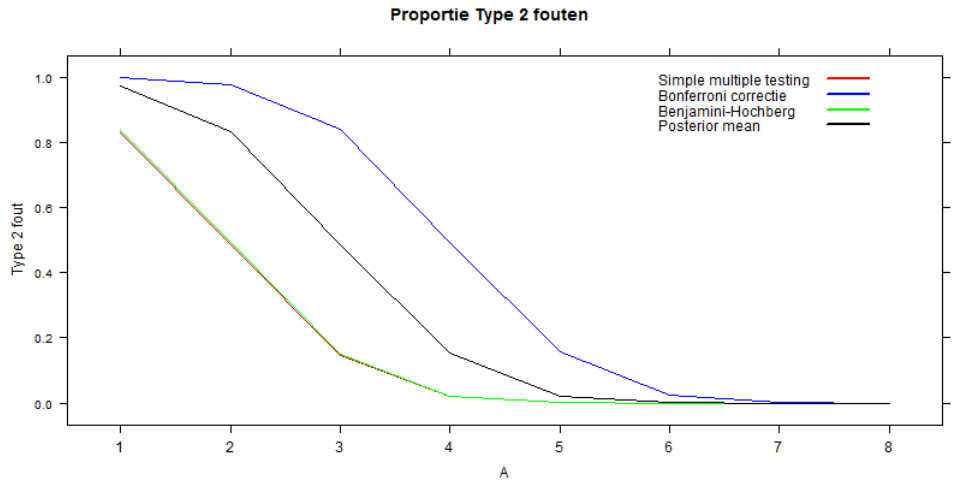
6.2 Horseshoe: Posterior mean

Nu we een geschikte threshold hebben gekozen kunnen we de resultaten van de Type 1 en 2 fouten bij de posterior mean vergelijken met de drie eerder gebruikte methoden.

We bepalen de proportie van Type 1 en Type 2 fouten voor $A = 1, 2, \ldots, 8$. Zoals gezegd nemen we voor de posterior mean een threshold gelijk aan $\frac{1}{2}$. Er worden 1000 simulaties uitgevoerd. θ_0 bevat opnieuw 1000 elementen, waarvan er 900 gelijk zijn aan nul.



Figuur 19: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$, met threshold $= \frac{1}{2}$ voor de posterior mean

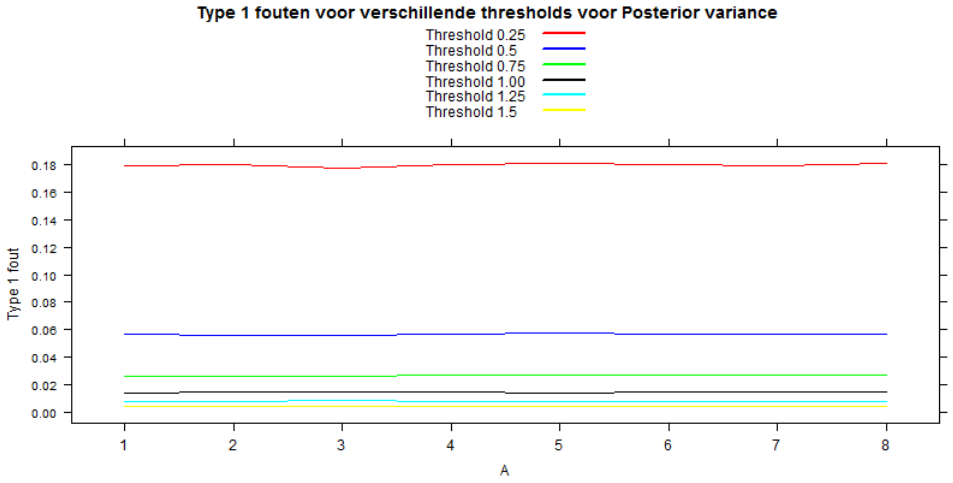


Figuur 20: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$, met threshold $= \frac{1}{2}$ voor de posterior mean

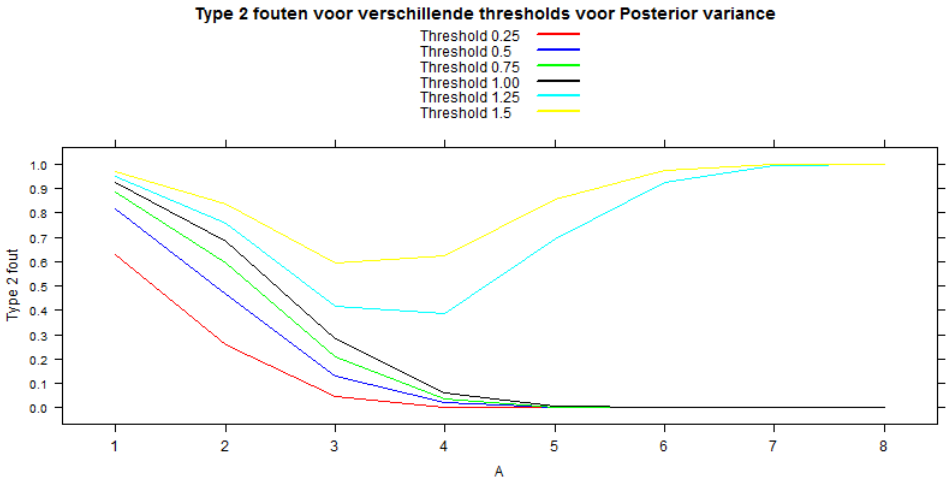
Bij de Type 1 fouten zien we dat de proporties vrijwel constant zijn voor verschillende waarden van A. Wat betreft de Type 2 fouten kunnen we zeggen dat de posterior mean niet beter is dan Simple multiple testing of de Benjamini-Hochberg methode. Dit is goed te zien voor $A \leq 4$.

6.3 Threshold Posterior variance

Naast het kiezen voor een geschikte threshold voor de posterior mean, moeten we ook een geschikte threshold bepalen voor de posterior variance. We bekijken in de onderstaande twee figuren het gedrag van de posterior variance met betrekking tot de Type 1 en 2 fouten voor zes verschillende thresholds.



Figuur 21: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$, met verschillende thresholds voor de posterior variance



Figuur 22: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$, met verschillende thresholds voor de posterior variance

Net zoals bij de posterior mean hangt bij de keuze voor een threshold voor de posterior variance in de praktijk af van de toepassing van het probleem. Daarom zullen we op een soortgelijke manier als bij de posterior mean kiezen voor een geschikte threshold voor de posterior variance.

Voor thresholds groter dan 1 zien we bij de Type 2 fouten een curve die we niet eerder hebben gezien. De fouten zijn een stuk hoger dan bij andere waarden voor de threshold.

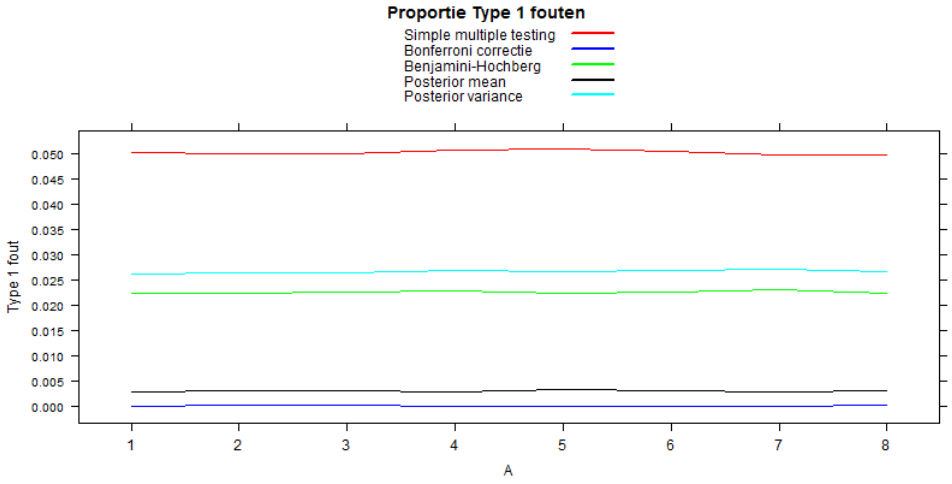
Bij een threshold van 0.25 zijn de Type 1 fouten het grootst. Bovendien is 0.25 wel wat aan de lage kant.

Bij ons probleem waar we de parameters ongelijk aan nul willen detecteren, lijkt een threshold gelijk aan 1 weer iets te hoog.

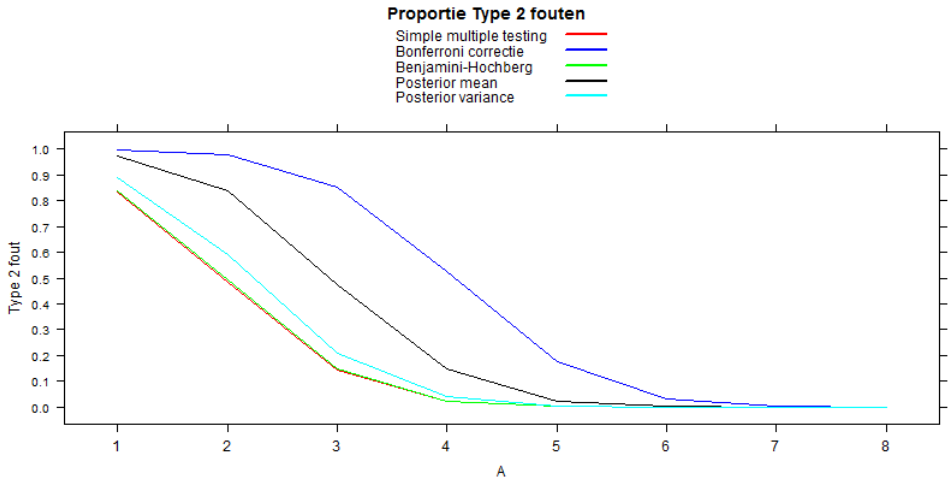
Dan blijven de waarden 0.5 en 0.75 over als keuze voor de threshold. Beide waarden zijn geschikt als threshold. We benadrukken nog maar eens dat de keuze in de praktijk af zal hangen van de toepassing van het probleem. In onze vervolgberekeingen kiezen wij voor een threshold gelijk aan 0.75.

6.4 Horseshoe: Posterior variance

Een threshold gelijk aan 0.75 geeft ons de volgende resultaten.



Figuur 23: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$, met threshold $= \frac{1}{2}$ voor de posterior mean en threshold $= \frac{3}{4}$ voor de posterior variance



Figuur 24: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$, met threshold $= \frac{1}{2}$ voor de posterior mean en threshold $= \frac{3}{4}$ voor de posterior variance

De posterior variance geeft ons samen met de Benjamini-Hochberg methode van alle vijf methoden de beste resultaten, waarbij de Benjamini-Hochberg methode het nog iets beter doet. Wat vooral opvalt is dat de posterior variance het bij de Type 2 fouten een stuk beter doet dan de posterior mean, ondanks dat de Type 1 fouten maar iets hoger zijn.

6.5 Combinatie van de posterior mean en posterior variance

De volgende stap is om de posterior mean en posterior variance te combineren. Hiervoor zijn verschillende mogelijkheden. We zouden bijvoorbeeld kunnen eisen dat een θ_i ongelijk is aan nul als zowel de posterior mean als de posterior variance zich boven de eigen threshold bevinden. Dit maakt de schatting echter conservatief. Hiermee bedoelen we dat we minder snel een θ_i zullen aanmerken als signaal. Door te eisen dat $\theta_i \neq 0$ als óf de posterior mean óf de posterior variance zich boven de eigen threshold bevindt, krijgen we een minder conservatieve methode.

Een *credible* interval is een Bayesiaanse versie van een betrouwbaarheidsinterval (zie [12]). Dit wordt gegeven door

$$\left[\mathbb{E}[\theta_i|y_i] \ - c_2 * \sqrt{Var[\theta_i|y_i]} \ , \ \mathbb{E}[\theta_i|y_i] \ + c_2 * \sqrt{Var[\theta_i|y_i]} \ \right].$$

Als dit interval groot is of als het centrum van dit interval ver van nul zit, dan hebben we te maken met een signaal. Een groot interval betekent namelijk een grote variantie, tenzij c_2 groot is. In de simulaties zullen we voor c_2 niet een te grote waarde kiezen. Een centrum ver van nul betekent dat $\mathbb{E}[\theta_i|y_i]$ ver van nu zit.

We kiezen voor de volgende combinatie van de posterior mean en posterio variance.

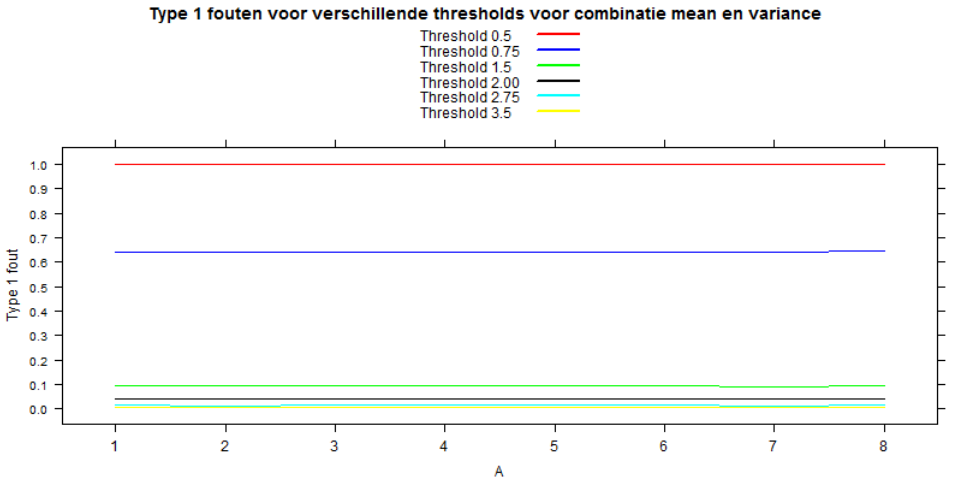
$$\left| \mathbb{E}[\theta_i|y_i] \right| + c_2 * \sqrt{Var[\theta_i|y_i]} \tag{19}$$

Dit correspondeert dan met de grootste afstand in het interval vanaf nul. Voor $c_2 > 0$ is (19) een bovengrens voor het interval. Als $\mathbb{E}[\theta_i|y_i] < 0$, dan krijgen we met (19) de absolute waarde van de ondergrens van het interval. In de simulaties kiezen we voor $c_2 > 0$.

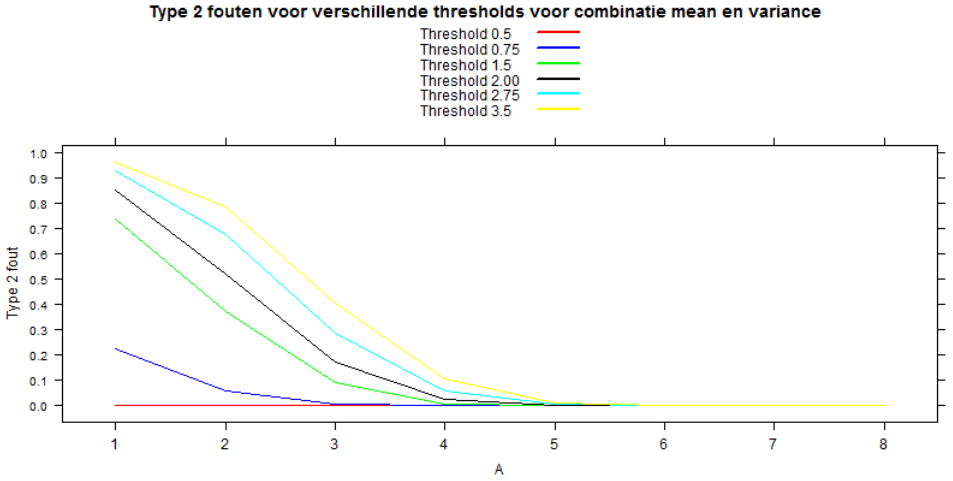
6.5.1 Threshold

De eerste stap is het kiezen van een waarde c_2. We kiezen voor $c_2 = 1.96$. Dit betekent dat 95% van de mogelijk waarden hoogstens 1.96 keer de standaardafwijking afwijkt van $\left| \mathbb{E}[\theta_i|y_i] \right|$.

De volgende stap is het kiezen van een geschikte threshold. We bekijken de resultaten voor de volgende zes thresholds: 0.5, 0.75, 1.50, 2.00, 2.75 en 3.5.



Figuur 25: $c_2 = 1.96$

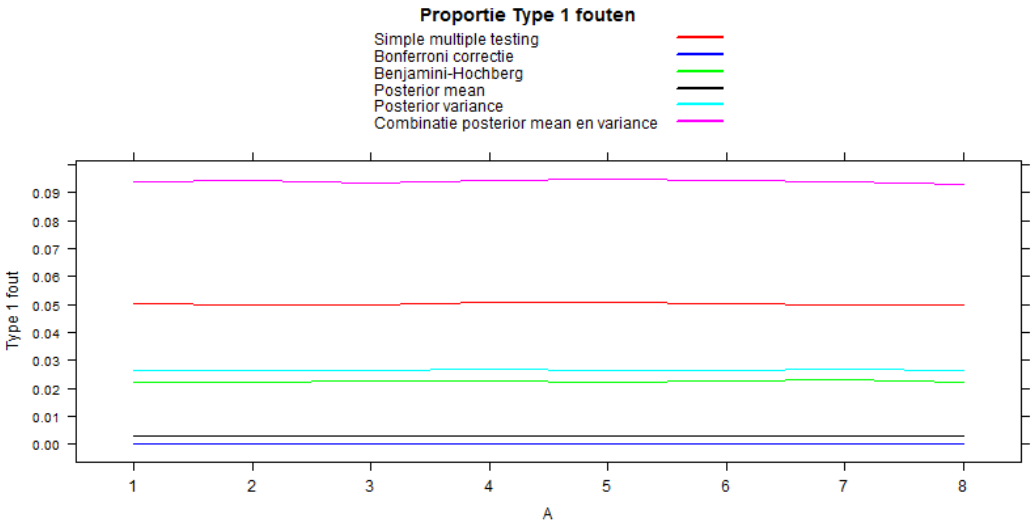


Figuur 26: $c_2 = 1.96$

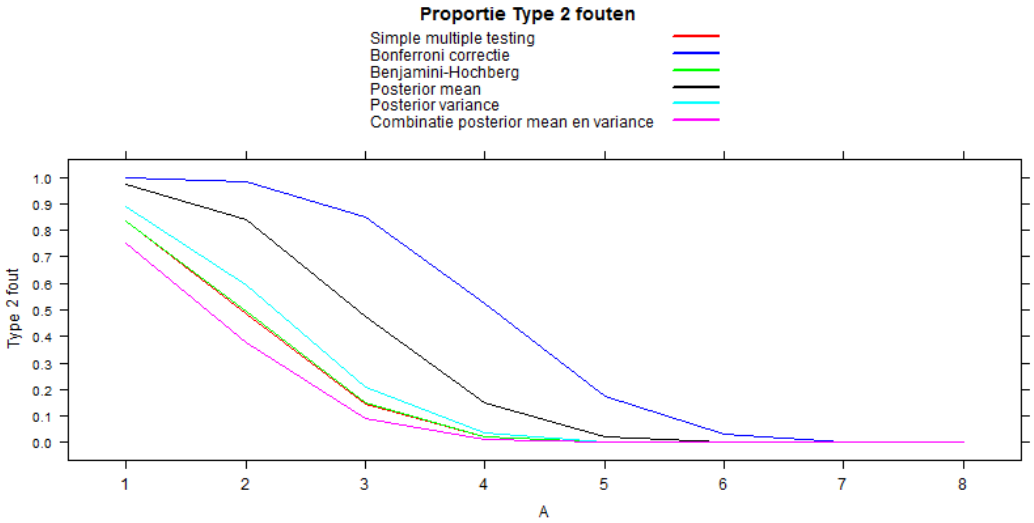
Hoe hoger de threshold, des te lager de Type 1 fouten. Een hogere threshold betekent immers dat het minder waarschijnlijk is dat de nulhypothese wordt verworpen. Aangezien de rode en blauwe lijnen bij de Type 1 fouten erg hoog liggen, zijn de corresponderende thresholds niet geschikt voor deze toetsingsprocedure. Op basis van de Type 2 fouten lijkt in dat geval een threshold van 1.5 de beste keuze, omdat voor hogere thresholds de fouten een stuk hoger liggen dan bij de methoden in Figuur 24.

6.5.2 Proportie Type 1 en 2 fouten

Een threshold van 1.5 geeft de volgende resultaten:



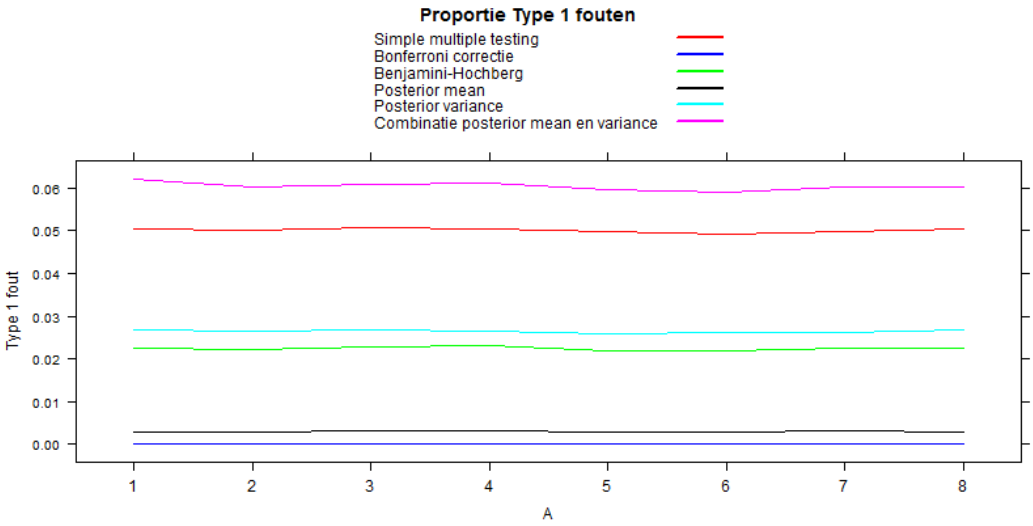
Figuur 27: $c_2 = 1.96$ en threshold 1.5 voor combinatie mean en variance



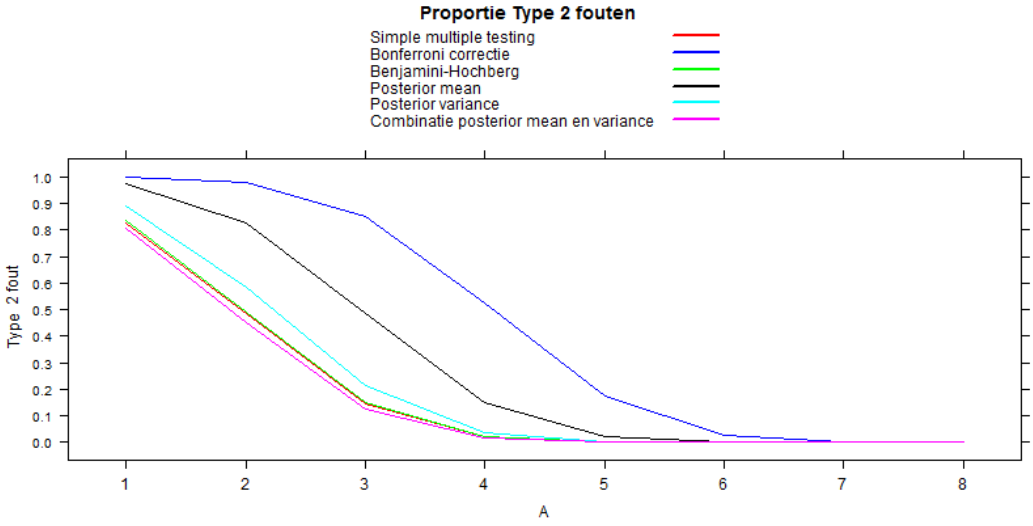
Figuur 28: $c_2 = 1.96$ en threshold 1.5 voor combinatie mean en variance

Vanwege de relatief grote Type 1 fouten bij een threshold van 1.5, zijn we niet helemaal tevreden. We proberen de threshold daarom een klein stuk te verhogen en hopen dat het verschil bij de Type 1 fouten met de overige methoden kleiner wordt. Tegelijkertijd is er voor de Type 2 fouten nog wat ruimte om deze te laten stijgen.

We bekijken wat voor resultaten we krijgen als de threshold gelijk is aan 1.75.



Figuur 29: $c_2 = 1.96$ en threshold 1.75 voor combinatie mean en variance



Figuur 30: $c_2 = 1.96$ en threshold 1.75 voor combinatie mean en variance

De Type 1 fouten zijn inderdaad (licht) afgenomen. Bovendien is de combinatie van de posterior mean en posterior variance nog steeds de beste methode als het gaat om de Type 2 fouten.

Op basis van deze twee laatste plots zullen wij normaliter bij deze methode een threshold gelijk aan 1.75 gebruiken.

6.6 Variaties in θ_0

Tot zover hebben we tijdens de simulaties bij θ_0 gekozen voor 100 elementen ongelijk aan nul, waarbij deze dezelfde waarden hadden. Daarnaast kozen we iedere keer voor een significantieniveau van $\alpha = 0.05$. We willen hier wat variatie in aanbrengen, zodat we kunnen zien of we grote afwijkingen krijgen in onze resultaten.

In dit hoofdstuk bekijken we de volgende variaties:

- We kiezen voor een ander significantieniveau.
- Voor de niet-nul elementen in θ_0 kiezen we naast de waarde A ook voor een waarde B.
- De elementen in θ_0 worden getrokken uit kansverdelingen (Normale verdeling, exponentiële verdeling en Gamma-verdeling).

Tenzij anders vermeld, gebruiken we net zoals in de vorige secties voor de posterior mean een threshold van 0.5, voor de posterior variance een threshold van 0.75 en voor de combinatie van de posterior mean en posterior variance een threshold van 1.75 met $c_2 = 1.96$.

- **$\alpha = 0.10$**

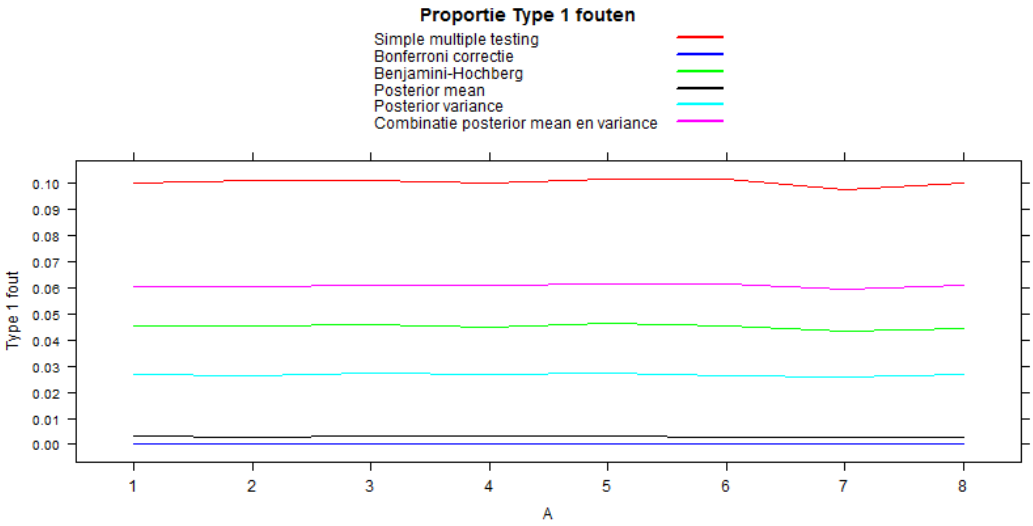
  We doen exact hetzelfde als in de vorige paragraaf, maar nu met een ander significantieniveau. Voor simple multiple testing is de kritieke waarde gelijk aan

  ```
  > qnorm(1-0.10/2)
  [1] 1.644854
  ```

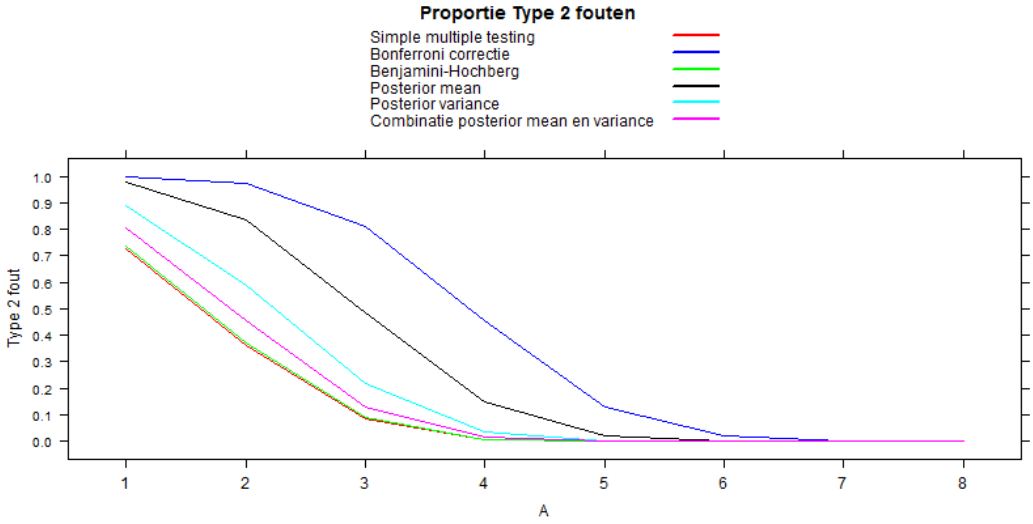
  en voor de Bonferroni correctie krijgen we

  ```
  > qnorm(1-(0.10/1000)/2)
  [1] 3.890592.
  ```

  De resultaten zijn als volgt.



Figuur 31: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$, met significantieniveau $\alpha = 0.10$



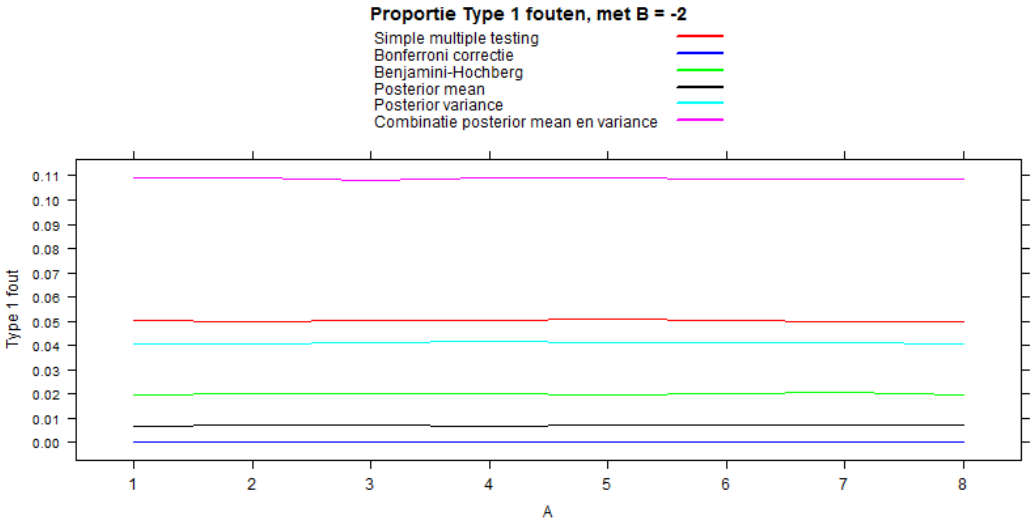
Figuur 32: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$, met significantieniveau $\alpha = 0.10$

Omdat we hebben gekozen voor $\alpha = 0.10$, is de kritieke waarde bij simple multiple testing lager, waardoor er meer nulhypothesen worden verworpen. Omdat α per definitie de kans is op het maken van een Type 1 fout, zijn de Type 1 fouten begrensd begrensd door 0.10. Als er meer nulhypothesen worden verworpen, zullen er minder Type 2 fouten worden gemaakt zoals we kunnen zien. Bij de Bonferroni correctie ligt de kritieke waarde slechts 0.17 lager, waardoor er nauwelijks veranderingen zijn.

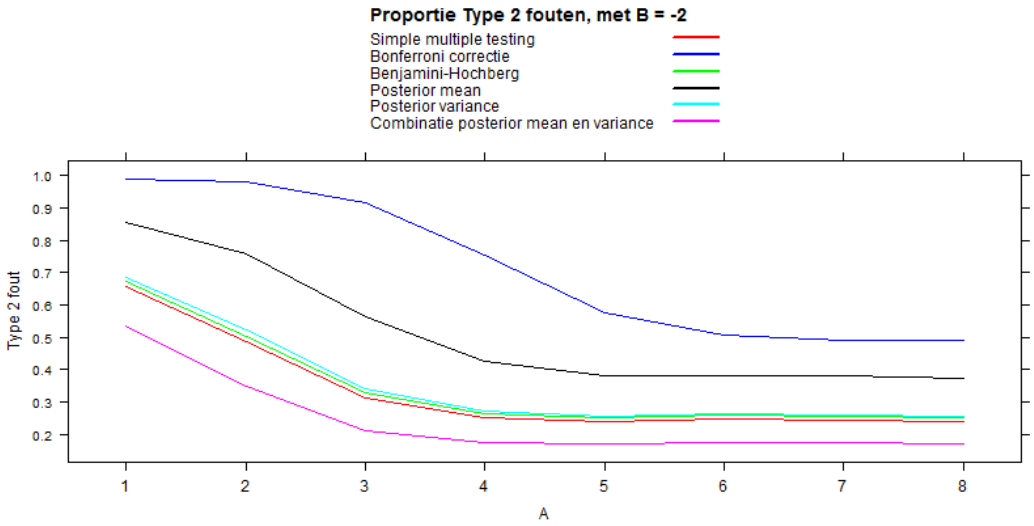
De grootste verandering die we hier zien ten opzichte van het geheel zijn de resultaten bij de Benjamini-Hochberg methode. De Type 1 fouten zijn groter geworden, terwijl het tegenovergestelde het geval is bij de Type 2 fouten. Als we nog eens de Benjamini-Hochberg methode in sectie 3.4 doornemen, dan betekent een hoger significantieniveau dat de drempelwaarden zullen toenemen. Dit heeft als gevolg dat er meer p-waarden kleiner zijn dan hun corresponderende drempelwaarden, waardoor er meer nulhypothesen worden verworpen. En ook hier geldt dat als er meer nulhypothesen worden verworpen, dan zullen er minder Type 2 fouten worden gemaakt. De Benjamini-Hochberg methode doet het bij beide type fouten beter dan bijvoorbeeld de combinatie van de posterior mean en posterior variance en lijkt in dit geval de beste optie te zijn.

- $\boldsymbol{B = -2}$

  We kiezen nu voor 200 elementen ongelijk aan nul in θ_0, waarvan er 100 de waarde $B = -2$ toegekend krijgen.



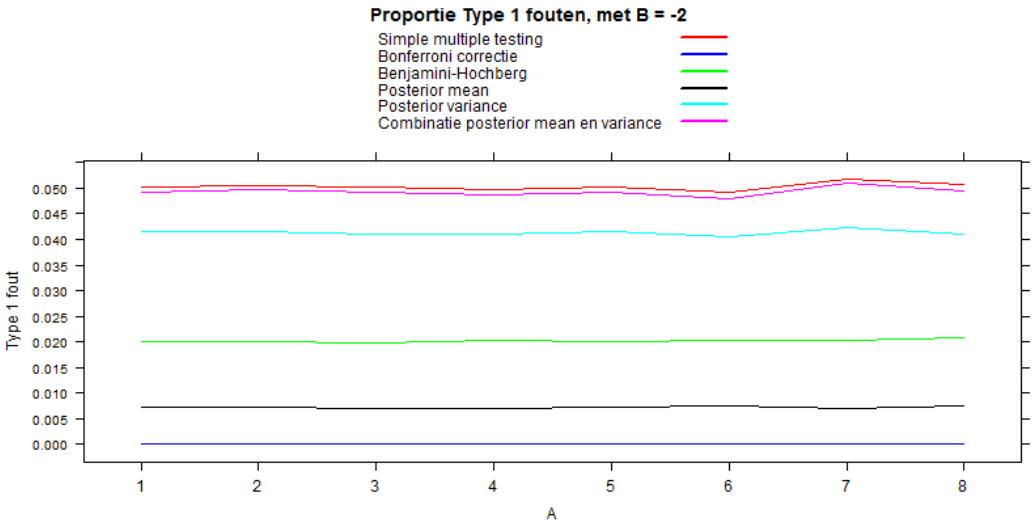
Figuur 33: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$ en $B = -2$



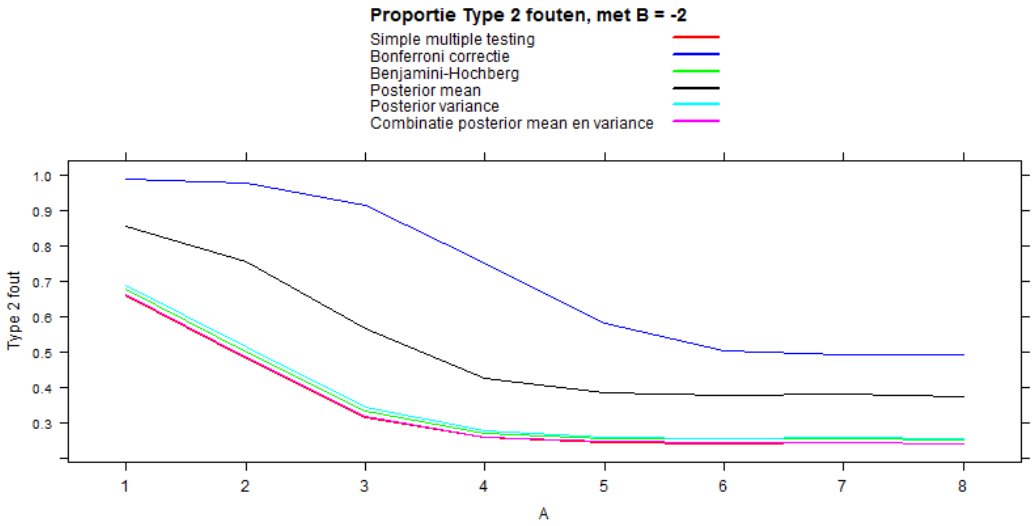
Figuur 34: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$ en $B = -2$

Bij de Type 1 fouten doen Benjamini-Hochberg, posterior mean en Bonferroni het erg goed. Van deze drie doet alleen Benjamini-Hochberg het ook goed bij de Type 2 fouten. Deze methode is net iets beter dan de posterior variance, al is het verschil erg klein.

De combinatie methode doet het daarentegen vooral bij de Type 1 fouten niet goed. Dit heeft dan wel weer als gevolg dat er minder Type 2 fouten gemaakt worden. Om dit verschil te verkleinen, proberen we in plaats van een threshold van 1.75 een threshold van 2.25 voor de combinatie van de posterior mean en posterior variance.



Figuur 35: Proportie Type 1 fouten voor $A = 1, 2, \ldots, 8$ en $B = -2$, met threshold 2.25 voor combinatie mean en variance



Figuur 36: Proportie Type 2 fouten voor $A = 1, 2, \ldots, 8$ en $B = -2$, met threshold 2.25 voor combinatie mean en variance

Een hogere threshold betekent dat een nulhypothese minder snel wordt verworpen. De combinatie is bij de Type 1 fouten net iets beter dan simple multiple testing.

Als de Type 1 fouten kleiner worden, dan worden de Type 2 fouten groter. Bij de Type 2 fouten is het niet duidelijk bij welke methode de proporties het laagst zijn. Om hier achter te komen, bekijken we de vectoren van proporties van Type 2 fouten in R.

Vec4 hoort bij simple multiple testing en vec12 bij de combinatie van de posterior mean en posterior variance.

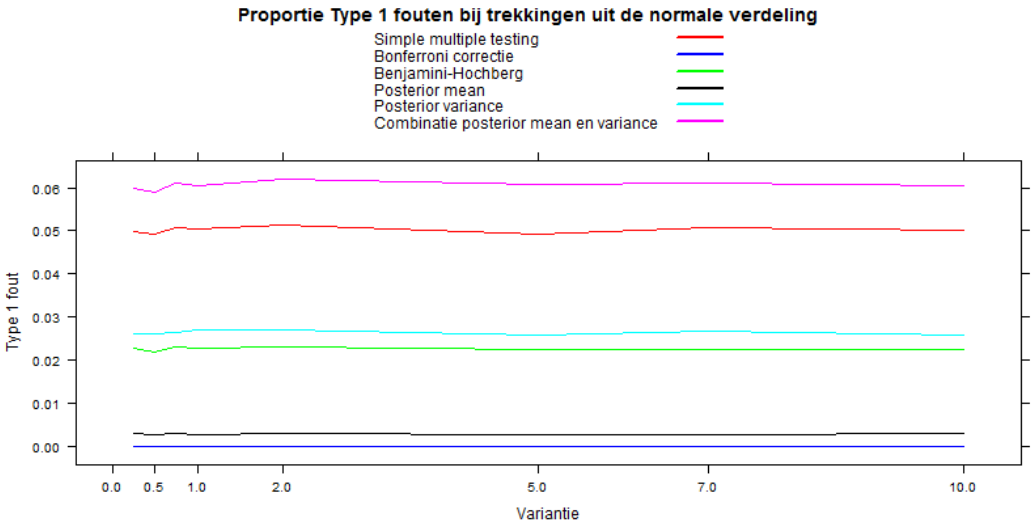
```
> vec4
[1] 0.65940 0.48260 0.31365 0.25690 0.24285 0.24025 0.24260 0.23840
> vec12
[1] 0.66195 0.48550 0.31700 0.25895 0.24520 0.24155 0.24460 0.24020
```

Het verschil wordt in de meeste gevallen pas gemaakt in de derde decimaal, waarbij simple multiple testing het net iets beter doet. Echter zitten de Benjamini-Hocherg methode en de posterior variance er weer net iets boven bij de Type 2 fouten.

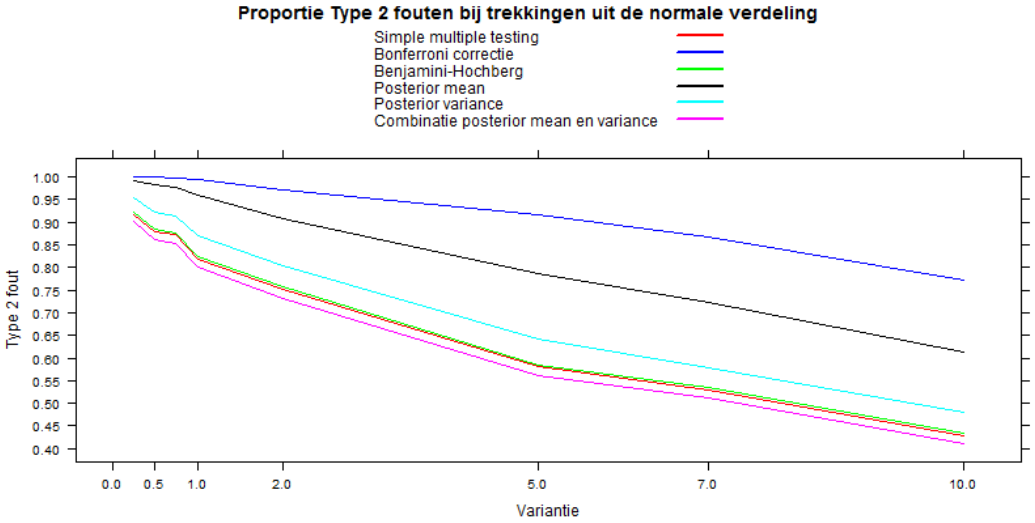
- **Trekkingen uit de normale verdeling.**

  We nemen een vector θ_0 ter lengte 1000, met 100 elementen ongelijk aan nul. Deze elementen bepalen we door acht verschillende trekkingen uit een normale verdeling met verwachting 0 en variantie gelijk aan elementen uit de vector $vec.A$ zoals hieronder weergegeven.

```
> vec.A
[1]  0.25  0.50  0.75  1.00  2.00  5.00  7.00 10.00
```



Figuur 37: Proportie Type 1 fouten voor 8 verschillende trekkingen uit de normale verdeling met verwachting 0 en variantie gelijk aan de elementen uit de vector vec.A



Figuur 38: Proportie Type 2 fouten voor 8 verschillende trekkingen uit de normale verdeling met verwachting 0 en variantie gelijk aan de elementen uit de vector vec.A

Bij de Type 1 fouten zien we vrijwel een gelijke trend als bij de vorige voorbeelden. De fouten zijn ongeveer net zo groot. De combinatie van de posterior mean en posterior variance doet het hier echter het minst goed. Dit zou kunnen veranderen als we de threshold opnieuw verhogen.

Bij de Type 2 fouten valt op dat de fouten een stuk hoger zijn. Dit gebeurt

als de oberservaties of schattingen zich onder een kritieke waarde of threshold bevinden. Kleine varianties hebben als gevolg dat we voornamelijk kleine waarden voor de trekkingen krijgen. Maar ook bij de twee grootst gekozen varianties krijgen we enkele trekkingen die net zo klein zijn. We zien hieronder 10 steekproeven voor varianties gelijk aan 0.75 en 7. De waarden kunnen bij dit aantal iedere keer sterk verschillen, maar wanneer we 100 trekkingen doen, zullen er waarschijnlijk ook wel voor een variantie gelijk aan 7 kleine waarden bij zitten zoals hieronder is weergegeven.

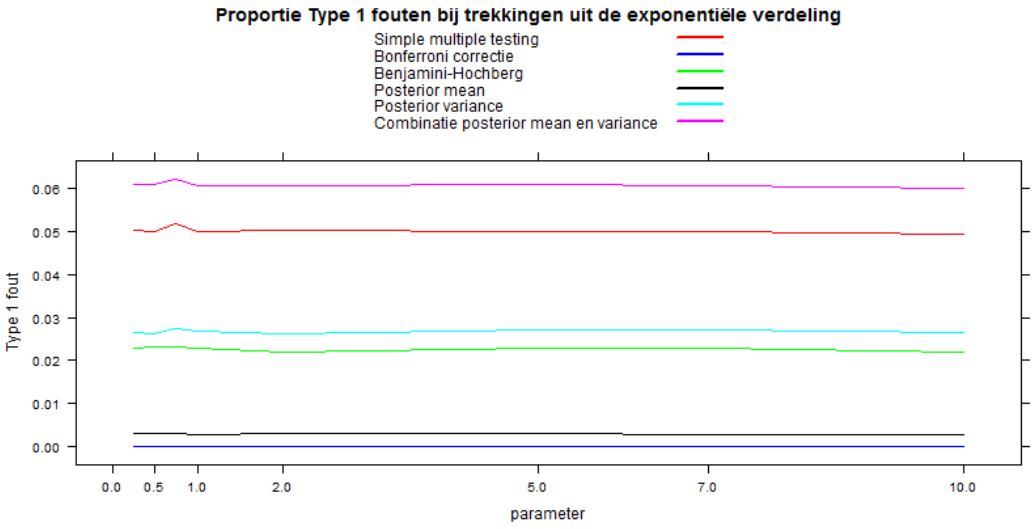
```
> rnorm(10,0,sqrt(0.75))
 [1] -0.09878815 -0.34986024 -0.58317221  0.20611175 -0.36963143
 [6]  0.13064928  1.68979456  0.45764795 -0.40181589 -1.47548418
> rnorm(10,0,sqrt(7))
 [1]  0.270713643 -2.033946700  1.973412164 -0.139509234  0.009256296
 [6] -3.603842921  2.924314491 -4.364482789  3.137518359  0.735810683
```

Dit verklaart de hoge proporties van Type 2 fouten bij hogere varianties. We zien nog wel dezelfde rangschikking van methoden op volgorde van fouten: de Bonferroni correctie werkt het minst goed, terwijl de combinatie van de posterior mean en posterior variance de minst slechte methode is. Het mag duidelijk zijn dat de combinatie beter werkt dan de Bonferroni correctie.

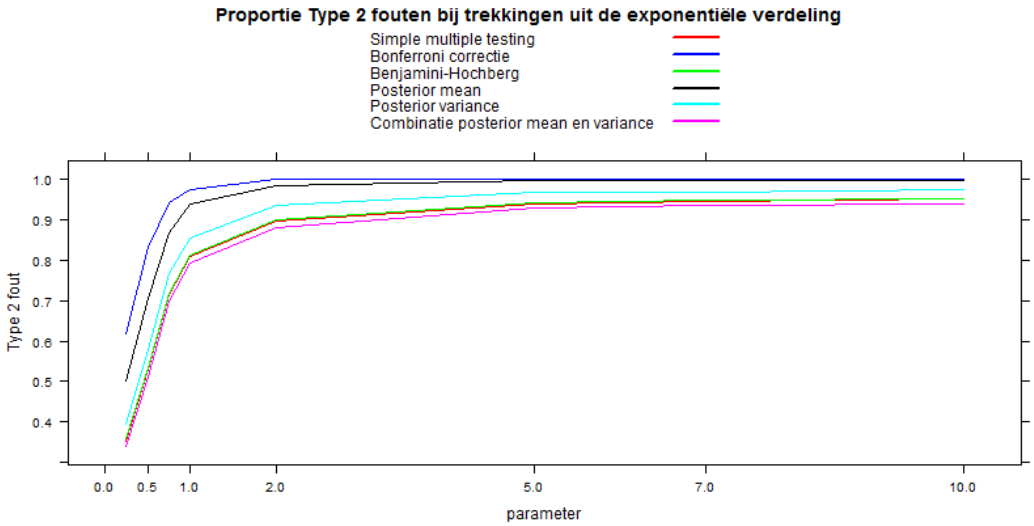
- **Trekkingen uit de exponentiële verdeling.**

  Nu doen we trekkingen uit de exponentiële verdeling. De random gegenereerde waarden komen uit een exp(λ)-verdeling. Er geldt dat $\lambda \in vec.A$, waarbij de vector $vec.A$ hetzelfde is als in het vorige voorbeeld met trekkingen uit de normale verdeling.

  θ_0 bevat 1000 elementen, waarvan 100 ongelijk zijn aan nul.



Figuur 39: Proportie Type 1 fouten voor trekkingen uit de exponentiële verdeling met acht verschillende parameters



Figuur 40: Proportie Type 2 fouten voor trekkingen uit de exponentiële verdeling met acht verschillende parameters

Hoe hoger de parameters van de exponentiële verdeling zijn, des te lager zijn de waarden van de trekkingen. We bekijken wat voorbeelden uit R.

```
> mean(rexp(100, rate = 0.25))
[1] 3.190732
> mean(rexp(100, rate = 0.5))
[1] 1.744672
> mean(rexp(100, rate = 0.75))
[1] 1.397679
```

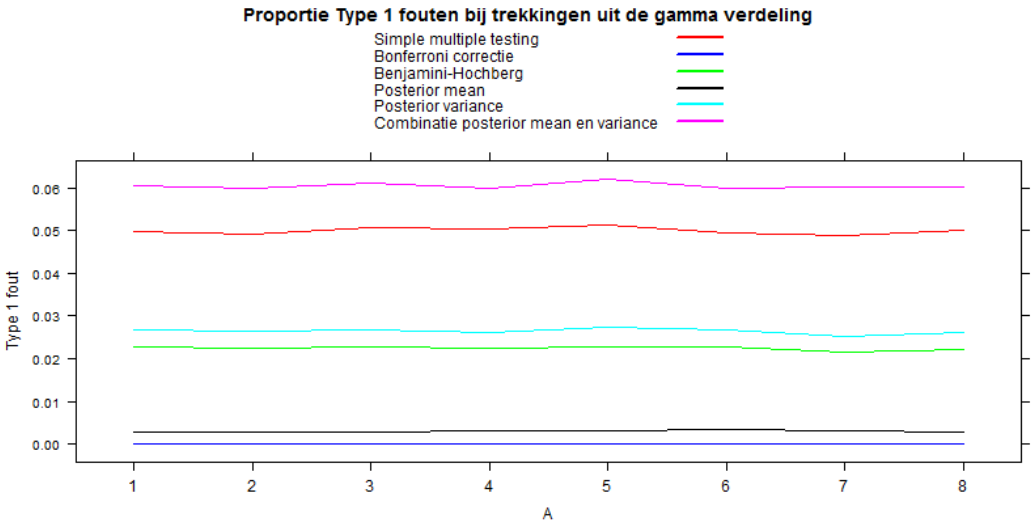
Dit geeft ons niet de werkelijke gemiddelden, aangezien er slechts 100 trekkingen zijn gedaan, maar wel een indicatie van de waarden. Het wordt hierdoor ook duidelijk waarom de proportie van Type 1 fouten relatief laag zijn. De random gegenereerde waarden uit de exponentiële verdeling bevinden zich voornamelijk onder de kritieke waarden of thresholds. De Type 1 fouten zijn vrijwel hetzelfde als bij de trekkingen uit de normale verdeling.

Omdat de waarden van de trekkingen lager zijn voor hogere λ, verklaart dit ook de curves bij de Type 2 fouten. Voor de parameters groter dan 0.75 zijn de proporties van fouten al groter dan of dicht bij 0.8. Dit komt doordat de waarden van de trekkingen zich onder de kritieke waarde of threshold bevinden.

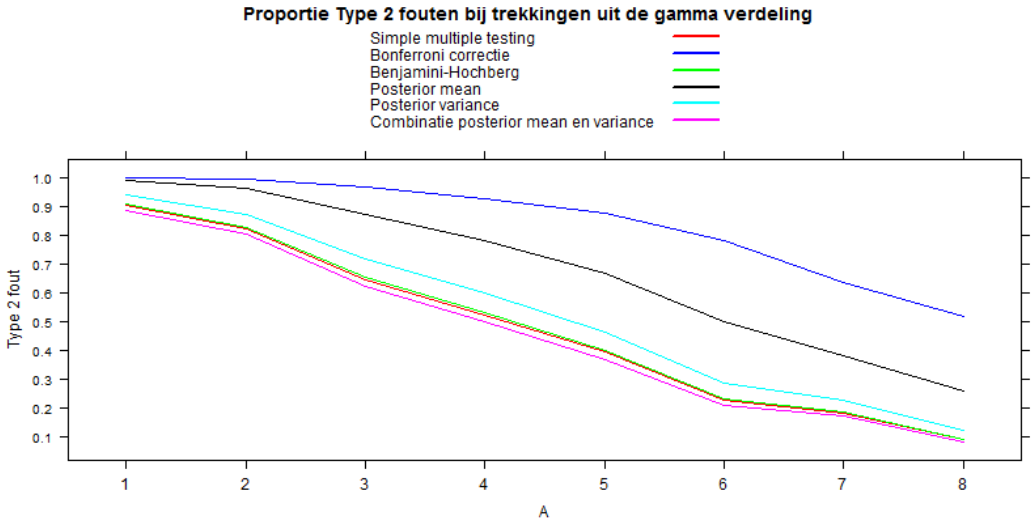
- **Trekkingen uit een gamma verdeling.**

  We doen trekkingen uit een gamma(A,2)-verdeling voor $A = 1, 2, \ldots, 8$. De schaalparameter is dus 2.

  Ook hier bevat θ_0 1000 elementen, waarvan 100 ongelijk zijn aan nul.



Figuur 41: Proportie Type 1 fouten voor trekkingen uit de gamma(A,2)-verdeling voor $A = 1, 2, \ldots, 8$



Figuur 42: Proportie Type 2 fouten voor trekkingen uit de gamma(A,2)-verdeling voor $A = 1, 2, \ldots, 8$

We zien hier hetzelfde gedrag als bij bijna alle andere voorbeelden. Wat noemenswaardig is, is het verschil tussen de combinatie van de mean en variance en de overige methoden. Zoals we in sectie 6.5.2 en in het voorbeeld in dit hoofstuk met $B = -2$ hebben kunnen zien, kan dit verschil worden verkleind door een hogere threshold te kiezen. In de praktijk hangt het maar net van de toepassing af welke threshold de voorkeur geniet.

7 Conclusie en vervolgonderzoek

Het doel bij deze scriptie is om te onderzoeken of de horseshoe prior een geschikte toetsingsprocedure is voor het meervoudig toetsen. 'Geschikt' wilt zeggen dat we een goede *trade-off* hebben tussen Type 1 en Type 2 fouten. Voordat we de horseshoe prior introduceerden, leek de Benjamini-Hochberg methode de beste keuze te zijn. Deze deed het namelijk bij de Type 1 fouten beter dan simple multiple testing. De Bonferroni correctie daarentegen richt zich op het verkleinen van de Type 1 fouten, maar houdt geen rekening met de Type 2 fouten en is derhalve geen geschikte methode voor een goede trade-off.

Bij de horsehoe prior begonnen we met het gebruiken van de posterior mean. Het gedrag hiervan hangt af van de threshold. Voor de threshold die wij hebben gekozen doet de posterior mean het goed bij de Type 1 fouten, omdat deze klein zijn. Echter gaat dit wel ten koste van de Type 2 fouten. In dat opzicht lijkt de posterior mean, met de threshold die we hebben gekozen, op een verbeterde versie van de Bonferroni correctie, aangezien het verschil bij de Type 1 fouten minimaal is. De posterior variance is vooral geschikt voor het karakteriseren van de onzekerheid van de Bayesiaanse procedure, maar heeft veel beperkingen. Het werd pas echt interessant toen we de posterior mean en posterior variance hebben gecombineerd. Het levert ons namelijk ook een goede balans op.

Welke methode de voorkeur geniet hangt af van de toepassing van het probleem. Willen we naast een goede trade-off ook de Type 1 fouten zo laag mogelijk houden, dan kiezen we voor de Benjamini-Hochberg methode. Willen we de Type 2 fouten laag houden en tegelijkertijd een goede trade-off, dan is de combinatie van de posterior mean en posterior variance de beste optie. Op basis van de simulaties zijn dit ook de twee beste methoden. Door te variëren met de treshold kunnen we de nadruk bovendien leggen op het verkleinen van een type fout. Dus de horseshoe prior is, op basis van onze simulaties, een geschikte toetsingsprocedure als we de posterior mean en posterior variance combineren zoals we dat hebben gedaan.

Ondanks dat we een eenvoudig sparse probleem $(*)$ hebben gebruikt voor het onderzoek, geeft ons dit wel inzichten in complexe toepassingen en statistische modellen. Dit kan in de praktijk van pas komen. Het onderzoek in deze scriptie is echter beperkt en de conclusies zijn gebasseerd op een klein aantal simulaties. Het bestuderen van een grotere verscheidenheid aan simulaties zou leiden tot genuanceerdere conclusies.

Voor een vervolgonderzoek zijn er verschillende mogelijkheden. Bij de horseshoe prior kozen we voor $\tau = \frac{\#\{\theta_i \neq 0\}}{n}$. Omdat we $\#\{\theta_i \neq 0\}$ echter van tevoren niet weten, kunnen we ervoor kiezen om τ afhankelijk van de data te kiezen. Dit kunnen we doen door een empirische Bayes benadering zie ([9], 2008). Hierbij wordt er een prior verdeling voor τ geschat op basis van de data. Vervolgens wordt er een posterior verdeling afgeleid.

Ook voor de threshold bij de horseshoe prior zijn er mogelijkheden. In eerste instantie hebben we bij de posterior mean, poster variance en de combinatie hiervan zelf een threshold gekozen. Vervolgens hebben we deze thresholds gebruikt in de simulaties in het vervolg. Er zou onderzocht kunnen worden of er bijvoorbeeld een betere manier is om de threshold te bepalen bij de posterior mean. Ook zou er gekeken kunnen worden naar een threshold die optimaal

is. We zouden bijvoorbeeld de threshold van de posterior mean kunnen laten afhangen van τ, nadat deze data afhankelijk is bepaald, maar ook van het significantieniveau. Op die manier hebben we de informatie van de data, maar ook een vooraf aantal ingecalculeerde Type 1 fouten die we verwachten te maken in de threshold verwerkt.

8 Referenties

[1] Iain M. Johnstone and Bernard W. Silverman, Needles And Straw in Haystacks: Emprirical Bayes Estimates Of Possibly Sparse Sequences, The Annals of Statistics, Vol. 32, No. 4 (2004), pp. 1594 - 1649

[2] John Thomas White and Subhashis Ghosal, Multiple testing approaches for removing background noise from images, in "Topics in NonParametric Statistics"(M. Akritas, S. N. Lahiri, and D. Politis, eds.), SpringerProceedings in Mathematics and Statstics, Vol 74, 2014, pp. 95-104

[3] Clements, N., Sarkar, S. and Guo, W., Astronomical transient detection controlling the false discovery rate, in Statistical Challenges in Modern Astronomy, edited by Eric D. Feigelson and G. Joseph Babu, Lecture Notes in Statistics, Vol. 209, Part 4, 2012, Springer-Verlag, pp. 383-396.

[4] John A. Rice: Mathematical Statistics and Data Analysis, third edition, Brooks/Cole Cengage Learning, 2007, International edition

[5] F. Bijma, M.A. Jonker, A.W. van der Vaart, Inleiding in de statistiek, *Epsilon Uitgaven*, 2013

[6] Yoav Benjamini and Yosef Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, No. 1 (1995), pp. 289-300

[7] R. Nowak (2010), Lecture notes 'Multiple Hypothesis Testing', beschikbaar via `http://nowak.ece.wisc.edu/ece830/ece830_lecture12.pdf` (geraadpleegd juni 2016)

[8] Larry Wasserman, Lecture notes 'A Few Random Topics', beschikbaar via `http://www.stat.cmu.edu/~larry/=stat705/Lecture17.pdf` (geraadpleegd juni 2016)

[9] Carlos M. Carvalho, Nicholas G. Polson, James G. Scott, The horseshoe estimator for sparse signals, Biometrika, Vol. 97, No. 2 (2010), pp. 465-480

[10] L. R. Pericchi and A. F. M. Smith, Exact and Approximate Posterior Moments for a Normal Location Parameter, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 54, No. 3 (1992), pp. 793-804

[11] S.L. van der Pas, B.J.K. Kleijn, A.W. van der Vaart, The Horseshoe Estimator: Posterior Concentration around Nearly Black Vectors, Electron. J. Statist. Volume 8, Number 2 (2014), pp. 2585-2618

[12] Stéphanie van der Pas, Botond Szabó, Aad van der Vaart, How many needles in the haystack? Adaptive inference and uncertainty quantification for the horseshoe, 2016, preprint beschikbaar via `http://arxiv.org/abs/1607.01892`

Appendix: R-code

Propoties Type 1 en 2 fouten

Met de volgende R-code kunnen de Propoties van Type 1 en 2 fouten uit de hoofdstukken 3 en 6 (eerste vijf paragrafen) worden verkregen.

Door het aanpassen van de kritieke waarden bij simple multiple testing en de Bonferroni correctie en het significantieniveau bij Benjamini-Hochberg kan ook het eerste voorbeeld uit sectie 6.6 worden gesimuleerd. Als we het tweede voorbeeld uit sectie 6.6 willen simuleren, dan definiëren we aan het begin dat $B = -2$ en veranderen we de uitdrukking voor 'theta' in de for-loop. Dit laatste is ook nodig voor trekkingen uit kansverdelingen.

```
############################################################

a <- 100                        #number of simulations
k <- 1000                       #length of theta_0
l <- 0.1*k                      #number of non-zeros

A <- 1
#B <- -2

n <- function(z, y, tau) {z^(-(1/2))*(1-z)*(1/(tau^2 + (1-tau^2)*z)) * exp((1/2)*z* y^2)};
d <- function(z, y,tau) {z^(-(1/2))*(1/(tau^2 + (1-tau^2)*z)) * exp((1/2)*z* y^2)};

pm <- function(yi,t){
    p <- integrate(n,lower = 0, upper = 1, y =yi, tau=t )$value;
    q <- integrate(d,lower = 0, upper = 1, y =yi, tau=t )$value;
    theta2 <- yi*(1-(p/q))
    return(theta2)
}

pm.vec <- Vectorize(pm)


int <- function(z,y,tau,k) {z^k * (1/(tau^2 +(1 - tau^2)*z)) * exp((1/2)*z* y^2)}

I <- function(yi,t,ki){
    int2 <- integrate(int,lower = 0, upper = 1, y =yi, tau=t, k=ki)$value;
    return(int2)
}

I.vec <- Vectorize(I)

tau <- (l/k)*sqrt(2*log(k/l))

st1 <- rep(0,a)    #matrix with type 1 errors simple multiple testing
bt1 <- rep(0,a)    #matrix with type 1 errors bonferroni
bht1 <- rep(0,a)   #matrix with type 1 errors benjamini hochberg
```

```
mt1 <- rep(0,a)    #matrix with type 1 errors posterior mean
vt1 <- rep(0,a)    #matrix with type 1 errors posterior variance
ct1 <- rep(0,a)    #matrix with type 1 errors combinations posterior mean and variance

vec1 <- rep(0,8)    #average type 1 errors for eight values of A simple mutiple testing
vec2 <- rep(0,8)    #average type 1 errors for eight values of A bonferroni
vec3 <- rep(0,8)    #average type 1 errors for eight values of A benjamini-hochberg
vec7 <- rep(0,8)    #average type 1 errors for eight values of A posterior mean
vec9 <- rep(0,8)    #average type 1 errors for eight values of A posterior variance
vec11 <- rep(0,8) #average type 1 errors for eight values of A combination mean & variance

st2 <- rep(0,a)    #matrix with type 2 errors simple multiple testing
bt2 <- rep(0,a)    #matrix with type 2 errors bonferroni
bht2 <- rep(0,a)   #matrix with type 2 errors benjamini hochberg
mt2 <- rep(0,a)    #matrix with type 2 errors posterior mean
vt2 <- rep(0,a)    #matrix with type 2 errors posterior variance
ct2 <- rep(0,a)    #matrix with type 2 errors combinations posterior mean and variance

vec4 <- rep(0,8)  #average type 2 errors for eight values of A simple mutiple testing
vec5 <- rep(0,8)  #average type 2 errors for eight values of A Bonferroni
vec6 <- rep(0,8)  #average type 2 errors for eight values of A benjamini-hochberg
vec8 <- rep(0,8)  #average type 2 errors for eight values of A posterior mean
vec10 <- rep(0,8) #average type 2 errors for eight values of A posterior variance
vec12 <- rep(0,8) #average type 2 errors for eight values of A combination mean & variance

while (A != 9){
    print(A)
    theta <- c(rep(0, k-l), rep(A,l) )  #parameters
   #theta <- c(rep(0, k-l), rep(B,l/2) , rep(A,l/2) ) #If B = -2
    for (j in 1:a){
        print(j)
        Z <- rnorm(k)              #noise
        Y <- theta + Z             #observations
        theta2 <- pm.vec(Y, tau)   #estimations of parameters with posterior mean

        p1 <- I.vec(Y,tau,0.5)
        p2 <- I.vec(Y,tau,-0.5)
        p3 <- I.vec(Y,tau,1.5)

        Var <- (p1 / p2) + Y^2 * ((p3/p2) - (p1/p2)^2)        #posterior variance

        pval <- 2*pnorm(abs(Y), lower.tail = F)               #p-values
        psort <- sort(pval)                                   #ordered p-values
        siglvl <- 0.05                 #significance level Benjamini-Hochberg
        threshold <- siglvl *(1:length(psort))/length(psort)  #threshold for p-values

        x <- 0;   #nominator mistake type 1 simple multiple testing
        b <- 0;   #nominator mistake type 1 Bonferroni
        bh1 <- 0; #nominator mistake type 1 Benjamini=Hochberg
        v <- 0;   #nominator mistake type 1 posterior mean
```

```
y <- 0;   #nominator mistake type 2 simple multiple testing
c <- 0;   #nominator mistake type 2 Bonferroni
bh2 <- 0; #nominator mistake type 2 Benjamini=Hochberg
w <- 0;   #nominator mistake type 2 posterior mean

#Proportion Type 1 errors: Simple multiple testing, Bonferroni, B-H
for (i in 1:(k-l)){
    if (abs(Y[i]) > 1.96){      #1.96 for 5% and 1.645 for 10%;
        x <- x+1}
    if (abs(Y[i]) > 4.06){      #k=100: 3.48 for 5% and 3.29 for 10%;
                                #k=1000: 4.06 for 5% and 3.89 for 10%
        b <- b+1}
    if (pval[i] < threshold[i]){
        bh1 <- bh1 +1}
}

#Proportion Type 2 errors: Simple multiple testing, Bonferroni, B-H
for (i in (k+1-l):k){
    if (abs(Y[i]) < 1.96){    #1.96 for 5% and 1.645 for 10%
        y <- y+1}
    if (abs(Y[i]) < 4.06){     #k=100: 3.48 for 5% and 3.29 for 10%
                               #k=1000: 4.06 for 5% and 3.89 for 10%
        c <- c+1}
    if (pval[i] > threshold[i]){
        bh2 <- bh2 +1}
}

res <- as.numeric(theta2/Y > 0.5)  #compare posterior mean with threshold
v <- sum(res[1:(k-l)])
w <- sum(res[(k-l+1):k]==0)

res2 <- as.numeric(Var > 0.75)     #compare posterior variance with threshold
f <- sum(res2[1:(k-l)])
g <- sum(res2[(k-l+1):k]==0)

res3 <- as.numeric(abs(theta2)+1.96*sqrt(Var) > 1.75) #combination mean & variance
t <- sum(res3[1:(k-l)])
u <- sum(res3[(k-l+1):k]==0)


st1[j] <- x/(k-l)
bt1[j] <- b/(k-l)
bht1[j] <- bh1/(k-l)
mt1[j] <- v/(k-l)
vt1[j] <- f/(k-l)
ct1[j] <- t/(k-l)

st2[j] <- y/(l)
bt2[j] <- c/(l)
```

```
        bht2[j] <- bh2/(l)
        mt2[j] <- w/(l)
        vt2[j] <- g/(l)
        ct2[j] <- u/(l)
    }

    vec1[A] <- mean(st1)
    vec2[A] <- mean(bt1)
    vec3[A] <- mean(bht1)
    vec7[A] <- mean(mt1)
    vec9[A] <- mean(vt1)
    vec11[A] <- mean(ct1)

    vec4[A] <- mean(st2)
    vec5[A] <- mean(bt2)
    vec6[A] <- mean(bht2)
    vec8[A] <- mean(mt2)
    vec10[A] <- mean(vt2)
    vec12[A] <- mean(ct2)

    A <- A+1
}

library(lattice)

xyplot(vec1 + vec2 + vec3 + vec7 + vec9 + vec11 ~1:8,
       xlab="A", ylab="Type 1 fout",
       main="Proportie Type 1 fouten", type = "l",
       col = c("red", "blue","green", "black", "cyan", "magenta"),
       key=list(text=list(c("Simple multiple testing", "Bonferroni correctie",
                            "Benjamini-Hochberg", "Posterior mean", "Posterior variance",
                            "Combinatie posterior mean en variance")),
                lines=list(lwd=c(2.5,2.5,2.5,2.5, 2.5, 2.5), type="l",
                           col=c("red", "blue","green", "black", "cyan", "magenta")),
                columns=1, divide=6), scales=list(x=list(tick.number=8, cex=1),
                                                  y=list(tick.number=10)))

xyplot(vec4 + vec5 + vec6 + vec8 + vec10 + vec12 ~1:8,
       xlab="A", ylab="Type 2 fout",
       main="Proportie Type 2 fouten", type = "l",
       col = c("red", "blue","green", "black", "cyan", "magenta"),
       key=list(text=list(c("Simple multiple testing", "Bonferroni correctie",
                            "Benjamini-Hochberg", "Posterior mean", "Posterior variance",
                            "Combinatie posterior mean en variance")),
                lines=list(lwd=c(2.5,2.5,2.5,2.5, 2.5, 2.5), type="l",
                           col=c("red", "blue","green", "black", "cyan", "magenta")),
                columns=1, divide=6), scales=list(x=list(tick.number=8, cex=1),
                                                  y=list(tick.number=10)))

#############################################################
```