Procescalculus: analyse van gegevens en statistische technieken

Procescalculus: analyse van gegevens en statistische technieken

In dit artikel, het zesde van een serie, wordt ingegaan op de verschillende gegevens die nodig zijn om het dynamische gedrag van een model met behulp van de procescalculus te onderzoeken en de bijbehorende statistische technieken. Gegevens zijn enerzijds onder te verdelen in exogene en endogene anderzijds in deterministische en stochastische gegevens. Ingegaan wordt op kansverdelingen; discrete en continue kansverdelingen worden beschreven. Het gebruik van histogrammen om gegevens te analyseren wordt getoond en het schatten van parameters van kansverdelingen wordt toegelicht. Random generatoren om gegevens te genereren, zoals aanwezig in de procescalculus, worden behandeld. Het valideren van modellen en de bijbehorende gegevens wordt toegelicht en het interpreteren van (waarnemings-) uitkomsten wordt vervolgens aangestipt. Enige opgaven zijn toegevoegd. Dit artikel wordt afgerond met een praktijkvoorbeeld.

Prof.dr.ir.J.Rooda is hoogleraar Werktuigbouwkunde aan de Technische Universiteit Eindhoven.
Dr.ir.J.H.A.Arentsen is gastdocent aan de faculteit der Werktuigbouwkunde van de Technische Universiteit Eindhoven.

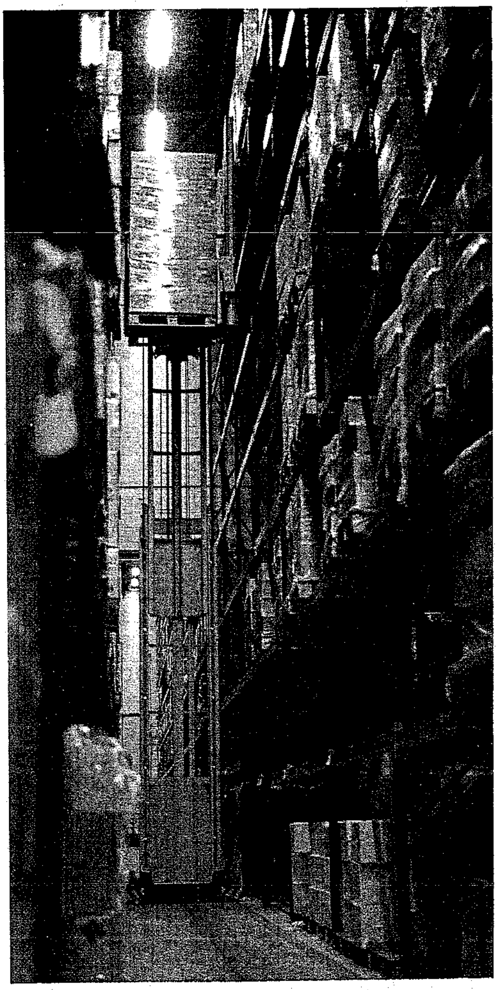
In [Rooda, 1991a] is beschreven, dat voor het verkrijgen van resultaten uit model-experimenten, modellen en gegevens vereist zijn. In vorige artikelen [Rooda, 1991b; Rooda, Arentsen, 1991; Rooda, Arentsen, Smit, 1992] is geïllustreerd hoe modellen van industriële systemen kunnen worden opgesteld. Het afbeelden van een werkelijkheid in een model, waarbij altijd vereenvoudigingen worden aangebracht, is een lastig intuïtief proces. Het verzamelen van de bij het model behorende gegevens blijkt eveneens een lastig en moeilijk karwei: de gegevens zijn strijdig, de gegevens zijn onvolledig, de gegevens zijn niet representatief, de gegevens zijn te omvangrijk, etc. Hierbij kan men denken aan invoergegevens als omsteltijden, bewerkingstijden, levertijden, te produceren typen en aantallen artikelen, en recepturen. En aan uitvoergegevens – model-resultaten – als doorlooptijd, bezettingsgraad, hoeveelheid onderhanden werk en voorraden. Tijdrovend is het verzamelen van invoergegevens middels tijdstudies en multi-moment opnames. Het verzamelen en interpreteren van gegevens van een fabriek ten behoeve van een op te stellen model kost al gauw enige maanden. Het interpreteren en pre-

senteren van uitkomsten dient eveneens nauwgezet door de modelbouwer te worden uitgevoerd. In dit artikel worden enige inleidende technieken beschreven hoe (invoer-) gegevens uit de praktijk en hoe (uitvoer-) gegevens uit modellen kunnen worden bewerkt.
Vaak is de ontwerper van industriële systemen daarnaast aangewezen op de statistiek om de ontbrekende gegevens met behulp van verantwoorde schattingen aan te vullen of om de gegevens met behulp van statistische technieken te vereenvoudigen. Daarvan worden de meest gebruikte statistische technieken behandeld. Tevens is aan het einde van het artikel een korte literatuurlijst toegevoegd die door de modelbouwer kan worden geraadpleegd.

Gegevens

Gegevens zijn te verdelen in zogenaamde exogene en endogene variabelen. In plaats van exogene en endogene variabelen worden de begrippen onafhankelijke en afhankelijke variabelen gebruikt. Men spreekt ook wel van ingangsvariabelen en uitgangsvariabelen.
Exogene (ingangs-) variabelen zijn stuurvariabelen en omgevingsvariabelen. Stuurvariabelen zijn variabelen waarmee men het model als het ware kan sturen of instellen. Een stuurvariabele is bijvoorbeeld het aantal produktiemiddelen dat men in een systeem wil installeren. Door het kiezen van andere instellingen van de stuurvariabelen verkrijgt men doorgaans een redelijk inzicht in het gedrag van het model. De omgevingsvariabelen zijn die variabelen die men uit de omgeving verkrijgt. Over het algemeen heeft men nauwelijks invloed op deze variabelen. Het aankomstpatroon van schepen in een zeehaven is een voorbeeld van een stochastische omgevingsvariabele.
Voorbeelden van endogene (uitgangs-) variabelen zijn onder andere doorlooptijden en bezettingsgraden.

Gegevens kunnen ook worden verdeeld in deterministische en stochastische variabelen. Een deterministische variabele is een variabele waarvan de waarde vastligt. Een model van een industrieel systeem kan echter zelden worden beschreven door middel van deterministische variabelen, omdat er bijna altijd toeval in het spel is. Dit toeval kan worden weergegeven door stochastische variabelen. Stochastische variabelen zijn variabelen die voldoen aan de regels van de stochastiek: een bepaalde kansverdeling bepaalt de waarde van een stochastische variabele. Hierbij is geen sprake van een enkele vaste waarde, maar van een patroon van waarden. Door het gebruik van stochastische variabelen vertoont het systeem een stochastisch gedrag. Deze stochastische variabelen, ook wel random variabelen genoemd, zijn soms afkomstig uit theoretische statistische verdelingen. De theorie van de statistische verdelingen wordt hierna voor de volledigheid beknopt weergegeven. De bewerkingstijden, zoals die in vorige artikelen zijn gebruikt, zijn voorbeelden van stochastische variabelen die afkomstig zijn van een uniforme verdeling. Bij zo'n verdeling is de kans op het optreden van elke gebeurtenis op een interval even waarschijnlijk.

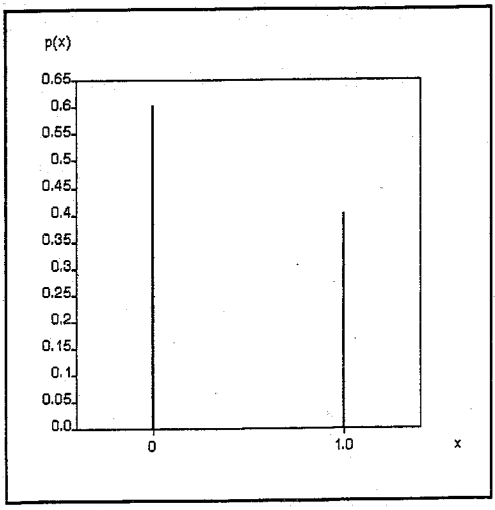


Statistische verdelingen

Statistische verdelingen kunnen worden onderverdeeld in discrete en continue verdelingen. Een bestelgrootte of het aantal defecten in een serie zijn waarden die afkomstig kunnen zijn van een discrete verdeling. Zo'n verdeling heeft een beperkt aantal, meestal gehele, waarden. Daarentegen zijn bewerkingstijden, transporttijden en omsteltijden waarden die afkomstig kunnen zijn van een continue verdeling.

Bij discrete en continue verdelingen speelt het begrip waarschijnlijkheid een rol. De waarschijnlijkheid geeft de verwachte frequentie aan van alle mogelijke waarden die de variabele kan aannemen. Alle verwachte frequenties tezamen bepalen de zogenaamde verdeling van de variabele. De verdeling wordt meestal

Fig.1. De punt-waarschijnlijkheden van de Bernouille-verdeling met μ=0,4 en σ^2=0,24



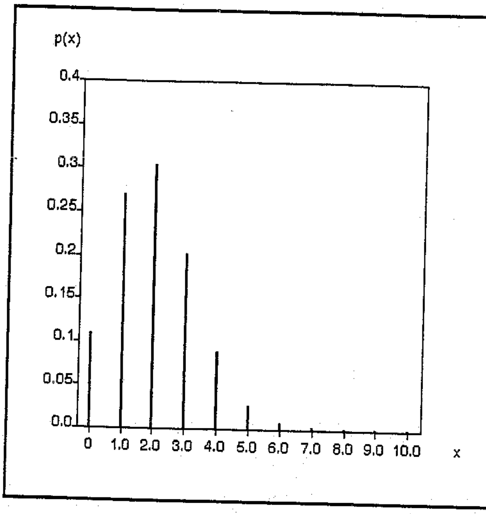


Fig.2. De punt-waarschijnlijkheden van de binominale verdeling met μ=2,0 en σ^2=1,6

weergegeven met behulp van een verdelingsfunctie.
Een functie F(x) wordt een cumulatieve verdelingsfunctie of kortweg verdelingsfunctie genoemd, indien [Zelen, Severo, 1965]:

- F(x) niet-dalend is, dat wil zeggen
  $F(a) \leq F(b)$ voor $a \leq b$
- F(x) overal continue vanaf rechts is, dat wil zeggen

$$F(x) = \lim_{\varepsilon \to 0+} F(x + \varepsilon)$$

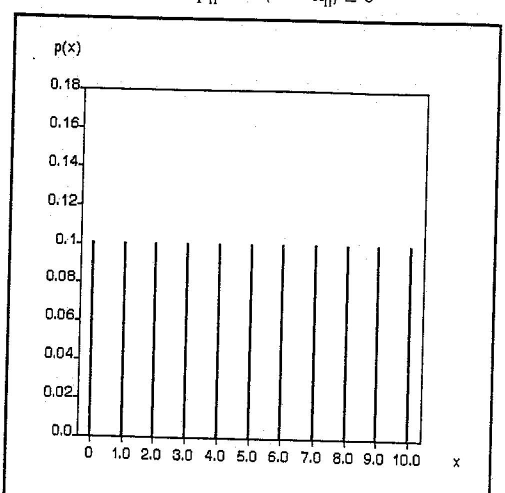
- $F(-\infty) = 0$, $F(\infty) = 1$

De kans op gebeurtenis $X \leq x$, waarin X een kansvariabele representeert, wordt weergegeven door F(x). Met andere woorden $Pr\{X \leq x\} = F(x)$.

Discrete verdelingen worden gekarakteriseerd door de kansvariabele X met een eindig aantal waarden $\ldots, x_{-1}, x_0, x_1, \ldots$ met een zogenaamde punt-waarschijnlijkheid

$p_n = Pr\{X = x_n\} \geq 0$

Fig.3.De punt-waarschijnlijkheden van de discreet-uniforme verdeling met μ=5,0 en σ^2=10,0



die slechts hoeven te voldoen aan de beperking

$$\sum_n p_n = 1$$

De bijbehorende verdelingsfunctie kan dan worden gegeven door

$$F(x) = Pr\{X \leq x\} = \sum_{x_n \leq x} p_n$$

waarbij de sommatie loopt over alle waarden van x waarvoor geldt $x_n < x$. De verzameling $\{x_n\}$ van waarden waarvoor $p_n \geq 0$ wordt het domein van de kansvariabele X genoemd.

Continue verdelingen worden gekarakteriseerd door een F(x) die absoluut continue is:

$$F(x) = Pr\{X \leq x\} = \int_{-\infty}^{x} f(t)dt$$

De afgeleide $f(x) = F'(x)$ wordt de dichtheidsfunctie genoemd. De waarden van x waarvoor geldt $f(x) > 0$ vormen het domein van de kansvariabele X.
Van iedere verdeling kan het gemiddelde en de variantie worden bepaald. Deze twee grootheden vormen met de punt-waarschijnlijkheden, dan wel de dichtheidsfunctie, de belangrijkste karakteristieke waarden. In tabel 1 worden de formules beschreven voor het berekenen van het gemiddelde en de variantie van discrete en continue verdelingen.

In industriële systemen spelen verschillende verdelingen een rol. De belangrijkste discrete en continue verdelingen zijn gegeven in de tabellen 2 en 3 [Zelen, Severo, 1965; Bratley, Fox, Schrage, 1987]. De punt-waarschijnlijkheden van de discrete verdelingen zijn weergegeven in figuur 1 tot en met 5. De dichtheidsfunctie van de continue verdelingen zijn weergegeven in de figuren 6 tot en met 11. Het toepassingsgebied van deze verdelingen is zeer uiteenlopend. Het voert hier te ver om op in te gaan. [Law, Kelton, 1982] geeft een overzicht van het toepassingsgebied van deze verdelingen.

Voordat bekend is volgens welke verdeling een variabele zich gedraagt is het eerst noodzakelijk dat de gegevens worden geanalyseerd. Deze analyse bestaat uit het vinden van een passing tussen deze gegevens en een theoretische verdeling. Een handig hulpmiddel hierbij zijn histogrammen.

Histogrammen

Een histogram wordt aan de hand van gegevens als volgt geconstrueerd:
De verzameling van gegevens wordt weergegeven door: $V = \{v_1, \ldots, v_n\}$. In tabel 4 is een verzameling van (n =) 100 bewerkingstijden van een machine weergegeven.
Het bereik van de waarnemingen is [0.3 , 9.5]. Een geschikt interval voor deze waarnemingen is het interval [0 , 10]. Dit interval wordt opgedeeld in 10 klassen. De breedte van iedere klasse bedraagt dan 1.
De verzameling van gegevens kan nu worden onderverdeeld in deze 10 klassen. Het aantal per klasse i wordt genoemd de absolute frequentie a_i. De (relatieve) frequentie per klasse i wordt nu gegeven door

$$f_i = \frac{a_i}{n}$$

In tabel 5 zijn het klasse interval,

8.3	4.1	2.0	3.3	3.3	3.6	4.3	3.1	3.5	1.9
2.0	1.7	1.1	1.3	2.0	2.8	1.4	5.6	5.0	1.0
1.9	0.9	7.4	2.6	2.5	5.2	2.2	2.5	1.1	2.2
3.9	5.3	1.2	6.7	2.9	1.0	3.9	5.4	3.9	4.3
2.3	5.8	4.7	9.5	3.4	3.0	1.9	5.1	2.6	1.8
1.2	0.7	3.1	2.4	3.5	1.7	2.1	2.7	1.8	2.1
2.2	3.1	2.8	3.1	0.6	3.1	1.9	3.9	0.3	7.8
3.0	2.1	0.8	2.2	1.5	2.5	0.3	3.0	3.2	1.3
2.8	6.0	3.2	3.7	1.1	1.0	8.2	2.4	2.2	3.9
2.5	1.2	4.7	1.7	2.2	4.8	3.2	1.8	4.3	3.4

Tabel 4. Bewerkingstijden op een machine in minuten.

de absolute frequentie en de (relatieve) frequentie weergegeven.

Klasse interval	Abs. freq. a_i	Freq. f_i
0 <= < 1	6	0.06
1 <= < 2	23	0.23
2 <= < 3	26	0.26
3 <= < 4	24	0.24
4 <= < 5	7	0.07
5 <= < 6	7	0.07
6 <= < 7	2	0.02
7 <= < 8	2	0.02
8 <= < 9	2	0.02
9 <= <10	1	0.01
	100	1.00

Tabel 5. Klasse intervallen, absolute en (relatieve) frequentie

In figuur 12 is het bijbehorende histogram weergegeven.
Er geldt dat

$\Sigma a_i = n$
$\Sigma f_i = 1$

De verzameling V kan nu worden gekarakteriseerd door:
de (relatieve) frequentie f_i,
het gemiddelde m:

$$m = \frac{\Sigma v_i}{n} = 3.03$$

en de (steekproef-) variantie s^2:

$$s^2 = \frac{\Sigma (m - v_i)^2}{n-1} = 3.30$$

De resultaten kunnen nu worden vergeleken met theoretische kansverdelingen. De vraag is dan of de gegevens uit een bepaalde verdeling afkomstig kunnen zijn. En indien dit het geval is welke parameters dan horen dan bij die verdeling.
Om vast te stellen of gegevens afkomstig zijn van een bepaalde verdeling wordt de vorm van het gevonden histogram vergeleken met de dichtheidsfunctie van een theoretische verdeling. De reden dat men de voorkeur geeft aan theoretische verdelingen is dat onder andere kleine afwijkingen in de waarnemingen worden geëlimineerd. Een eerste indicatie voor de vorm wordt gevonden in de aard van de gegevens. Zo zijn de tijdsintervallen tussen twee aankomsten van bijvoorbeeld schepen vaak Erlang verdeeld [Unctad, 1978] (figuren 8 en 6). Bewerkingstijden zijn vaak gamma verdeeld [Law, Kelton, 1982] (figuur)7.
De eerder gepresenteerde vorm van het histogram, in figuur 12, lijkt redelijk overeen te komen met een gamma-verdeling (figuur 7).
Vervolgens wordt vastgesteld wat de waarde van de parameter(s) van de verdeling is die het beste past bij het opgestelde histogram. Meestal is het voldoende om hierbij gebruik te maken van het gemiddelde en de variantie van het histogram. En deze te vergelijken met het gemiddelde en de variantie van een bepaalde verdeling. In figuur 13 is het histogram met de best passende gamma-verdeling aangegeven.
Tenslotte wordt getoetst of de verdeling voldoende met het histogram overeenkomt. Dit houdt in dat wordt gekeken of het histogram afkomstig kan zijn van een bepaalde verdeling. De χ^2-test is de bekendste toets [Mitrani, 1982]. De toets verloopt als volgt. De zogenaamde nulhypothese H_0 luidt: de verdeling van de bewerkingstijden is gamma verdeeld. Tabel 6 toont de klasse-intervallen, de absolute frequentie a_i, de absolute te verwachten frequentie verkregen door integratie van de dichtheidsfunctie van de gamma-verdeling e_i, en het aandeel in de toetsingsgrootheid

$$\frac{(\alpha_i - e_i)^2}{e_i}$$

De berekende waarde voor de toetsingsgrootheid bedraagt $\chi^2 = 4{,}71$. Het aantal vrijheidsgraden bedraagt 7 – 1 – 2 = 4. Bij een onbetrouwbaarheidsdrempel van $\alpha = 0{,}05$ bedraagt de kritieke waarde van de toetsingsgrootheid $\chi^2(0{,}05\ ;\ 4) = 9{,}49$ [Pears-

Discrete verdelingen	Continue verdelingen
$\mu = \sum_s x_s p_s$	$\mu = \int_{-\infty}^{+\infty} x\, f(x)\, dx$
$\sigma^2 = \sum_s (x_s - \mu)^2 p_s$	$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx$

Tabel 1. Gemiddelde en variantie van discrete en continue verdelingen

Klasse interval	a_i	e_i	$\frac{(\alpha_i - e_i)^2}{e_i}$
< 1	6	8,78	0,8785
1 <= < 2	23	24,05	0,0458
2 <= < 3	26	24,46	0,0969
3 <= < 4	24	18,00	2,0032
4 <= < 5	7	11,30	1,6368
5 <= < 6	7	6,46	0,0451
6 <= < 7	2	3,47	
7 <= < 8	2	1,78	
8 <= < 9	2	0,89	0,072
9 <= <10	1	0,43	
10	0	0,20	
			4,71

Tabel 6. Tableau voor de χ^2 toets

Tabel 5. Klasse intervallen, absolute en (relatieve) frequentie

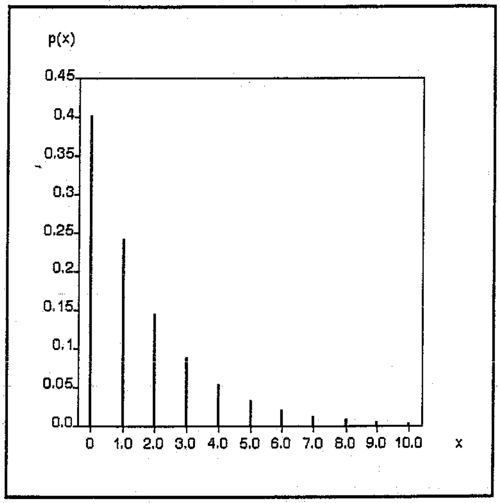


Fig.4.De puntwaarsdchijnlijkheden van de geometrische verdeling met μ=1,5 en σ^2=3.75

Tabel 2. Verschillende discrete verdelingen

naam	domein	punt-waarschijnlijkheid	parameter restricties	gemiddelde	variantie
Bernoulli	x ε {0,1}	1 - p als x = 0 p als x = 1	0 < p < 1	p	p (1-p)
binominaal	x ε {0,1,...,n}	$\binom{n}{x} p^x (1-p)^{n-x}$ met: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$	0 < p < 1	np	np (1-p)
discreet uniform	x ε {m,m+1,..., n-1, n}	$\frac{1}{n-m+1}$	m<n	$\frac{m+n}{2}$	$\frac{(n-m+1)^2-1}{12}$
geometrisch	x ε {0,1,...,∞}	$p(1-p)^x$	0 < p < 1	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	x ε {0,1,...,∞}	$\frac{e^{-m} m^x}{x!}$	0 < m < ∞	m	m

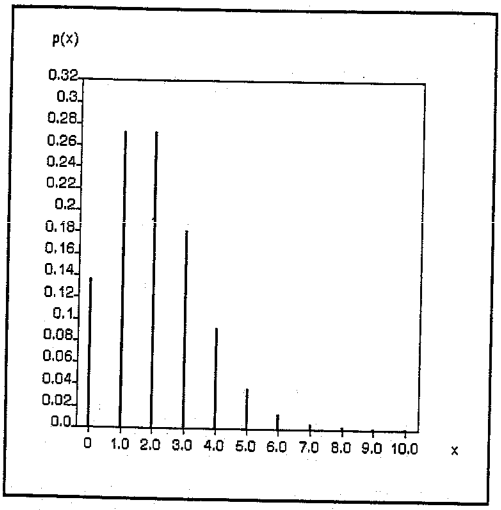


Fig.5.De punt-waarschijnlijkheden van de Poisson-verdeling met μ=2,0

on, Hartley, 1954]. Dus H_0 wordt niet verworpen omdat de berekende toetsingsgrootheid $\chi^2 = 4{,}71 < 9{,}49$. De hypothese, dat de verdeling van de bewerkingstijden gamma is verdeeld, wordt niet verworpen.

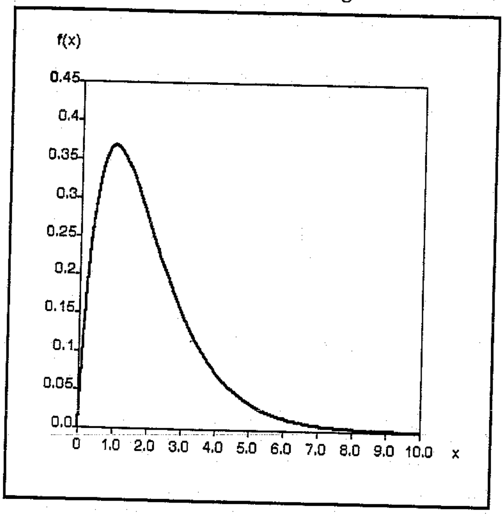
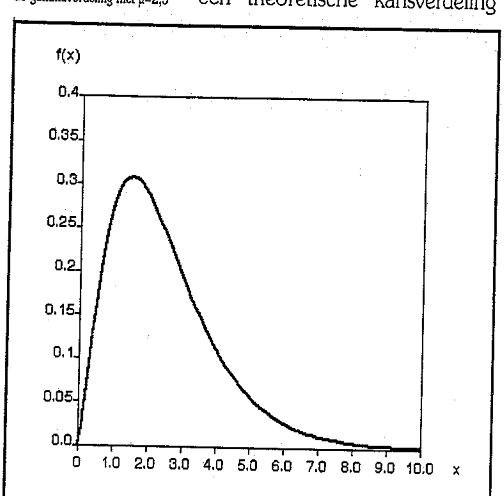


Fig.6.De dichtheidsfunctie van de Erlang-verdeling met μ=2,0 en σ^2=2,0

Fig.7.De dichtheidsfunctie van de gammaverdeling met μ=2,5



Indien blijkt dat een verzameling getallen niet te herleiden is tot een theoretische kansverdeling dan is de enige mogelijkheid om gebruik te blijven maken van de originele getallen. Deze getallen worden verwerkt tot een zogenaamde empirische kansverdeling. Hier wordt in dit artikel niet verder op ingegaan (zie Opgave 4).

Het model maakt nu op de volgende wijze gebruik van de verzamelde gegevens. Iedere keer dat het model behoefte heeft aan een gegeven, bijvoorbeeld een bewerkingstijd, dan wordt een getal uit een verdeling getrokken. Dit vindt plaats met behulp van zogenaamde random-generatoren. Deze zullen hierna worden behandeld.

Random generatoren

In het voorgaande is summier aangegeven hoe men aan de hand van gegevens een bepaalde verdeling kan bepalen. Aan de hand van trekkingen uit een verdeling kan vervolgens een bepaalde rij (random) getallen worden verkregen, die door het model worden gebruikt. Door het gooien van een dobbelsteen verkrijgt men een rij random getallen met waarden van 1 tot en met 6, namelijk trekkingen uit een discrete-uniforme verdeling. Een apparaat voor het trekken van lottoballen is zo ook geschikt om random getallen te trekken. In de modelbouw worden deze trekkingen doorgaans verkregen met behulp van een (pseudo-) random generator. Dit is een algoritme dat door een computer wordt uitgevoerd.

Een (pseudo-) random generator of kortweg generator genereert een rij getallen in het interval (0,N]. Hierbij dienen de gegenereerde getallen evenredig verspreid binnen een bepaald (deel-) interval op te treden. In de literatuur vindt men vele algoritmen voor het genereren van random getallen [Knuth, 1969; Press et al, 1986]. Andere verdelingen kunnen in principe worden verkregen door bewerking van de gevonden trekking. Zo wordt een uniforme kansverdeling u met onder- en bovengrens van respectievelijk 5 en 15 gevonden door toepassing van de betrekking:

$$u = a + (b - a)\,x$$

waarin a = 5, b = 15 en x de trekking uit een (0,1] verdeling.

Binnen de procescalculus is eveneens een random generator aanwezig. Deze generator levert getallen op in het interval (0,1]. Het aanmaken van een lijst c, het creëren van een (0,1] generator r, en het 100 keer uitvoeren van een trekking en het bewaren van deze trekking in de geordende verzameling, vindt plaats door bijvoorbeeld de volgende opdrachten:

```
c ← OrderedCollection new.
r ← Random new.
100 timesRepeat: [trekking ←
r next. c add: trekking]
```

Daarnaast zijn de meest voorkomende discrete en continue kansverdelingen aanwezig, zoals aangegeven in de tabellen 2 en 3. Het creëren van bijvoorbeeld een Erlang-verdeling e, met k = 2 en gemiddelde = 1.0, zie tabel 3, gebeurt door de opdracht:

```
e ← Erlang k: 2 mean: 1.0
```

Het trekken van een waarde, trekking, uit deze verdeling gebeurt door de opdracht:

```
trekking ← e next
```

Een model van een bewerkingsmachine, die produkten ontvangt, deze produkten bewerkt volgens tijden overeenkomstig de uit het voorbeeld gevonden gamma-verdeling en deze bewerkte produkten verzendt, kan dan als volgt worden beschreven:

```
Processor BewerkingsMachine
initializeTasks
  bewerkingsTijd ← Gamma
  mean: 3.03 variance: 3.30
  body
  produkt ← self receiveFrom:
  'in'.
  self workDuring: bewerkings-
  Tijd next.
  self send: produkt to: 'uit'
```

Modelvalidatie

Met behulp van gegevens is het mogelijk om met het model een

experiment uit te voeren. De vraag rijst nu of het model en de gegevens een goede beschrijving van de meestal nog te realiseren werkelijkheid zijn.
Een veel gevolgde werkwijze is die waarbij een experiment wordt uitgevoerd met het model en met historische gegevens. Hierbij kunnen de gegevens zijn voorbewerkt zoals boven beschreven. De uitkomsten worden vergeleken met overeenkomstige historische resultaten. Bij afwijkingen worden het model en/of de gegevens aangepast. Dit proces wordt zo vaak herhaald totdat voldoende overeenstemming is bereikt.
Bij afwezigheid van historische gegevens probeert men door verandering van (fictieve) gegevens een indruk te verkrijgen over de validiteit van het model.
Vervolgens wijzigt men het model of de gegevens. Hierbij kan men denken aan aantal produktiemiddelen, bewerkingstijden, te vervaardigen produkten, omsteltijden en besturingsstrategieën.
Indien sprake is van een stochastisch model, zoals de meeste modellen van industriële systemen zijn, dan is een enkele waarneming niet voldoende om een uitspraak te doen. Het is vergelijkbaar met een enkele worp van een dobbelsteen: indien een 6 verschijnt, dan kan men moeilijk volhouden dat bij ieder experiment een 6 bovenkomt. Meestal is men tevreden indien men na enige experimenten een uitspraak kan doen over de gemiddelde waarde van een uitgangsgrootheid, zoals de doorlooptijd.

naam	domein	dichtheids-functie	parameter restricties	gemiddelde	variantie
Erlang	$0 \leqslant x < \infty$	$ae^{-ax}\frac{(ax)^{k-1}}{(k-1)!}$	$0 < a < \infty$ $k \varepsilon \{1,2,\ldots,\infty\}$	$\frac{k}{a}$	$\frac{k}{a^2}$
gamma	$0 \leqslant x < \infty$	$ae^{-ax}\frac{(ax)^{p-1}}{\Gamma(p)}$ *	$0<a<\infty$ $0<p<\infty$	$\frac{p}{a}$	$\frac{p}{a^2}$
negatief exponentieel	$0 \leqslant x < \infty$	ae^{-ax}	$0<a<\infty$	$\frac{1}{a}$	$\frac{1}{a^2}$
normaal	$-\infty<x<\infty$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$	$-\infty<m<\infty$ $0<\sigma<\infty$	m	σ^2
uniform	$m-\frac{h}{2} \leqslant x \leqslant m+\frac{h}{2}$	$\frac{1}{h}$	$-\infty<m<\infty$ $0<h<\infty$	m	$\frac{h^2}{12}$
Weibull	$0 \leqslant x < \infty$	$ac(ax)^{c-1}e^{-(ax)^c}$	$0<a<\infty$ $0<c<\infty$	$\frac{1}{a}\Gamma\left(\frac{c+1}{c}\right)$ *	$\frac{1}{a^2}\left(\Gamma\left(\frac{c+2}{c}\right)-\left[\Gamma\left(\frac{c+1}{c}\right)\right]^2\right)$ *

Tabel 3. Verschillende continue verdelingen

$$* \; \Gamma(z) = \int_0^\infty e^{-t}\, t^{z-1}\, dt \text{ met } 0<z<\infty$$

Het is statistisch lastig om vast te stellen of een model valide (geldig) is. Meestal is het slechts mogelijk om een bepaald gevoel over de geldigheid te verkrijgen. Hierna wordt aangegeven hoe men uitgaande van een gevalideerd model (waarnemings-) uitkomsten en resultaten kan verkrijgen.

Interpretatie van waarnemingsuitkomsten

Een werkwijze om een bepaald resultaat te verkrijgen is de volgende:
De experimenteerperiode verdeelt men in tweeën. Gedurende het eerst deel probeert men het model in een evenwichtstoestand te brengen, de zogenaamde steady-state. Het tweede deel wordt opgesplitst in enige sub-delen. Van ieder sub-deel verzamelt men de uitkomsten. Aan de hand van deze uitkomsten wordt een gemiddelde en een variantie bepaald. Vervolgens wordt aangenomen dat de uitkomsten normaal verdeeld zijn en dat het resultaat van ieder sub-deel onafhankelijk is van het vorige. Met behulp van een (Student's-t) toets bepaalt men dan een betrouw-

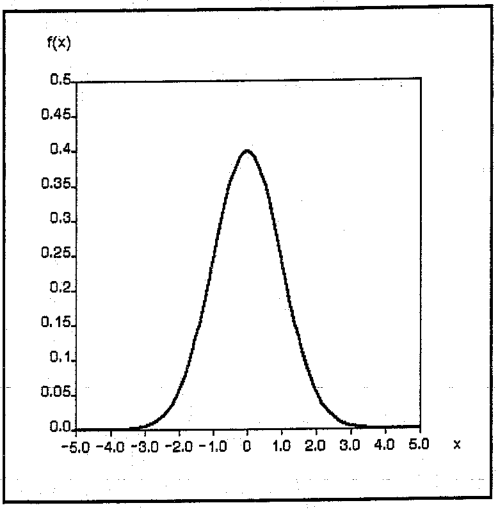


Fig.9.De dichtheidsfunctie van de normale verdeling met $\mu=0$ en $\sigma^2=1,0$

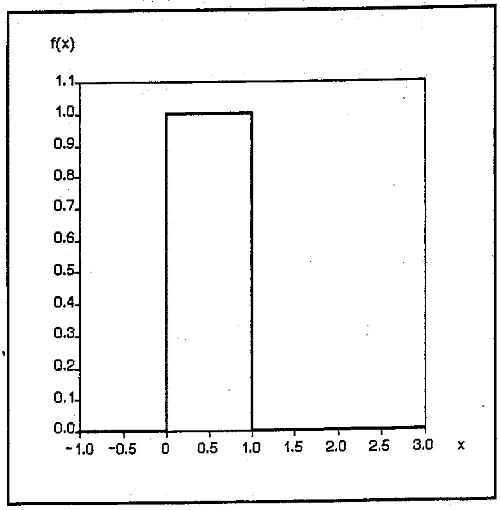


Fig.10.De dichtheidsfunctie van de uniforme verdeling met $\mu=0,5$ en $\sigma^2=0,08$

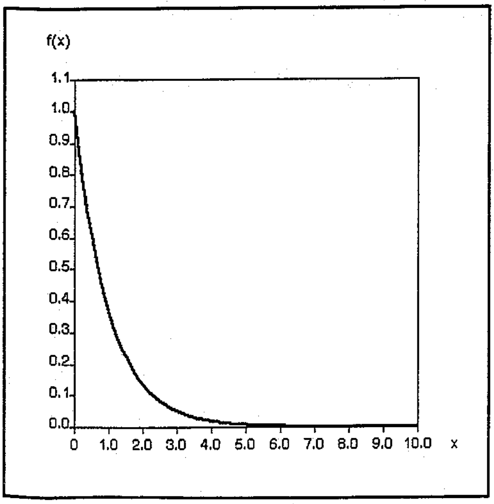


Fig.8.De dichtheidsfunctie van de negatief exponentiële verdeling met $\mu=1,0$

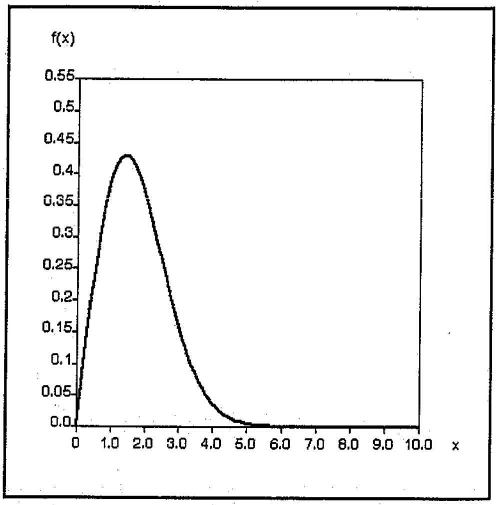


Fig.11.De dichtheidsfunctie van de Weibull-verdeling met μ=1,77 en σ^2=0.86

baarheidsinterval van de absoluut verkregen uitkomsten. Een resultaat kan luiden: de gemiddelde doorlooptijd ligt tussen de 18 en 24 dagen met een 95%-zekerheid.
Vaak is dit de enige werkwijze. Deze werkwijze heeft echter nadelen. Soms treedt geen steady-state op. Ook is soms moeilijk vast te stellen wanneer de steady-state begint. Daarnaast is niet duidelijk hoe groot een subdeel dient te zijn. Tevens is het twijfelachtig of de waarnemingsuitkomsten wel normaal zijn verdeeld. Bovendien kan men twijfelen aan de onafhankelijkheid van de (deel-) uitkomsten.
Soms is het mogelijk om een andere werkwijze te volgen. Deze werkwijze is om uitkomsten van twee onafhankelijke experimenten met elkaar te vergelijken, waarbij van bijvoorbeeld twee verschillende strategieën wordt uitgegaan. De uitkomsten van ieder experiment bestaan uit een verdeling. Vervolgens kan men deze verdelingen met elkaar vergelijken. Een uitspraak kan dan luiden: indien een strategie, gebaseerd op het idee dat een order met de kortste bewerkingstijd het eerst wordt bewerkt (SPT – shortest processing time), vergeleken wordt met een andere strategie, gebaseerd op het idee dat wie het eerst komt het eerst maalt (FIFO – first in first out), dan blijkt dat de eerste strategie een gemiddeld 10% kortere doorlooptijd oplevert.

In sommige gevallen heeft men te maken met modellen waarbij slechts een enkele variabele als parameter optreedt. Men is dan geïnteresseerd in het effect van een verandering van de variabele, in deze context ook wel factor genoemd, voor de uitkomsten. Hierbij kan men bijvoorbeeld denken aan het verkorten van de bewerkingstijd van een enkele machine. In veel gevallen is er echter sprake van meerdere variabelen. Zonder kennis van proefopzetten, regressie- en variantie-analyse zullen de experimenten veel uitkomsten produceren: ze zullen niet tot verantwoorde uitspraken leiden. Het voert te ver om in dit artikel hier nader op in te gaan. Er wordt hier verwezen naar met name [Kleynen, 1987].

Nabeschouwing

In dit artikel is inzicht gegeven in het verwerken van gegevens die kunnen worden gebruikt bij het bepalen van het dynamische gedrag van modellen van industriële systemen. Verschillende statistische technieken zijn beknopt behandeld, zoals de theorie van de statistische verdelingen, het gebruik van histogrammen, het schatten van verdelingen en parameters, het gebruik van random generatoren, modelvalidatie en de interpretatie van waarnemingsuitkomsten.
Bij het schrijven van dit artikel is gebruik gemaakt van Van Driesum [1991].
Dit artikel is te beschouwen als een inleiding op de grote hoeveelheid literatuur die beschikbaar is op dit gebied. Met name de volgende literatuur wordt aanbevolen: [Bratley, Fox, Schrage, 1987; Fishman, 1978; Kleynen, 1974, 1987; Law, Kelton, 1982].

In de voorgaande artikelen in de serie over de procescalculus is gebruik gemaakt van een object-georiënteerde modelleertechniek. Dit kwam tot uiting in de beschrijving van de verschillende processoren. In deze artikelen werd verondersteld dat de betekenis van deze beschrijvingen uit de context duidelijk zou worden. In een volgend artikel zal speciale aandacht worden geschonken aan deze object-georiënteerde modelleertechniek.

Toegift
Foto 1 toont een deel van het magazijn van de farmaceutische groothandel OPG te Utrecht. Het assortiment omvat circa 18500 verschillende produkten.
Het is de bedoeling om voor deze produkten een nieuw distributiecentrum te ontwerpen. In dit distributiecentrum zullen de produkten worden in-, op- en uitgeslagen. Als indelingscriteria worden die eigenschappen gebruikt die van belang zijn voor de keuze van het magazijnsysteem, namelijk volume, omzetsnelheid en voorraad.
Omdat het praktisch onmogelijk is om 18500 produkten in een relatief korte tijd te inventariseren zijn voor elk indelingscriterium drie klassen geformuleerd. Op deze wijze ontstaat een klasse pro-

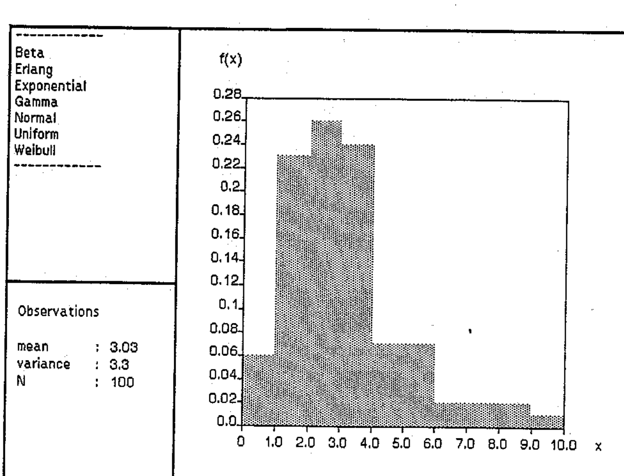


Fig.12.Het histogram van de bewerkingstijden

dukten met hoge omzetsnelheid, groot volume en hoge voorraad, een klasse met hoge omzetsnelheid, groot volume en middelmatige voorraad, etc. Het totale assortiment wordt hierdoor ingedeeld in $3^3 = 27$ klassen. Uit het assortiment is nu twee keer een steekproef van circa 280 produkten getrokken. Van ieder van deze produkten zijn de indelingscriteria vastgesteld. Elk produkt is op basis hiervan in één van de 27 klassen ingedeeld. De twee steekproeven zijn met elkaar vergeleken. Hierbij is geconstateerd dat er geen significante verschillen tussen de aantallen per klasse optraden. Ook na een gevoeligheidsanalyse ten opzichte van de (arbitrair) gekozen klassegrenzen zijn geen aanpassingen noodzakelijk gebleken. De resultaten van de verkregen steekproeven worden als empirische kansverdelingen gebruikt bij het modelleren van het nieuwe distributiecentrum.
Deze studie is uitgevoerd door van E.J.A. van Melick en van F.G.J. van Kempen.

Opgaven

1. Schets de punt-waarschijnlijkheidsfunctie en de dichtheidsfunctie van een discrete uniforme verdeling. Bepaal het gemiddelde en de variantie. Neem hierbij een zuivere dobbelsteen in gedachten.

2. Schets de punt-waarschijnlijkheidsfunctie en de bijbehorende dichtheidsfunctie van een binomiale verdeling. Bepaal het gemiddelde en de variantie. a) Neem n = 7 en p = 0,4. b) Neem n = 7 en p =0,01.

3. Schets de punt-waarschijnlijkheidsfunctie en de bijbehorende dichtheidsfunctie van een Poisson verdeling. Bepaal het gemiddelde en de variantie. Neem n = 7 en p = 0,01. Vergelijk deze figuur met het antwoord uit 2b.

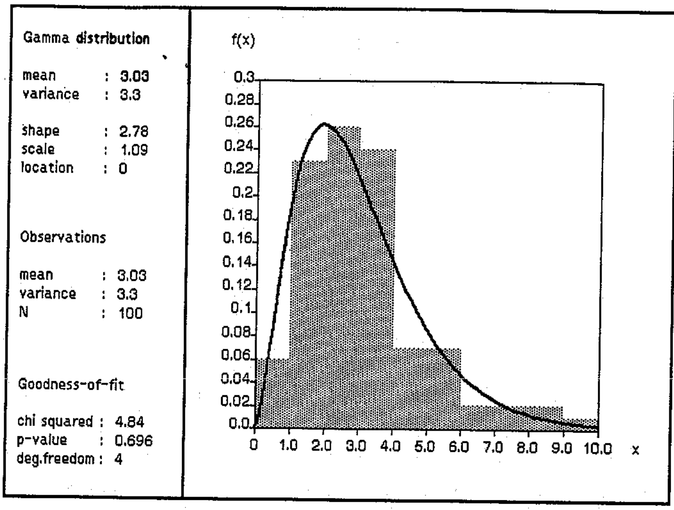


Fig.13.Het histogram van de bewerkingstijden met de best passende gammaverdeling

4. Van een histogram is bekend dat hier geen theoretische kansverdeling bij past (stel zelf een histogram samen). Dit histogram bestaat uit k aaneengesloten intervallen, $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$ zodat het j-de interval bevat n_j waarnemingen, waarbij $n_1 + n_2 + \ldots + n_k = n$. g(x) wordt nu gedefinieerd door: $g(a_0) = 0$ en $g(a_j) = (n_1 + n_2 + \ldots + n_j)/n$ voor $j = 1,2,\ldots,k$. Teken de functie g. g blijkt een continue, gedeeltelijk lineaire empirische verdelingsfunctie te zijn. Stel de waarde van x $(a_0 < x \leq a_k)$ vast, indien een waarde van g met behulp van een random-generator wordt getrokken. (Deze werkwijze kan worden gebruikt voor het trekken van getallen uit continue empirische verdelingen)

Literatuur

Bratley P., Fox B.L., Schrage L.E., 1987, A guide to simulation, 2nd edition, Springer Verlag New York.

Driesum G.J. van, 1991, Statistische aspecten van produktiesystemen, Afstudeerverslag, Sectie Automatisering van de Produktie, Faculteit Werktuigbouwkunde, Technische Universiteit Eindhoven.

Fishman G.S., 1978, Principles of discrete event simulation, John Wiley & Sons, New York.

Kleynen J.P.C., 1974, Statistical techniques in simulation, Part 1 and 2, Statistics: textbooks and monographs, volume 9, Marcel Dekker Inc., New York.

Kleynen J.P.C., 1987, Statistical tools for simulation practioners, Statistics: textbooks and monographs, volume 76, Marcel Dekker Inc., New York.

Knuth D., 1969, The art of computer programming, Volume 2, Addison-Wesley Publ. Comp., Reading (Ma).

Law A.M., Kelton W.D., 1982, Simulation modelling and analysis, McGraw-Hill Book Comp., New York.

Mitrani I., 1982, Simulation techniques for discrete event systems, Cambridge University Press, Cambridge.

Pearson E.S., Hartley H.O., 1954, Biometrika Tables for statisticians, Cambridge University Press, Cambridge.

Numerical recipes, Cambridge University Press, Cambridge.

Shannon R.E., 1975, Systems simulation, the art and science, Prentice Hall, Englewood Cliffs (NJ).

Unctad, 1978, Port development, A handbook for planners in developing countries, United Nations, New York.

Zelen M., Severo N.C., 1965, Probability Functions, in: Abramowitz M., Stegun I., 1965, Handbook of mathematical functions, Dover Publications, New York Englewood Cliffs (NJ).

Rooda J.E., 1991a, Procescalculus 2: Systemen, modellen en geschiedenis van de procescalculus, I² Werktuigbouwkunde 7(8), 36 - 39.

Rooda J.E., 1991b, Procescalculus 3: Definities en begrippen, I² Werktuigbouwkunde 7(10), 35 - 40.

Rooda J.E., Arentsen J.H.A., 1991, Procescalculus 4: Modelleren van flow-produktie fabrieken, Mechanische Technologie 1(1), 10 - 20.

Rooda J.E., Arentsen J.H.A., Smit G.H., 1992, Procescalculus 5: Modelleren van job-produktie fabrieken, Mechanische Technologie 2(2), 36 - 45.