

EEN "METAMODEL" IS EEN BENADERING

Tom A.B. Snijders *

Samenvatting

Als reactie op Rotmans et al. (1990) wordt gesteld dat de controle, die in dat artikel op de kwaliteit van de benadering van het simulatiemodel door het gelineariseerde model wordt uitgevoerd, tamelijk zuinig is; en dat het gebruik van statistische toetsen van de coëfficiënten in het gelineariseerde model nergens op slaat.

Het artikel van Rotmans, van Ham en Kleijnen (1990) geeft een interessante vorm van gevoeligheidsanalyse van ingewikkelde simulatiemodellen. In navolging van Kleijnen (1987) wordt hierbij een regressiemodel van het feitelijke simulatiemodel een metamodel genoemd. Dat klinkt mooi, maar klopt de zaak een beetje op. Waar het om draait, is dat een benadering van het simulatiemodel wordt gemaakt met gebruik van het instrumentarium van de regressie-analyse en statistische proefopzetten. Het benaderend model is lineair als functie van bepaalde verstandig te kiezen variabelen, die functies zijn van de voor het simulatiemodel in te stellen variabelen, de inputs: als variabelen voor het lineaire model gebruikt men ten eerste de inputs zelf, en in tweede instantie ook bijvoorbeeld hun kruisprodukten (interactie-effecten van de inputs), enz.; ik heb het dan liever over een gelineariseerd model dan over een metamodel. Door middel van coëfficiëntschattingen en t-toetsingsgrootheden wordt de gevoeligheid van het model ten aanzien van wijzigingen in de diverse inputs onderzocht.

* Vakgroep Statistiek en Meettheorie, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, tel. 050-636188; ook Vakgroep Empirisch Theoretische Sociologie, Rijksuniversiteit Utrecht.

Deze benadering van gevoeligheidsanalyse via proefopzetten en regressie lijkt me zeer nuttig en vruchtbaar. Uit polemische overwegingen wil ik in dit commentaar echter liever kritiek dan lof geven. Die kritiek richt zich op twee punten.

1. Dat het metamodel een benadering voor het simulatiemodel is, wordt te weinig benadrukt; de controle van de kwaliteit van de benadering is mager.
2. Op een wat flauwe manier wordt op sommige plaatsen in het artikel heel zorgvuldig gesteld dat een statistisch model voor de aanpassingsfouten $Y - \hat{Y}$ misschien niet gerechtvaardigd is en dat men dan op andere argumenten kan terugvallen, maar wordt elders druk met Student's t-toetsingsgrootheden gewerkt.

Ad 1. Boven formule (3) wordt gesteld, dat dient te worden 'getoetst of het metamodel valide is'. Een definitie van het begrip 'valide' wordt niet gegeven; de lezer mag aannemen, dat de auteurs doelen op de kwaliteit van benadering (2). Aan het eind van par. 3.2 wordt gesteld dat gevalideerd is d.m.v. kruisvalidatie en extra runs. In par. 4.2.4 kijken de auteurs daarbij naar de grootte der relatieve residuen. In par. 4.3.2 bekijken ze ook of het teken van de regressie-coëfficiënten met de theoretische verwachting overeenkomt.

Met het oog op het design dat gebruikt is om inputwaarden te kiezen voor de simulaties, waaruit de coëfficiënten van het gelineariseerde model worden geschat, vind ik deze 'validering' aan de magere kant. Het simulatiemodel wordt gedraaid voor inputwaarden uit een fractioneel factoriëel 2^{k-p} design. Een volledige fractionele proefopzet zou teveel runs vereisen. Het aantal runs, 2^{k-p} , wordt zo gekozen, dat alle in het model opgenomen effecten schatbaar zijn. Zo worden er in par. 4.2.3 32 runs gemaakt om 22 parameters te schatten. Tien residuele vrijheidsgraden lijkt mij wel wat weinig voor een model waar je niet bij voorbaat heel erg zeker van kunt zijn; kruisvalidatie zonder extra data (d.w.z. extra runs) helpt daar niet bij. Bij de uiteindelijk gekozen modellen worden wel wat extra runs ter validatie gedraaid: het totaal is daarmee 27 runs voor 15 effecten in par. 4.2.4 en 140 runs voor 88 effecten in par. 4.3.3. Voor een betrekkelijk zwak gefundeerd regressiemodel blijft het aantal residuele vrijheidsgraden laag t.o.v. het aantal parameters,

zodat er een wissel wordt getrokken op het geloof dat de modelbouwer en de lezer hebben in de specificatie van het regressiemodel. Als er bij een dergelijk simulatiemodel wel meer runs zouden worden gedraaid, zou ik vanwege de mogelijkheden van niet-lineariteit wel wat voelen voor het voorstel van Easterling (1989, p. 427) om ook veel designpunten in het inwendige van de hyperkubus te nemen; liever dan een klein aantal er binnen en ook nog een paar er buiten (wat Rotmans et al. doen in par. 4.2.4).

Ad 2. Om een benaderend model aan te passen aan deterministische data kan de kleinste kwadraten methoden een goede manier zijn. Zijn niet-robustheid, elders vaak een nadeel, is hier een voordeel: één enkele grote afwijking is al heel ernstig. Rotmans et al. (1990, einde par. 2) stellen: 'Deze wiskundige analyse kan worden uitgebreid met een statistisch model voor de aanpassingsfouten $e = Y - \hat{Y}$.' En boven Tabel 4.2, waarin de laatste kolom Student t-waarden geeft: 'Indien geen statistisch submodel voor de aanpassingsfouten e wordt aangenomen, dan kan de laatste kolom verwaarloosd worden'. De auteurs laten het blijkbaar aan de lezer over of die zo'n statistisch submodel wenst aan te nemen, en wassen zelf hun handen in onschuld. Dat vind ik flauw, omdat ze zich van de verantwoordelijkheid afmaken maar wel in de tekst met het begrip "significantie" en in tabellen met toetsingsgrootheden schermen.

Hoe zou een kansmodel voor benaderingsfouten van een gelineariseerd t.o.v. een simulatiemodel kunnen worden gemotiveerd?

- a. Bayesiaans of empirisch Bayesiaans: de modelbouwer heeft bepaalde a priori ideeën (zie bv. Sacks et al., 1989, blz. 412 e.v.);
- b. de designpunten komen door loting tot stand (zie Kleijnen, 1987, 1990);
- c. de factoren die tot afwijkingen tussen simulatiemodel en gelineariseerd model leiden zijn zo grillig afhankelijk van kleine variaties in de inputfactoren ("klein" in verhouding tot de afstand tussen de designpunten) dat het lijkt alsof de afwijkingen "toevallig" zijn met de gewenste onafhankelijkheidsstructuur.

In het onderhavige artikel wordt geen van deze motiveringen genoemd, laat staan onderbouwd. Er wordt verwezen naar Kleijnen (1990) en naar Sacks et al. (1989), zodat mogelijke motiveringen wel indirect genoemd worden; maar een onderbouwing van de houdbaarheid van tenminste een van

deze overwegingen in het onderhavige geval wordt niet gegeven. Het lijkt me ook twijfelachtig, of zo'n onderbouwing hier mogelijk is. In elk geval lijkt het nodig, als er al aan een toevalsmodel wordt gedacht, een ingewikkelder onafhankelijkheids-struktuur te postuleren dan onafhankelijke verstoringen; Rotmans et al. noemen al dat Sacks et al. (1989) dergelijke modellen gebruiken, zonder hier verder op in te gaan. De bewijslast voor het gebruik van een kansmodel ligt, hoe dan ook, bij de auteurs; voordat ze een bevredigende argumentatie hebben gegeven behoren ze, mijns inziens, het gebruiken van op kansmodellen gebaseerde statistische uitspraken over onzekerheid achterwege te laten.

In het bijzonder begrijp ik niet, hoe hier de Student t-waarden geïnterpreteerd kunnen worden. Wat is hier een standaardfout, waar van replicatie geen sprake is? In Tabel 4.6 is zo'n mooi design gebruikt, dat alle t-waarden 2.79 keer zo klein zijn als de effektschattingen. Die effektschattingen, evenals informatie over de grootte der residuen, zijn waardevol; de Student t-waarden overbodig. Het doel is hier een gevoeligheidsanalyse; daarvoor zijn de effektschattingen bruikbaar. Statistische significantie en t-waarden hebben betrekking op toetsing; deze vindt hier echter plaats in een deterministisch model, waarbij statistische begrippen verwarring in plaats van verheldering leveren.

Literatuur

- J.P.C. Kleijnen (1987), "Statistical tools for simulation practitioners", New York: Marcel Dekker.
- J.P.C. Kleijnen (1990), "Statistics and deterministic simulation models: Why not?", Tilburg: KUB, Report FEW 435.
- J. Rotmans, G. van Ham en J.P.C. Kleijnen (1990), "Gevoeligheidsanalyse via metamodelen en proefopzetten, toegepast op een simulatiemodel voor het broeikasprobleem", KM, dit nummer.
- J. Sacks, W.J. Welch, T.J. Mitchell en H.P. Wynn (1989), "Design and analysis of computer experiments", *Statistical Science* 4, 409-435 (incl. discussie).
- R.G. Easterling (1989), "Comment" on J. Sacks et al. (1989), *Statistical Science* 4, 425-427.

Ontvangen: 19-09-1990

Geaccepteerd: 19-09-1990