

STATISTIEK EN WIEGEDOOD

Aart F. de Vos *)

Samenvatting

Eind oktober 1987 was er enige commotie over de bekendmaking van prof. G.A. de Jonge dat er sterke aanwijzingen zijn voor een verband tussen buikligging bij babies en wiegedood. Opvallend in de presentatie van en de reacties op het onderzoek was de terughoudendheid t.a.v. het "bewezen zijn" van het verband. Een reactie in de NRC stelde zelfs dat een dergelijk bewijs onmogelijk zou zijn. Dit stuk grijpt deze discussie aan om aan een ruim publiek te laten zien hoe de Bayesiaanse statistische beslissingstheorie in staat is diverse argumenten en twijfels in een analyse te incorporeren. Dit om de lezer te overtuigen van de flexibiliteit van de Bayesiaanse analyse en het grote risico dat hij/zij loopt door een baby op de buik te slapen te leggen.

*)Faculteit der Economische Wetenschappen en Econometrie
Vakgroep Econometrie
Vrije Universiteit
Postbus 7161
1007 MC Amsterdam

Statistiek en wiegedood.

Op 22 oktober 1987 kwam de uitspraak van prof. G.A. de Jonge over zijn aanwijzingen voor een verband tussen buikligging en wiegedood in het nieuws. Hij had gevonden dat bij 150 gevallen van wiegedood in 85% van de gevallen sprake was van buikligging. In een controlegroep van 326 kinderen werd in 61% van de gevallen de buikligging als de gebruikelijke slaaphouding aangetroffen. Voorts voerde hij aan dat buikligging in de mode kwam in het begin van de jaren zeventig, en dat sindsdien het aantal gevallen van wiegedood was toegenomen van 0,42 tot 1,42 per 1000. De cijfers werden in deze vorm gepresenteerd, met de kantekening dat het "sluitend bewijs" voor een oorzakelijk verband nog niet is geleverd, maar dat er voldoende redenen waren om ouders vooralsnog aan te raden babies zo min mogelijk op de buik te leggen en om voorts een groot onderzoek te organiseren om zekerheid te krijgen.

Dit bericht maakte veel emoties los, waarvan een deel veroorzaakt werd door de ontstane onzekerheid over de te volgen gedragslijn. Deze onzekerheid werd o.a. reeds de volgende dag verwoord door NRC redacteur E.J. Boer. Hij beschuldigde de Jonge ervan voornamelijk onrust te hebben veroorzaakt. De vermoedens van de Jonge zouden op onvoldoende gegevens berusten. Boer ging zelfs nog veel verder: "experimenten bij onderzoek wiegedood lossen niets op" heette zijn stukje.

Opvallend in deze stukken was de afwezigheid van duidelijke statistische interpretatie van de gegevens. Voorts lopen er twee vragen door elkaar heen. Boer had in zekere zin gelijk dat het verband tussen buikligging en wiegedood niet "bewezen" is. De Jonge richtte zich op de vraag: wat kan men concluderen uit zijn gegevens, ook als het niet gaat om een bewezen verband en welke beslissingen vloeien daaruit voort? Dit artikel richt zich op de beantwoording van de laatste vraag. Om deze te beantwoorden maak ik gebruik van een hier te lande nog (te) weinig beoefende tak van de statistiek: de statistische beslissingstheorie. Primair is dit artikel een illustratie van deze vorm van omgaan met gegevens. Mijn conclusie zal zijn dat de gegevens voldoende reden zijn om buikligging vooralsnog ernstig af te raden, en zelfs om het verband als min of meer bewezen te beschouwen. Voorts poog ik aan te geven waarom ik dit soort analyse voor vele problemen zinniger acht dan meer klassieke statistische analyses.

Mijn oorspronkelijke bedoeling om dit stuk voor een breed publiek (de NRC) te schrijven bleek te hoog gegrepen. De statistisch onderlegde lezer voelt zich wellicht door de begintekst, die nog grotendeels op de tekst voor de krant berust, op een te laag niveau aangesproken, maar dat wordt hopelijk goed gemaakt door de voetnoten en appendices.

Orthodoxe statistiek versus statistische beslissingstheorie.

Boer stelt dat de uitspraken van de Jonge op onvoldoende rigoreus onderzoek berusten en dus onnodig paniek zaaien. Met deze houding bevindt hij zich in het goede gezelschap van wat ik zou willen noemen de orthodoxe statistiek. Uitgangspunt van de orthodoxen is dat statistiek ten dienste moet staan van het zoveel mogelijk tegengaan van niet volledig verantwoorde uitspraken. Pas wanneer het vrijwel zeker is dat bepaalde uitkomsten niet op toeval berusten mag een hypothese in de openbaarheid gebracht. Centraal staat -de naam is al kenmerkend- de "kans op een fout van de eerste soort": de kans op het ten onrechte verwerpen van de hypothese dat er niets aan de hand is, in dit geval eenvoudiger uit te leggen als "de kans om ten onrechte paniek te zaaien". Deze kans moet klein zijn, kleiner dan 1% houdt men graag aan. Er is echter ook nog een "kans op een fout van de tweede soort": ten onrechte zeggen dat er niets aan de hand is. En de consequenties van zo'n fout kunnen veel ernstiger zijn. De discussie over de buikligging geeft wel een heel duidelijk voorbeeld. De Jonge mag -zoals ik zal laten zien- op grond van zijn bevindingen beweren dat door geen babies meer op de buik te leggen naar schatting 100 van de 250 gevallen van wiegedood per jaar voorkomen kunnen worden. De keerzijde(!), het risico van een fout van de eerste soort, bestaat uit het op de zij of de rug leggen van alle babies zonder noodzaak, een aanmerkelijk minder ernstig gebeuren zolang niet blijkt dat dit andere risico's met zich meebrengt.

Een evenwichtige afweging van de diverse soorten fouten is mogelijk wanneer men zich baseert op de "liberale" statistische beslissingstheorie. Deze richt zich niet zozeer op het aantonen van "statistische waarheden" maar op het nemen van beslissingen in onzekere situaties met beperkte informatie. Net als bij religies zijn er juist onder liberalen vele variaties mogelijk, afhankelijk van de mate waarin

men het centrale leerstuk laat vallen. Dit leerstuk, het orthodoxe dogma, is dat alleen "objectieve" feiten tellen. Hoe liberaler, hoe verder men bereid is te gaan in de subjectieve inschatting van de omstandigheden waardoor het mogelijk wordt ook uit niet geheel verantwoorde experimenten conclusies te trekken. Deze wijze van statistiek bedrijven komt veel beter overeen met de dagelijkse praktijk van het omgaan met informatie dan de klassiek statistische. Economische beslissingen worden meestal genomen onder griezelig grote onzekerheid, maar ook vele medische beslissingen zijn gebaseerd op onvolledige kennis en subjectieve inschattingen. Vaak kan het niet anders. Ook in dit geval niet, zoals hieronder uiteengezet wordt. De na de beschrijving van de meest orthodoxe procedure volgende analyse van de gegevens van de Jonge moge aantonen hoe bruikbaar liberale beoefening van de statistiek kan zijn.

De onmogelijkheid van waterdicht onderzoek.

Tegen vrijwel ieder statistisch onderzoek zijn bedenkingen in te brengen. Wil men al deze bedenkingen -ze komen nog uitgebreid aan de orde- uitsluiten dan moet men zeer omvangrijke onderzoeken opzetten, vooral wanneer het om zeldzame gebeurtenissen gaat.

Een voorbeeld van dit soort onderzoek vormt het in 1954 in de VS gehouden onderzoek naar het effect van inenting tegen polio. Honderdduizenden werden ingeënt met het vaccin en met een placebo, wie wat kreeg werd door loting bepaald. Het bleek dat de kans op polio na inenting 0,4 per 1000 lager was (Solomon 1966 p.324). De enorme omvang van het onderzoek was nodig omdat het om hele kleine kansen gaat. Het bewijs van de Jonge's vondst zou dus geleverd moeten worden door bij loting te bepalen welke helft van alle 175.000 babies van Nederland een jaar lang op de buik gelegd moest worden en welke niet. Een dergelijk onderzoek is onuitvoerbaar. Bovendien is het de vraag of dit onderzoek ethisch verantwoord is wanneer men eenmaal vermoedt dat buikligging riskant is. De 60 gevallen van polio in het Amerikaanse onderzoek die de "pech" hadden een placebo te treffen hadden als enige troost dat het niet geheel zeker was dat ze het niet gekregen zouden hebben als ze het vaccin hadden geloot.

Er zijn veel eenvoudiger methoden om statistische verbanden vast te stellen, maar dan blijft men altijd zitten met de vragen. Toegespitst op de wiegedood: de vraag of er misschien redenen kunnen zijn waardoor babies die aan iets onbekends lijden dat uiteindelijk tot wiegedood leidt vaker op de buik gelegd worden. Dat kan: ze zijn misschien erg onrustig, wat bij buikligging minder het geval pleegt te zijn. Een dergelijk verband maakt het onderzoek waardeloos. Wanneer men redelijk aannemelijk kan maken dat er geen alternatieve verklaring voor het statistisch verband tussen buikligging en wiegedood bestaat is het mogelijk om uitspraken te doen op basis van een onderzoek als dat van de Jonge. Bij mijn weten zijn er geen aanwijzingen dat babies die aan wiegedood sterven daarvoor onrustiger waren dan andere babies, dus de eerder geformuleerde hypothese is onaannemelijk. Mijn analyse gaat in eerste instantie uit van een verklaring middels een direct oorzakelijk verband (minder zuurstof). In een latere fase wordt aangegeven hoe twijfels hieraan kunnen worden geïncorporeerd.

De gegevens van de Jonge.

De belangrijkste gegevens zijn de volgende: er zijn 150 gevallen van wiegedood onderzocht. In 127 (85%) van de gevallen was er sprake van buikligging. In een -misschien niet helemaal representatieve-controlesteekproef van 326 babies werd bij 61% buikligging geconstateerd.

Andere gegevens als de parallele toename van het aantal gevallen van wiegedood en buikligging zijn minder informatief. Ze worden in een latere fase behandeld.

Om te beginnen is de informatie over de percentages buikligging bij wiegedood en in de controlesteekproef te gebruiken om te schatten hoeveel maal groter de kans op wiegedood is bij buikligging dan bij andere ligging, de in dit stuk centraal staande "risicoverhouding". Die schatting valt in dit geval veel hoger uit dan de meeste mensen denken. Voor diegenen die hun statistiekkennis willen ophalen geef ik de formule voor de berekening:

$$\frac{P(w|b)}{P(w|-b)} = \frac{P(b|w)P(w)}{P(b)} \cdot \frac{P(-b)}{P(-b|w)P(w)}$$

hierin staat P voor de kans, w voor wiegedood, b voor buikligging, de - voor "niet" en | voor "gegeven de gebeurtenis dat". Bijvoorbeeld $P(w|-b)$ is de kans op wiegedood als een baby niet op de buik ligt. De formule is een dubbele toepassing van de "regel van Bayes", misschien wel de belangrijkste regel uit de kansrekening en bovendien de basis van de "Bayesiaanse" (liberale) statistiek. Bij de uitwerking valt $P(w)$ eruit en de cijfers leren:

$$\frac{P(w|b)}{P(w|-b)} = \frac{0.85}{0.15} \cdot \frac{0.39}{0.61} = 3.62$$

de geschatte risicoverhouding is dus 3.62. Dit zou ongeveer neerkomen op een kans op wiegedood bij buikligging van 2 per 1000 tegen 0,55 per 1000 bij andere ligging.

De volgende vraag is hoe betrouwbaar deze schattingen zijn. Deze vraag is alleen te beantwoorden wanneer we bereid zijn aanvullende veronderstellingen maken. Onder geschikte veronderstellingen zijn "modelmatig" uitspraken te doen. En daar beginnen de controverses.

Een modelmatige klassiek-statistische aanpak vergt nogal wat (een overzicht geeft Kleinbaum e.a. (1982)): behalve de reeds genoemde noodzaak van een oorzakelijk verband moeten de cijfers representatief zijn, de Jonge moet de eerste zijn die het verband heeft onderzocht en eigenlijk moet hij zijn onderzoek speciaal hebben opgezet om het verband tussen buikligging en de kans op wiegedood te onderzoeken. Als aan de laatste voorwaarden niet is voldaan bestaat het risico dat het hier gaat om één experiment uit een hele verzameling experimenten, in dat geval is er altijd wel iets "significant".

Het zal duidelijk zijn dat de huidige onderzoeksresultaten niet aan alle voorwaarden voldoen. Toch is het nuttig om na te gaan wat men had kunnen zeggen over de risicoverhouding ware dit wel het geval geweest. Dit kan met klassieke methoden, maar vrijwel identieke resultaten zijn ook volgens de lijnen van de Bayesiaanse statistiek te bereiken door met zgn. niet-informatieve voorverdelingen te werken (het standaardwerk op dit gebied is Box & Tiao (1973)). Voordelen van de Bayesiaanse aanpak zijn dat a) de uitspraken in een vorm gegoten worden die iedereen kan

begrijpen¹⁾ en b) een soepele overgang mogelijk is naar een meer liberale benadering waarin rekening gehouden wordt met twijfels aan de omstandigheden (klassieke methoden kennen deze overgang niet of nauwelijks).

De kansverdeling van de risicoverhouding.

Typerend voor de Bayesiaanse aanpak is de presentatie van de analyse middels een kansverdeling van de grootte waarin men geïnteresseerd is. In dit geval kozen wij de risicoverhouding daarvoor. Op grond van in de appendix verantwoorde berekeningen is grafiek 1 geconstrueerd waarin de twee met symbolen aangegeven functies aangeven wat men onder optimale omstandigheden op grond van de cijfers van de Jonge zou kunnen zeggen: de kansverdeling (de bult) en de cumulatieve dichtheid (de stijgende lijn) van de risicoverhouding. De cumulatieve dichtheid is het meest informatief: de risicoverhouding is a) zeker groter dan 1,5; b) met 98% zekerheid groter dan 2; c) waarschijnlijk groter dan 3 (kans is 70%) en d) misschien wel groter dan 5 (kans 7%). Kortom de data suggereren zeer sterk een forse invloed van buikligging.

Het is al gesteld dat dit soort resultaten ook op klassieke wijze verkregen kunnen worden. Een kleine uitweiding over de moeizame wijze waarop dat pleegt te gaan is hier op zijn plaats. In de eerste plaats wordt de "p-waarde" berekend deze pleegt geïnterpreteerd te worden als de kans om de hypothese dat de ligging niets uitmaakt ten onrechte te verwerpen (en zelfs deze voor leken onbegrijpelijke uitspraak is niet geheel correct). De p-waarde is hier heel klein: 0.00000008; zij correspondeert met de kans dat de risicoverhouding kleiner dan 1 is in de grafiek. Vervolgens geeft de klassieke aanpak betrouwbaarheidsintervallen: grenswaarden die met een bepaalde willekeurig te kiezen waarschijnlijkheid (99% bijvoorbeeld) aan weerszijden van de ware risicoverhouding vallen (zij hangen af van de data). Ook hier is de interpretatie moeizaam, de methodologie verwarrend (Kleinbaum e.a. (1982, ch.15) geeft voor een eenvoudige "case-control" studie als deze vier verschillende recepten) en het resultaat ook te

1) Een aardige beschrijving van de ervaringen met de Bayesiaanse presentatie (met niet-informatieve voorverdelingen) in de praktijk is te vinden in Racine e.a. (1985).

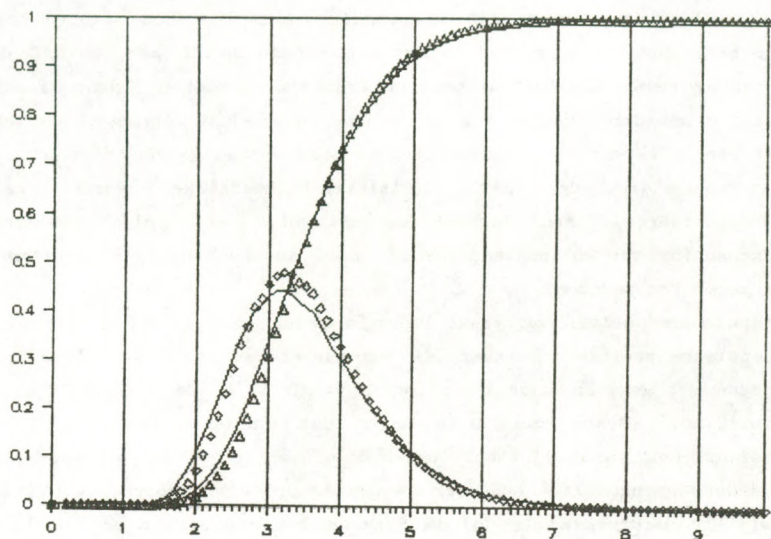
verkrijgen door in de Bayesiaanse kansverdeling een gebied te nemen waarbinnen 99% van de kansverdeling ligt, bijvoorbeeld de waarden 1,8 en 6,8 met aan weerszijden 0,5% van de kansmassa. Het zal duidelijk zijn dat een volledige kansverdeling informatiever is.

Halvering van de informatie in de controlesteekproef.

De hierboven gedane uitspraken moeten gecorrigeerd worden voor het feit dat de omstandigheden waarschijnlijk niet optimaal waren. Allereerst is er het risico dat de controlesteekproef niet representatief is. Het trekken van een representatieve steekproef is buitengewoon lastig zo niet voor dit geval onmogelijk. Overigens bleek mij bij navraag dat de Jonge zeer concientieus te werk is gegaan op dit punt (het betreft een zgn. "gematchte steekproef").

Het contrast tussen orthodoxe en liberale beoefening van de statistiek wordt duidelijk wanneer we het onderzoek niet terzijde schuiven met "de controlesteekproef is misschien niet representatief" maar laten zien dat het niet geheel representatief zijn nauwelijks van invloed is op de conclusies. De informatie uit een niet geheel representatieve steekproef die zo is opgezet dat er geen argumenten zijn om aan te nemen dat er systematisch te veel of te weinig buikliggers in voorkomen kan (in liberale ogen) worden benaderd door net te doen alsof niet in een steekproef van 326 er 199 gevallen van buikligging zijn aangetroffen, maar 100 in een steekproef van 163 (wel nog steeds 61% omdat er geen redenen zijn anders te veronderstellen). Er is dan rekening gehouden met het risico van een heel slordige steekproef (zie appendix F voor de rechtvaardiging van deze procedure). Het effect is verrassend: de getrokken lijnen in grafiek 1 vervangen de symbolen en er is haast geen verschil te zien, de conclusies moeten een beetje worden afgezwakt.

Op analoge wijze kan men omgaan met andere mogelijke bedenkingen tegen het onderzoek. De Bayesiaanse statistiek doet dit op zeer eigen wijze. De kansuitspraken als hierboven afgeleid worden niet letterlijk genomen, maar de informatie wordt gebruikt om de scepsis en oordelen van hen die moeten beslissen te beïnvloeden. Daartoe is het nodig dat scepsis en "vooroordeel" worden vormgegeven middels de zgn. voorverdeling of prior. Soms is het mogelijk deze prior te baseren op objectieve informatie. Men zou bijvoorbeeld kunnen nagaan hoeveel wetenschappers zich al met het



Grafiek 1: kansverdeling risicoverhouding op basis van informatie uit de steekproeven (kansverdeling en cumulatieve verdeling).

Symbolen: informatie on gecorrigeerd

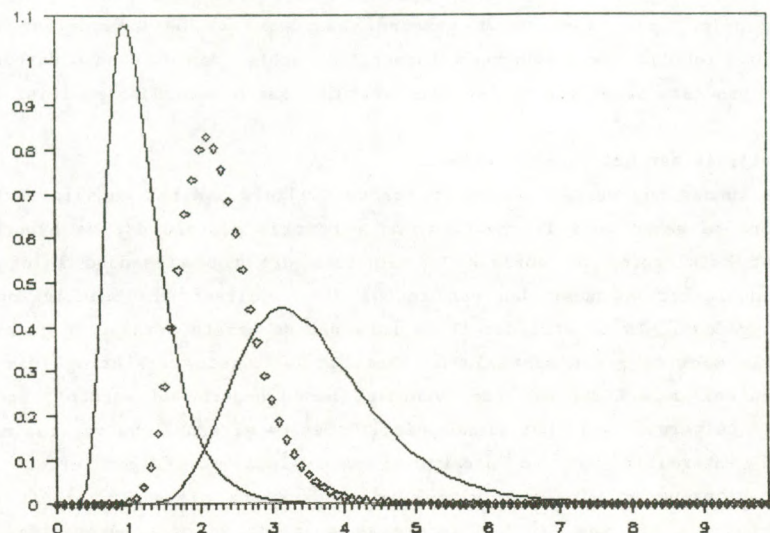
Lijnen: na correctie voor eventuele niet-representativiteit controlesteekproef.

verschijnsel wiegedood hebben beziggehouden, al of niet op een verband met buikligging komend. Er zijn onderzoekers die het verband hebben geopperd (de Jonge maakt in een ingezonden brief aan de NRC dd 6 november gewag van Australische en Franse onderzoeken). Maar misschien leert een nadere analyse van de totale onderzoekinspanning in de wereld dat een verband toch onwaarschijnlijk moet worden geacht. Een dergelijk onderzoek is erg lastig (mislukte onderzoeken worden zelden gerapporteerd!), men ontkomt meestal niet aan enige subjectieve inschatting van de omstandigheden. Deze inschatting kan variëren van persoon tot persoon.

Niet in de positie verkerend om mijn prior hard te maken heb ik een sceptische positie ingenomen. Ik heb aangenomen dat mijn twijfels goed worden weergegeven door de prior in grafiek 2: de hoge bult aan de linkerkant. Deze komt erop neer dat in het "vooordeel" de risicoverhouding vrij zeker tussen de 0,5 en de 2 ligt: aan weerszijden van deze grenzen ligt ongeveer 5% van de prior. In woorden kan ik deze twijfels uitdrukken als: a) ik vermoed dat het effect van de ligging klein is en b) ik kan mij haast niet voorstellen dat als de kans op wiegedood bij buikligging meer dan twee maal zo groot is, dit niet eerder overtuigend zou zijn aangetoond²⁾.

Het is wederom de regel van Bayes die aangeeft hoe het vooroordeel moet worden bijgesteld op basis van de informatie uit de steekproef. Vandaar Bayesiaanse statistiek. Het resultaat ziet men in grafiek 2. De informatie van de Jonge (met correctie voor niet representativiteit van de steekproef) is krachtig genoeg om een vooroordeel als het mijne om te buigen tot de met symbolen aangegeven kansverdeling, de "naverdeling". Deze geeft aan dat ook na het formuleren van twijfels middels de prior, de conclusie is dat buikligging vrijwel zeker een ongunstig effect heeft en dat de risicoverhouding vermoedelijk boven de 2 ligt (kans 55%). Eenvoudig kunnen ook de consequenties worden berekend voor andere

2) Eerder is ook het punt geopperd dat de hypothese slechts een van de vele hypothesen is die onderzocht zijn. In de Bayesiaanse benadering blijkt een dergelijke "data instigated hypothesis" niet van invloed te zijn op de analyse. De onderzoeker moet wel als hij een hypothese eerst niet en later wel in zijn analyse betreft zich realiseren dat hij de hypothese blijkbaar onaannemelijk achtte en dit in de prior verwerken. Zie Leamer (1978, sec. 9.4).



Grafiek 2: kansverdeling risicoverhouding
 links: voorverdeling
 rechts: gecorr. steekproefinformatie uit grafiek 1
 midden, symbolen: naverdeling.

priors, waarbij uiteraard geldt dat hoe sceptischer men staat, hoe meer de prior naar links wordt gekozen, waardoor ook de naverdeling naar links schuift. De steekproefinformatie is echter van dien aard dat zelfs de grootste scepticus zal worden overtuigd dat het verband positief is.

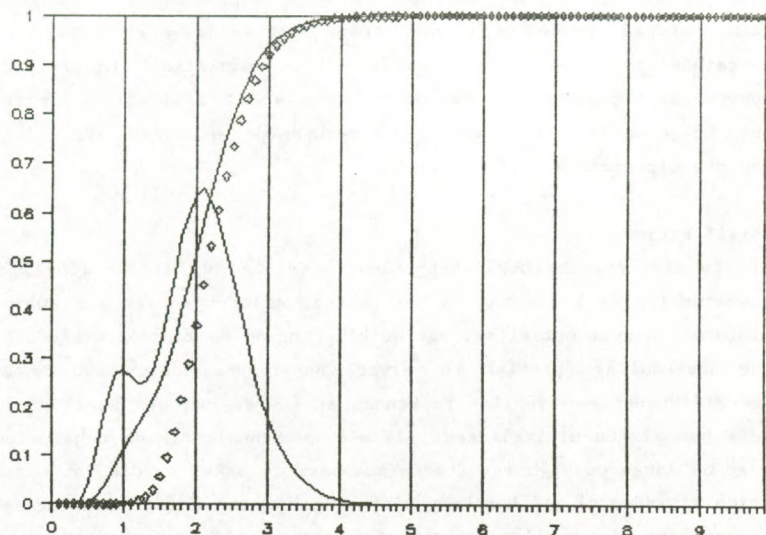
Twijfels aan het causale verband.

We kunnen nog verder gaan door ook de twijfels aan het causale verband vorm te geven door inschatting. Al eerder is gesteld dat de hypothese dat buikligging de oorzaak is van iets dat tot wiegedood leidt mij plausibeler voorkomt dan een factor die resulteert in buikligging én wiegedood. Als ik stel dat ik de kans dat de eerste verklaring juist is drie maal zo groot acht als de kans dat de tweede verklaring juist is (en ook deze uitspraak kan eventueel nader onderbouwd worden), dan is dit te verwerken in het eindoordeel. Onder de alternatieve verklaring is de informatie van de steekproef waardeloos: over het effect van buikligging op wiegedood vertellen de data ons niets dus blijft onze prior wat hij was. In het andere geval geldt de oude naverdeling. De kansverdeling van de risicoverhouding is dan eenvoudig een gewogen gemiddelde van voor- en naverdeling: $1/4$ maal de eerste plus $3/4$ maal de tweede (zie Appendix H), weergegeven in grafiek 3.

De stijging van het aantal wiegedoden

Nog zijn we niet klaar. Er is ook nog de informatie dat het aantal wiegedoden per 1000 nuljarigen gestegen is van 0,42 in 1969 tot 0,83 in 1973 en 1,42 in 1983, terwijl buikligging begin jaren 70 in de mode kwam. Dat bewijst niets maar spoort beter met de hypothese dan andere verklaringen. De 0,42 van 1969 ligt verdacht dicht bij de schatting van 0,55 bij niet-buikligging die uit de steekproeven naar voren kwam. Het is moeilijk deze informatie te verwerken, maar niet onmogelijk.

In appendix I is uitgewerkt hoe een prior die weergeeft welk geloof men hecht aan andere verklaringen dan de toename van het aantal gevallen van buikligging kan worden omgezet in uitspraken over de risicoverhouding. In grafiek 4 staat het resultaat: een kansverdeling met een maximum bij 4 (de toename van wiegedood is erg groot en andere verklaringen voldoen voorlopig niet), maar met een grote spreiding (het gevolg van de toch grote ruimte die aan andere verklaringen is gelaten). Deze informatie is



Grafiek 3: kansverdeling risicoverhouding na correctie voor
mogelijk niet-causaal verband.

symbolen: cumulatieve naverdeling uit grafiek 2

lijnen: dichtheid en cumulatieve verdeling na correctie

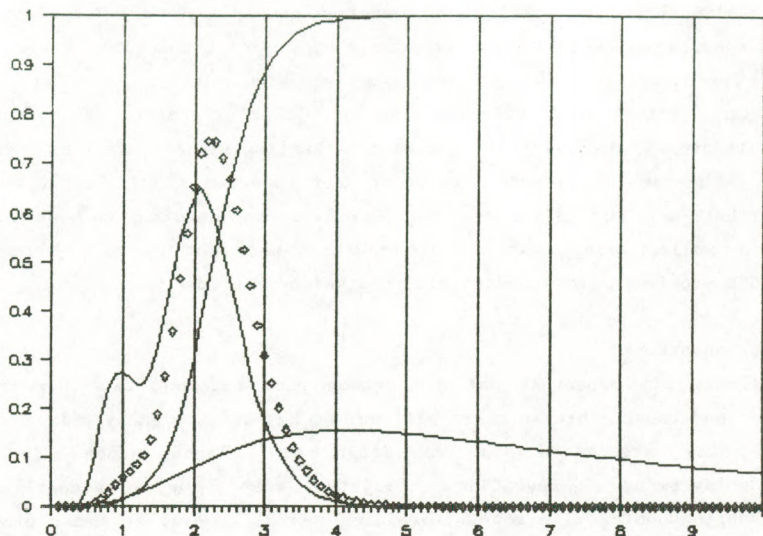
van geheel andere aard dan die uit de steekproeven (onafhankelijk) en kan gevoegd worden bij het reeds bestaande beeld, wat tot de uiteindelijke naverdeling leidt die (ook cumulatief) in grafiek 4 is opgenomen. Opvallend is dat de informatie uit de stijging de twijfels aan het causale verband ongeveer compenseerd: de naverdeling lijkt sterk op die uit grafiek 2.

Beslissingen.

In de statistische beslissingstheorie (zie bijv Lindley (1971)) is de naverdeling de basis van de beslissing. Alle objectieve en subjectieve informatie over het effect van buikligging op de kans op wiegedood is in de naverdeling (grafiek 4) vervat. Beslissers moeten aan de diverse mogelijkheden een verlies toekennen en kunnen dan een beslissing nemen die hun risico minimaliseert. Ik wil twee beslissingen behandelen: die van de Jonge om zijn resultaten openbaar te maken en die van ouders die zich afvragen of zij hun baby niet langer op de buik moeten leggen.

Allereerst de beslissing van de Jonge. Alles wijst erop dat zijn beslissing om de resultaten openbaar te maken gerechtvaardigd is. De kans dat hij onnodig paniek zaait doordat later blijkt er geen oorzakelijk verband is, is ook bij een gezonde dosis scepsis (vervat in de prior) erg klein (2%, geheel toe te schrijven aan de mogelijkheid dat het verband niet causaal is). Het risico dat hij zou lopen door bekendmaking van zijn resultaten enige jaren op te schorten is wel groot: ook na het inbouwen van de scepcis blijft een risicoverhouding van 2 nog een voorzichtige schatting, met als consequentie bij 60% buikligging ongeveer 100 onnodige gevallen van wiegedood per jaar. Ook wanneer maar een deel hiervan gehaald wordt is zijn beslissing volstrekt gerechtvaardigd. Daarbij zij vermeld dat ik de indruk heb dat de door mij gekozen subjectieve inschattingen alle zeer voorzichtig zijn: waarschijnlijk is reeds meer te rechtvaardigen dan mijn naverdeling.

De enige resterende twijfel is of met name rugligging niet andere risico's dan wiegedood met zich meebrengt: een bekende hypothese is verstikkingsgevaar. Dit schijnt buitengewoon klein te zijn, maar het valt te hopen dat op zeer korte termijn bevestigd wordt dit zo is (een onderzoek als dat naar wiegedood is vermoedelijk voldoende). Vooral nog is reeds geconstateerd dat zijligging het veiligst lijkt.



Grafiek 4: kansverdeling risicoverhouding na inbouw informatie
over het gestegen aantal wiegedoden.

lage dichtheid rechts: kansverdeling op basis van stijging
wiegedoden (afgeknot).

linker dichtheid naverdeling uit grafiek 3

symbols: naverdeling, dichtheid.

stijgende lijn: naverdeling cumulatief.

De beslissing van ouders ligt vermoedelijk gecompliceerder. Wiegedood is een spookbeeld, maar de kans erop is klein. Het is ook slechts $1/6$ van alle doodsoorzaken. Wat te doen als een kind niet op zijn zij wil liggen? Hoe moet je de afweging maken als je er vrij zeker van bent dat buikligging bevorderlijk is voor de ontwikkeling van een kind, maar weet dat je de kans op wiegedood daardoor doet toenemen met $1/1000$? Niemand weet het, maar dat is nog geen reden om de mensen onwetend te houden, al is het alleen maar opdat in die gevallen waar het op de zij leggen zonder problemen kan onnodige risico's vermeden worden.

Nader onderzoek.

Tenslotte zij benadrukt dat dit betoog niet betekent dat zorgvuldig nader onderzoek achterwege kan blijven. De Bayesiaanse analyse laat zien dat alle argumenten van de Jonge een plaats hebben in de oordeelsvorming. Bovendien blijkt waar de belangrijkste aanknopingspunten zijn voor uitbreiding van de kennis. En ieder brokje kennis verbetert het eindresultaat in de zin van een naverdeling die minder in het ongewisse laat. Gezien het feit dat een afdoende antwoord wel eens onmogelijk zou kunnen zijn is dat een hoopvol perspectief. Het proces heeft echter meer analogie met het verzamelen van bewijsmateriaal in de rechtspraak³⁾ dan met statistische "kookboekreceptuur".

De stelling van Boer "Experimenten met buikligging lossen niets op" moet dus vervangen worden door "Ieder onderzoek naar de effecten van buikligging brengt de waarheid een stukje dichterbij".

3) Twee buitengewoon instructieve verhandelingen over het gebruik van de regel van Bayes in de rechtspraak zijn Darroch (1987) en Aitken(1987).

Appendices: mathematische verantwoording en toelichting.

Zij $P(b)$ de kans op buikligging

$P(w|b)$ de kans op wiegedood bij buikligging

$P(w|-b)$ de kans op wiegedood bij niet buikligging

De basis van de Bayesiaanse Statistiek (zie bijv Box en Tiao(1973)) is de herziening van verdelingen van parameters middels informatie vervat in de aannemelijkheid (likelihoed) van data middels:

naverdeling $\propto$ voorverdeling $\times$ likelihoed.

waarin $\propto$ staat voor "is evenredig met", d.w.z. vermenigvuldigingsfactoren die niet van de parametres afhangen zijn weggelaten. Om de naverdeling te berekenen moet eenvoudig die constante worden berekend die het geheel weer een kansverdeling doet zijn (integraal moet 1 zijn).

Verder maken wij impliciet gebruik van:

De volgorde waarin informatie (de likelihoeds) wordt toegevoegd is niet van belang voor het eindresultaat.

A.De likelihoed van de controlesteekproef

Deze is, met weglating van constanten (d.w.z. factoren onafhankelijk van $P(b)$)

$$L(s_c) \propto P(b)^x (1-P(b))^{s-x}$$

met s =onderzochte gevallen, x =aantal gevallen van buikligging. (bij de Jonge: $s=326, x=199$)

B.De likelihoed van de steekproef van wiegedoden

Dit gaat analoog: met weglating van constanten geldt

$$L(s_w) \propto P(b|w)^k (1-P(b|w))^{n-k}$$

met k het aantal gevallen van buikligging in steekproef ter grootte n . Bij de Jonge: $k=127, n=150$.

C.De likelihoed van beide steekproeven gezamenlijk.

Beide steekproefuitkomsten zijn gegeven de kansen onafhankelijk :

$$L(s_x, s_k) = P(x, k | P(b), P(w|b), P(w|-b), s, n) \\ \propto P(b)^x (1-P(b))^{s-x} P(b|w)^k (1-P(b|w))^{n-k}$$

$$N.B. P(w) = P(b)P(w|b) + (1-P(b))P(w|-b)$$

en $P(b|w)$ volgt uit $P(b), P(w|b)$ en $P(w)$ immers

$$\frac{P(w|b)}{P(w|-b)} = \frac{P(b|w)P(w)}{P(b)} \cdot \frac{P(-b)}{P(-b|w)P(w)}$$

D. De likelihood van de risicoverhouding.

Werken met de risicoverhouding is vooral elegant omdat geldt:

$$r = \frac{P(w|b)}{P(w|-b)} = \frac{P(b|w)}{1-P(b|w)} \cdot \frac{1-P(b)}{P(b)}$$

dus we zijn geïnteresseerd in het product van twee "odd ratio's" ($p/(1-p)$ is de odd ratio van p).

De eerder afgeleide likelihoods kunnen getransformeerd worden naar likelihoods van odd ratio's middels het volgende Lemma:

Lemma:

$$\text{als } f(p) \propto p^a(1-p)^b$$

$$\text{en } r = p/(1-p)$$

$$\text{dan } f(r) \propto r^b/(1+r)^{a-b+2}$$

bewijs: $p = r/(1+r)$ en $1-p = 1/(1+r)$ dus

$$f(r) \propto dp/dr \cdot r^a(1+r)^{-a}(1+r)^{-b}$$

$$dp/dr = (1+r)^{-2}$$

$$f(r) \propto r^a(1+r)^{-(a+b+2)}$$

$$= (r/(1+r))^a/(1+r)^{b+2} \quad (\text{reken iets nauwkeuriger})$$

einde bewijs.

Nu geldt: de likelihood van r is de likelihood van het product van twee onafhankelijke ratio's. Deze is gelijk aan het product van de likelihoods van de ratios:

$$L(r|s_x, s_k) \propto L(s_x) \cdot L(s_k)$$

Geven wij de parameters van de "superpopulatie" (de populatie waar alle steekproeven uit getrokken zijn) aan als:

$$c = P(b)/(1-P(b))$$

$$t = P(b|w)/(1-P(b|w))$$

dan krijgen we met behulp van het lemma:

$$L(c) \propto c^x/(1+c)^{s+2} \propto$$

$$L(t) \propto t^k/(1+t)^{n+2}$$

er geldt $r = c/t$

dus (het rekenen met likelihoods is analoog aan het rekenen met kansdichtheden)

$$L(r) \propto \int (t \cdot L_c(rt) L_t(t)) dt \\ \propto \int ((rt)^x / (1+rt)^{s+2} t^k / (1+t)^{n+2}) dt$$

Deze integraal kan numeriek berekend worden.

Een alternatief eenvoudig algoritme dat wij gevolgd hebben gaat direct van c en t uit:

-verdeel het relevante bereik van c, t en d in kleine stapjes

-doorloop alle combinaties van c en t

-bereken steeds $L(c) \cdot L(t)$

-ga na in welk interval $r=c/t$ valt en tel daar de dichtheid $L(c) \cdot L(t)$ bij op.

Op deze wijze werkt men direct, zonder Jacobiaan, er zijn echter wel enige numerieke onnauwkeurigheden.

E. De kansverdelingen van de risicoverhouding bij niet informatieve prior.

Om de kansdichtheid van r weer te geven moet de likelihood vermenigvuldigd met een voorverdeling. In eerste instantie zijn we uitgegaan van een zgn. niet-informatieve voorverdeling. Dat is een voorverdeling die maakt dat de naverdeling vrijwel uitsluitend door de likelihood bepaald wordt. Helaas is hierover veel discussie mogelijk, zonder dat het voor het eindresultaat iets uitmaakt. Box & Tiao⁵⁾ gaan uit van een theorie met de charmante eigenschap dat de Bayesiaanse resultaten op veel punten overeenstemmen met klassieke uitspraken, echt overtuigend is het echter niet.

Ik heb een eenvoudige weg gekozen: een prior voor r , met als aardige eigenschap dat het dezelfde prior levert voor $1/r$ is:

$$f_0(r) \propto (1+r)^{-2}$$

er geldt o.a.

$$f_0(r) = 1/(1+r); \text{ dus } P(r > 2) = P(r < 1/2) = 2/3 \text{ etc.}$$

Door $f_0(r)$ met de reeds berekende $L(r)$ te vermenigvuldigen krijgen we nu een kansverdeling van r , gebaseerd op een niet informatieve voorverdeling. De uitspraken die kunnen worden gedaan (schattingen en betrouwbaarheidsintervallen), corresponderen sterk met klassieke uitspraken, staan alleen anders (begrijpelijker) geformuleerd.

In grafiek 1 zijn de resultaten van deze berekeningen weergegeven, voor de gegevens van de Jonge en voor de "wegens niet representativiteit gehalveerde" controlesteekproef.

F. Rechtvaardiging van het halveren van steekproefinformatie.

Niet-representativiteit kan natuurlijk vele vormen aannemen. Om mijn procedure te illustreren kies ik de volgende mogelijkheid:

Stel dat het percentage buikligging varieert per gemeente (afhankelijk van de opstelling van het consultatiebureau), laten we zeggen met een standaardafwijking van 0,1 bij een landelijk gemiddelde $\bar{b}=0,6$.

Men neemt nu een steekproef van s gemeenten en binnen iedere gemeente een steekproef van n personen. Zij $b|g$ de ware fractie buikligging per gemeente. Per gemeente treft men B gevallen van buikligging aan. Er geldt

$$E(B|g) = s(b|g). \text{ Var}(b|g) = 0,01. E(B) = EE(B|g) = \bar{b} = 0,6$$

$$\text{Var}(B|g) = s(b|g)(1-(b|g))$$

$$\begin{aligned} \text{Var}(B) &= \text{Var}E(B|g) + E\text{Var}(B|g) = s^2 \cdot 0,01 + s \cdot \bar{b} - s(\bar{b}^2 + 0,01) \\ &= (s^2 - s) \cdot 0,01 + s \cdot (\bar{b} - \bar{b}^2) \end{aligned}$$

Dit is de variantie van het aantal gevallen per gemeente, de totale variantie is a maal zo groot. Bij een totale steekproefomvang van n geldt $a=n/s$ en

$$\text{Var}(\Sigma B_g) = (n/s) \cdot \text{Var}(B) = n(s-1) \cdot 0,01 + n(0,6 - 0,6^2)$$

De laatste term is de variantie die gehanteerd wordt als er een landelijke steekproef wordt gehouden, de mogelijke variatie over gemeenten pleegt dan verwaarloosd te worden. Blijkens bovenstaand resultaat is daarvoor eigenlijk vereist dat in iedere gemeente 1 persoon wordt getrokken.

De correctie om de helft van de informatie terzijde te schuiven correspondeert met het werken met dubbele variantie. In dit voorbeeld is dat gerechtvaardigd wanneer

$$\text{Var}(\Sigma B_g) = 2n(0,6 - 0,6^2) \text{ dus } (s-1) \cdot 0,01 = (0,6 - 0,6^2)$$

dus als $s=25$. Met andere woorden: wanneer niet een steekproef is genomen van 300 in verschillende gemeenten, maar een steekproef van 12 gemeenten en in iedere gemeente 25 personen, dan is het gerechtvaardigd om de steekproefinformatie te halveren. Mooier zou overigens zijn om de variatie over gemeenten te modelleren. In het algemeen geldt dat door modelleren het onderzoek en de wijze van steekproeftrekken te verbeteren valt (de Vos(1987)).

G. De kansverdelingen van de risicoverhouding bij een informatieve prior.

De overgang naar een informatieve prior is eenvoudig. Door te experimenteren met functies van de vorm:

$$f(r) \propto r^a / (1+r)^b$$

heeft men ruime mogelijkheden om verwachting en variantie te variëren, terwijl de berekeningen heel eenvoudig blijven. Kiest men

$$b=2a+2$$

dan is het resultaat symmetrisch in r en $1/r$: er wordt van te voren niets over de richting van de risicoverhouding ingestopt, men spreekt alleen twijfel uit aan extremen. Hoe groter a , hoe sterker de prior rond 1 is geconcentreerd.

Ik heb gekozen $a=12$, $b=26$, de prior in grafiek 2 die na vermenigvuldiging met de likelihood de daar eveneens getekende naverdeling oplevert.

H. Van kansverhouding naar risicoverhouding.

Grafiek 3 is eenvoudigweg berekend middels:

$$f(r) = P(H_0)f(r|H_0) + P(H_1)f(r|H_1)$$

met

H_1 = de hypothese "op de buik leggen bepaalt het risico"

$$P(H_1)=3/4$$

$P(H_1)f(r|H_1)$ is de naverdeling.

H_0 = de hypothese "iets bepaalt het op de buik leggen en het risico"

$$P(H_0)=1/4$$

$P(H_0)f(r|H_0)$ is de voorverdeling, immers de gegevens zeggen niets over het risico.

I. Het incorporeren van de informatie over de toename van wiegedood.

We gebruiken de volgende gegevens:

1969: 10% buikligging, wiegedood 0,46 per 1000

1984: 60% buikligging, wiegedood 1,42 per 1000

De verhouding tussen de percentages wiegedood voorzover niet verklaard uit de toename van buikligging (een functie van de risicoverhouding) is ons uitgangspunt. Dit is:

$$o|r = \frac{1,42}{0,46} \times \frac{0,1r + 0,9}{0,6r + 0,4}$$

Aangezien geen van de factoren die met wiegedood verband houden een duidelijke stijging heeft vertoond nemen wij voor o een prior met 5% boven de 2 en 5% onder de 0,5. Een onderbouwing zou kunnen zijn de

kansverdeling van het verloop van diverse soortgelijke ziekten in diverse perioden en diverse landen.

De $f(o)$ is dezelfde functie als de prior uit grafiek 2 (toevallig kwam dit zo uit). De resulterende kansverdeling van r is:

$$f(r) = (0,6r+0,4)^{-2} \cdot f(o(r))$$

waarbij de eerste term (op een constante na) de Jacobiaan van de transformatie $|do/dr|$. De resulterende $f(r)$ staat in grafiek 4 (afgekapt bij 10, wat gezien de overige informatie niet uitmaakt). De informatie is onafhankelijk van de andere informatie, dus de naverdeling ontstaat uit vermenigvuldiging met de naverdeling uit grafiek 3.

Referenties.

- Aitken, C.G.C.(1987) "Attempting to Measure the Value of Evidence" The Professional Statistician (1987) Vol 6, Issue 3.
- Box, G.E.P., and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading.
- Darroch, J.(1987) "Probability and Criminal Trials" The Professional Statistician (1987) Vol 6, Issue 3:
- Kleinbaum,D.G., L.L. Kupper and H.M.Morgenstern(1982) "Epidemiologic Research" Van Nostrand Reinhold Company.
- Leamer, E.E. (1978), *Specification Searches, Ad-hoc Inference with Nonexperimental Data*, Wiley, New York.
- Lindley, D.V. (1971), *Making Decisions*, Wiley, New York.
- Racine,A; A.P.Grieve, H. Fluhter and A.F.M.Smith: "Bayesian Inference in Practice, Experiences in the Pharmaceutical Industry".(with discussion). Applied Statistics Vol 35 pp 93-150.
- Solomon,H. (1966) "Jurimetrics" In F.N. David ed. "Research papers in statistics" Wiley.
- de Vos, A.F. (1987) "On randomization, modeling and Experimental Design" Onderzoekverslag, Vrije Universiteit.

Ontvangen: 30-11-1987

Geaccepteerd: 14-12-1987