



BSc verslag TECHNISCHE WISKUNDE

”Het schatten van de Minimale Effectieve Dosis in een Dosis-Respons Model”

JOS STEIJN

Technische Universiteit Delft

Begeleider

Dr.ir. F. van der Meulen

Overige commissieleden

Drs. E. van Elderen

Dr. L. van Iersel

Juli, 2016

Delft

INHOUDSOPGAVE

1. Introductie	3
2. Keuze Methode	6
3. Lineair model	6
3.1. Methode	6
3.2. Methode testen	10
3.3. Andere parametrisatie	14
4. Een niet-parametrisch Model	17
4.1. Keuze van F	17
4.2. A priori verdelingen	17
4.3. Hulpmiddelen voor het testen van de methode	19
4.4. Volledige Methode	21
4.5. Simulaties	23
5. Werkelijke Dataset	27
5.1. Dataset	27
5.2. Toepassen Niet-parametrisch Model	27

1. INTRODUCTIE

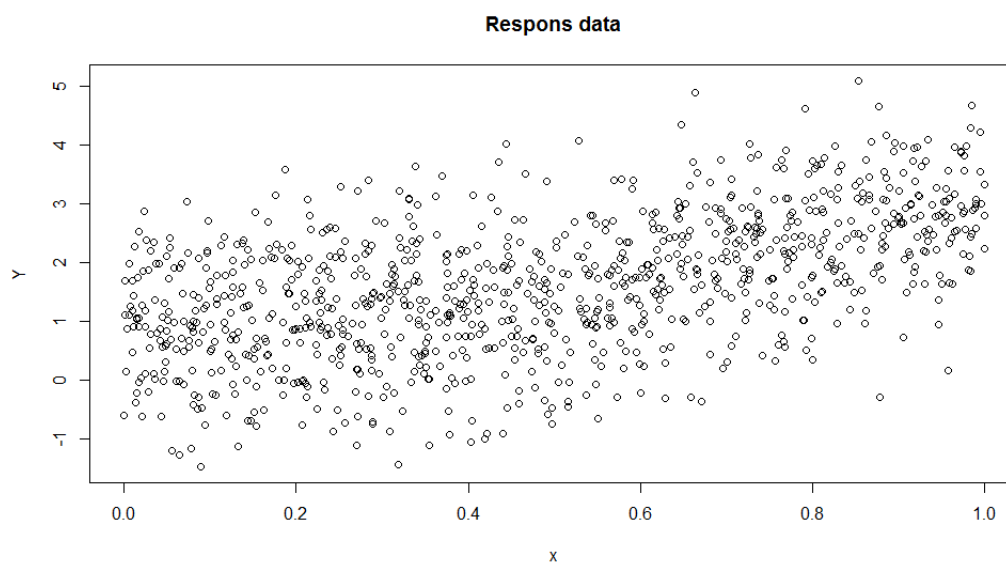
In dose-response analyses is het van belang na te gaan wat het effect is van het toedienen van een medicijn op de patiënt. Soms is het effect binair (genezen of niet), en kan de vraag zijn om na te gaan wat de kans op genezing is voor een patiënt als deze een dosis x van het medicijn krijgt. Soms is het meer kwantitatief, en wordt een meting gedaan die als respons kan worden gezien. Binnen een bepaalde range van doses, kan vaak worden aangenomen dat het effect van de dosis een stijgende functie van die dosis is. Wanneer we de respons noteren als (stochastische variabele) y , is een eenvoudig model dat

$$y|x \sim N(\mu(x), \sigma^2),$$

waarin μ een monotoon stijgende functie is en $\sigma^2 > 0$ onbekend. De Minimal Effective Dose (MED) is, voor een vaste vooraf gekozen waarde $\Delta > 0$, de laagste dosis waarvoor effect Δ optreedt:

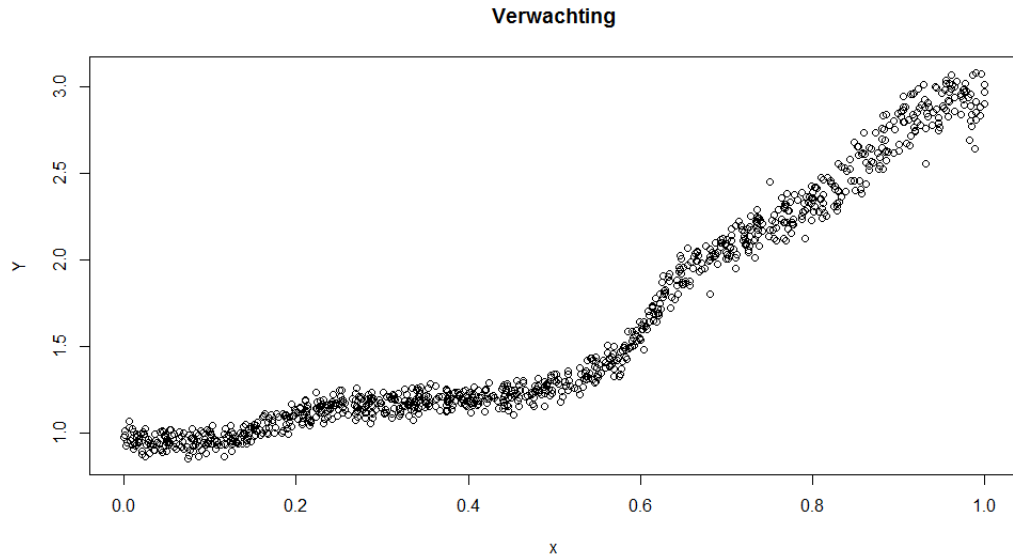
$$MED = \min\{x > 0 : \mu(x) \geq \mu(0) + \Delta\}$$

Het doel is nu om op basis van een dataset, verkregen of zelf gecreëerd, bestaande uit paren (x_i, y_i) deze MED te schatten. In een plaatje kan dit worden weergegeven als, gegeven een dataset bestaande uit (x_i, y_i) , zoals



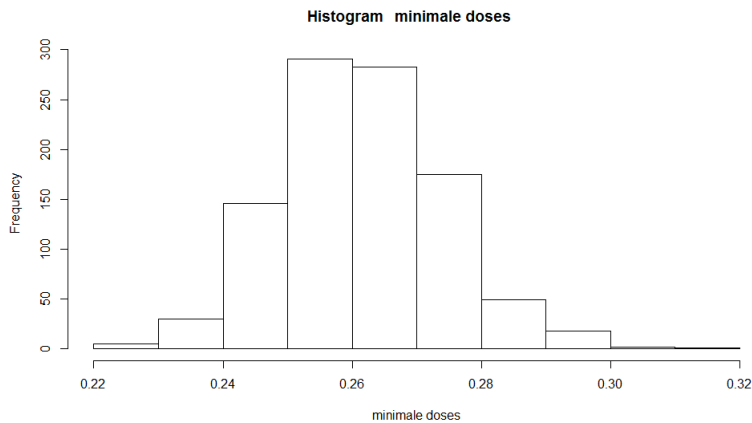
FIGUUR 1. Een dataset bestaande uit paren (x_i, y_i)

, genereren we verwachtingen $(x_i, \mu_j(x_i))$ en variantie waarmee de data afwijkt van deze verwachtingen σ_i^2 .



FIGUUR 2. De geconstrueerde verwachting bestaande uit paren $(x_i, \mu(x_i))$

Hieruit kunnen we dan de MED berekenen:



FIGUUR 3. De bij de verwachtingen horende MED's

Om dit te bereiken zullen we verschillende keuzes moeten maken in onze aanpak en aannames in ons model. We kiezen er eerst voor om $\mu(x)$ vrij simpel uit te drukken als

$$\mu(x_i) = \beta_0 + \beta_1 x_i,$$

waarbij β_0 en β_1 parameters groter dan 0 zijn. Op basis van de data gaan we dan verdelingen vinden waaruit we trekkingen kunnen doen voor β_0, β_1 en σ^2 . Met deze

trekkingen kunnen μ en uiteindelijk ook de MED benaderd worden.

Hierna is gekeken naar een niet-parametrische keuze van $\mu(x)$. De keuze viel op

$$\mu(x) = \beta_0 + \beta_1 \mu^0(x),$$

met

$$\mu^0(x) = \int_{\Xi} F(x, \xi) P(d\xi) = \sum_{j=1}^J w_j F(x, \xi_j).$$

Hierin is $\xi_j = (m_j, v_j)$ een vector. J is een constante. Er is voor gekozen om J en v_j vast te nemen om het model eenvoudig te houden. Net als bij het lineaire model proberen we op basis van data verdelingen te vinden waaruit we trekkingen kunnen doen voor $\beta_0, \beta_1, \sigma^2$, en nu ook w_j en m_j . Er zal blijken dat dit voor het niet-parametrische model met meer moeite gaat dan voor het lineaire model. Met deze trekkingen kunnen we dan weer een benadering maken voor $\mu(x)$ en een MED vinden.

Uiteindelijk zullen we het niet-parametrische model ook toepassen op een werkelijke dataset i.p.v. een dataset die we zelf hebben gecreëerd.

2. KEUZE METHODE

Er zijn verschillende methodes die we kunnen gebruiken om het vraagstuk aan te pakken. De uitdaging is om μ te schatten, nadat we een aanname hebben gedaan over de vorm van deze functie.

We kunnen ervoor kiezen om μ te schatten met behulp van de methode van maximum likelihoodschatters. We verwachten dat deze methode prima werkt zolang als we de vorm van μ niet te moeilijk nemen, bijvoorbeeld $\mu(x_i) = \beta_0 + \beta_1 x_i$.

Voor ingewikkeldere, niet-parametrische vormen van μ levert het gebruik van maximum likelihoodschatters echter problemen op. Met het gebruik van maximum likelihoodschatters is het lastig om, als we de MED hebben berekend, de onzekerheid van deze MED te kwantificeren. Daarom betekent het kiezen voor de methode van maximum likelihoodschatters dat we ons moeten beperken in de keuze van μ .

Daarom hebben we gekozen voor een Bayesiaanse aanpak. De Bayesiaanse methode werkt flexibel met het wat uitgebreidere model waar we later in dit verslag naar zullen kijken, en werkt, hopelijk, ook goed voor een lineaire μ .

De Bayesiaanse methode berust op de regel van Bayes. Die zegt dat:

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

Deze theorie is ook toe te passen op dichtheden. De posterior verdeling waaruit we, gebaseerd op de data, trekkingen doen voor de verschillende parameters is evenredig met het product van de kansdichtheden van de 'prior' (een aanname die vooraf wordt gedaan op de verdeling van de parameters) en de 'likelihood' (de 'waarschijnlijkheid' van de data gegeven de parameters). Wiskundig wordt dit geformuleerd als:

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\boldsymbol{\theta}) \times f(\mathbf{y}|\boldsymbol{\theta})$$

Dit is een stukje theorie dat we nog veel zullen gebruiken in het vervolg van dit verslag.

3. LINEAIR MODEL

3.1. Methode. In de intro werd deze eenvoudige vorm voor μ al besproken. Laat \mathbf{y} een vector ter grootte n die gemeten responswaardes bevat voorstellen, en laat \mathbf{x} de vector ter grootte n zijn die de doses behorende bij deze responswaardes bevat. Elke $y_i \in \mathbf{y}$ is dus gekoppeld aan $x_i \in \mathbf{x}$, voor $i \in \{1, \dots, n\}$. Nu nemen we aan dat deze data als volgt is opgebouwd.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \mu(x_i) + \epsilon_i$$

Met $\beta_0, \beta_1 \in \mathbb{R}_{\geq 0}$ en $\epsilon_i \sim N(0, \phi^{-1})$, $\epsilon_i \in \boldsymbol{\epsilon}$, met $\phi = \sigma^{-2}$ als variantie. μ is nu dus gedefinieerd als zijnde

$$\mu(x_i) := \beta_0 + \beta_1 x_i$$

Verder definiëren we nu

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

We veronderstellen dat deze a priori gegeven ϕ is verdeeld als:

$$\boldsymbol{\beta}|\phi \sim N(\boldsymbol{\mu}_0, \phi^{-1} \mathbf{C}_0^{-1}).$$

en $n_0 \sigma_0^2 \phi \sim \chi_{n_0}^2$.

Nu we deze definities en aannames gedaan hebben, gaan we proberen de MED te bepalen. Eerder was hadden we de MED gedefinieerd als

$$MED = \min\{x > 0 : \mu(x) \geq \mu(0) + \Delta\}$$

We hebben nu verondersteld dat $\mu(x) = \beta_0 + \beta_1 x$. Dan geldt dat

$$\begin{aligned} & \mu(x) \geq \mu(0) + \Delta \\ \iff & \beta_0 + \beta_1 x \geq \beta_0 + \Delta \\ \iff & x \geq \frac{\Delta}{\beta_1} \end{aligned}$$

Er geldt dus dat we eigenlijk alleen β_1 moeten bepalen, want Δ kunnen we zelf kiezen, en de MED is dan gelijk aan $\frac{\Delta}{\beta_1}$ (want de MED is de minimale x zodanig dat $\mu(x) \geq \mu(0) + \Delta$).

Hoewel we dus voor de berekening van de MED slechts β_1 nodig hebben, zullen we toch ook β_0 en ϕ berekenen. Dit doen we omdat het een kleine moeite is voor de extra validatie van de methode, en omdat het interessant is om een idee te hebben van de variantie/onzekerheid van de berekende MED.

Om dit te berekenen stellen we eerst de prior verdeling en likelihood functie op. Hiermee kunnen we de posterior dichtheid berekenen, die we dan gaan proberen om te schrijven naar een dusdanige vorm dat we hier trekkingen voor β_0, β_1 en ϕ uit kunnen doen. Aldus:

We stellen de dichtheid van de prior verdeling op:

$$\begin{aligned} p(\boldsymbol{\beta}, \phi) &= (2\pi)^{-p/2} |\phi \mathbf{C}_0|^{1/2} \exp\left\{-\frac{\phi}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{C}_0 (\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right\} \\ &\times \frac{(n_0 \sigma_0^2 / 2)^{n_0/2}}{\Gamma(n_0/2)} \phi^{(n_0/2)-1} \exp\left\{\frac{n_0 \sigma_0^2}{2} \phi\right\} \\ &\propto \phi^{[(n_0+p)/2]-1} \exp\left\{-\frac{\phi}{2} [n_0 \sigma_0^2 + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{C}_0 (\boldsymbol{\beta} - \boldsymbol{\mu}_0)]\right\} \end{aligned}$$

We gaan nu de marginale verdeling van $\boldsymbol{\beta}$ bekijken. Deze is evenredig met $p(\boldsymbol{\beta}, \phi)$ gedeeld door $p(\phi|\boldsymbol{\beta})$. We vinden zo dat de dichtheid van $p(\boldsymbol{\beta})$ wordt gegeven door

$$p(\boldsymbol{\beta}) \propto [n_0 \sigma_0^2 + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{C}_0 (\boldsymbol{\beta} - \boldsymbol{\mu}_0)]^{-(n_0+p)/2}$$

Wat correspondeert met de $t_{n_0}(\boldsymbol{\mu}_0, \sigma_0^2 \mathbf{C}_0^{-1})$ verdeling. De normalisatie constante is

$$\frac{\Gamma[(n_0+p)/2]}{\Gamma(n_0/2) n_0^{p/2}} (n_0 \sigma_0^2)^{n_0/2} |\mathbf{C}_0|^{1/2}$$

Voor verdere berekeningen definiëren we de modelmatrix \mathbf{X} als:

$$\mathbf{X} := \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Nu kunnen we de likelihood van de data gegeven $\boldsymbol{\beta}$ en ϕ uitdrukken als

$$\phi^{n/2} \times \exp\left\{-\frac{\phi}{2} [n_0 \sigma_0^2 [S_e + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]]\right\}$$

, met $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ en dit heeft dezelfde vorm als de a priori dichtheid. We kunnen de posterior dichtheid opstellen. Deze gaat er als volgt uit zien:

$$p(\boldsymbol{\beta}, \phi | \mathbf{y}) \propto \phi^{((n+n_0+p)/2)-1} \times \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + S_e + (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{C}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]\right\}$$

We gaan proberen deze nu naar een wat eenvoudigere vorm om te schrijven. Dit doen we als volgt:

$$\begin{aligned} & (\boldsymbol{\beta} - \boldsymbol{\mu}_0)' \mathbf{C}_0(\boldsymbol{\beta} - \boldsymbol{\mu}_0) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{\beta} - \boldsymbol{\mu}_1)' \mathbf{C}_1(\boldsymbol{\beta} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_0' \mathbf{C}_0 \boldsymbol{\mu}_0 + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_1' \mathbf{C}_1 \boldsymbol{\mu}_1 \end{aligned}$$

Met

$$\begin{aligned} \boldsymbol{\mu}_1 &= \mathbf{C}_1^{-1}(\mathbf{C}_0 \boldsymbol{\mu}_0 + \mathbf{X}'\mathbf{y}) \\ \mathbf{C}_1 &= \mathbf{C}_0 + \mathbf{X}'\mathbf{X} \end{aligned}$$

De posterior ziet er nu dus als volgt uit:

$$\phi^{((n+n_0+p)/2)-1} \times \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + S_e + (\boldsymbol{\beta} - \boldsymbol{\mu}_1)' \mathbf{C}_1(\boldsymbol{\beta} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_0' \mathbf{C}_0 \boldsymbol{\mu}_0 + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_1' \mathbf{C}_1 \boldsymbol{\mu}_1]\right\}$$

We schrijven:

$$\begin{aligned} & S_e + (\boldsymbol{\beta} - \boldsymbol{\mu}_1)' \mathbf{C}_1(\boldsymbol{\beta} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_0' \mathbf{C}_0 \boldsymbol{\mu}_0 + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_1' \mathbf{C}_1 \boldsymbol{\mu}_1 \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_0' \mathbf{C}_0 \boldsymbol{\mu}_0 \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_0' \mathbf{C}_0 \boldsymbol{\mu}_0 + \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\mu}_1' (\mathbf{C}_0 \boldsymbol{\mu}_0 + \mathbf{X}'\mathbf{y}) \\ &= (\mathbf{y} - \mathbf{X} \boldsymbol{\mu}_1)' \mathbf{y} + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)' \mathbf{C}_0 \boldsymbol{\mu}_0 \end{aligned}$$

Hiermee volgt dat de posterior verdeling nu kan worden omschreven naar

$$p(\boldsymbol{\beta}, \phi | \mathbf{y}) \propto \phi^{p/2} \exp\left\{-\frac{\phi}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_1)' \mathbf{C}_1(\boldsymbol{\beta} - \boldsymbol{\mu}_1)\right\} \phi^{(n_1/2)-1} \phi^{(n_1/2)-1} \exp\left\{-\frac{\phi}{2} n_1 \sigma_1^2\right\}$$

Met:

$$\begin{aligned} n_1 &= n + n_0 \\ n_1 \sigma_1^2 &= N_0 \sigma_0^2 + (\mathbf{y} - \mathbf{X} \boldsymbol{\mu}_1)' \mathbf{y} + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)' \mathbf{C}_0 \boldsymbol{\mu}_0 \end{aligned}$$

De posterior verdeling is nu van dezelfde vorm als die van de prior, en we kunnen hier dus de posterior verdelingen voor β_0 , β_1 en ϕ uit halen. Er geldt voor de β 's dat ze verdeeld zijn met een t-verdeling. De t-verdeling is voor grote sample sizes zo goed als gelijk aan de normale verdeling. Aldus:

$$\boldsymbol{\beta} | \mathbf{y} \sim N\left(\boldsymbol{\mu}_1, \frac{n_1}{n_1 - 2} \sigma_1^2 \mathbf{C}_1^{-1}\right)$$

Voor ϕ vinden we de posteriori verdeling

$$n_1 \sigma_1^2 \phi | \mathbf{y} \sim \chi_{n_1}^2$$

met verwachtingswaarde σ_1^{-2} .

Voor het vervolg van het verslag wordt er nu nieuwe notatie geïntroduceerd. Stel dat er voor β_0 en β_1 een m aantal trekkingen gedaan worden uit de eerder genoemde verdeling, dan noemen we de getrokken waardes $\beta_{0,j}$ en $\beta_{1,j}$ voor $j \in \{1, \dots, m\}$. We definiëren μ_j als zijnde de door ons geconstrueerde verwachtingfunctie gebaseerd op $\beta_{0,j}$ en $\beta_{1,j}$, dus

$$\mu_j(x_i) := \beta_{0,j} + \beta_{1,j} x_i$$

We hadden eerder al gezien dat de MED gelijk was in de minimale x zodanig dat

$x \geq \frac{\Delta}{\beta_1}$. Als we dus een m aantal $b_{1,j}$ hebben getrokken, vinden we een m aantal schattingen voor de MED die zijn gelijk aan $\frac{\Delta}{\beta_{1,j}}$, voor $j \in \{1, \dots, m\}$.

3.2. Methode testen. Na het uitwerken van onze Bayesiaanse methode willen we deze natuurlijk ook geen testen om na te gaan dat de methode daadwerkelijk werkt. Hiertoe gaan we data creëren, om vervolgens met de eerder uitgewerkte Bayesiaanse methode de parameters waarmee we onze data hebben gecreëerd te schatten en dan, als het goed is, te verifiëren dat die geschatte parameters in de buurt komen van de vooraf vastgezette waarden.

We zetten de parameters in ons model als volgt vast:

$$\boldsymbol{\beta} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\phi = 1$$

Vervolgens creëren we hiermee onze data:

$$\epsilon \sim N(0, 1)$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

voor $i \in \{1, \dots, n\}$. Verder hebben we $\Delta = 0,5$ gezet voor het berekenen van de MED. De vector \mathbf{y} stelt dus nu onze data voor. Met deze data kunnen we trekkingen gaan doen uit de posterior verdelingen. Een aantal a priori parameters in onze methode, te weten n_0 , μ_0 , σ_0 en C_0 , moeten nog ingevuld worden. Als onze methode goed werkt kunnen hier vrij extreme waardes invullen, alleen zullen we dan moeten compenseren met een grote sample size, zodat de methode kan compenseren. Dit wordt duidelijk uit het volgende:

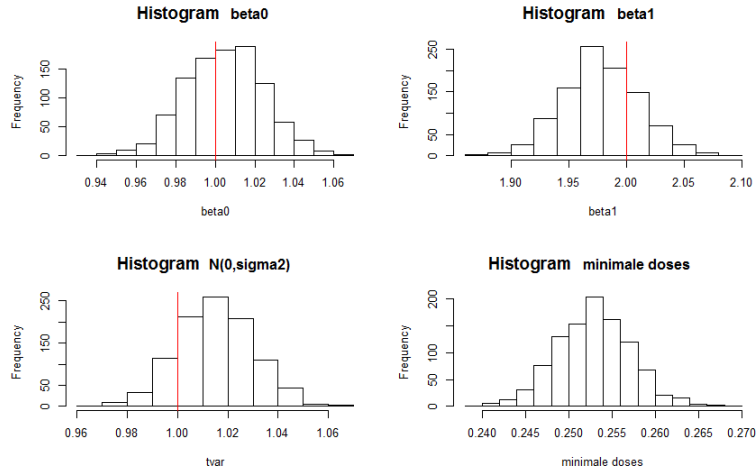
We schrijven de eerdergenoemde regel van Bayes om:

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{y}) &\propto && f(\boldsymbol{\theta}) \times f(\mathbf{y}|\boldsymbol{\theta}) \\ &\propto && f(\boldsymbol{\theta}) \times f(y_1|\boldsymbol{\theta}) \times f(y_2|\boldsymbol{\theta}) \times \dots \times f(y_n|\boldsymbol{\theta}) \\ &\propto && \exp\{\log(f(\boldsymbol{\theta})) + \sum_{i=1}^n \log(f(y_i|\boldsymbol{\theta}))\} \end{aligned}$$

Dit laat zien dat bij grote hoeveelheden data, het effect van de prior $f(\boldsymbol{\theta})$ op de posterior in het niet valt vergeleken met het effect van de likelihood. Daarom is de verwachting dat er voor extremere a priori waardes (dat wil zeggen waardes die ver af liggen van de waardes waarmee we de data hebben gegenereerd) een grotere sample size nodig is, om toch een goede uitkomst te hebben voor de posterior trekkingen.

Allereerst hebben we gekeken naar wat vrij simpel gekozen a priori waardes (zie bijschrift onder de resultaten). Simpel, omdat de paramters nu a priori gelijk zijn aan de paramters die we gebruikt hebben om de data te genereren.

De volgende resultaten zijn verkregen:

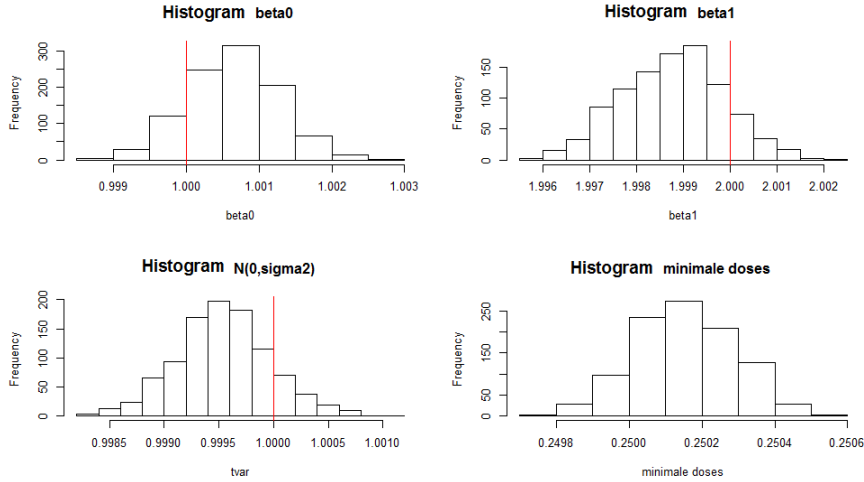


FIGUUR 4. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 1, \boldsymbol{\mu}_0 = (1, 2), \sigma_0 = 1, \mathbf{C}_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^4$. Aantal geplote resultaten: 10^3 .

De verdelingen lijken redelijk in de buurt te komen met de waardes die we vooraf vastgezet hadden ($\beta_0 = 1, \beta_1 = 2$ en $\phi = 1$), maar het midden van het histogram lijkt toch net iets naast de waarde die we willen krijgen te liggen. Dat is te verklaren: de data is natuurlijk licht natuurlijk nooit perfect rond $\mu(x_i)$, gezien de data is opgebouwd als $y_i = \mu(x_i) + \epsilon_i$, met $\epsilon_i \sim N(0, \sigma^2)$.

Verder zien we dat de MED rond 0,253 lijkt te liggen en dat de resultaten voor deze MED niet al te gespreid zijn. Dat was ook te verwachten, gezien de MED gelijk was aan $\frac{\Delta}{\beta_1}$ in dit model, en Δ hebben we vast gekozen, en de spreiding van de $\beta_{1,i}$ ook niet al te groot was.

We kijken naar wat het effect is van het vergroten van de sample size. We vergroten hiertoe de sample size naar 10^7 . De andere factoren zijn hetzelfde gebleven. De gevonden resultaten zijn:

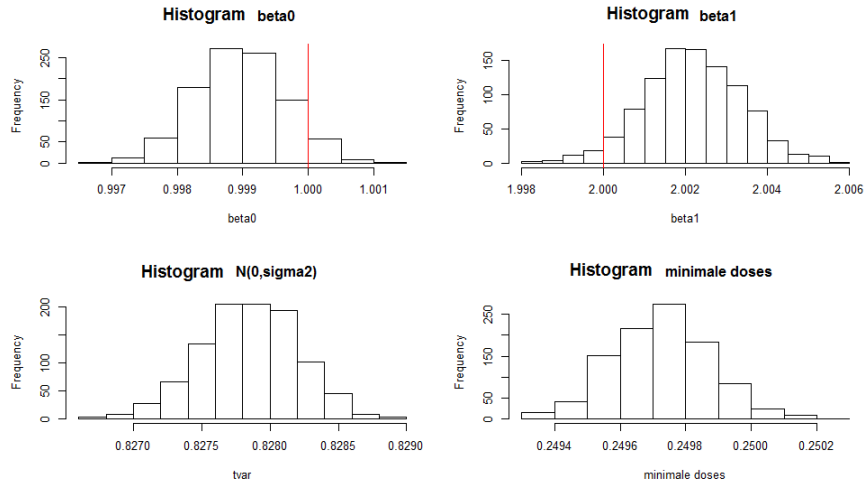


FIGUUR 5. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 1, \mu_0 = (1, 2), \sigma_0 = 1, C_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^7$. Aantal geplote resultaten: 10^3 .

We zien dat het erg lijkt op wat we eerder krijgen. Nog steeds komt het midden van het histogram erg dicht bij de werkelijke waarde, maar zit er net naast. Dit verschil is wel kleiner geworden, wat ook te verwachten is bij deze grotere sample size. Een grotere sample size gaat namelijk de variantie tegen in het gemiddelde van de data.

De gevonden MED lijken ook erg op het resultaat wat we eerder hadden gekregen. Het midden van het histogram lijkt iets te zijn verschoven, wat ook logisch is want het midden van het histogram van de gevonden $\beta_{1,i}$ is dat ook.

In de eerdere tests waren de a priori in te vullen variabelen vrij simpel gekozen. We kiezen ze nu wat extremer, en kijken wat dit doet met de uitkomsten ons model.



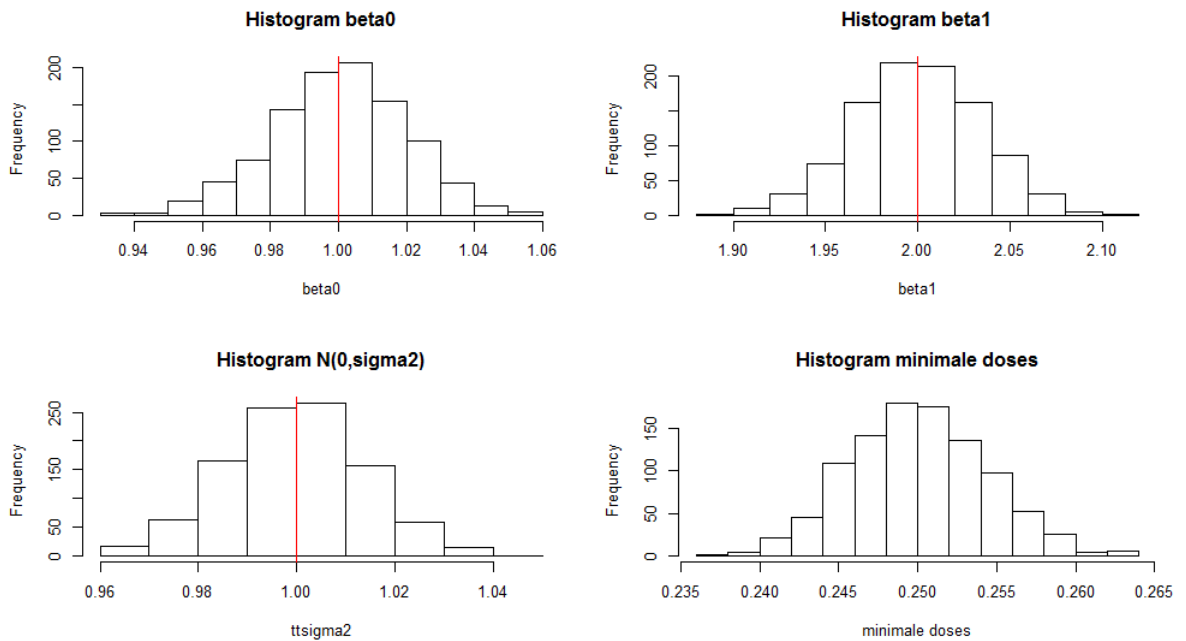
FIGUUR 6. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 50, \mu_0 = (5, 64), \sigma_0 = 200, \mathbf{C}_0 = 20 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^7$. Aantal geplote resultaten: 10^3 .

We zien dat de resultaten voor β vrijwel hetzelfde blijven. De verdeling van ϕ lijkt deze vreemde waarden wel slecht te verdragen. De reden hiervoor is niet zo snel te bepalen, maar het mag duidelijk zijn dat bij extremere a priori waarden σ^2 een grotere sample size nodig heeft dan β voordat het naar de juiste waarde convergeert.

Gelukkig blijft de MED, ook bij deze extremere a priori waarden, rond dezelfde waarde liggen. De methode lijkt dus in ieder geval consistent te werken voor ons eindresultaat.

Om te laten zien dat we met onze methode de parameters wel heel precies kunnen bepalen op basis van de data, nemen we de opstelling die als eerst werd doorgenomen er weer bij. De parameters voor deze opstelling zijn: $n_0 = 1$, $\boldsymbol{\mu}_0 = (1, 2)$, $\sigma_0 = 1$, $\mathbf{C}_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^4$.

Wat we nu echter gaan doen, is dat we nadat we 10^3 trekkingen hebben gedaan voor β_0 , β_1 en σ^2 , het gemiddelde van deze trekkingen voor elk van deze drie parameters bewaren, en weer opnieuw data gaan creëren. Op basis van de nieuwe data doen weer trekkingen voor β_0 , β_1 en σ^2 , en we bewaren weer de gemiddeldes. Als we dit proces 10^3 maal doorlopen, vinden we



FIGUUR 7. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 1$, $\boldsymbol{\mu}_0 = (1, 2)$, $\sigma_0 = 1$, $\mathbf{C}_0 = \times I$ (I de identiteitsmatrix). Sample size: $n = 10^4$. Aantal datasets gebruikt: 10^3 . Aantal geplote gemiddeldes van trekkingen: 10^3 .

We zien dat de resultaten nu gecentreerd zijn rond de werkelijke waarden, zoals we hadden verwacht. We kunnen dus concluderen dat onze methode goed werkt.

3.3. Andere parametrisatie. Als onderdeel van dit onderzoek is er ook nog gekeken naar een andere parametrisatie van μ . In het boek waaruit de theorie voor het lineaire model komt, "Statistical Inference, An Integrated Approach", wordt een voorbeeld gegeven waarin niet $\mu(x_i) = \beta_0 + \beta_1(x_i)$, maar $\mu(x_i) = \beta_0 + \beta_1(x_i - \bar{x})$ wordt genomen. Deze parametrisatie is wat eenvoudiger, om redenen die in de afleiding van de posterior zullen blijken:

We nemen β_0 en β_1 a priori onafhankelijk gegeven een ϕ , verdeeld volgens:

$$\beta_0|\phi \sim N(\mu_0, (c_0\phi)^{-1})$$

$$\beta_1|\phi \sim N(\mu_1, (c_1\phi)^{-1})$$

Met deze nieuwe keuze van $\mu(x)$ is de design matrix \mathbf{X} veranderd. Daarvoor geldt nu:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{bmatrix}$$

Hierdoor is de matrix $\mathbf{X}'\mathbf{X}$ nu een diagonaalmatrix geworden. Nu kunnen we de kwadratische vorm $(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{X}'\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$ omschrijven tot $S(\beta_0)$ en $S(\beta_1)$, met

$$S(\beta_0) = n(\beta_0 - \bar{y})^2$$

$$S(\beta_1) = \sum_{i=1}^n (x_i - \bar{x})^2 (\beta_1 - \hat{\beta}_1)^2$$

De likelihood functie wordt in dit model gegeven door:

$$l(\beta_0, \beta_1, \phi; \mathbf{y}) \sim \phi^{n/2} \exp\left\{-\frac{\phi}{2}[S_e + S(\beta_0) + S(\beta_1)]\right\}$$

Dit kunnen we combineren met de priors om de volgende posterior te krijgen:

$$p(\beta_0, \beta_1, \phi|\mathbf{y}) \sim \phi^{(n+n_0+2)/2-1} \exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + S_e + c_0(\beta_0 - \mu_0)^2 + S(\beta_0) + c_1(\beta_1 - \mu_1)^2 + S(\beta_1)]\right\}$$

Hieruit kunnen we de marginale dichtheden voor β_0, β_1 en ϕ halen. Voor β_0 en β_1 geldt dat ze verdeeld zijn met een student-t verdeling met n_1 vrijheidsgraden, met verwachting μ_j^* en variantie σ_1^2/c_j , $j = \{0, 1\}$. De posterior verdeling voor ϕ wordt gegeven door $n_1\sigma_1^2\phi|\mathbf{y} \sim \chi_{n_1}^2$ met

$$n_1 = n_0 + n$$

$$c_0^* = c_0 + n$$

$$c_1^* = c_1 + \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\mu_0^* = \frac{c_0\mu_0 + n\bar{y}}{c_0 + n}$$

$$\mu_1^* = \frac{c_1\mu_1 + \sum_{i=1}^n (x_i - \bar{x})\hat{\beta}_1}{c_1 + \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$n_1\sigma_1^2 = \sum_{i=1}^n y_i^2 - n\mu_0^*\hat{\beta}_0 - \mu_1^*\hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})^2 + c_0(\beta_0 - \mu_0)^2 + c_1(\beta_1 - \mu_1)^2$$

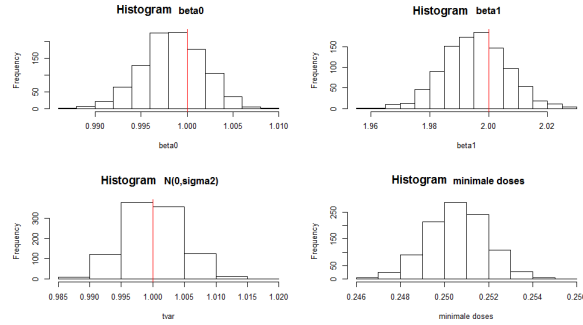
met

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

,zoals ook in eerdere modellen het geval was.

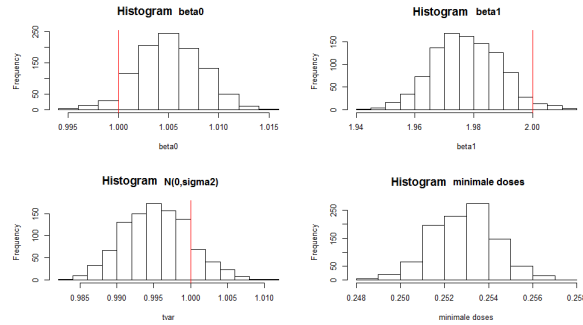
We hebben op basis van de gecreëerde data 10^3 trekkingen uit de posterior gedaan. Dit gaf wisselende resultaten. Hieronder zijn twee voorbeelden van resultaten die we hebben verkregen:

Een voorbeeld van een dataset (ter grootte 10^5) waarin de berekende verdelingen goed in de buurt van de gewenste waarden lijken te liggen:



FIGUUR 8. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 1, \boldsymbol{\mu}_0 = (1, 2), \sigma_0 = 1, \mathbf{C}_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^5$. Aantal geplote resultaten: 10^3 .

Een voorbeeld van een andere dataset (ook ter grootte 10^5) waarin de berekende verdelingen net af lijken te wijken van de gewenste waarden:



FIGUUR 9. Resultaten voor $\beta_0, \beta_1, \sigma^2$ en de minimale effective doses. Parameters voor deze opstelling zijn: $n_0 = 1, \boldsymbol{\mu}_0 = (1, 2), \sigma_0 = 1, \mathbf{C}_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^5$. Aantal geplote resultaten: 10^3 .

Deze eenvoudigere, in het boek uitgewerkte $\mu(x)$ geeft dus resultaten vergelijkbaar met die we vonden met de parametrisatie $\mu(x) = \beta_0 + \beta_1 x$.

Er kan worden nagegaan dat bij het gebruik van andere (klassieke) methodes, vergelijkbare resultaten worden gevonden. Daarom zijn we nog steeds tevreden met onze huidige methode van Bayesiaans schatten.

4. EEN NIET-PARAMETRISCH MODEL

Tot nu toe hebben we de minimum effective dose $\mu(x)$ een lineaire functie laten zijn. We gaan dit nu veranderen en stellen

$$\mu(x) = \beta_0 + \beta_1 \mu^0(x)$$

Nu moet $\mu^0(x)$ gekozen worden.

In het artikel *Bayesian Nonparametric Estimation of Continuous Monotone Functions with Applications to Dose-Response Analysis* Van Björn Bornkamp en Katja Ickstadt uit Maart 2009 wordt voorgesteld om μ^0 te zien als

$$\mu^0(x) = \int_{\Xi} F(x, \xi) P(d\xi) = \sum_{j=1}^J w_j F(x, \xi_j)$$

Hierin stelt $F(x, \xi_j)$ een parametrische verdelingsfunctie van op $[0, 1]$ begrensde continue variabelen voor met parameters $\xi \in \Xi$. w_j stelt het gewicht voor dat aan elke $F(x, \xi_j)$ wordt toegekend. Dus voor w_j geldt $w_j \geq 0$ en $\sum_{j=1}^J w_j = 1$. We stellen dat deze informatie bevat wordt in een vector $\theta := (J, w_1, w_2, \dots, \xi_1, \xi_2, \dots)$. Hierin stelt J het index getal voor.

Er is, hier en in het artikel, gekozen voor deze μ^0 omdat dit model erg flexibel is. Er zijn nu geen voorwaarden die vooraf gelegd moeten worden op de structuur van μ^0 , anders dan dat het een monotone functie is. Verder geldt nu dat de a priori variantie van ξ_j onafhankelijk is van J en de w_j . Hoewel dit nu nog niet helemaal duidelijk zal zijn, komt dit erg van pas later in onze berekeningen.

4.1. Keuze van F. We willen onze F wel zo kiezen dat elke continue verdelingsfunctie kan worden opgebouwd uit een combinatie van $F(\cdot, \xi_1), F(\cdot, \xi_2), \dots$. Een veelgebruikte verdelingsfunctie met deze eigenschap is de beta verdeling. Er is echter niet gekozen voor deze verdeling, omdat deze erg computationeel intensief wordt geacht.

In het eerder genoemde artikel en in dit verslag wordt daarom F gekozen als zijnde

$$F(x, \xi) = \begin{cases} m(\frac{x}{m})^v & 0 \leq x \leq m \\ 1 - (1 - m)(\frac{1-x}{1-m})^v & m \leq x \leq 1 \end{cases}$$

Hierbij hangt F dus van de parameter $\xi = (m, v) \in [0, 1] \times \mathbb{R}^+$ af. We kunnen elke verdelingsfunctie maken met een combinatie van bovengenoemde $F(\cdot, \xi_j)$, dus aan die voorwaarde is voldaan.

We kiezen ervoor om in het vervolg van het verslag de v vast te houden. Dit om het model, in ieder geval voorlopiger, nog wat eenvoudiger te houden, en omdat we dan waarschijnlijk alsnog prima resultaten kunnen vinden.

4.2. A priori verdelingen. De keuze van F brengt met zich mee dat we naast β en σ^2 nu ook met $\theta = (J, w_1, w_2, \dots, \xi_1, \xi_2, \dots)$ rekening moeten houden in de berekening van onze a priori en posteriori verdelingen. We gebruiken de normaal inverse gamme verdeling om de priors voor β, σ^2 conditioneel op θ uit te drukken. Dit ziet er als volgt uit:

$$p(\beta, \sigma^2 | \theta) \sim (\sigma^2)^{-(d+4)/2} \times \exp[-\{(\beta - \mathbf{m})' \mathbf{V}^{-1} (\beta - \mathbf{m}) + a\} / (2\sigma^2)]$$

Deze verdeling hangt af van parameters a, d de 2×2 matrix \mathbf{V} en een tweedimensionale vector \mathbf{m} . Als we zwakke prior informatie willen tonen dan laten we de a priori variantie naar oneindig gaan. Dit kan worden gedaan door $\mathbf{V}^{-1} \rightarrow 0$.

Als we dan zeggen $a = 0$ en $d = -2$ dan krijgen we de niet-informatieve prior verdeling

$$p(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{\theta}) \sim \sigma^{-2}$$

Merk op dat zowel voor de informatieve prior als de niet-informatieve geldt dat $\boldsymbol{\beta}$ en σ^2 niet van $\boldsymbol{\theta}$ afhankelijk zijn.

Net zoals in het lineaire model werken we weer met een vector $\mathbf{y} = (y_1, \dots, y_n)$ die de data voorstelt. De likelihood van deze data wordt gegeven door

$$f(\mathbf{y} | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) \sim (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\},$$

welke afhankelijk is van $\boldsymbol{\theta}$ door de design matrix \mathbf{X} .

Op deze manier kunnen we dus, door het kiezen van de priors voor $\boldsymbol{\beta}$ en σ^2 als zijnde normaal inverse gamma verdelingen, de a posteriori verdeling $p(\boldsymbol{\theta} | \mathbf{y})$ berekenen. Na veel rekenwerk (zie *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*, 2nd edition van A. O'Hagan and J. Forster (2004), p. 305-308 voor de volledige uitwerking) vinden we de posterior verdeling:

$$p(\boldsymbol{\theta} | \mathbf{y}) \sim \det\{(\mathbf{V}^{-1} + \mathbf{X}'\mathbf{X})^{-1}\}^{\frac{1}{2}} \times [a + (n-2)\hat{\sigma}^2 + (\mathbf{m} - \hat{\boldsymbol{\beta}})'(\mathbf{V} + (\mathbf{X}'\mathbf{X})^{-1})^{-1}(\mathbf{m} - \hat{\boldsymbol{\beta}})]^{-\frac{n+d}{2}} p(\boldsymbol{\theta})$$

Met:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ (n-2)\hat{\sigma}^2 &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned}$$

Met de niet-informatieve prior verdeling ($\mathbf{V}^{-1} \rightarrow 0$) geldt dat deze gelijk is aan:

$$p(\boldsymbol{\theta} | \mathbf{y}) = \det\{(\mathbf{X}'\mathbf{X})^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}^2)^{-\frac{n-2}{2}} p(\boldsymbol{\theta})$$

Nu moet er een manier gevonden worden om uit deze verdeling trekkingen te gaan doen.

4.3. Hulpmiddelen voor het testen van de methode. Net zoals we bij het simpelere lineaire model hebben gedaan, willen we ook deze wat uitgebreidere methode gaan testen. De β en σ^2 kunnen op vergelijkbare wijze als in het lineaire model getest worden, met als enige verandering dat de design matrix \mathbf{X} van θ afhankelijk is. Het trekken uit de posterior van $\theta|\mathbf{y}$ is wel nieuw en vormt nog een behoorlijke opgave.

In de theorie hadden we de keuze van $\theta = (J, w_1, w_2, \dots, w_J, \xi_1, \xi_2, \dots, \xi_J)'$ met $\xi_i = (m_j, v_i)$ besproken. Voor het gemak zetten we nu J en alle v_i vast. Dit maakt onze berekeningen ene stuk eenvoudiger en het idee achter het probleem blijft hetzelfde. De w_j en m_j zijn dus de getallen die we met onze posterior trekkingen willen gaan benaderen.

Om trekkingen te doen uit de verdeling van $\theta|\mathbf{y}$ wordt het Metropolis Hasting algoritme gebruikt. Dit gaat als volgt. We initialiseren een J aantal w_j en m_j willekeurig op zo'n manier dat $\forall w_j \geq 0$ en $\sum_{j=1}^J w_j = 1$, en $\forall m_j \in [0, 1]$. Dit doen we door aan te nemen dat de w_j volgens de Dirichlet verdeling verdeeld zijn, en de m_j onafhankelijk uit de uniforme verdeling begrensd tussen 0 en 1 te trekken.

Nu de w_j en m_j geïnitieerd zijn, kunnen ze worden aangepast in een poging om zo, gegeven de data, waarschijnlijkeres waarden te vinden. Voor m_j doen we dit door $vN(0, 1)$ erbij op tellen, waarbij v een naar eigen inzicht vast te stellen constante is. Merk op dat m_j begrensd moest blijven tussen 0 en 1. Om hieraan te blijven voldoen wordt er een log-transformatie gedaan alvorens de $vN(0, 1)$ wordt opgeteld. Daarna wordt de inverse van die log-transformatie weer toegepast. Hierdoor blijft m_j begrensd tussen 0 en 1 na het optellen van $vN(0, 1)$. Dit proces ziet er als volgt uit::

$$\begin{aligned}\hat{m}_j &= \log\left(\frac{m_j}{1 - m_j}\right) \\ \hat{m}_j' &= \hat{m}_j + vN(0, 1) \\ m_j' &= \frac{1}{1 + e^{-\hat{m}_j'}}\end{aligned}$$

Waarin v een naar eigen inzicht vast te stellen constante is en m_j' de uiteindelijk nieuwe voorgestelde vervanger voor m_j is. Op deze manier blijven de m_j en m_j' begrensd tussen 0 en 1 en kunnen we toch een kleine verstoring toevoegen.

Nu het vinden van een voorgestelde vervanger w_j' van w_j . Voor de w_j geldt dat $\forall w_j$ geldt dat $1 \geq w_j \geq 0$ en $\sum_{j=1}^J w_j = 1$. Het doel is om een w_j' te vinden die hier ook aan voldoet. Dit kan niet op dezelfde manier als hoe voor m_j een m_j' werd gevonden, want met die methode voldoen de w_j' niet aan $\sum_{j=1}^J w_j' = 1$. Hier moet dus een oplossing voor worden gevonden.

Houden we het model echter simpel en stellen we $J = 2$, dan kunnen we een w_1 intialiseren en w_2 gelijknemen aan $1 - w_1$. Dan kan een voorgestelde w_1' worden gevonden op dezelfde manier als hoe voor een m_j een m_j' wordt gevonden (zie methode hierboven), en w_2' wordt gelijkgesteld aan $1 - w_1'$. Deze methode werkt niet voor grotere J (waarbij w_1', \dots, w_{J-1}' worden gevonden zoals bij $J = 2$ alleen w_1' wordt gevonden, en w_J' gelijk wordt gesteld aan $1 - \sum_{j=1}^{J-1} w_j'$) omdat als de som van w_1' tot w_{J-1}' gelijk wordt aan, of groter wordt dan, 1, w_J' gelijk aan of kleiner dan 0 kan worden, en we willen dat de w_j' begrensd blijven tussen 0 en 1.

Om dit probleem op te lossen, en het model voor $J > 2$ te laten werken, is de parametrisatie veranderd. In plaats van

$$\mu^0(x) = \sum_{j=1}^J w_j F(x, \xi_j)$$

hebben we de parametrisatie

$$\mu^0(x) = \sum_{j=1}^J \frac{u_j}{\sum_{i=1}^J u_i} F(x, \xi_j)$$

,waarbij alle $u_i \geq 0$, genomen. We nemen nu als a priori verdeling voor de u_j

$$u_1, \dots, u_J \stackrel{\text{ind}}{\sim} \text{Exp}(\lambda),$$

waarbij de λ vast is genomen. Deze u_j kunnen we nu updaten. Dat wordt vrij simpel gedaan met

$$\log(u'_j) = \log(u_j) + vN(0, 1)$$

Waarin v een naar eigen inzicht vast te stellen constante is en u'_j de uiteindelijk nieuwe voorgestelde vervanger voor u_j is.

Nu we een methode hebben om onze m_j en w_j/u_j te updaten, moeten we deze met een bepaalde kans gaan accepteren/afwijzen. We stellen dat we de nieuwe θ accepteren met kans

$$(1) \quad \min\left(1, \frac{p(\theta'|\mathbf{y})q(\theta|\theta')}{p(\theta|\mathbf{y})q(\theta'|\theta)}\right)$$

Merk op dat de design-matrix \mathbf{X} onderhevig is aan verandering van θ . Daarom definiëren we

$$\mathbf{X}_\theta := \begin{bmatrix} 1 & \sum_{j=1}^J w_j F(x_1, \xi_j) \\ 1 & \sum_{j=1}^J w_j F(x_2, \xi_j) \\ \vdots & \vdots \\ 1 & \sum_{j=1}^J w_j F(x_n, \xi_j) \end{bmatrix}, \quad \mathbf{X}_{\theta'} := \begin{bmatrix} 1 & \sum_{j=1}^J w'_j F(x_1, \xi'_j) \\ 1 & \sum_{j=1}^J w'_j F(x_2, \xi'_j) \\ \vdots & \vdots \\ 1 & \sum_{j=1}^J w'_j F(x_n, \xi'_j) \end{bmatrix},$$

met $\xi'_j := (v, m'_j)$.

Met deze definities kan de breuk uit 1 worden doorgerekend tot:

$$\frac{p(\theta'|\mathbf{y})q(\theta|\theta')}{p(\theta|\mathbf{y})q(\theta'|\theta)} = \frac{\{\det(\mathbf{X}_{\theta'}' \mathbf{X}_{\theta'})^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}^2)^{-\frac{n-2}{2}} \prod_{j=1}^J m'_j (1-m'_j) u'_j}{\{\det(\mathbf{X}_\theta' \mathbf{X}_\theta)^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}^2)^{-\frac{n-2}{2}} \prod_{j=1}^J m_j (1-m_j) u_j}$$

Dit is voor de parametrisatie die we gekozen hebben voor grote J . In het simpelere geval met $J = 2$ is deze breuk iets eenvoudiger en wordt:

$$\frac{\{\det(\mathbf{X}_{\theta'}' \mathbf{X}_{\theta'})^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}_{\theta'}^2)^{-\frac{n-2}{2}} w'_1 (1-w'_1) \prod_{j=1}^J m'_j (1-m'_j)}{\{\det(\mathbf{X}_\theta' \mathbf{X}_\theta)^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}_\theta^2)^{-\frac{n-2}{2}} w_1 (1-w_1) \prod_{j=1}^J m_j (1-m_j)}$$

Let op in het verschil tussen $\mathbf{X}_{\theta'}$ en \mathbf{X}_θ . De design-matrix is onderhevig aan het verschil in de θ , en zit ook verwerkt in de $\hat{\sigma}^2$.

Met deze kans kunnen we nu de voorgestelde θ' aannemen of verwerpen. Hierna start het hele proces opnieuw en wordt er weer een nieuwe θ' gecreërd, vergeleken met θ en weer opnieuw aangenomen of verworpen.

4.4. Volledige Methode. Voor het overzicht is de volledige methode, van het data genereren tot trekken van verwachtingen uit de gevonden verdeling, hier uitgeschreven.

Neem $\mu(x) = \beta_0 + \beta_1 \mu_0(x)$, met $\mu_0(x) = \sum_{j=1}^J \frac{u_j}{\sum_{j=1}^J u_j} F(x, \xi_j)$, met $\xi_j = (m_j, v_j)$. Houd alle v_j gelijk en vast. Houd J vast.

Creëer een dataset \mathbf{y} . Kies hiervoor zelf $\beta_0^*, \beta_1^*, \sigma^{2*}, \theta^*$ en maak een n aantal $y_i = \beta_0^* + \beta_1^* \mu^{0*}(x_i) + N(0, \sigma^{2*})$, $\forall i \in \{1, \dots, J\}$ (μ^{0*} stelt de μ^0 met de parameters van θ^* voor). De x_i zijn onafhankelijk uniform verdeeld op $[0, 1]$.

- (1) Initialiseer een \mathbf{m} en \mathbf{u} . Voor de $m_j \in \mathbf{m}$ moet gelden dat $1 \geq m_j \geq 0, \forall j \in \{1, \dots, J\}$. Voor de u_j kan gewoon een a priori verdeling gekozen worden, bijvoorbeeld $u_j \stackrel{\text{ind}}{\sim} \text{Exp}(\lambda), \forall j \in \{1, \dots, J\}$.
- (2) Schrijf de m_j en u_j om en update ze. Schrijf voor de m_j :
 - (a) $\hat{m}_j = \log\left(\frac{m_j}{1-m_j}\right)$
 - (b) $\hat{m}_j' = \hat{m}_j + vN(0, 1)$
 - (c) $m_j' = \frac{1}{1+e^{-\hat{m}_j'}}$

Waarin m_j' de geupdate m_j is. v is een zelf in te vullen constante, die de grootte van de verschillen tussen de iteraties bepaalt. De u_j kunnen simpelweg geupdate worden met:

$$\log(u_j') = \log(u_j) + vN(0, 1)$$

Dus

$$u_j' = u_j e^{vN(0,1)}$$

Waarin v weer een zelf in te vullen constante is. Deze hoeft niet hetzelfde te zijn als de v die gebruikt is bij het updaten van de m_j , maar dat kan wel.

- (3) Bereken

$$A = \frac{p(\theta' | \mathbf{y}) q(\theta | \theta')}{p(\theta | \mathbf{y}) q(\theta' | \theta)} = \frac{\{\det(\mathbf{X}_{\theta'}' \mathbf{X}_{\theta'})^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}^2)^{\frac{-n-2}{2}} \prod_{j=1}^J m_j' (1-m_j') u_j'}{\{\det(\mathbf{X}_{\theta}' \mathbf{X}_{\theta})^{-1}\}^{\frac{1}{2}} ((n-2)\hat{\sigma}^2)^{\frac{-n-2}{2}} \prod_{j=1}^J m_j (1-m_j) u_j}$$

Accepteer de update en noem $\theta' = \theta$ met $\min(1, A)$.

- (4) Herhaal stappen (2)-(3) een l aantal keer, met l naar eigen inzicht in te vullen.

Bovenstaand proces levert een $\theta = (J, u_1, u_2, \dots, u_J, \xi_1, \xi_2, \dots, \xi_J)$ (met $\xi_j = (v, m_j)$), waarop we een een functie μ_{θ}^0 baseren, gedefinieerd als zijnde:

$$\mu_{\theta}^0(x) = \sum_{j=1}^J \frac{u_j}{\sum_{i=1}^J u_j} F(x, \xi_j)$$

Maak de design-matrix \mathbf{X}_θ met de berekende $\mu_\theta^0(x)$ die uit het bovenstaande proces is gekomen. De design matrix wordt

$$\mathbf{X}_\theta = \begin{bmatrix} 1 & \mu_\theta^0(x_1) \\ \vdots & \vdots \\ 1 & \mu_\theta^0(x_n) \end{bmatrix}$$

Kies n_0 , σ_0 , $\boldsymbol{\mu}_0$ en \mathbf{C}_0 naar eigen inzicht. Trek $\boldsymbol{\beta}$ uit de verdeling:

$$\boldsymbol{\beta} | \mathbf{y} \sim N(\boldsymbol{\mu}_1, \frac{n_1}{n_1 - 2} \sigma_1^2 \mathbf{C}_1^{-1})$$

Met:

$$\boldsymbol{\mu}_1 = \mathbf{C}_1^{-1}(\mathbf{C}_0 \boldsymbol{\mu}_0 + \mathbf{X}_\theta' \mathbf{y})$$

$$\mathbf{C}_1 = \mathbf{C}_0 + \mathbf{X}_\theta' \mathbf{X}_\theta$$

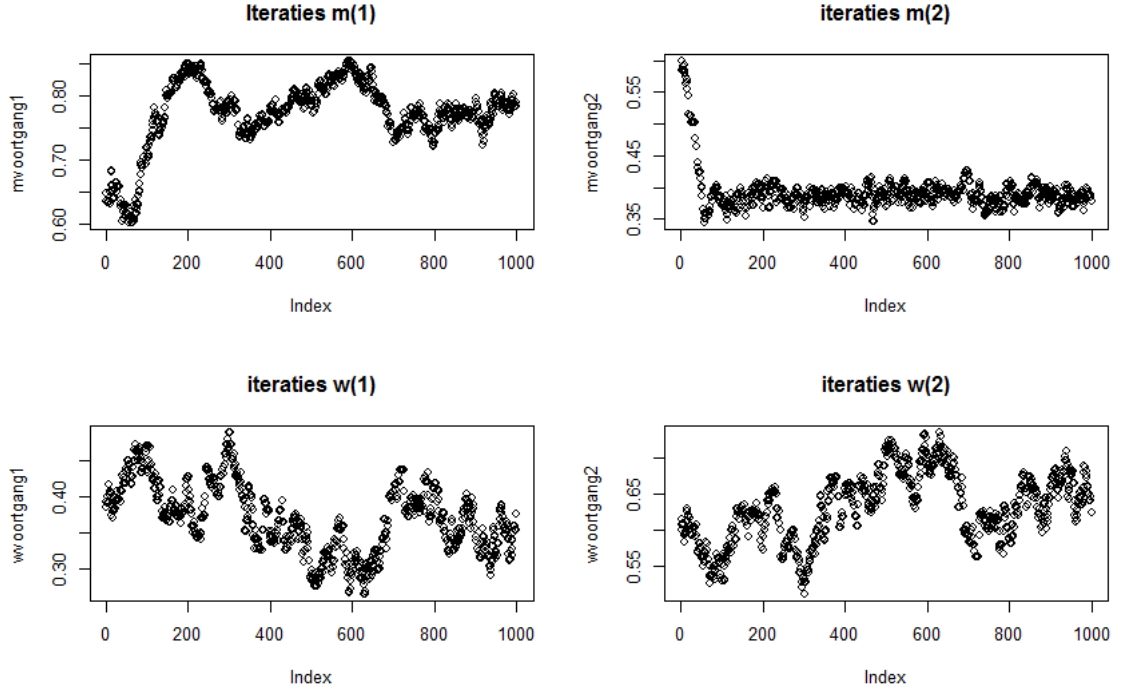
$$n_1 = n + n_0$$

$$n_1 \sigma_1^2 = n_0 \sigma_0^2 + (\mathbf{y} - \mathbf{X} \boldsymbol{\mu}_1)' \mathbf{y} + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)' \mathbf{C}_0 \boldsymbol{\mu}_0$$

Als de prior gegeven door n_0 , σ_0 , $\boldsymbol{\mu}_0$ en \mathbf{C}_0 ver afwijkt van de waardes die gebruikt zijn om de data te creëren volgens $y_i = \beta_0^* + \beta_1^* \mu^{0*}(x_i) + N(0, \sigma^{2*})$, zal er een grotere sample size nodig zijn om goede trekkingen te krijgen. Dat will zeggen, trekkingen die gelijken op de waardes waarmee de data gecreëerd is.

Nu kan $\mu(x) = \beta_0 + \beta_1 \mu_\theta^0(x)$ berekend worden. Deze kan vergeleken worden met onze data \mathbf{y} of liever met $\beta_0^* + \beta_1^* \mu^{0*}(x)$ (omdat we aannemen dat de data is opgebouwd als $y_i = \beta_0^* + \beta_1^* \mu^{0*}(x_i) + N(0, \sigma^{2*})$).

4.5. **Simulaties.** In een eerste test situatie hebben we gewerkt met $J = 2$ en $v = 10$. Er is een dataset ter grootte van 10^3 genomen. De data is gesimuleerd met $\mathbf{m} = (\frac{4}{5}, \frac{2}{5})$ en $\mathbf{w} = (\frac{1}{3}, \frac{2}{3})$. Er zijn 10^3 iteraties van het algoritme uitgevoerd. Dit gaf de volgende resultaten:



De waarden lijken vrij goed te convergeren naar de echte waarden. We hebben de simulatie nog een paar kee uitgevoerd en met dezelfde \mathbf{m} en \mathbf{w} , en het aantal iteraties van het algoritme en de v gevarieerd. Vervolgens hebben we van de iteraties vanaf iteratie 200 het gemiddelde gepakt. Dit gaf de volgende resultaten:

$v = 10, 10^3$ iteraties	$\mathbf{m} = (0,79, 0,37), \mathbf{w} = (0,38, 0,62)$
$v = 10, 10^4$ iteraties	$\mathbf{m} = (0,77, 0,39), \mathbf{w} = (0,34, 0,66)$
$v = 30, 10^3$ iteraties	$\mathbf{m} = (0,80, 0,40), \mathbf{w} = (0,33, 0,67)$
$v = 30, 10^4$ iteraties	$\mathbf{m} = (0,82, 0,40), \mathbf{w} = (0,35, 0,65)$

Deze resultaten zijn min of meer wat we verwachtten. Meer iteraties zorgt ervoor dat de gemiddeldes die we vinden voor \mathbf{m} en \mathbf{w} dichter bij de echte waarden liggen, en de keuze van v lijkt niet bijzonder veel uit te maken.

Nu beschouwen we de situatie met grotere J . We nemen $J = 10$. We zouden weer plaatjes van de w_j en m_j uitgezet tegen het aantal iteraties kunnen maken. Daarom komen dan echter wel wat problemen bij kijken. Omdat w_j en m_j begrensd zijn tussen 0 en 1, en zeker omdat de w_j ook nog eens tot 1 moeten optellen, is het lastig om te zien of de berekende waarden al rond één van de vooraf gestelde waarden schommelen, of nog maar willekeurig ergens heen gaan. Verder is het zo dat omdat de waarden best dicht op elkaar liggen, de m_j en w_j op bepaalde iteraties onderling kunnen wisselen. Stel bijvoorbeeld dat de berekende m'_1 in eerste instantie naar de vooraf gestelde $m_1 = 0,5$ convergeert, en een berekende m'_2 naar een vooraf gestelde $m_2 = 0,6$. Dan kan het gebeuren dat de m'_1 na een bepaald aantal iteraties bij toeval wat doorschiet richting 0,6, en de m'_2 in reactie weer naar 0,5 schiet.

Met $J = 10$ is de kans redelijk groot dat het eerder beschrevene met enige regelmaat optreedt. Daarom besluiten we niet aan de hand van grafieken te proberen af te lezen of onze methode werkt. We kunnen wel de w_j en m_j gewoon berekenen en de hierop gebaseerde $\mu^0(x)$ en dus $\mu(x)$ berekenen, en vergelijken met de werkelijke $\mu^0(x)$ en $\mu(x)$. Dan weten we of het eindresultaat goed is, en dat is waar we het uiteindelijk voor doen.

We nemen $J = 10$ en laten 10^4 iteraties van het eerder opgestelde Metropolis Hastings algoritme voor het bepalen van θ lopen. Voor het genereren van de data hebben we \mathbf{w} en \mathbf{m} gekozen als:

$$\mathbf{w} = (0, 18, 0, 16, \dots, 0, 04, 0, 02)'$$

$$\mathbf{m} = (1, \frac{9}{10}, \dots, \frac{2}{10}, \frac{1}{10})'$$

Vervolgens hebben we deze op basis van de data proberen te benaderen. Op basis van deze benaderingen hebben we een $\mu^0(x)$ opgesteld. Deze $\mu^0(x)$ hebben we in de design matrix X verwerkt, en hiermee konden we β bepalen en dus $\mu(x)$. Voor het berekenen van de β hebben we dezelfde techniek gebruikt als in het lineaire model; hierin werden eerst een aantal waarden (te weten n_0 , σ_0 , μ_0 en C_0) a priori vastgelegd. Net zoals bij het lineaire model, geldt ook hier dat als deze waarden a priori zó worden gekozen dat de prior sterk afwijkt van de waarden waarmee de data is gegenereerd, een grotere sample size nodig is om te compenseren. Omdat we met een sample size ter grootte 10^3 gewerkt hebben, zijn we dus weer behoorlijk vrij in de keuze van de a priori variabelen (zie onderschrift v.d. plots voor de genomen waarden).

Vervolgens trekken we β uit de verdeling:

$$\beta | \mathbf{y} \sim N(\boldsymbol{\mu}_1, \frac{n_1}{n_1 - 2} \sigma_1^2 C_1^{-1})$$

Met:

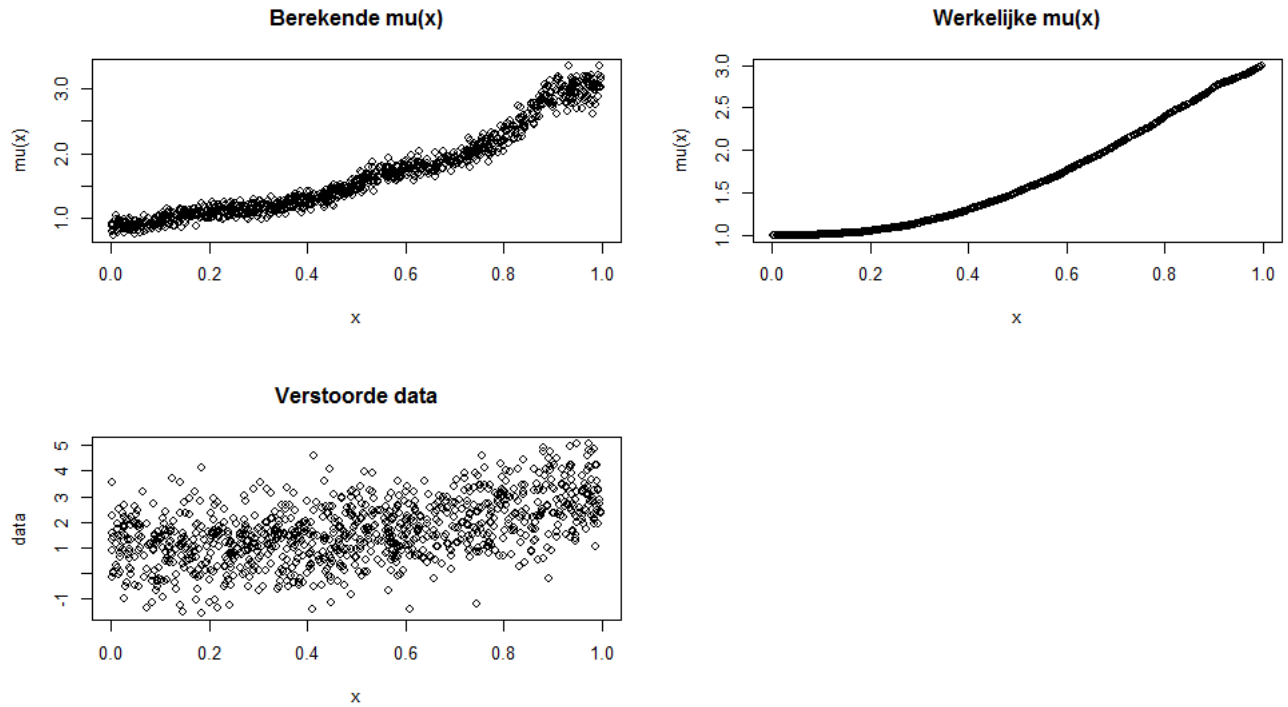
$$\boldsymbol{\mu}_1 = C_1^{-1}(C_0 \boldsymbol{\mu}_0 + \mathbf{X}' \mathbf{y})$$

$$C_1 = C_0 + \mathbf{X}' \mathbf{X}$$

$$n_1 = n + n_0$$

$$n_1 \sigma_1^2 = n_0 \sigma_0^2 + (\mathbf{y} - \mathbf{X} \boldsymbol{\mu}_1)' \mathbf{y} + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)' C_0 \boldsymbol{\mu}_0$$

We hebben de volgende resultaten verkregen:



FIGUUR 10. Linksboven: de door ons berekende $\mu_i(x_i)$. Parameters voor de berekende μ_i zijn: $n_0 = 10$, $\boldsymbol{\mu}_0 = (7, 12)$, $\sigma_0 = 8$, $\mathbf{C}_0 = 2 \times I$ (I de identiteitsmatrix). Sample size: $n = 10^3$. v is vastgezet op 10. Aantal geplote resultaten: 10^3 . Rechtsboven, de daadwerkelijke $\mu(x_i)$, met de \boldsymbol{w} en \boldsymbol{m} zoals eerder gespecificeerd. v vastgezet op 10. $\beta_0 = 1$ en $\beta_1 = 2$. Linksonder: De data \boldsymbol{y} gegeven door $y_i = \mu(x_i) + \epsilon_i$.

In de weergegeven grafiek lijkt de door ons berekende $\mu(x)$ in ieder geval al redelijk in de buurt te komen van de werkelijke regressie functie. We kijken iets beter naar de resultaten en vergelijken verschillende waarden van de door ons berekende $\mu(x)$ met wat de waarde zou moeten zijn in die punten:

	berekende $\mu(x)$	$\beta_0^* + \beta_1^* \mu^{0*}(x_i)$
$x = 0,28$	1,06	1,10
$x = 0,48$	1,47	1,45
$x = 0,77$	2,14	2,27

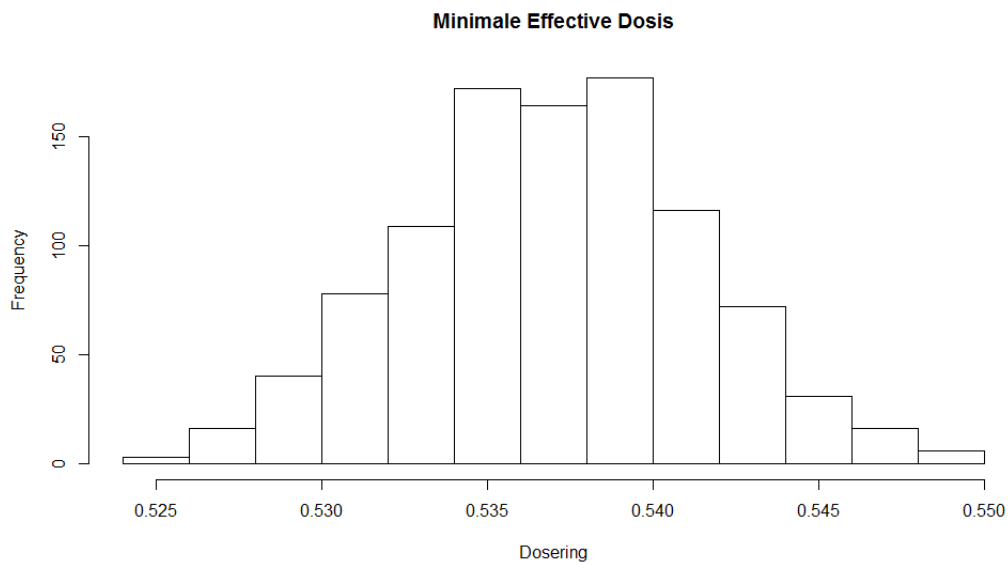
We zien, niet alleen in de punten die hier afgebeeld zijn, maar ook in andere punten die we bekijken, dat rond $x = 0.5$ onze benadering van $\mu(x)$ redelijk goed is, maar dat hij minder is in de buurt van 0 en 1. De verklaring hiervoor ligt in de definitie van $F(x_i, \boldsymbol{\xi}_j)$. Als de x_i boven/onder de m_j ligt, gaat de berekende waarde snel naar 1/0. Dit heeft als gevolg dat onze schatting bij een x_i die dicht in de buurt van 0 of 1 ligt, sneller afwijkt van de gewenste waarde.

Nu we de $\mu_i(x_i)$ hebben berekend, willen we nog de laatste stap maken naar de

berekening van de MED. We schrijven

$$\begin{aligned} & \mu(x) \geq \mu(0) + \Delta \\ \iff & \beta_0 + \beta_1 \mu^0(x) \geq \beta_0 + \beta_1 \mu^0(0) + \Delta \\ \iff & \mu^0(x) \geq \frac{\Delta}{\beta_1} \end{aligned}$$

We hebben de functie $\mu^0(x)$ opgesteld, Δ is een door ons zelf te bepalen waarde en we hebben $\beta_{1,i}$ gegenereerd. We kunnen dus x_i bepalen als schatting voor de MED. Dit heeft ons de volgende resultaten geleverd:



FIGUUR 11. MED bij $\Delta = \frac{1}{2}$

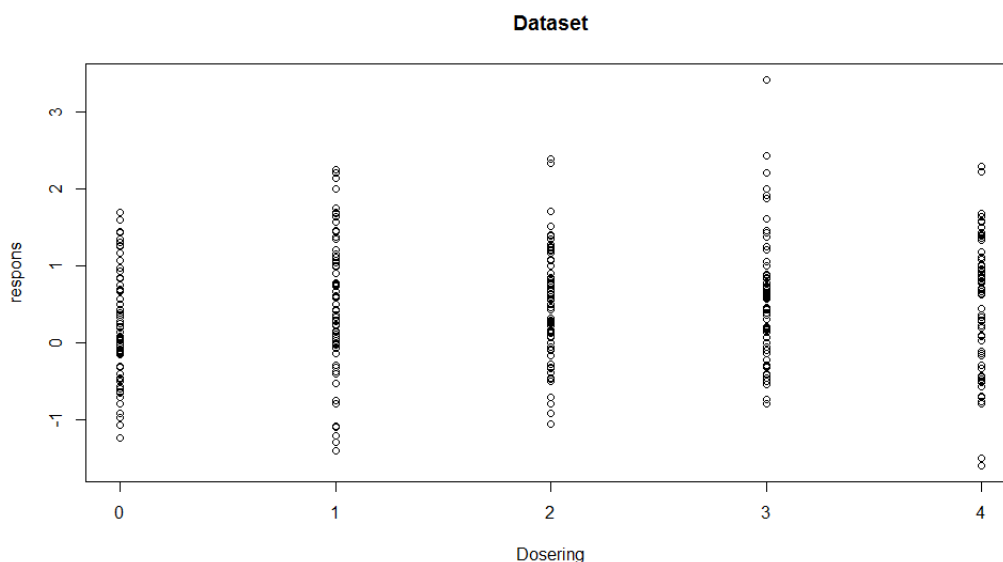
Het gemiddelde blijkt op 0,53 te liggen. We gaan na dat met de door ons geconstrueerde $\mu^0(x)$ geldt dat $\mu^0(0,53) = 0,26$ (afgerond). Dit ligt in lijn met onze verwachtingen, want we hebben $\Delta = 0,5$ en $\beta_1 = 2$ gezet, wat dus $\frac{\Delta}{\beta_1}$ op 0,25 zou doen uitkomen.

5. WERKELIJKE DATASET

5.1. **Dataset.** Tot nu toe zijn de modellen alleen nog maar getest met zelf gemaakte data. Data die zo gemaakt was dat er makkelijk kon worden nagegaan of onze methode goed werkte. Nu bekend is hoe goed onze methode werkt, is het interessant om hem ook toe te passen op een daadwerkelijke dataset.

De gebruikte dataset is de IBS dataset uit de MCPMod package in R. Deze dataset is een dose-response analyse waarbij vijf verschillende doses zijn gebruikt, te weten doseringen 0, 1, 2, 3 en 4. Er zijn 369 metingen gedaan bij verschillende doseringen, en de respons is gegeven als een getal tussen $-1,6$ en $3,4$. De data ziet er als volgt uit:

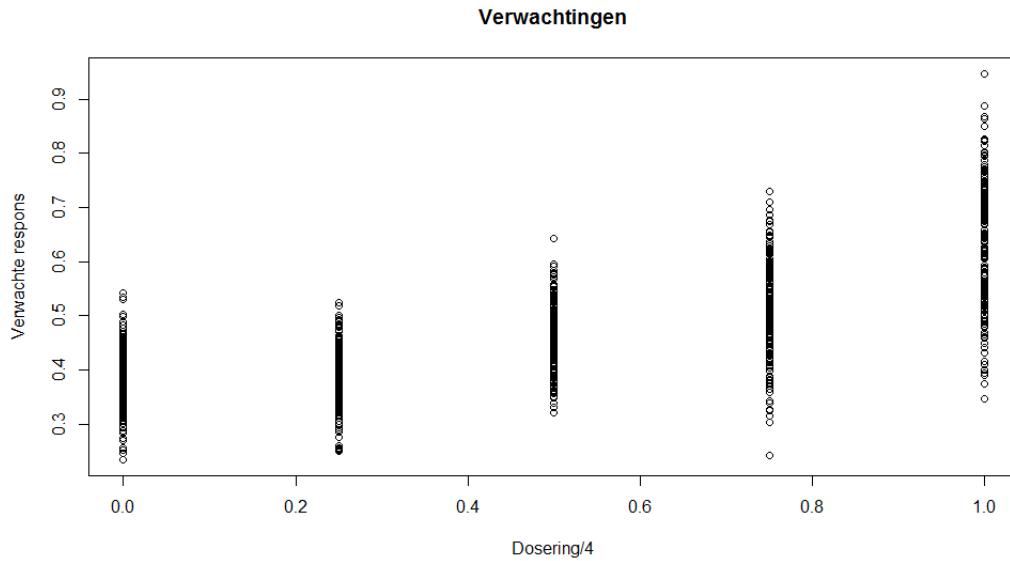
Voordat we onze methode toepassen op deze nieuwe dataset moeten we wel



FIGUUR 12. Dataset bestaande uit 369 respons metingen bij een totaal van vijf verschillende doseringen

even opletten. In onze methode hebben we $\mu(x_i) = \beta_0 + \beta_1 \mu^0(x_i)$ gesteld, met $\mu^0(x_i) = \sum_{j=1}^J w_j F(x_i, \xi_j)$. Hierin was $F(x_i, \xi_j)$ een functie afhankelijk van of x_i groter of kleiner was dan de $m_j \in \xi_j$. Hierin waren de m_j begrensd tussen 0 en 1, onder de aanname dat x_i dat ook was. De doseringen van 0, 1, 2, 3 en 4 moeten dus veranderd worden in $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ en 1.

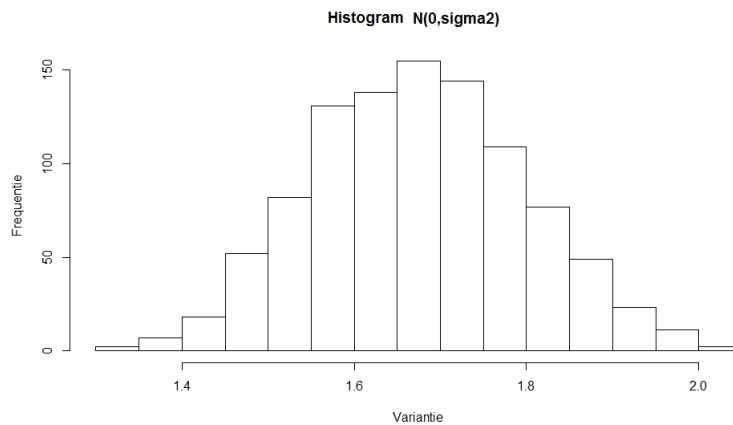
5.2. **Toepassen Niet-parametrisch Model.** We noemen onze x_i de verschillende doseringen $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ en 1, en verbinden daaraan onze y_i bestaande uit de verschillende respons-waardes. We nemen aan dat voor de y_i geldt dat $y_i = \mu(x_i) + \epsilon_i$, met $\epsilon_i \sim N(0, \phi^{-1})$. Vervolgens verwerken we deze gegevens met de laatst beschreven methode waarin we stelden dat $\mu(x_i) = \beta_0 + \beta_1 \mu^0(x_i)$, met $\mu^0(x_i) = \sum_{j=1}^J w_j F(x_i, \xi_j)$ en $J = 10$. Hiermee gaan we nu verwachtingen $\mu_j(x_i)$ en de variantie $\phi_i^{-1} = \sigma_i^2$ berekenen.



FIGUUR 13. Gevonden verwachtingen bij de 369 doses uit de dataset

Deze uitkomst was enigszins te verwachten. De gemeten responses bij de data liggen verder uit elkaar dan bij de verwachtingen. Dit komt doordat we aannemen dat in de data nog een bepaalde error zit. De grootte van de variantie zal laten zien dat het dan dus logisch is dat de data gespreider is.

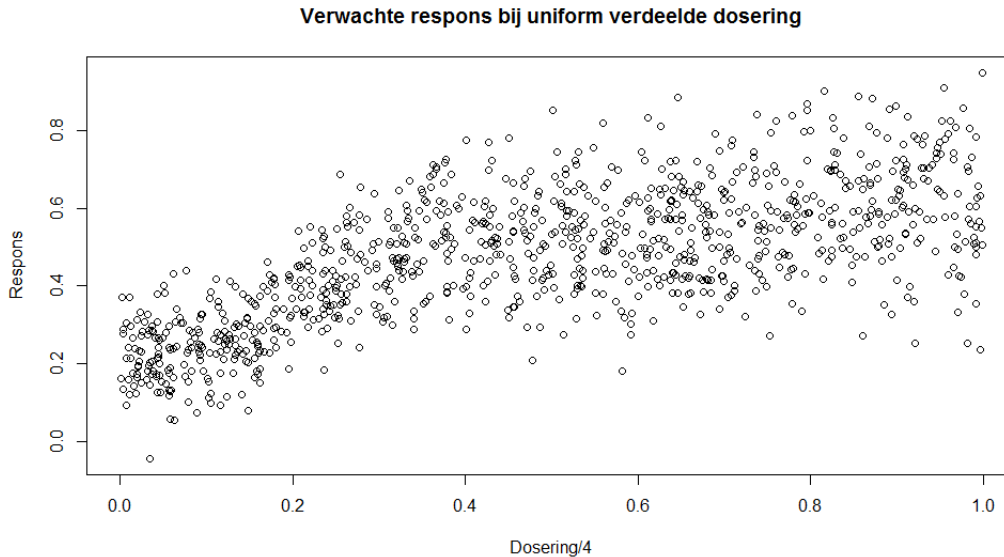
De gevonden varianties $\phi^{-1} = \sigma^2$ zien er als volgt uit:



FIGUUR 14. Berekende varianties

Het gemiddelde ligt op 1,70. Deze variantie lijkt redelijk in lijn te zijn met wat we zouden verwachten, gezien onze data en de door ons berekende $\mu_j(x_i)$; de data had een veel wijdere spreiding dan de $\mu_j(x_i)$, wat veroorzaakt wordt door de error $\epsilon_i \sim N(0, \phi^{-1})$.

Nu hebben we slechts verwachtinge geconstrueerd bij de doses uit de dataset. We gaan nu verwachtingen construeren voor doses uniform verspreid op het interval tussen $[0, 1]$, dus in plaats van de $\mu_j(x_i)$ te genereren voor $x_i \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ genereren we ze voor $x_i \in [0, 1]$. Dit geeft ons het volgende resultaat:



FIGUUR 15. Gegeneerde $\mu_j(x_i)$ bij 1000 gegeneerde x_i , met $x_i \in [0, 1]$.

Bij deze verwachtingen hoort nog steeds dezelfde variantie als eerder. De grafiek van de $\mu_j(x_i)$ bij de uniform verdeelde doseringen levert ons verder geen grote verrassingen. Wat we in deze plot wel duidelijk kunnen zien is dat de respons sterk stijgt tussen $x = 0,2$ en $x = 0,4$, en dat vanaf $x = 0,4$ er een grotere spreiding in de verwachtingen lijkt te zitten. Als we kijken naar de verwachtingen die we bij $x_i \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ hadden geconstrueerd, lijkt dat ook logisch. Daar zie je een grote verandering tussen de verwachtingen op $x = \frac{1}{4}$ en $x = \frac{1}{2}$, en het gelijkblijven, of misschien zelfs enigszins dalen, van de verwachtingen tussen $x = \frac{1}{2}$ en $x = \frac{3}{4}$ verklaart de grotere spreiding in dit gebied bij de uniform verdeelde x_i .

We moeten nog de laatste stap maken en de MED berekenen. Herhaling: de MED was gedefinieerd als

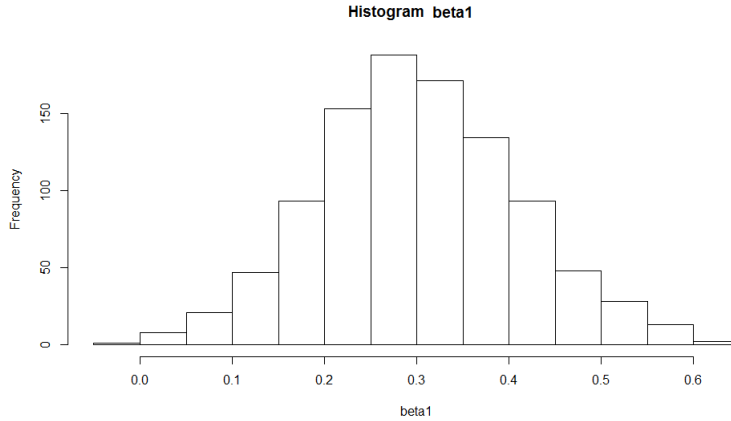
$$MED = \min\{x > 0 : \mu(x) \geq \mu(0) + \Delta\}$$

Merk op dat toen we zelf de data genereerden we $\beta_0 = 1$ en $\beta_1 = 2$ hadden gekozen, met als resultaat dat de door ons berekende $\mu_j(x_i)$ ongeveer tussen 1 en 3 zaten. Toen was de keuze van $\Delta = \frac{1}{2}$ dus best wel redelijk als verschil tussen de functiewaarde van μ in 0 en in x .

De verwachtingen $\mu_j(x_i)$ zitten nu echter tussen 0,2 en 0,8. Een waarde van $\frac{1}{2}$ voor Δ zou de x_i opzoeken die tegen 1 aan liggen. Besef ook dat we, zoals eerder uitgelegd, naar de x zoeken zo dat

$$\mu^0(x) \geq \frac{\Delta}{\beta_1}$$

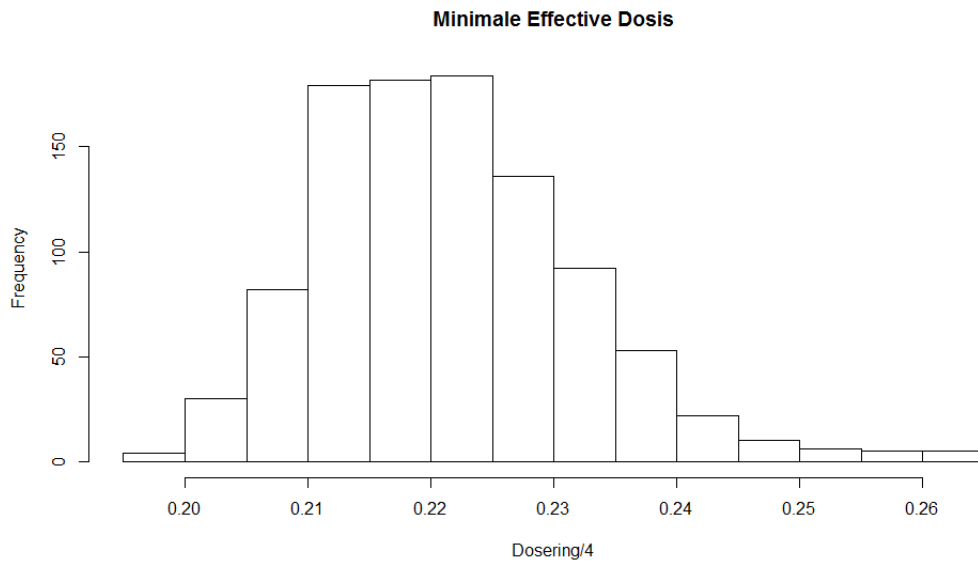
Als we de getrokken $\beta_{1,i}$ plotten krijgen we:



FIGUUR 16. Gegeneerde $\beta_{1,i}$

De $\mu^0(x)$ is wel begrensd tussen 0 en 1, en de keuze van $\Delta = 0,5$ zou met deze $\beta_{1,i}$ er vaak tot leiden dat $\frac{\Delta}{\beta_1} \geq 1$, wat ons model dus niet aan kan. Daarom kiezen we ervoor om de Δ te verlagen naar 0,05. Dan blijft $\frac{\Delta}{\beta_1}$ keurig, meestal tussen 0 en 1. Omdat de $\beta_{1,i}$ natuurlijk uiteindelijk toch uit een verdeling worden getrokken, hebben we wel een paar uitkomsten van $\frac{\Delta}{\beta_1}$ die buiten het interval $[0, 1]$ vallen (minder dan 1% van de berekende $\frac{\Delta}{\beta_1}$) in onze testopstelling. Deze verwijderen we voor we de MED gaan berekenen.

Nu kan de MED berekend worden. We vinden:



FIGUUR 17. De geschatte MED's bij $\Delta = 0,05$

Het gemiddelde ligt op 0,25. Het is ons dus gelukt om de MED te schatten met een zelf bepaalde Δ van 0,05.