Aantekeningen bij het smoothing algoritme van het AG

W.J. Willemse

2 november 2006

Inleiding

Deze tekst geeft enkele aantekeningen die ik gemaakt heb bij de bestudering van het van Broekhoven-algoritme. Dit algoritme is de afgelopen jaren gebruikt bij het afronden van de ruwe sterftequotiënten tot gladde overlevingstafels die gebruikt worden bij het vaststellen van verzekeringstarieven en bij de waardering van verzekeringsverplichtingen. Het algoritme is beschreven in de appendix van (van Broekhoven, 2002).

Karakterisering van het algoritme

Het algoritme gaat uit van ruwe sterftequotiënten *QR(x)* die getransformeerd worden door

$$\hat{f}(x) = \ln(-\ln(1 - QR(x)))$$

Het algoritme bepaalt vervolgens

$$\min_{a,b,c} \sum_{k=-m}^{m} \left[f(x+k) - \hat{f}(x+k) \right]^2$$

waarbij verondersteld wordt dat m = 5 en $f(x) = a + bx + cx^2$. De resulterende *f(x)* levert de gladgestreken overlevingstafels via

$$q(x) = 1 - \exp(-\exp(f(x))).$$

Het algoritme van van Broekhoven is een voorbeeld van het gladstrijken van gegevens door toepassing van een kernel. Kernel smoothing behoort, net als spline smoothing, tot de groep van niet-parametrische methoden om schattingen te maken. Meer specifiek is het algoritme ook bekend als ‘local polynomial kernel smoothing’ en ‘generalised moving weighted average’. Gavin, Haberman en Verrall hebben enkele artikelen over de toepassing van kernel smoothing op sterftestatistiek gepubliceerd, zie Gavin, Haberman, Verrall (1993, 1994, 1995). Verder is Copas en Haberman (1983) te noemen.

Met kernel smoothing kan een eenvoudige en bruikbare schatting gemaakt worden van een onbekende functie op basis van een aantal waarnemingen waarin een ruisterm verwerkt is. Gegeven een aantal waarnemingen $\left(1,\hat{f}(1)\right),\ldots,\left(n,\hat{f}(n)\right)$ die elk voldoen aan

$$\hat{f}(k)=f(k)+\varepsilon_k$$

voor k = 1, …, n. Met kernel smoothing kan de onbekende regressiefunctie *f(.)* gevonden worden.

Als we elk punt schatten met een regressiefunctie door in een beperkt deel van de dataverzameling een *p*-de graads polynoom te schatten door middel van een kernel gewogen kleinste kwadratenmethode dan spreken we over 'local polynomial kernel smoothing' of 'generalised moving weighted average'.

Stel dat we een kernel met bandbreedte *h* gebruiken en we de waarde van *f(.)* willen bepalen op punt *x*. In dit punt *x* wordt de schatter $\hat{f}(x;p,h)$ wordt gevonden door de polynoom $a_0+a_1k+\ldots+a_pk^p$ op de datapunten $\left(k,\hat{f}(k)\right)$ te fitten door gebruik te maken van de kleinste kwadratenmethode met kernel gewichten $K_h(k-x)$.

De waarde van $\hat{f}(x;p,h)$ is de waarde waarbij de volgende formule geminimaliseerd wordt.

$$\sum_{k=1}^{n}\left[\hat{f}(k)-a_0-\ldots-a_p(k)^p\right]^2 K_h(k-x)$$

Hieronder wordt gedemonstreerd dat het algoritme van van Broekhoven feitelijk neerkomt op de hierboven gegeven formule. Het algoritme gaat uit van de volgende te minimaliseren functie.

$$\min_{a,b,c}\sum_{k=-m}^{m}\left[f(x+k)-\hat{f}(x+k)\right]^2$$

waarbij verondersteld wordt dat m = 5 en $f(x)=a+bx+cx^2$.

Welke impliciete kernel wordt hier gebruikt? We herschrijven de te minimaliseren functie als volgt

$$\sum_{k=-m}^{m}\left[f(x+k)-\hat{f}(x+k)\right]^2=\sum_{k=x-m}^{x+m}\left[f(k)-\hat{f}(k)\right]^2=\sum_{k=1}^{n}\left[f(k)-\hat{f}(k)\right]^2 \mathrm{I}_{[-m,m]}(k-x)$$

waarbij I(.) de indicatorfunctie is. Hiermee hebben we, op naamgeving van de parameters na, een gelijke formule als bovenstaande kernel regressie met een polynoom. Aangezien de functie geminimaliseerd wordt, kan deze met een constante ½ m vermenigvuldigd worden zonder dat de uitkomst van de minimaliseringoperatie wijzigt. We delen de parameter van de indicatorfunctie vervolgens door dezelfde constante zodat de het gebied waarbinnen de indicatorfunctie een positieve waarde geeft loopt van [-½, ½].

$$\sum_{k=1}^{n}\left[f(k)-\hat{f}(k)\right]^{2}\mathrm{I}_{[-m,m]}(k-x)=\sum_{k=1}^{n}\left[f(k)-\hat{f}(k)\right]^{2}\frac{1}{2\mathrm{m}}\mathrm{I}_{\left[-\frac{1}{2},\frac{1}{2}\right]}\left(\frac{k-x}{2m}\right)$$

Substitutie levert

$$\sum_{k=1}^{n}\left[f(k)-\hat{f}(k)\right]^{2}\mathrm{I}_{[-m,m]}(k-x)=\sum_{k=1}^{n}\left[f(k)-\hat{f}(k)\right]^{2}K_{2m}(k-x)$$

met $K_h(x)=\frac{1}{h}K\left(\frac{x}{h}\right)$ en als kernel de zogenaamde indicator kernel $K(x)=\mathrm{I}_{\left[-\frac{1}{2},\frac{1}{2}\right]}(x)$.

De methode van kernel smoothing wordt veelvuldig gebruikt, zowel binnen het levenactuariaat als binnen andere statistische domeinen.

Uitgangspunten van het algoritme

De fundamentele uitgangspunten van het algoritme van van Broekhoeven zijn dus

- De dataverzameling wordt gladgestreken met als basis een indicator kernel.
- De regressie wordt met een tweedegraads polynoom uitgevoerd.
- De bandbreedte h (= $2m$) is 10.

De bandbreedte van de kernel is van groot belang. Een lage waarde impliceert dat de kernel smal zal zijn en als gevolg daarvan wordt de kleinste kwadratenmethode toegepast op een klein aantal observaties. Dit resulteert in een ruwe benadering van *f(.)*. Als een hoge waarde gebruikt wordt dan leidt dit tot een gladde benadering, aangezien de lokale regressie gebaseerd is op een groot aantal datapunten. In het limiet geval, als h naar oneindig gaat, zal de kleinste kwadratenmethode toegepast worden op het gehele gebied en wordt de p-de graads polynoom op de hele dataverzameling toegepast. In (Gavin, Haberman en Verrall, 1994) wordt specifiek ingegaan op de keuze van de bandbreedte bij kernel smoothing. Ongeacht de keuze van h is het effect van het gladstrijken dat bergen minder hoog worden en dalen minder diep. Als we bijvoorbeeld naar de zogenaamde ‘accident hump’ bij mannen kijken en blijkt dat met kernel smoothing de berg breder en minder hoog wordt. Hetzelfde zal waar te nemen zijn bij kindersterfte.

In Gavin et al. (1993) wordt opgemerkt dat “... For the kernel graduation, a bandwidth of 10 has been chosen as this corresponds approximately to Spencer’s 21-term formula”. Spencer’s formule met 21 parameters is volgens dit artikel een van de meest gebruikte en succesvolle ‘moving weighted average formulas’.

Een andere keuze die van groot belang is, is de keuze voor de kernel. De standaardkeuze is de Gaussische kernel, ofwel normale kernel met *K(x)* = N(0, 1). Een andere gebruikte kernel is de indicator kernel, die resulteert in een bewegend raam door de data, hetgeen enigszins lijkt op ‘block thresholding’.

De genoemde uitgangspunten zijn van groot belang voor de uiteindelijke uitkomst. Deze uitgangspunten moeten dus voldoende onderbouwd worden.

Oplossing van het algoritme

Toepassing van de indicator kernel levert betrekkelijk eenvoudige resultaten op.

Bij een lineaire regressie geldt

$$f(x;1,h) = \frac{1}{h+1} \sum_{k=-\frac{1}{2}h}^{\frac{1}{2}h} \hat{f}(x+k)$$

In dit geval worden dus voor alle waarden dezelfde gewichten gebruikt.

Voor een tweedegraads polynoom kan de oplossing gevonden worden door de afgeleide te bepalen naar de parameters van de formule. Het resultaat is een stelsel van 3 vergelijkingen die vervolgens op nul gesteld worden. Na enig rekenwerk, zie de appendix, kan voor een regressie met een tweedegraads polynoom, met p = 3, de volgende formule gevonden worden.

$$f(x;2,h) = \sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) \hat{f}(x+k)$$

Hier zien we dat de waarde van *f(x)* uitgedrukt kan worden als een gewogen som van de ruwe $\hat{f}(x)$. In dit geval geldt dat als $\hat{f}(x) = a + bx + cx^2$ dat dan ook geldt $f(x;2,h) = a + bx + cx^2$.

Verder geldt

$$\sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) = 1 \quad \text{en} \quad \sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) k^2 = 0$$

Vanwege deze eigenschappen wordt de kernel in de som de ‘optimal smoothing kernel’ genoemd.

Voor h = 10 geldt

$$f(x;2,10) = \sum_{k=-5}^{5} \left(\frac{89}{429} - \frac{5}{429} k^2 \right) \hat{f}(x+k)$$

Aan de hand van deze formule kan dus eenvoudig de gladde benadering uit de ruwe dataverzameling afgeleid worden.

Toepassing van deze formule voor h = 0, 2, 4, 6, 8, 10 en 12 levert de volgende gewichten op.

f(x-6)	f(x-5)	f(x-4)	f(x-3)	f(x-2)	f(x-1)	f(x)	f(x+1)	f(x+2)	f(x+3)	f(x+4)	f(x+5)	f(x+6)
						1						
					0	1	0					
				-3/35	12/35	17/35	12/35	-3/35				
			-3/105	12/105	27/105	32/105	27/105	12/105	-13/105			
		-21/231	14/231	39/231	54/231	59/231	54/231	39/231	14/231	-21/231		
	-36/429	9/429	44/429	69/429	84/429	89/429	84/429	69/429	44/429	9/429	-36/429	
-11/143	0/143	9/143	16/143,	21/143	24/143	25/143	24/143	21/143	16/143	9/143	0/143	-11/143

De toepassing van het algoritme wordt in het laatste voorbericht van de AG-tafels gerechtvaardigd vanuit de gedachte dat de sterftewet van Gompertz overeenkomt met een lineaire *f(k)*. Aangezien niet voor alle leeftijdsintervallen de waargenomen sterftequotiënten lineair zijn wordt uitgegaan van een kwadratische functie.

Indien de data afkomstig is uit een Gompertz verdeelde populatie dan zullen de afgeronde gegevens identiek zijn aan de ruwe gegevens. Dit kan eenvoudig ingezien worden indien we een lineaire functie invullen in $f(x;\, 2,\, 10)$. Een Gompertz verdeelde populatie wordt dus niet door het algoritme vervormd. Echter, dataverzamelingen die niet afkomstig zijn uit een Gompertz verdeling maar uit een andere verdeling zullen op een of andere manier vervormd worden. Als gevolg van de toegepaste transformatie zal bijvoorbeeld een Makeham verdeelde populatie niet een populatie met dezelfde verdeling tot gevolg hebben. Aangezien algemeen wordt aangenomen dat de sterftewet van Makeham door de extra parameter s een betere benadering geeft dan de sterftewet van Gompertz zou het interessant zijn een kernel te vinden die wel een Makeham verdeelde populatie kan reproduceren.

Parametrische en niet-parametrische methoden

In veel statistische onderzoeken worden veronderstellingen gebruikt over de vorm van de kansverdeling van populaties. Er bestaan echter ook methoden die slechts minimale veronderstellingen over de populatie vereisen. Deze methoden staan bekend onder de naam niet-parametrische of distributievrije methoden.

Niet-parametrische methoden zijn er niet op gericht om de complexiteit van het dataverzameling te verminderen. Dat wil zeggen dat de methode er niet toe leidt dat we minder gegevens nodig hebben om dezelfde dataverzameling, of een goede schatting daarvan, te reproduceren. Bij parametrische methoden is dat duidelijk anders omdat die methoden erop gericht zijn om de dataverzameling te representeren in een beperkt aantal parameters die als het ware de dataverzameling zo efficiënt mogelijk samenvat. Niet-parametrische methoden zijn dus sterk datagedreven. Aan de andere kant zijn parametrische methoden zijn kennisgedreven. De laatste zijn daarom beter geschikt voor extrapolaties omdat het eenvoudiger is om met een beperkt aantal parameters een trend te modelleren dan met een grote dataverzameling.

Toetsing

Overlevingstafels worden gebruikt om de toekomstige sterftekansen te voorspellen. Als verwacht wordt dat historische gegevens voldoende representatief zijn voor de toekomst dan vormen deze gegevens uiteraard een basis voor voorspellingen. Het primaire doel is echter niet een beschrijving te geven van historische gegevens maar om toekomstige kansen nauwkeurig te voorspellen. Het ligt daarom voor de hand ons af te vragen in hoeverre de resulterende overlevingstafel naderhand getoetst kan worden aan de hand van waargenomen sterftekansen.

Bij parametrische modellen is toetsing redelijk voor de hand liggend. Het is mogelijk om betrouwbaarheidsintervallen vast te stellen van de gebruikte parameters en op grond daarvan eventueel te concluderen dat de schatting van de parameters op grond van waargenomen gegevens goed dan wel foutief was geweest. Verder kunnen verschillende modellen gebruikt worden om inzicht te krijgen in de gevoeligheid van modelveronderstellingen. Bij niet-parametrische methoden, die geen veronderstellingen doen ten aanzien van de onderliggende kansdistributies, is dat niet op dezelfde manier mogelijk. Toch bestaan er methoden om hiervoor betrouwbaarheidsintervallen op te stellen. In veel elementaire statistiekboeken zoals (Dudewicz en Mishra, 1988) zijn methoden voor statistische inferentie beschreven. Dat neemt niet weg dat toetsing van niet-parametrische modellen veel lastiger is dan parametrische modellen en dat is onmiskenbaar een nadeel van niet-parametrische methoden. Deze methoden benaderen de waargenomen cijfers weliswaar, maar of de methode ook de toekomstige cijfers goed voorspelt is daarmee niet beantwoord.

De betekenis van de soms gebruikte risicodecompositie in stochastische variatie, parameter en model onzekerheid en structurele onzekerheid is in deze context enigszins problematisch.

Het algoritme op de grenspunten

Hoe werkt dit algoritme uit op de grenspunten? Bij een leeftijd dicht bij 0 of dicht bij het einde van de distributie kan het gladstrijken niet over de gehele *-k* tot *k* gebeuren. Het algoritme kan wel toegepast worden, maar moet dan op een beperkter bereik gebaseerd worden. Berekend kan worden dat voor de grenspunten voor *h=10* geldt

$$f(0;2{,}5)=\sum_{k=0}^{5}\left(\frac{46-33k-5k^2}{56}\right)\hat{f}(x+k)$$

$$f(1;2{,}5)=\sum_{k=-1}^{5}\left(\frac{4-k}{14}\right)\hat{f}(x+k)$$

$$f(2;2{,}5)=\sum_{k=-2}^{5}\left(\frac{13+k-k^2}{56}\right)\hat{f}(x+k)$$

$$f(3;2{,}5)=\sum_{k=-3}^{5}\left(\frac{1072+93k-85k^2}{4620}\right)\hat{f}(x+k)$$

$$f(4;2{,}5)=\sum_{k=-4}^{5}\left(\frac{74+3k-5k^2}{330}\right)\hat{f}(x+k)$$

Voor $x \geq 5$ geldt

$$f(x;2{,}5)=\sum_{k=-5}^{5}\left(\frac{89}{429}-\frac{5}{429}k^2\right)\hat{f}(x+k)$$

We kunnen deze vaststellen door

$X^T X = Y$ waarbij Y een 3x3-matrix is met $y_{ij}=\sum_{b=-m}^{m}\sum_{a=0}^{i+j-2}\binom{i+j-2}{a}k^{i+j-2-a}b^a$.

als

$$X = \begin{bmatrix} 1 & x-m & (x-m)^2 \\ 1 & x-m+1 & (x-m+1)^2 \\ ... & ... & \\ 1 & x+m-1 & (x+m-1)^2 \\ 1 & x+m & (x+m)^2 \end{bmatrix}$$

Voor de grensgevallen wordt voor $0 \le x < m$ geevalueerd

$$y_{ij} = \sum_{b=-m}^{x} \sum_{a=0}^{i+j-2} \binom{i+j-2}{a} k^{i+j-2-a} b^a$$

Voor de grensgevallen aan het andere eind, voor $w\text{-}m \le x < w$

$$y_{ij} = \sum_{b=-m}^{m} \sum_{a=0}^{i+j-2} \binom{i+j-2}{a} k^{i+j-2-a} b^a$$

Een tweede probleem is waarop we de kernel toepassen. Er zijn lege cellen en daar komt geen berg. Maar dat betekent dat daar de wijdte te klein is.

Literatuur

Broekhoven, H. van (2002), Market value of liabilities mortality risk: a practical model, North American Actuarial Journal, volume 6, nr. 2, pp. 95-106

Copas, J.B., Haberman, S. (1983), Non-parametric graduation using kernel methods, Journal of the Institute of Actuaries, vol. 110, pp 135-156, http://www.actuaries.org.uk/files/pdf/library/JIA-110/0135-0156.pdf

Dudewicz, E.J., Mishra, S.N. (1988), Modern mathematical statistics, John Wiley & Sons

Gavin, J., Haberman S., Verrall, R. (1995), Graduation by kernel and adaptive kernel methods with a boundary correction, pp. 173-209, http://library.soa.org/library/tsa/1990-95/TSA95V478.pdf

Gavin, J., Haberman, S., Verrall, R. (1993), Moving weighted average graduation using kernel estimation, Insurance: Mathematics and Economics 12, pp. 113-126, http://www.sciencedirect.com/science/article/B6V8N-459J63T-1/2/2c493e699c3e80dbf3d89c847041457f

Gavin, J., Haberman, S., Verrall, R. (1994), On the choice of bandwidth for kernel graduation, Journal of the Institute of Actuaries, 121, pp. 119-134, http://www.actuaries.org.uk/files/pdf/library/JIA-121/0119-0134.pdf

Appendix: bewijs van regressie met een tweedegraads polynoom

We bewijzen dat

$$\min_{a,b,c} \sum_{k=-m}^{m} \left[f(x+k) - \hat{f}(x+k) \right]^2 \qquad (1)$$

voor p = 2 het volgende oplevert

$$f(x;2,h) = \sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) \hat{f}(x+k)$$

Allereerst wordt opgemerkt dat (1) hetzelfde is als

$$\min_{a,b,c} \sum_{k=-m}^{m} \left[f(x+k)^2 - 2f(x+k)\hat{f}(x+k) \right] \qquad (2)$$

Substitutie levert

$$\min_{a,b,c} \sum_{k=-m}^{m} \left[\left(a + b(x+k) + c(x+k)^2\right)^2 - 2\left(a + b(x+k) + c(x+k)^2\right)\hat{f}(x+k) \right]$$

Differentiëren naar a, b en c levert 3 vergelijkingen met drie onbekenden

$$\sum_{k=-m}^{m} 2\left(a + b(x+k) + c(x+k)^2\right) - 2\sum_{k=-m}^{m} \hat{f}(x+k) = 0$$
$$\sum_{k=-m}^{m} 2\left(a + b(x+k) + c(x+k)^2\right)(x+k) - 2\sum_{k=-m}^{m} \hat{f}(x+k)(x+k) = 0$$
$$\sum_{k=-m}^{m} 2\left(a + b(x+k) + c(x+k)^2\right)(x+k)^2 - 2\sum_{k=-m}^{m} \hat{f}(x+k)(x+k)^2 = 0$$

De oplossing van dit stelsel is

$$a = \frac{\sum_{k=-m}^{m} \hat{f}(x+k)}{2m+1} - \left(\frac{1}{3} nc(n+1) + x(b+cx)\right)$$

$$b = \frac{\sum_{k=-m}^{m} k\hat{f}(x+k)}{\frac{1}{3} m(m+1)(2m+1)} - 2cx$$

$$c = \frac{45 \sum_{k=-m}^{m} k^2 \hat{f}(x+k) - 15n(n+1) \sum_{k=-m}^{m} \hat{f}(x+k)}{b(2n+3)(2n+1)(2n-1)(n+1)}$$

De kernel kan nu gevonden worden door deze oplossing in te vullen in

$$f(x;2,h) = a + bx + cx^2$$

Dit levert, met $h = 2\ m$,

$$f(x;2,h) = \sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) \hat{f}(x+k)$$

Allereerst wordt opgemerkt dat (1) hetzelfde is als

$$\min_{a,b,c} \sum_{k=\max(0-m,0)}^{\min(\omega,m)} \left[f(x+k)^2 - 2 f(x+k)\hat{f}(x+k) \right]$$

(2)

Substitutie levert

$$\min_{a,b,c} \sum_{k=\max(-m,0)}^{\min(m,\omega)} \left[\left(a+b(x+k)+c(x+k)^2\right)^2 - 2\left(a+b(x+k)+c(x+k)^2\right)\hat{f}(x+k) \right]$$

Differentiëren naar a, b en c levert 3 vergelijkingen met drie onbekenden

$$\sum_{k=\max(-m,0)}^{\min(m,\omega)} 2\left(a+b(x+k)+c(x+k)^2\right) - 2\sum_{k=\max(-m,0)}^{\min(m,\omega)} \hat{f}(x+k) = 0$$

$$\sum_{k=\max(-m,0)}^{\min(m,\omega)} 2\left(a+b(x+k)+c(x+k)^2\right)(x+k) - 2\sum_{k=\max(-m,0)}^{\min(m,\omega)} \hat{f}(x+k)(x+k) = 0$$

$$\sum_{k=\max(-m,0)}^{\min(m,\omega)} 2\left(a+b(x+k)+c(x+k)^2\right)(x+k)^2 - 2\sum_{k=\max(-m,0)}^{\min(m,\omega)} \hat{f}(x+k)(x+k)^2 = 0$$

De oplossing van dit stelsel is

$$a = \frac{\sum_{k=-m}^{m} \hat{f}(x+k)}{2m+1} - \left(\frac{1}{3}nc(n+1) + x(b+cx)\right)$$

$$b = \frac{\sum_{k=-m}^{m} k\hat{f}(x+k)}{\frac{1}{3}m(m+1)(2m+1)} - 2cx$$

$$c = \frac{45\sum_{k=-m}^{m} k^2 \hat{f}(x+k) - 15n(n+1)\sum_{k=-m}^{m} \hat{f}(x+k)}{b(2n+3)(2n+1)(2n-1)(n+1)}$$

De kernel kan nu gevonden worden door deze oplossing in te vullen in

$f(x;2,h) = a + bx + cx^2$

Dit levert, met $h = 2\ m$,

$$f(x;2,h) = \sum_{k=-m}^{m} \left(\frac{3\left(\frac{3}{4}h^2 + \frac{3}{2}h - 1\right) - 15k^2}{(h+3)(h+1)(h-1)} \right) \hat{f}(x+k)$$