



**RWS INFORMATIE**

## **Handreiking Bayesiaanse update**

Het aanpassen van faaldata op basis van metingen

Datum	6 september 2017
Status	Definitief (versie 1.0.1)



## Colofon

Naam Standaard	Handreiking Bayesiaanse Update
Beschrijving:	De Handreiking Bayesiaanse Update beschrijft hoe faaldata bijgewerkt kan worden op basis van metingen.
Status:	Definitief
Datum	6 september 2017
Versienummer:	1.0.1
Soort:	Handreiking
Verantwoordelijke PE:	Jean-Luc Beguin
Gebruik in proces:	OAM, AenO
Netwerk:	HVWN, HWS en HWN
Object:	Alle RWS-infrastructuur
Hoofdkennisveld:	Assetmanagement en Methodiek
Kennisveld:	Risicogestuurd Beheer en Onderhoud (RGOB)
Informatie:	<a href="mailto:Peter.blanker@rws.nl">Peter.blanker@rws.nl</a>
Verantwoordelijke afdeling:	RWS GPO – ICO
WW RWS Nummer:	# 5500

## Overzicht wijzigingen

Versie	Datum	Wijzigingen
0.9		Eerste versie aangeboden aan WWA0 (nu WW RWS)
1.0	20 dec 2016	Versie opgenomen in WWA0. Geen commentaar EG TM en EG CM welke tot inhoudelijke wijzigingen hebben geleid van versie 0.9
1.0.1	6 sep 2017	Uniformeringslag over alle ProBO (deel)methoden, geen inhoudelijke aanpassingen.



## Inhoud

<b>1</b>	<b>Inleiding 7</b>
1.1	Context 7
1.2	Doel 7
1.3	Gebruikersgroep 8
1.4	Koppeling aan proces 8
1.5	Relatie met andere kaders en handreikingen 8
1.6	Samenvatting en leeswijzer 8
<b>2</b>	<b>De regel van Bayes 10</b>
2.1	De regel van Bayes in een discrete kansruimte 10
2.2	De regel van Bayes voor kansverdelingen 11
<b>3</b>	<b>De Bayesiaanse update van faaldata 13</b>
3.1	Lognormale verdeling 13
3.1.1	Algemeen 13
3.1.2	De ligging van de mediaan 13
3.1.3	De ligging van overige grenzen 14
3.1.4	De errorfactor 14
3.1.5	De ligging van de modus, het maximum van de kansdichtheid 15
3.2	Velddata 15
3.2.1	De likelihood 15
3.3	De update 17
<b>4</b>	<b>Implementatie van de Bayesiaanse update 18</b>
4.1	Het vinden van het gemiddelde en de normalisatieconstante 18
4.2	Het vinden van de percentielen 19
4.3	Controles 20
4.3.1	A priori controles 20
4.3.2	A posteriori controles 20
4.4	Samenvatting 20
4.5	Tot slot 21
<b>5</b>	<b>Enkele speciale situaties 23</b>
5.1	Geen faalmomenten of geen tijdsduur 23
5.2	Velddata "gelijk" aan het gemiddelde van de generieke data 23
<b>Referenties 26</b>	
<b>Bijlage A De likelihoodfunctie 27</b>	



# 1 Inleiding

## 1.1 Context

In 2010 is door Rijkswaterstaat besloten tot het beheerst invoeren van risico-gestuurd beheer en onderhoud (RGO) binnen asset management (AM). Met RGO worden alle risico's voor het functioneren van een object in kaart gebracht, waardoor deze op een transparante en weloverwogen manier beheerst kunnen worden. Dit in tegenstelling tot traditioneel onderhoud dat veelal conditie-gestuurd is, gericht op het handhaven van een bepaald technisch niveau.

Het doel van RGO is om de risico's in het functioneren van de drie netwerken via beheer- en onderhoudsacties zodanig te beheersen, dat de afgesproken prestaties worden geleverd tegen minimale (levensduur)kosten. RGO maakt de relatie tussen de netwerkprestatie en onderhoud expliciet. In 2013 besloot het bestuur RWS tot een verdere doorontwikkeling van RGO om volledig in control te komen middels een vervolgtraject RGO, gevolgd door een herijking ervan in 2016.

Binnen Rijkswaterstaat is daartoe in 2016 de handreiking Prestatiegestuurde Risicoanalyses (PRA) opgesteld om het risicogestuurd denken toepasbaar te maken voor alle infrastructurele assets, die Rijkswaterstaat in beheer heeft. Deze handreiking integreert en vervangt daarmee de Leidraad RAMS en de Leidraad risicogestuurd beheer en onderhoud.

Prestatiegestuurde risicoanalyse (PRA) is een belangrijk instrument. De PRA brengt de balans in beeld tussen de prestaties van een object, de risico's die de prestaties beïnvloeden en de kosten van het in stand houden van de prestatie. Met hulp van PRA's kan Rijkswaterstaat onderbouwde beslissingen nemen bij aanleg, beheer en onderhoud.

In aanvulling op deze handreiking zijn verschillende methodes inhoudelijk verder uitgewerkt en vastgelegd in aparte handreikingen. Zo ook deze standaard die beschrijft hoe faaldata aan de hand van metingen aangepast kan worden.

## 1.2 Doel

Het voorspellen van de betrouwbaarheid en beschikbaarheid van een systeem gebeurt aan de hand van het uitvoeren van een (kwantitatieve) betrouwbaarheids- en beschikbaarheidsanalyse.

De manier waarop zo'n analyse moet worden uitgevoerd is vastgelegd in het kader "Verificatiemethode Betrouwbaarheid en Beschikbaarheid" [1]. In dat document wordt geëist dat een opdrachtnemer de kans op falen van de gebruikte of de ontworpen hardware kwantificeert. Daartoe worden suggesties gegeven van internationaal aanvaarde databases. Ook Rijkswaterstaat beheert een database waarin faaldata, specifiek geschikt voor zijn assets, zijn verzameld.

Maar de beste storingsdata om een object te beschrijven zijn de gegevens die zijn verzameld bij het object zelf. Soms kan op basis van uitsluitend deze gegevens een faalfrequentie worden bepaald met voldoende zekerheid. Vaak is echter onvoldoende objectspecifieke informatie voorhanden om een faalfrequentie te bepalen. In dat geval kan met behulp van het Bayesiaanse update proces de

faalkans die uit generieke databases afkomstig is, worden geüpdatet met de gemeten data.

De procedure waarmee de nieuwe, gemeten, objectspecifieke data verwerkt wordt in de initiële generieke data, het Bayesiaanse updateproces, wordt in deze handreiking beschreven. Bij de handreiking hoort een Excel Werkboek [2], waarmee de update wordt uitgevoerd.

### **1.3 Gebruikersgroep**

De procedure kan in twee processen gebruikt worden en heeft dus twee doelgroepen.

Ten eerste kunnen, met behulp van deze procedure, de door Rijkswaterstaat of zijn opdrachtnemers gemeten storingsdata worden verwerkt in de eerder genoemde faaldatabase. De (specialistische) beheerder van deze database is daarmee de gebruiker van de methode. Hij zorgt hiermee voor actuele data, die bij aanleg kan worden gebruikt.

Ten tweede kan de methode worden gebruikt in de gebruiksfase. Met name voor opdrachtnemers van Rijkswaterstaat die objectbeheerders zijn, bijvoorbeeld in een DBFM-contract, is het zinvol te eisen dat periodiek een update van de gebruikte faaldata wordt gemaakt. Indien onderdelen van een object relatief vaak, of juist weinig falen, is deze procedure een juridisch afdwingbare mogelijkheid om de (evident onjuiste) aangenomen storingsverwachtingen aan te passen. In dit geval zal de contractant niet alleen zijn storingen meten en bijhouden, hij zal ook periodiek deze data moeten verwerken tot een nieuwe storingsverwachting en daarmee zijn risicoanalyse moeten aanpassen. Deze procedure kan natuurlijk ook gebruikt worden door de faalkansmanagers van Rijkswaterstaat zelf, om faaldata van een specifiek object aan te passen aan de metingen.

### **1.4 Koppeling aan proces**

Deze handreiking valt in het proces Aanleg en Onderhoud, deelproces Ontwerp, effecten en techniek. Activiteitencluster Opstellen Systeemspecificatie (eisen, ontwerp en ontwerpnotitie).

### **1.5 Relatie met andere kaders en handreikingen**

Deze handreiking heeft een directe relatie met de handreiking Prestatiegestuurde Risico Analyses [3]. Hierin wordt nut en gebruik van Bayesiaanse updaten in de bredere context van risicogestuurd onderhouden geplaatst.

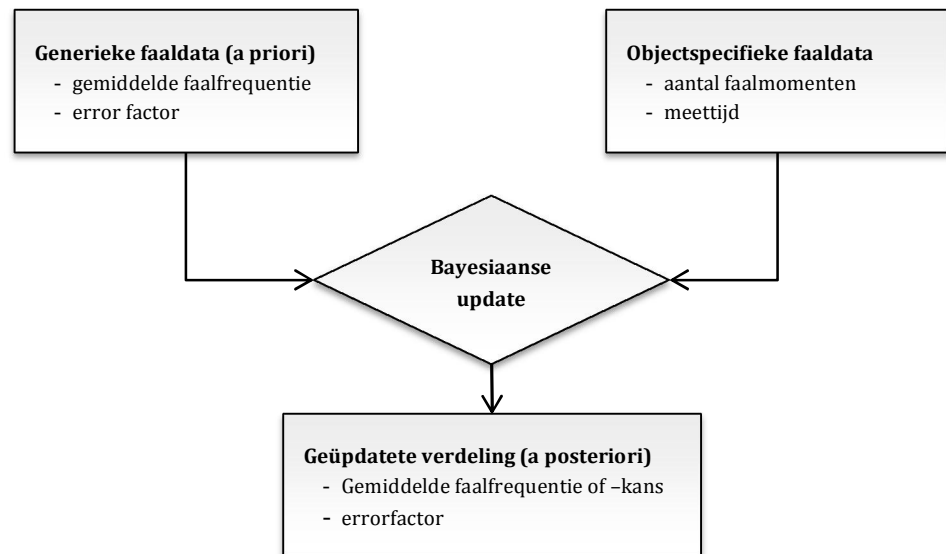
Deze handreiking is van toepassing op alle onderdelen van het HWN, HVWN, HWS waarvoor kwantitatieve eisen gesteld zijn of moeten worden in termen van betrouwbaarheid en beschikbaarheid.

### **1.6 Samenvatting en leeswijzer**

De verdeling van generieke faalfrequentie voor vergelijkbare objecten wordt gecombineerd met de objectspecifieke gegevens. De verdeling van de generieke faalfrequentie (de a-priori verdeling) wordt dan geüpdatet met het aantal opgetreden storingen in een bepaalde periode.

Op deze manier ontstaat de a-posteriori verdeling (zie Figuur 1).





Figuur 1 Principe van de Bayesiaanse update methodiek

Dit document legt de theoretische achtergrond van het combineren van generieke faaldata met velddata met behulp van het Bayesiaanse updateproces vast. Het Bayesiaans update-proces is geïmplementeerd in een Excel Werkboek [2] dat als hulpmiddel kan worden gebruikt om generieke faaldata en velddata te combineren.

De algemene theorie van de Bayesiaanse update wordt uiteengezet. De theorie wordt uitgelegd aan de hand van discrete kansen en vervolgens voor kansverdelingen. Daarna wordt deze theorie verder uitgewerkt specifiek voor faaldata, door het gebruik van een lognormale verdeling voor generieke faaldata en de Poisson verdeling voor eigen waarnemingen. Tot slot wordt gekeken naar situaties waarbij een Bayesiaanse update gedaan wordt met velddata van nul faalmomenten of nul uur en naar het niet intuïtieve resultaat bij het updaten van een verdeling met velddata, waarvan het gemiddelde gelijk is aan het gemiddelde van de generieke data.

De uitgelegde implementatie voor de update van faaldata is grotendeels gelijk aan de wijze waarop dit in het Excel Werkboek wordt gedaan. In dit document wordt aangegeven waar de verschillende onderdelen van de implementatie in het Excel Werkboek zijn te vinden.

In hoofdstukken 2 en 3 wordt de theorie uiteengezet. In hoofdstuk 4 wordt de implementatie uiteengezet. De beschrijving van de implementatie van het Bayesiaanse updateproces in dat hoofdstuk komt overeen met hoe dit in de macro in het eerdergenoemde Excel Werkboek is toegepast.

Tot slot worden enkele valkuilen bij het gebruik van de Bayesiaanse update belicht.

## 2 De regel van Bayes

In dit hoofdstuk wordt de regel van Bayes uitgelegd, als voorbereiding op de toepassing van deze regel op faaldata. Eerst zal de regel van Bayes worden afgeleid in een discrete kansruimte vanwege het intuïtieve karakter van de afleiding. Vervolgens wordt de regel van Bayes toegepast voor kansverdelingen.

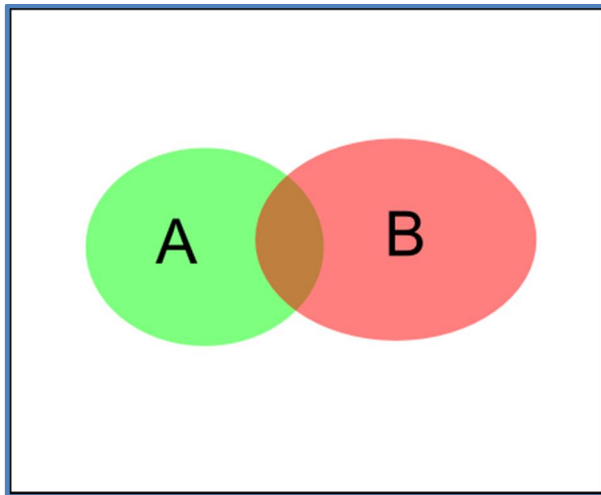
### 2.1 De regel van Bayes in een discrete kansruimte

Een kansexperiment kent verschillende uitkomsten, gebeurtenissen genaamd. De verschillende gebeurtenissen hebben elk een kans van optreden. Soms kan de informatie van de ene gebeurtenis van invloed zijn op de kans op de andere gebeurtenis. Men spreekt dan van afhankelijkheid.

Zie het Venn diagram in

Figuur 2, waar een willekeurig punt binnen het vierkant vier opties heeft:

- Alleen binnen ovaal A: alleen gebeurtenis A treedt op
- Alleen binnen ovaal B: alleen gebeurtenis B treedt op
- In de overlap van A en B ( $A \cap B$ ): beide gebeurtenissen treden op
- Buiten A en buiten B: gebeurtenissen A en B treden beide niet op



Figuur 2 Venn diagram met verzamelingen A en B

Uit de figuur volgt dat de kans dat, als de informatie aanwezig is dat een willekeurig punt zich in A bevindt, het zich ook in B bevindt gelijk is aan het deel van A dat door B overlapt wordt, gedeeld door de oppervlakte van A. Met andere woorden:

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

Het combineren van de uitdrukkingen, gebruik makende van het feit dat  $P(B \cap A)$  gelijk is aan  $P(A \cap B)$  geeft de regel van Bayes:

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

Oftewel, de kans dat een willekeurig punt zich in A bevindt, als bekend is dat het punt zich in B bevindt, is gelijk aan de kans dat een punt zich in B bevindt als al bekend is dat die zich in A bevindt, vermenigvuldigd met het quotiënt van de kans op A en de kans op B.

## 2.2 De regel van Bayes voor kansverdelingen

Omdat de kans op A de te bepalen kans is en de kans op B een constante, kan de regel van Bayes ook geschreven worden als een afhankelijkheid met een constante  $\alpha$  die overeenkomt met  $1/P(B)$ . Voor een opdeling van A in verschillende klassen  $A_i$  geldt nu:

$$P(A_i|B) = \alpha \cdot P(B|A_i)P(A_i)$$

Voor het Bayesiaans updaten van kansverdelingen is het nodig om een kansverdeling  $f(A)$  met eigen observaties B te updaten. Vervangen van een kans  $P(x)$  door een kansverdeling  $f(x)$  levert:

$$f(A|B) = \alpha \cdot f(B|A)f(A)$$

De kansverdeling van B gegeven A is evenredig aan de waarschijnlijkheid of likelihood van A gegeven B. Zie ook Bijlage A. Vanwege de overgang op de likelihood wordt in onderstaande formule  $\alpha$  vervangen door  $\alpha'$ . De likelihood is hier ook te interpreteren als "hoe waarschijnlijk is het dat mijn observatie (B) uit kansverdeling A komt".

$$f(A|B) = \alpha' \cdot L(A|B)f(A)$$

Hier is  $L(A|B)$  de aannemelijkheid, of *likelihood*, van A gegeven B. De likelihood geeft aan hoe goed de parameters A passen bij het resultaat B. In het kader van het updaten van kansverdelingen is  $f(A)$  de a-priori verdeling van A en  $f(A|B)$  de kansverdeling van A gegeven B, oftewel de a-posteriori verdeling van A. Voor een update rest nog het bepalen van de constante  $\alpha'$ . Zoals altijd is de integraal van de kansverdeling per definitie 1, als over het gehele bereik van de functie geïntegreerd wordt. Dus

$$\int f(A|B)dA = \int \alpha' \cdot L(A|B)f(A)dA = 1$$

$$\alpha' = \frac{1}{\int_0^\infty L(A|B)f(A)dA}$$

Dit geeft de Bayesiaanse update van een generieke kansverdeling  $f(A)$  met velddata B:

$$f(A|B) = \frac{L(A|B)f(A)}{\int_0^\infty L(A|B)f(A)dA}$$

Deze interpretatie geeft een uitleg van de regel van Bayes:

*De kansverdeling na update is de a priori kansverdeling, gecorrigeerd voor de waarnemingen. Dit wordt gedaan door in de teller de a priori kansverdeling te vermenigvuldigen met de likelihood functie van de updatedata, de noemer kan geïnterpreteerd worden als een normalisatie.*

### 3 De Bayesiaanse update van faaldata

In dit hoofdstuk wordt de gevonden uitdrukking voor de Bayesiaanse update uitgewerkt voor de toepassing van de update op faaldata. Hiervoor worden de gebruikte kansverdeling voor  $f(A)$  en de likelihood functie  $L(A|B)$  en hun eigenschappen toegelicht.

#### 3.1 Lognormale verdeling

In het algemeen wordt de lognormale verdeling beschouwd als een goede manier om de spreiding van de gemiddelde faalkansen te beschrijven. Voor de generieke of a priori data wordt de lognormale verdeling gebruikt.

##### 3.1.1 Algemeen

De lognormale verdeling is de verdeling van een variabele waarvan het logaritme normaal verdeeld is: als  $x$  normaal verdeeld is, is  $e^x$  lognormaal verdeeld. De verdeling heeft, net als de normaalverdeling, twee parameters. De parameters  $M$  en  $S$  zijn het gemiddelde en de spreiding van de onderliggende normale verdeling.  $\mu$  ( $\mu$ ) en sigma ( $\sigma$ ) worden ook vaak gebruikt maar om verwarring te voorkomen worden deze hier niet gebruikt. Met de parameters  $M$  en  $S$  is de dichtheidsfunctie van de lognormale verdeling gegeven als [4]:

$$f(x) = \frac{1}{xS\sqrt{2\pi}} e^{\left[-(\ln x - M)^2 / 2S^2\right]}$$

En het gemiddelde en de variantie als:

$$\text{Gem} = e^{(M+S^2/2)}$$

$$\text{Var} = e^{2M+2S^2}(e^{S^2} - 1) = \text{Gem}^2(e^{S^2} - 1)$$

En  $M$  en  $S$ :

$$M = \ln(\text{Gem}) - \frac{S^2}{2}$$

$$S = \sqrt{\ln\left(1 + \frac{\text{Var}}{\text{Gem}^2}\right)}$$

De cumulatieve kansverdeling wordt gegeven door

$$D(x) = \int_0^x f(x') dx' = \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{\ln x - M}{S\sqrt{2}}\right) \right]$$

Met  $\text{erf}$  de errorfunctie:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz$$

##### 3.1.2 De ligging van de mediaan

De mediaan  $x_{0.5}$  ligt zodanig dat de helft van de waarde van de kansdichtheid links van de mediaan ligt:

$$\int_0^{x_{0.5}} P(x') dx' = \frac{1}{2}$$

Invullen van de cumulatieve dichtheidsfunctie geeft:

$$\frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\ln x_{0.5} - M}{S\sqrt{2}} \right) \right] = \frac{1}{2}$$

$$\operatorname{erf} \left( \frac{\ln x_{0.5} - M}{S\sqrt{2}} \right) = 0$$

De inverse errorfunctie is enkel numeriek te benaderen. En  $\operatorname{erf}(0)=0$ .

$$\frac{\ln x_{0.5} - M}{S\sqrt{2}} = 0$$

$$x_{0.5} = e^M$$

### 3.1.3

*De ligging van overige grenzen*

Op dezelfde manier als de mediaan wordt deze gegeven (met  $Z$  het percentiel) door

$$\frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\ln x - M}{S\sqrt{2}} \right) \right] = Z$$

$$\operatorname{erf} \left( \frac{\ln x - M}{S\sqrt{2}} \right) = 2Z - 1$$

met of  $Z=0.05$ :

$$\operatorname{erf} \left( \frac{\ln x_{0.05} - M}{S\sqrt{2}} \right) = -0.9$$

Dit numeriek benaderen geeft:

$$\frac{\ln x_{0.05} - M}{S\sqrt{2}} \approx -1.163$$

$$\frac{\ln x_{0.05} - M}{S} \approx -1.645$$

$$x_{0.05} \approx e^{-1.645 \cdot S + M}$$

En met  $Z=0.95$

$$\operatorname{erf} \left( \frac{\ln x_{0.95} - M}{S\sqrt{2}} \right) = 0.9$$

$$\frac{\ln x_{0.95} - M}{S\sqrt{2}} \approx 1.163$$

$$x_{0.95} \approx e^{1.645 \cdot S + M}$$

### 3.1.4

*De errorfactor*

Om de onzekerheid van de faalfrequentie (de spreiding van de mogelijke faalfrequenties) eenvoudig te beschrijven, wordt bij faaldata-analyse vaak gebruik gemaakt van de errorfactor. Deze is gedefinieerd als de wortel uit de verhouding tussen de 95% en 5% grenzen van een kansverdeling en aangeduid met EF:

$$EF = \sqrt{\frac{x_{0.95}}{x_{0.05}}}$$

De EF kan uiteraard niet kleiner zijn dan 1. Hoe dichter bij 1, hoe "smaller" de verdeling waarmee de waarde nauwkeuriger bekend is. Een heel hoge errorfactor geeft aan dat de verdeling erg 'vlak' is en dus weinig informatie bevat.

### 3.1.5 *De ligging van de modus, het maximum van de kansdichtheid*

Het maximum van een functie ligt op punt  $x$  zodat geldt<sup>1</sup>:

$$\begin{aligned}\frac{df(x)}{dx} &= \frac{d}{dx} \left\{ \frac{1}{xS\sqrt{2\pi}} e^{\left[-(\ln x - M)^2 / 2S^2\right]} \right\} = 0 \\ \frac{df(x)}{dx} &= \frac{d}{dx} \left\{ \frac{1}{xS\sqrt{2\pi}} \right\} e^{\left[-(\ln x - M)^2 / 2S^2\right]} + \frac{1}{xS\sqrt{2\pi}} \frac{d}{dx} \left\{ e^{\left[-(\ln x - M)^2 / 2S^2\right]} \right\} \\ \frac{df(x)}{dx} &= \frac{-1}{x^2 S\sqrt{2\pi}} e^{\left[-(\ln x - M)^2 / 2S^2\right]} + \frac{1}{xS\sqrt{2\pi}} \cdot \frac{\ln x - M}{-xS^2} e^{\left[-(\ln x - M)^2 / 2S^2\right]} = 0 \\ \frac{1}{x^2 S\sqrt{2\pi}} e^{\left[-(\ln x - M)^2 / 2S^2\right]} &= \frac{1}{xS\sqrt{2\pi}} \cdot \frac{\ln x - M}{-xS^2} e^{\left[-(\ln x - M)^2 / 2S^2\right]} \\ 1 &= \frac{\ln x - M}{-S^2} \\ M - S^2 &= \ln x \\ x_{\text{mode}} &= e^{M - S^2}\end{aligned}$$

## 3.2 **Velddata**

Voor een discreet proces als het falen van componenten wordt de Poisson verdeling gebruikt. De Poisson verdeling wordt beschreven door een enkele parameter  $\alpha$ , de verwachte waarde. Om de link met faaldata te leggen, wordt het verwachte aantal failures genoteerd als  $\lambda t$  (faalfrequentie x tijdsduur). De kansdichtheid voor  $k$  faalmomenten wordt gegeven door [5]:

$$P(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Deze kansverdeling geeft de kans op  $k$  faalmomenten, gegeven een verwacht aantal faalmomenten van  $\lambda t$ .

### 3.2.1 *De likelihood*

In deze sectie zal de likelihood functie voor dit probleem afgeleid worden. Een intuïtieve afleiding kan gevonden worden in Bijlage A. De likelihood van  $\lambda$  is de waarschijnlijkheid van de waarde van  $\lambda$ , gegeven dat  $k$  faalmomenten en

<sup>1</sup> Het gevonden punt kan in het geval van een lognormale verdeling geen minimum zijn.

tijdsinterval  $t$  waargenomen zijn. Als er  $k_i$  faalmomenten zijn opgetreden in tijdsinterval  $t_i$ , wordt de likelihood van  $\lambda$  gegeven door:

$$L(\lambda|t, k) = \prod_{i=1}^N \frac{(\lambda t_i)^{k_i} e^{-\lambda t_i}}{k_i!}$$

$$L(\lambda|t, k) = \frac{(\prod t_i)^{\sum k_i} \lambda^{\sum k_i} e^{-\lambda \sum t_i}}{\prod k_i!}$$

De som van alle faalmomenten,  $\sum k_i$ , is het totaal aantal faalmomenten en zal worden aangeduid met  $N$ . De som van alle tijdsintervallen,  $\sum t_i$ , is de totale meettijd  $T$ . De likelihood wordt nu:

$$L(\lambda|t, k) = \frac{(\prod t_i)^N \lambda^N e^{-\lambda T}}{\prod k_i!}$$

Beide producten hangen niet af van  $\lambda$  en worden dus samengenomen in een constante. Daarnaast wordt de variabele  $\lambda$  vanaf nu met  $x$  aangeduid. De likelihood van gemiddelde faalfrequentie  $x$ , gegeven het aantal faalmomenten  $N$  in tijd  $T$  wordt dus:

$$L(x|N, T) = \beta x^N e^{-xT}$$

Met

$$\beta = \frac{(\prod t_i)^N}{\prod k_i!}$$

De top van de likelihood is het meest waarschijnlijke punt (de meest waarschijnlijke faalkans) en dit ligt op punt  $x$  waarvoor geldt dat (grafisch blijkt dat dit een top is):

$$\frac{\partial L(x)}{\partial x} = 0$$

$$\frac{d[\beta x^N e^{-xT}]}{dx} = \frac{d[x^N]}{dx} \beta e^{-xT} + \beta x^N \frac{d[e^{-xT}]}{d[-xT]} \frac{d[-xT]}{dx} = 0$$

$$\beta N x^{N-1} e^{-xT} - \beta x^N e^{-xT} T = 0$$

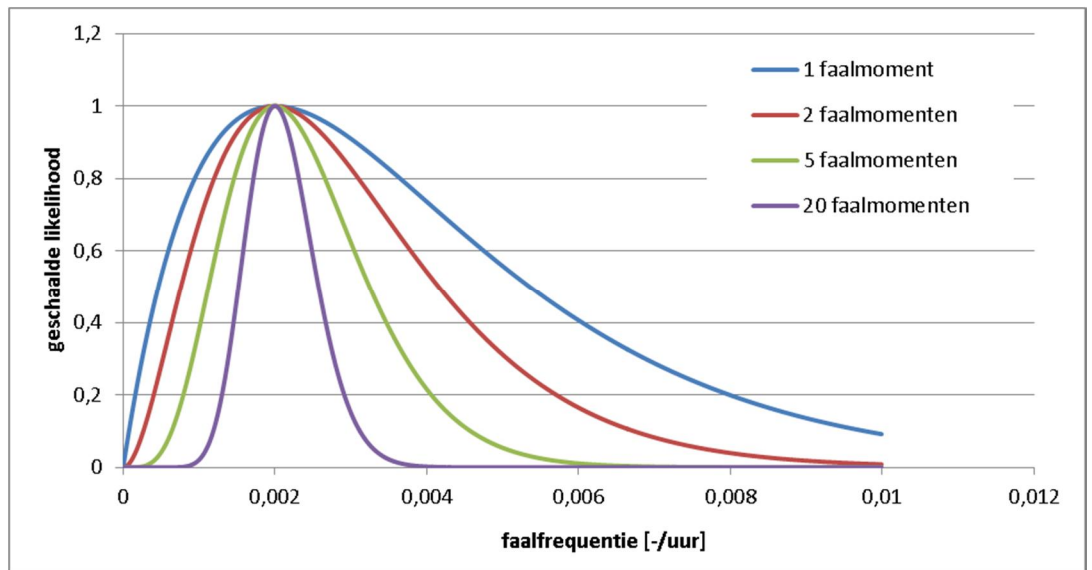
$$\beta N x^{N-1} e^{-xT} = \beta x^N e^{-xT} T$$

$$N x^{-1} = T$$

$$x = \frac{N}{T}$$

Met andere woorden: de meest waarschijnlijke faalkans is het aantal faalmomenten gedeeld door de meettijd. De likelihood functie geeft aan hoe waarschijnlijk een andere faalkans zou zijn.





Figuur 3 De (geschaalde) likelihood van de gemiddelde faalfrequentie bij N faalmomenten in N\*500 uur.

Figuur 3 Geeft de likelihood van de gemiddelde faalfrequentie bij N faalmomenten in N \* 500 uur als fractie van de maximale likelihood. Te zien is dat de meest waarschijnlijke waarde voor de faalfrequentie  $2 \cdot 10^{-3}$  per uur is, maar dat  $4 \cdot 10^{-3}$  per uur relatief ook een behoorlijke kans maakt bij twee of minder waargenomen faalmomenten.

### 3.3

#### De update

Voor de verdeling van de generieke data wordt de lognormale verdeling aangenomen. Deze zal enkel met  $f_{\text{prior}}$  worden aangeduid. De verdeling na update, de a posteriori verdeling  $f_{\text{post}}$ , wordt nu gegeven door:

$$f_{\text{post}}(x) = \frac{\beta x^N e^{-xT} f_{\text{prior}}(x)}{\int \beta x^N e^{-xT} f_{\text{prior}}(x) dx}$$

$$f_{\text{post}}(x) = \frac{x^N e^{-xT} f_{\text{prior}}(x)}{\int x^N e^{-xT} f_{\text{prior}}(x) dx}$$

Doordat hier een lognormale verdeling wordt vermenigvuldigd met  $x^N e^{-xT}$ , is de a posteriori verdeling geen lognormale verdeling.

## 4 Implementatie van de Bayesiaanse update

In dit hoofdstuk wordt de implementatie van de theorie behandeld. Tijdens het uitvoeren van de update wordt de nieuwe verdeling berekend. Vervolgens worden hiervan de 95% boven- en 5% ondergrens bepaald.

De koppeling naar de implementatie in het Excel Werkboek is aangegeven in de tabellen aan het einde van iedere sectie.

### 4.1 Het vinden van het gemiddelde en de normalisatieconstante

De Bayesiaanse update is gegeven door:

$$f_{\text{post}}(x) = \frac{x^N e^{-xT} f_{\text{prior}}(x)}{\int x^N e^{-xT} f_{\text{prior}}(x) dx}$$

De noemer van deze deling is constant en zal in het vervolg met  $C$  aangeduid worden. Deze noemer wordt berekend volgens:

$$C = \int_0^{x_{\max}} x^N e^{-xT} \cdot f_{\text{prior}}(x) dx$$

Omdat niet geïntegreerd kan worden tot  $x = \infty$  wordt voor  $x_{\max}$  een factor ( $\gamma \cdot EF$ ) maal de 95% grens gekozen:

$$x_{\max} = \gamma EF \cdot x_{0.95} = \gamma EF_{\text{prior}}^2 e^{\text{Gem}_{\text{prior}}}$$

Waarbij  $\gamma$  instelbaar is maar standaard waarde 2 heeft en  $\text{gem}_{\text{prior}}$  is het gemiddelde van de a priori kansverdeling. Een integraal moet, voor een numerieke berekening, immers een bovengrens kleiner dan oneindig hebben. Als de bovengrens  $x_{\max}$  zo gedefinieerd wordt, wordt de meest relevante data meegenomen. Aangezien de verwachtingswaarde van de a posteriori verdeling gedefinieerd wordt als:

$$E[f_{\text{post}}] = \int_0^{\infty} x f_{\text{post}}(x) dx$$

Is het gemiddelde te vinden als:

$$\text{Gem}[f_{\text{post}}] = \frac{1}{C} \int_0^{x_{\max}} x^{N+1} e^{-xT} \cdot f_{\text{prior}}(x) dx$$

Deze integraal en die voor  $C$  kunnen met ieder gewenst algoritme opgelost worden. In het Excel Werkboek is gekozen voor een variatie op het Romberg- algoritme [6]. In grote lijnen werkt dit algoritme als volgt. Het gehele bereik wordt opgedeeld in  $n$  intervallen, op basis van deze opdeling wordt een Riemann som<sup>2</sup> uitgerekend. Vervolgens wordt het interval opgedeeld in  $3n$  intervallen en met een nieuwe Riemann som wordt de integraal opnieuw berekend. Als het verschil tussen de twee sommen onder een bepaald getal, het convergentie criterium of tolerantie, komt, wordt deze waarde voor de integraal geaccepteerd. Merk op dat een tolerantie van  $10^{-6}$  dus niet betekent dat de echte oplossing binnen de  $10^{-6}$  van het berekende

<sup>2</sup> Een methode die de oppervlakte onder een kromme benadert als een serie rechthoeken en de oppervlakken van die rechthoeken gebruikt om de integraal te benaderen.

antwoord ligt. Deze tolerantie is aan te passen. Om te voorkomen dat dit proces oneindig lang doorgaat is er een maximaal aantal stappen van standaard 15 ingesteld. De verwerking hiervan is bijna letterlijk terug te vinden in het Excel Werkboek, zie Tabel 1.

Tabel 1. Implementatie van paragraaf 4.1 in het Excel Werkboek

omschrijving	implementatie plaats en methode	opmerkingen
de berekening van C	in de macro: functie 'func'	variabele 'functie'=2
de andere integraal voor het gemiddelde	in de macro: functie 'func'	variabele 'functie'=1
de integralen uit deze sectie	in de macro:functie 'midpnt' en subroutine 'calc_mean'	midpnt berekent de integraal calc_mean kapt hem af
maximaal aantal stappen	op sheet 'parameters'	
tolerantie	op sheet 'parameters'	ook wel afkapcriterium
$\gamma$ , voor de berekening van $X_{\max}$	op sheet 'parameters'	

## 4.2 Het vinden van de percentielen

De z-percentielen van de a posteriori verdeling zijn gegeven door:

$$\int_0^{x_z} f_{\text{post}}(x) dx = z$$

Dus bijvoorbeeld voor de ligging van  $x_{0.05}$ , het 5% percentiel (de 5% ondergrens):

$$\int_0^{x_{0.05}} f_{\text{post}}(x) dx = 0.05$$

Omdat het eerder gebruikte integratiealgoritme niet bruikbaar is, is deze integraal geïmplementeerd als een Riemann som met stapgrootte  $h$ :

$$h \sum_{j=0}^J f_{\text{post}}(x_j) = 0.05$$

$$x_{0.05} = Jh$$

Met:

$$x_j = h \cdot j$$

$$h = \text{gem}[f_{\text{post}}] \cdot 10^\eta$$

Standaard is  $\eta$  gelijk aan -6 maar dit kan aangepast worden. Hier wordt de waarde van de a posteriori verdeling uitgerekend voor  $h$ ,  $2h$  etc. Zo wordt een cumulatieve verdeling gemaakt. Zodra deze 0.05 is, is de 5% grens bereikt, evenzo voor de 50% en 95% grenzen. Ook de berekeningen in deze sectie zijn vrijwel letterlijk terug te vinden in het Excel Werkboek, zie Tabel 2.

Tabel 2. Implementatie van paragraaf 4.2 in het Excel Werkboek

omschrijving	implementatie plaats en methode	opmerkingen
de berekening van $f_{\text{post}}$	in de macro: functie 'func'	variabele 'functie'=1, vervolgens moet gedeeld worden door C, maar die is al bekend
de integraal	in de macro: subroutine 'calc_lim'	
$\eta$ , voor de berekening van de stapgrootte	op sheet 'parameters'	

### 4.3 Controles

#### 4.3.1 *A priori controles*

Voordat de update berekend wordt, wordt gecontroleerd of de generieke data van toepassing kan zijn op de ingevoerde velddata. Hiervoor wordt gecontroleerd of beide verdelingen genoeg overlappen. En gecontroleerd wordt of de verdeling 'informatie' bevat en bestaat, de errorfactor moet tussen 1 en de 30 liggen. Genoeg overlappen is gedefinieerd als: de 95% bovengrens van de generieke data ligt boven de 5% ondergrens van de velddata als het gemiddelde van de velddata hoger ligt, en omgekeerd. De ondergrens van de velddata wordt bepaald op basis van de  $\chi$ -kwadraat verdeling:

$$x_{0.05} = \frac{\chi^2(2N, 0.05)}{2T}$$

En de bovengrens is gedefinieerd als

$$x_{0.95} = \frac{\chi^2(2N + 2, 0.95)}{2T}$$

#### 4.3.2 *A posteriori controles*

Allereerst moet, de a posteriori verdeling bestaan: de 95% grens moet boven de 5% grens van de verdeling liggen. Vervolgens wordt de berekening van de integralen beschouwd. Zo kan het voorkomen dat een integraal niet convergeert: het verschil tussen twee opeenvolgende ordes wordt niet kleiner. Zelfs als dit wel het geval is, kan het zijn dat de integraal zo traag convergeert dat de iteratie-limiet bereikt wordt voordat een oplossing voldoet aan het convergentie- criterium. Deze controles worden uitgevoerd op beide integralen. De oplossing wordt getoond, eventueel met een foutmelding.

### 4.4 Samenvatting

De berekende integralen voor de a posteriori verdeling zijn:

$$\text{Gem}[f_{\text{post}}] = \frac{1}{C} \int_0^{x_{\text{max}}} x^{N+1} e^{-x \cdot T} \cdot f_{\text{prior}}(x) dx$$

$$C = \int_0^{x_{\text{max}}} x^N e^{-x \cdot T} \cdot f_{\text{prior}}(x) dx$$

Met

$$x_{\max} = \gamma EF \cdot P_{0.95} = \gamma EF_{\text{input}}^2 e^{\text{Gem}_{\text{input}}}$$

En voor de Z-percentielen:

$$h \sum_{i=0}^J f_{\text{post}}(x_i) = Z$$

$$x_z = Jh$$

Met

$$x_i = h \cdot i$$

$$h = \text{Gem}[f_{\text{post}}] \cdot 10^n$$

Voor de implementatielocatie van het bovenstaande zie paragraaf 4.1 en 4.2. De vergelijkingen zijn bijna letterlijk terug te vinden in de macro van het Excel Werkboek. De volgende berekeningsparameters kunnen veranderd worden vanuit de spreadsheet:

- De bovengrens van de integratie. Een verhoging van de bovengrens (door het aanpassen van  $\gamma$ ) kan leiden tot een beter resultaat maar zorgt er ook voor dat de berekeningen veel langer duren. Dit zal voornamelijk helpen bij vlakke verdelingen (hoge errorfactoren).
- Het maximale aantal stappen.. Verhoging van het maximale aantal stappen kan een langdurige berekening tot gevolg hebben en helpt alleen als de integraal niet snel genoeg convergeert.
- De tolerantie: de maximale relatieve verbetering van de integraal. Een lagere waarde geeft een beter gemiddelde maar het helpt doorgaans weinig.
- De stapgrootte: de stapgrootte voor het bepalen van de grenzen (door het aanpassen van  $\eta$ ) zal voornamelijk helpen bij spitse verdelingen (lage errorfactoren).

## 4.5

### Tot slot

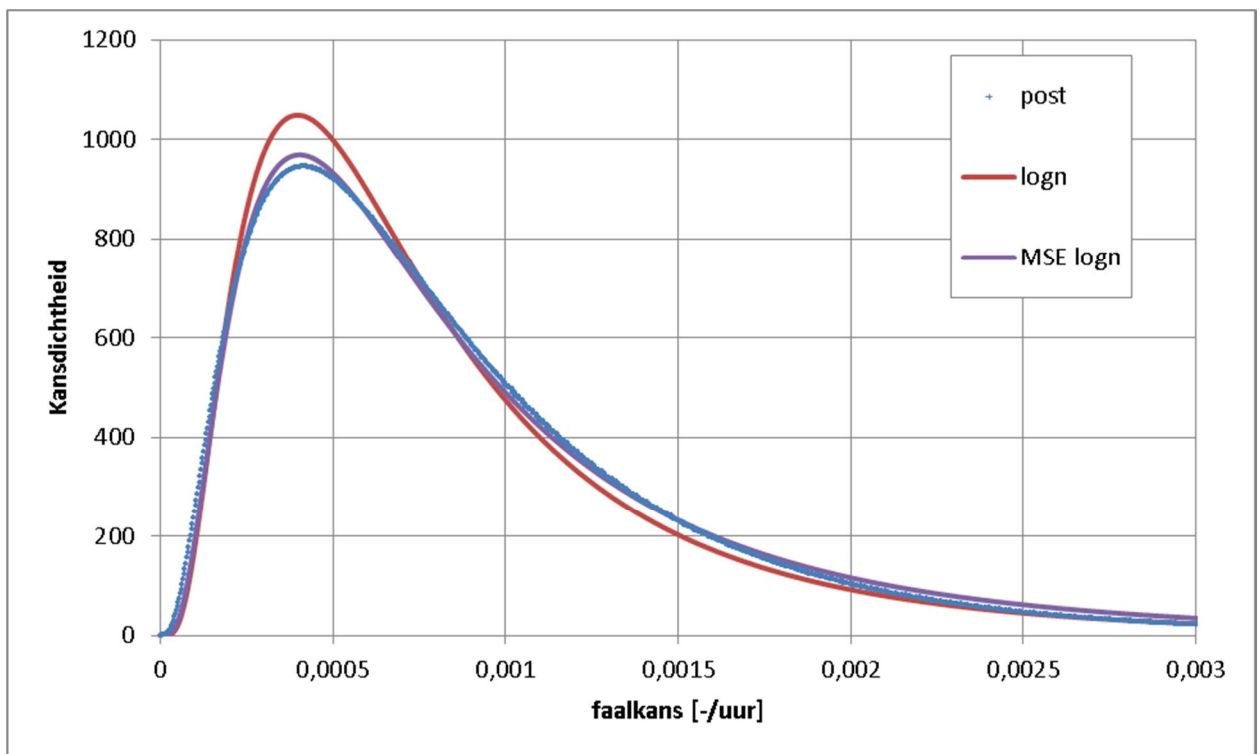
Zoals genoemd is de uitkomst van de update geen lognormale verdeling. Het resultaat is een verdeling die eventueel benaderd kan worden met een lognormale verdeling. De uitkomst van de berekening, het gemiddelde en de grenzen, kan dus gebruikt worden om een lognormale verdeling te beschrijven. Deze verdeling zal de data echter niet perfect beschrijven.

Er zijn veel mogelijkheden om een fit te maken van een lognormale verdeling op de a-posteriori data. Enkele voor de hand liggende lognormale verdelingen worden gevonden door de volgende combinaties van gegevens gelijk te houden aan de a posteriori data:

1. Gemiddelde en error factor
2. Gemiddelde en 95% grens
3. Gemiddelde en 5% grens
4. Gemiddelde en mediaan

Daarnaast zijn er nog veel manieren om een fit te maken, bijvoorbeeld op basis van de ongewogen minimum squared error.

In Figuur 4 geven de blauwe kruisjes de datapunten van een geüpdatete verdeling weer. Daarnaast zijn twee lognormale verdelingen opgenomen: de rode lijn geeft de lognormale verdeling op basis van het gemiddelde en de error factor, de paarse lijn geeft de 'beste' lognormale verdeling op basis van de ongewogen minimum squared error. De paarse lijn is niet te genereren op basis van de data in het Excel Werkboek. Deze is berekend met een code die niet in het Excel Werkboek beschikbaar is. Ook deze fit is niet geheel correct; er is dus geen lognormale verdeling die de data exact beschrijft.



Figuur 4 Twee 'fits' van de output data: de kruisjes geven de output data weer, de rode lijn geeft de lognormale verdeling op basis van het gemiddelde en de EF bepaald uit de 5 en 95% grenzen, de paarse lijn geeft de 'beste' lognormale verdeling op basis van de minimum squared error

## 5 Enkele speciale situaties

### 5.1 Geen faalmomenten of geen tijdsduur

Stel: willekeurige faaldata wordt geüpdatet met de volgende velddata:

- $N$  faalmomenten in 0 uur;
- 0 faalmomenten in  $T$  uur;
- 0 faalmomenten in 0 uur.

$N$  faalmomenten in 0 uur is onmogelijk, de berekening kan niet.

0 faalmomenten in  $T$  uur is mogelijk. Dit is alleen numeriek lastig: het maximum van de likelihood ligt op de plek waar de lognormale verdeling 0 is ( $x = 0$ ). Hierdoor worden de waarden van de integralen erg laag. De numerieke benadering kan dan worden gevonden door het aantal integratiestappen te vergroten.

0 faalmomenten in 0 uur wordt niet berekend want dit is geen update.

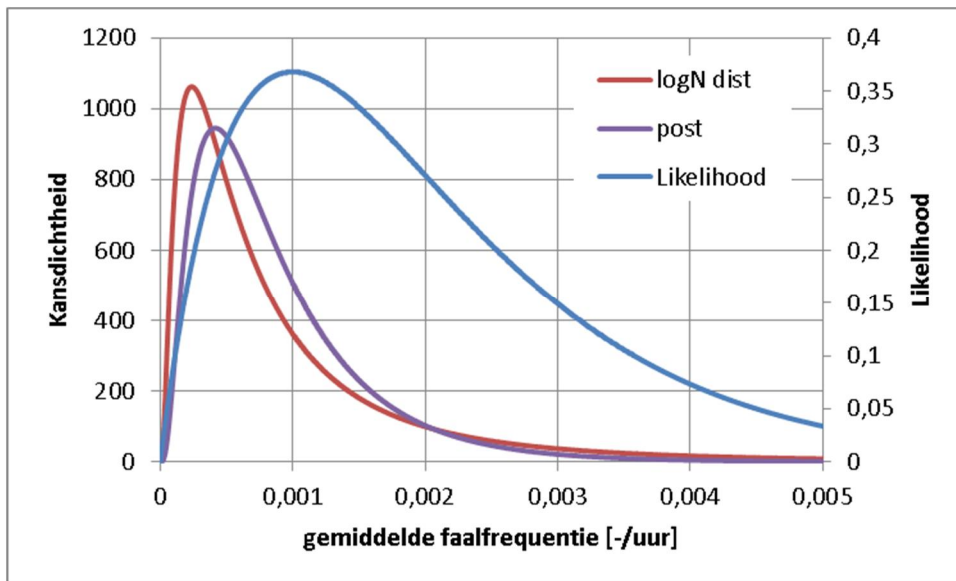
### 5.2 Velddata "gelijk" aan het gemiddelde van de generieke data

Bijvoorbeeld: een update van een a priori verdeling met een gemiddelde  $10^{-3}$  /uur en  $EF = 5$  met 1 failure in 1000 uur geeft een a posteriori verdeling met een gemiddelde faalfrequentie van  $8,9 \cdot 10^{-4}$  /uur en  $EF = 3,3$ . Wellicht verwacht men een a posteriori gemiddelde van  $10^{-3}$  /uur. Waardoor is de berekende waarde lager?

In Figuur 5 worden de verdeling van de generieke data ('logN dist') en de likelihood functie voor dit probleem getoond. Merk op dat de modus van de lognormale verdeling van de generieke data lager ligt dan het gemiddelde.

Tijdens de Bayesiaanse update worden beide verdelingen elementsgewijs met elkaar vermenigvuldigd<sup>3</sup>, zie Figuur 5. Te zien is dat de a posteriori ('post') verdeling 'spitser' is dan de a priori verdeling ('logN dist') en dat de staart 'dunner' is. Na  $x = 0.002$  is de a posteriori lager dan de a priori verdeling. Dit blijkt te compenseren voor de verschuiving van de modus naar rechts waardoor het gemiddelde uiteindelijk lager uitkomt. Kort uitgedrukt: de waargenomen velddata maken de hoge waarden van de a priori verdeling minder waarschijnlijk, waardoor het gemiddelde iets lager wordt.

<sup>3</sup> Dus  $f(x) = f_1(x) \cdot f_2(x)$  voor elke waarde van  $x$



Figuur 5 Het resultaat ('post') van de update van de functies uit sectie 5.2, de a posteriori lijn beschrijft de daadwerkelijke datapunten; de waarde van de functies is niet van belang, enkel de vorm



**Lijst van tabellen**

Tabel 1. Implementatie van paragraaf 4.1 in het Excel Werkboek	19
Tabel 2. Implementatie van paragraaf 4.2 in het Excel Werkboek	20

**Lijst van figuren**

Figuur 1	Principe van de Bayesiaanse update methodiek	9
Figuur 2	Venn diagram met verzamelingen <i>A</i> en <i>B</i>	10
Figuur 3	De (geschaalde) likelihood van de gemiddelde faalfrequentie bij <i>N</i> faalmomenten in <i>N</i> *500 uur.	17
Figuur 4	Twee 'fits' van de output data: de kruisjes geven de output data weer, de rode lijn geeft de lognormale verdeling op basis van het gemiddelde en de EF bepaald uit de 5 en 95% grenzen, de blauwe lijn geeft de 'beste' lognormale verdeling op basis van de minimum squared error	22
Figuur 5	Het resultaat ('post') van de update van de functies uit sectie 5.2, de a posteriori lijn beschrijft de daadwerkelijke datapunten; de waarde van de functies is niet van belang, enkel de vorm	24

## Referenties

- [1] RWS  
*Verificatiemethode Betrouwbaarheid en Beschikbaarheid, versie 1.0.7 (6 september 2017)*  
WW RWS-nummer 1567
  
- [2] RWS  
*RWS Bayesian Update 20 december 2016, Product bij Handreiking Bayesiaanse update, definitief, versie 1.0*  
Excel spreadsheet dd. 20-12-2016
  
- [3] RWS  
*Handreiking Prestatiegestuurde Risico Analyses, versie 1.0.0 (september 2016)*  
WW RWS-nummer 5333
  
- [4] Weisstein, E. (sd)  
*Log Normal Distribution*  
MathWorld--A Wolfram Web Resource [Online]:  
<http://mathworld.wolfram.com/LogNormalDistribution.html>  
Gebruikt op 14-10-2014
  
- [5] Weisstein, E. (sd)  
*Exponential Distribution*  
MathWorld--A Wolfram Web Resource [Online]:  
<http://mathworld.wolfram.com/ExponentialDistribution.html>  
Gebruikt op 14-10-2014
  
- [6] Weisstein, E. (sd)  
*Romberg Integration*  
MathWorld--A Wolfram Web Resource [Online]:  
<http://mathworld.wolfram.com/RombergIntegration.html>  
Gebruikt op 21-11-2014

## Bijlage A De likelihoodfunctie

Hierna wordt op een eenvoudige manier uitgelegd wat de likelihoodfunctie uit velddata voorstelt.

In paragraaf 3.2 is aangegeven dat de kans op een bepaald aantal failures in een tijdsduur bij een constante gemiddelde faalfrequentie bepaald kan worden met het Poisson proces.

De kans op  $k$  faalmomenten in een tijdsduur  $t$  wordt gegeven door:

$$P(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

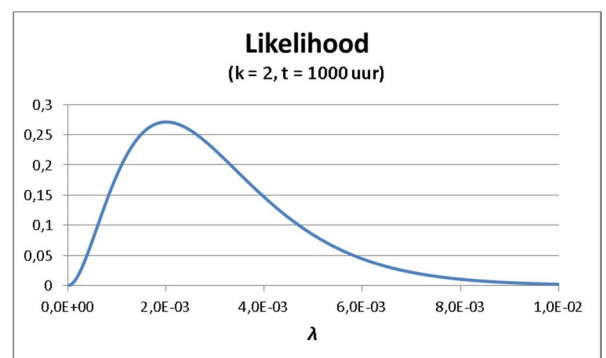
Met:  $\lambda$  = gemiddelde faalfrequentie.  
 $t$  = tijdsduur

Vanuit de velddata liggen  $k$  en  $t$  vast. Voor elke  $\lambda$  kan nu worden bepaald wat de waarschijnlijkheid (likelihood) is dat precies  $k$  failures optreden in tijdsduur  $t$ .

Als  $k = 2$  en  $t = 1000$  uur volgt een functie die de waarschijnlijkheid van  $\lambda$  beschrijft voor 2 failures in 1000 uur:

$$L(\lambda) = \frac{(1000 \cdot \lambda)^2 e^{-1000 \cdot \lambda}}{2!}$$

$\lambda$ [1/uur]	$L$
0,001	0,1839
0,002	0,2707
0,003	0,2240
0,004	0,1465



Uit de tabel en de grafiek blijkt:

- Bij 2 failures in 1000 uur is de waarschijnlijkheid dat de onderliggende  $\lambda$  0,003/uur is ongeveer 82% van de waarschijnlijkheid dat  $\lambda$  0,002/uur is.
- De meest waarschijnlijke waarde voor  $\lambda$  is 0,002/uur. Dit is gelijk aan  $(k/t)$ . Dit is niet het gemiddelde van alle mogelijke waarden voor  $\lambda$ .

Alleen de relatieve waarden van de likelihood functie zijn van belang, bijvoorbeeld: op plaats 1 is de likelihood twee keer zo groot als op plaats 2. Daarom is het niet interessant om te weten welke constanten er in de likelihood functie voorkomen. De likelihood functie:

$$L(\lambda) = \frac{(1000 \cdot \lambda)^2 e^{-1000 \cdot \lambda}}{2!} = \frac{1000^2 \lambda^2 e^{-1000 \cdot \lambda}}{2!}$$

Is daarom dus ook te schrijven als:

$$L(\lambda) = \lambda^2 e^{-1000 \cdot \lambda}$$

Of, gebruik makend van  $k$  en  $t$ :

$$L(\lambda) = \lambda^k e^{-t \cdot \lambda}$$