STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Rapport S 296[a]

Wachttijden

door

C.J. van Tooren

juli 1965

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non - profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Inleiding.

Een ieder kent uit eigen ervaring situaties waarin hij moet wachten. Als wij een winkel binnengaan, en er blijken meer klanten dan verkoopsters te zijn, dan worden wij niet onmiddellijk geholpen. Over het algemeen zal men hieruit niet de conclusie trekken dat het aantal verkoopsters te klein is. Omdat de aankomst van de klanten niet volgens een vast schema verloopt, blijft immers altijd de mogelijkheid bestaan dat verscheidene klanten ongeveer tegelijk binnenkomen. Zou men het aantal verkoopsters zo groot kiezen dat een klant vrijwel nooit hoeft te wachten, dan zou van de kant van de verkoopsters zeer veel gewacht moeten worden.

Iets anders ligt de zaak bij een tandarts die volgens afspraak behandelt, en waar de aankomsten dus wèl volgens een vast schema verlopen. Incidenteeel zal het voorkomen dat men hier toch moet wachten, bijv. omdat de tijd die de tandarts voor een bepaalde patiënt nodig heeft, wel eens langer kan zijn dan werd verwacht. Als de tandarts de afspraken met zijn patiënten zo zou maken dat deze nooit hoeven te wachten, zou hij zelf een veel groter deel van de dag werkeloos moeten doorbrengen, en zijn tijd dus veel inefficienter gebruiken.

Deze variabiliteit van de benodigde bedieningstijd heeft natuurlijk ook in het voorbeeld van de winkel invloed op de tijd dat men eventueel moet wachten.

De essentiële kenmerken van bovenbeschreven situaties blijkt men in veel meer gevallen terug te vinden dan men aanvankelijk zou vermoeden. Talrijke praktijkproblemen zijn terug te brengen tot situaties waar vraagelementen (klanten) die een bepaalde dienst verlangen, aankomen bij een bedieningsinstallatie (loket), waar rijvorming kan optreden. In de volgende tabel zijn enkele van dergelijke situaties opgesomd.

	systeem	aankomende vraag-elementen	rij	bedienings-installatie (loket)	aard van de dienst
1)	winkel	klanten	wachtende klanten	toonbank, verkoopsters	verkoop artikelen
2)	luchthaven	vlieg-tuigen	rondcirke-lende vlieg-tuigen	landingsbaan	landen
3)	haven	schepen	voor anker liggende schepen	kade, lossers	lossen
4)	telefoon-dienst	oproepen	nog op uitvoe-ring wachtende gesprekken	telefoon-installatie	gesprek
5)	fabriek	defect rakende machines	defecte machines	reparateur	reparatie
6)	taxi-standplaats	passagiers	wachtende passagiers	taxi's	vervoer passagier
7)	taxi-standplaats	taxi's	wachtende taxi's	passagiers	rit
8)	kruispunt met verkeers-lichten	auto's	wachtende auto's	kruispunt en de weg erna	door laten rijden
9)	stad	brand-meldingen	brandende gebouwen	brandweer	blussen
10)	leger	berichten	nog te decoderen berichten	decodeurs	decodering
11)	restaurant	gasten	gasten die nog niets bestelden	ober	bestelling opnemen
12)	kantoor + magazijn	geprodu-ceerde goederen	voorraad	orders	groothandel
13)	kantoor + magazijn	orders	achterstallige orders	voorraad	levering artikel

In bijna al deze situaties zal worden geprobeerd het wachten zoveel mogelijk te beperken. Bij voorraadproblemen zoals geformuleerd in voorbeeld 12, gaat het er echter juist om, een voorraad (dus een rij) in stand te houden. Deze mag enerzijds niet uitgeput raken, en anderzijds niet te groot worden.

Een van de meest voor de hand liggende maatregelen om opstoppingen tegen te gaan is het opvoeren van de bedieningscapaciteit. Tegenover de kostenbesparing die uit het verhelpen van de opstopping voortvloeit, staat echter de kostenstijging die in het algemeen gepaard gaat met de uitbreiding van de bedieningsinstallatie. Met behulp van de wachttijdtheorie kan men deze beide effecten quantitatief vergelijken, en een tussenweg trachten te vinden waarin noch het wachten van de vraagelementen, noch het exploiteren van de bedieningsinstallatie extreem hoge kosten met zich meebrengt. Het is noodzakelijk om hierbij gebruik te maken van wiskunde, vooral omdat de samenhang tussen de verschillende factoren die het gedrag van het systeem bepalen dermate gecompliceerd is, dat de resultaten vaak in het geheel niet aansluiten bij onze intuïtie. Het kan bijv. gebeuren dat, bij gelijk blijvende bedieningscapaciteit, een kleine stijging in de toeloop der klanten een enorme invloed heeft op hun wachttijd. Een andere reden waarom de wachttijdtheorie wordt gebruikt is de omstandigheid dat experimenteren met de betrokken systemen vaak zeer kostbaar is.

De eerste resultaten van de theorie bereikt door ERLANG in 1908, kwamen ten goede aan het telefoonwezen. Vóór de tweede wereldoorlog zijn de toepassingen voornamelijk tot dit terrein beperkt gebleven. Tijdens de oorlog werden alle communicatie- en transportmiddelen zoveel intensiever gebruikt, dat de vraag het aanbod op zijn minst dicht benaderde, een situatie waar de kans op opstoppingen, zoals gedemonstreerd zal worden, ontstellend toeneemt. Sindsdien is de theorie een bredere basis gegeven (vooral door de Engelsen en later de Fransen), en tevens een breder toepassingsgebied (speciaal door de Amerikanen).

2. Wiskundige beschrijving van wachttijdproblemen.

De beschrijving van het beslissingsprobleem bij wacht-

tijdsituaties in de terminologie van Hoofdstuk I is zeer wel mogelijk. De historische groei is echter zodanig geweest dat de wachttijdtheorie in feite neerkwam op toegepaste waarschijnlijkheidsrekening, d.w.z. het eigenlijke beslissingsaspect bleef op de achtergrond. De alternatieven bij een concrete beslissing zijn meestal zo gering in aantal dat de consequenties van alle alternatieven worden doorgerekend, en naar aanleiding van de resultaten een keuze wordt gemaakt. Daarom zal hier worden nagelaten de algemene terminologie in extenso in te voeren.

Het essentiële begrip "wiskundig model" is reeds in Hoofdstuk I ingevoerd en toegelicht. In het wiskundige model van een wachttijdprobleem worden, met behulp van een stelsel vergelijkingen, de kwantificeerbare grootheden die in het proces een rol spelen in een onderlinge samenhang geplaatst, welke de fundamentele betrekkingen tussen deze grootheden weergeeft. Zoals al uit § 1 van dit hoofdstuk is gebleken, is het een essentieel kenmerk van wachttijdsituaties dat sommige factoren buiten onze controle staan. Het fluctuerende karakter van dergelijke factoren kan in het wiskundige model vaak goed worden beschreven met behulp van stochastische variabelen, d.w.z. variabelen die een kansverdeling bezitten.

Teneinde de modellen hanteerbaar te doen zijn, zullen alleen processen beschreven worden die de eigenschap hebben <u>stationair</u> te zijn. Dit begrip is zeer moeilijk en laat zich in woorden niet scherp vatten. Daarom zal hier worden volstaan met slechts een indruk van de strekking van het begrip stationair te geven.

Men stelle zich een loket voor dat zojuist geopend is. De bediende wacht op de eerste klant. Met kans één treft deze bij aankomst nul klanten in het systeem aan. De bediende

begint onmiddellijk deze eerste klant te helpen. De tweede klant die arriveert zal 0 of 1 klanten in het systeem aantreffen, al naar gelang de bediening van de eerste klant bij zijn aankomst reeds voltooid is of niet. Evenzo zal de derde klant 0,1 of 2 klanten in het systeem aantreffen, waarbij elk van deze mogelijkheden een bepaalde kans op realisatie heeft. Zo voortgaande zal men voor de n^e klant kunnen uitrekenen met welke kans hij bij aankomst resp. 0,1,2,...,n-1 klanten in het systeem zal aantreffen. Kortom bij iedere klant hoort een kansverdeling voor het aantal in het systeem aangetroffen klanten. Dit aantal wordt de toestand van het systeem genoemd. Opgemerkt dient te worden dat men bij de berekening van deze kansverdeling geen informatie verwerkt over de feitelijke toestanden, welke voorafgaande klanten hebben ontmoet. De kansverdelingen worden dus berekend aan het begin van het proces. Indien nu blijkt dat deze kansverdelingen van de toestand, aangetroffen door de n^e en de volgende klanten, bij stijgende n steeds meer op elkaar gaan lijken (convergentie van de kansverdeling), spreekt men in het limietgeval van een stationair proces.

Gedraagt een proces zich als bevond het zich in het zojuist beschreven limietgeval, en nummert men de nog aan te komen klanten, dan is de kansverdeling van de toestand van het systeem onafhankelijk van het rangnummer dat deze klanten bij aankomst in het systeem krijgen.

De toestand van het systeem verandert nu wel in de tijd, maar de wijze van veranderen wordt onafhankelijk van de tijd.

In de wachttijdtheorie doet men bij het berekenen van de kansverdeling van de toestand, waarin de eerstvolgende klant het systeem zal aantreffen, alsof aan het loket reeds een onbeperkt groot aantal klanten zijn verwerkt, terwijl men over de <u>feitelijke</u> loop van dat proces geen informatie heeft.

Indien de kansverdeling van de toestand waarin een klant het systeem zal aantreffen gelijk is aan die welke behoort bij een klant met een onbegrensd hoog rangnummer, dan zegt men dat de klant het systeem aantreft in de stationaire toestand. Deze benaming is enigszins misleidend omdat het stationair zijn niet op de toestand betrekking heeft maar op de kansverdeling van de toestand.

Stel dat door de keuze van een gedetermineerde begin-toestand niet gestart wordt in de stationaire toestand, zoals in het zojuist geschetste geval toen gestart werd met toestand 0. Strikt genomen komt het systeem nu pas in de stationaire toestand na het verstrijken van oneindig veel tijd. Gebleken is echter dat vaak een systeem, dat na onbegrensde tijd in de stationaire toestand kan komen, zich reeds na een korte aanlooptijd met redelijke precisie gedraagt als ware het in de stationaire toestand (snelle convergentie der kansverdelingen naar die welke behoort bij de stationaire toestand). De gebruikte idealisering zal hier dus resultaten geven die niet ver van de werkelijkheid afliggen. Deze procedure wordt in de statistiek en de waarschijnlijkheidsrekeningen veel gebruikt: Men benadert een eindig geval met het dikwijls beter te hanteren limietgeval.

In de gevallen waar de processen in het praktijkprobleem duidelijk niet stationair zijn, zoals bij het optreden van spitsuren, en waar het essentieel is in het model met deze omstandigheid rekening te houden, is het vaak voldoende om de beschouwde periode in stukken te hakken, die ieder kort genoeg zijn om het proces hierin stationair te onderstellen.

Het skelet van het model voor de wachttijdsituatie is zeer eenvoudig (zie fig. 1).

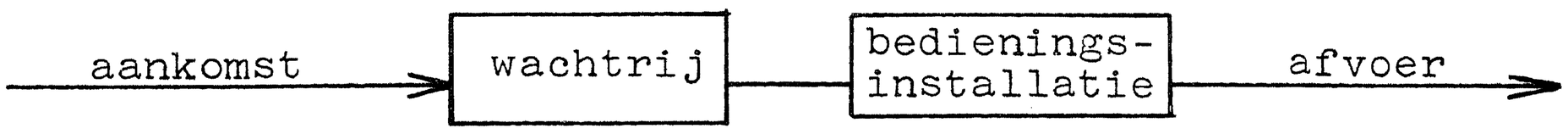


fig. 1

De factoren die het systeem beheersen kunnen dus als volgt worden onderscheiden

1. Het patroon van de aanvoer der vraagelementen
2. Het patroon van de benodigde bedieningstijd
3. Het selectiepatroon van de vraagelementen uit de rij
4. De grootte en de inrichting van de bedieningsinstallatie

Als over al deze factoren gedetailleerde gegevens voorhanden zijn, kunnen uit het wiskundig model verschillende grootheden worden berekend, zoals

1. de rijlengte
2. het aantal vraagelementen in het systeem
3. de wachttijd van een vraagelement
4. de tijd die een vraagelement in het systeem doorbrengt
5. de leegloop (de fractie van de tijd dat er geen vraagelementen in het systeem zijn)
6. de "treinlengte" (de tijd dat de bedieningsinstallatie onafgebroken in bedrijf is).

De grootheden in dit lijstje zijn natuurlijk niet alle even interessant voor de verschillende bij een wachttijdprobleem betrokken personen. Zo zal de architect van een te bouwen V.V.V. kantoortje voor de afmetingen van zijn ontwerp hoofdzakelijk geinteresseerd zijn in de kansverdeling van het aantal vraagelementen in het systeem, een machinereparateur zal meer belang stellen in de leegloop, terwijl de klanten van een zelfbedieningswinkel zich voornamelijk om hun wachttijd zullen bekommeren.

Soms is het voldoende om alleen de verwachting van dergelijke grootheden te berekenen, maar het kan voorkomen dat het beslist nodig is om meer gegevens over de kansverdeling te kennen. Als voorbeeld beschouwen we een taxistandplaats. De rij wachtende passagiers kan nu ook negatief worden: er staat dan een rij wachtende taxi's. Indien zowel de taxi's

als de passagiers op aselecte wijze bij de standplaats aankomen met hetzelfde gemiddelde aantal aankomsten per tijdseenheid, dan zal het gemiddelde aantal wachtende passagiers 0 zijn. Men kan evenwel bewijzen dat de variantie van de rijlengte op den duur willekeurig grote waarden aanneemt. Bovendien nadert de kans het aantal eenheden in beide rijen kleiner is dan een willekeurig eindig getal M, naar nul. De verwachting alleen geeft hier dus geen adequate informatie over het systeem.

3. Dobbelsteenvoorbeeld.

Het voorbeeld dat wij in deze paragraaf zullen behandelen, vertoond enkele essentieele trekken van een eenvoudige wachttijd-situatie. Het wiskundig model ervan sluit dan ook nauw aan bij de gebruikelijke modellen van vele praktijkproblemen.

Aan een reeks van worpen met twee dobbelstenen geven we de volgende interpretatie:

a) gooien we 9 ogen of meer, dan komt er een klant binnen
b) gooien we 6 ogen of minder, dan is een bediening voltooid, indien een klant aanwezig was.
c) gooien we 7 of 8, dan gebeurt er niets
d) het resultaat van de i^e worp heeft betrekking op de gebeurtenis op tijdstip i (de lengte van het tijdsinterval tussen twee opeenvolgende worpen is dus de tijdseenheid, en kan op bijv. 1 minuut worden gesteld)

Men kan eenvoudig inzien dat de kans op een aankomst op een bepaald tijdstip dan $\frac{10}{36}$ is, terwijl de kans dat de bedieningstijd van een klant die wordt bediend, op een bepaald tijstip wordt beëindigd, gelijk is aan $\frac{15}{36}$. De verhouding $\frac{10}{36} : \frac{15}{36} = \frac{2}{3}$ wordt de bezettingsgraad van het systeem genoemd.

Voeren we een reeks van dergelijke worpen uit, dan zien we dat vaak geen klanten in het systeem aanwezig zijn (zodat dus de bedieningsinstallatie "wacht"), terwijl het anderzijds voorkomt dat er een rij klanten staat te wachten.

Hetzelfde experiment kan ook worden gedaan voor een andere bezettingsgraad, bijv. door de worpen met de twee dobbelstenen als volgt te interpreteren:

a) 8 of meer betekent een aankomst
b) 4 t/m 7 betekent een vertrek
c) 3 of minder betekent geen aankomst of vertrek

De bezettingsgraad is nu $\frac{15}{36} : \frac{18}{36} = \frac{5}{6}$ In dit geval zal het veel minder vaak voorkomen dat de bedieningsinstallatie moet wachten, terwijl echter zeer lange rijen kunnen optreden.

4. Gedetailleerde beschrijving van het wachttijdmodel

In § 2 is de structuur geschetst van het te gebruiken model voor wachttijdsystemen. In detail zal thans worden ingegaan op de daar onderscheiden aspecten van het model, speciaal op de wijze waarop men de in de werkelijkheid waargenomen gedragingen wiskundig tracht te beschrijven, en dit ook met succes doet.

A. Aankomstproces

Van belang is allereerst het aankomstproces. In § 1 is er reeds op gewezen dat, naast onzekerheid over de voor de dienstverlening benodigde tijd, het juist de gehele of gedeeltelijke oncontroleerbaarheid van de aankomst der vraagelementen is die de oorzaak vormt der opstoppingsverschijnselen. De volgende verschillende aankomstpatronen kunnen worden onderscheiden

A.I. Onderling onafhankelijke aankomsten.

De vraagelementen behoren tot een oneindig groot gedachte populatie en komen onafhankelijk van elkaar bij het dienstverlenend systeem aan.

A.I.1. Poissonverdeling - exponentiële verdeling.

Wanneer in deze aankomsten, afgezien van een constant gemiddelde dat men over een lange tijdsduur moet meten, voor het overige op het eerste gezicht volstrekte willekeur heerst, geeft het zogenaamde Poissonproces dikwijls een goede beschrijving van het gebeuren. Men spreekt van een Poissonproces wanneer aan een aantal veronderstellingen is voldaan die in de practijk gemakkelijk te verifiëren zijn, te weten:

a) In een eindig tijdsinterval komen slechts een eindig aantal aankomsten voor.
b) Op ieder tijdstip treedt hoogstens één aankomst op.
c) De kans op een bepaald aantal aankomsten in een tijdsinterval is slechts afhankelijk van de lengte van dit interval.
d) De aantallen aankomsten in disjuncte (d.w.z. buiten elkaar liggende) tijdsintervallen zijn onderling onafhankelijke stochastische variabelen.

Men kan bewijzen dat onder deze veronderstellingen de kans op k aankomsten gedurende een tijdsinterval ter lengte van T tijdseenheden gegeven wordt door:

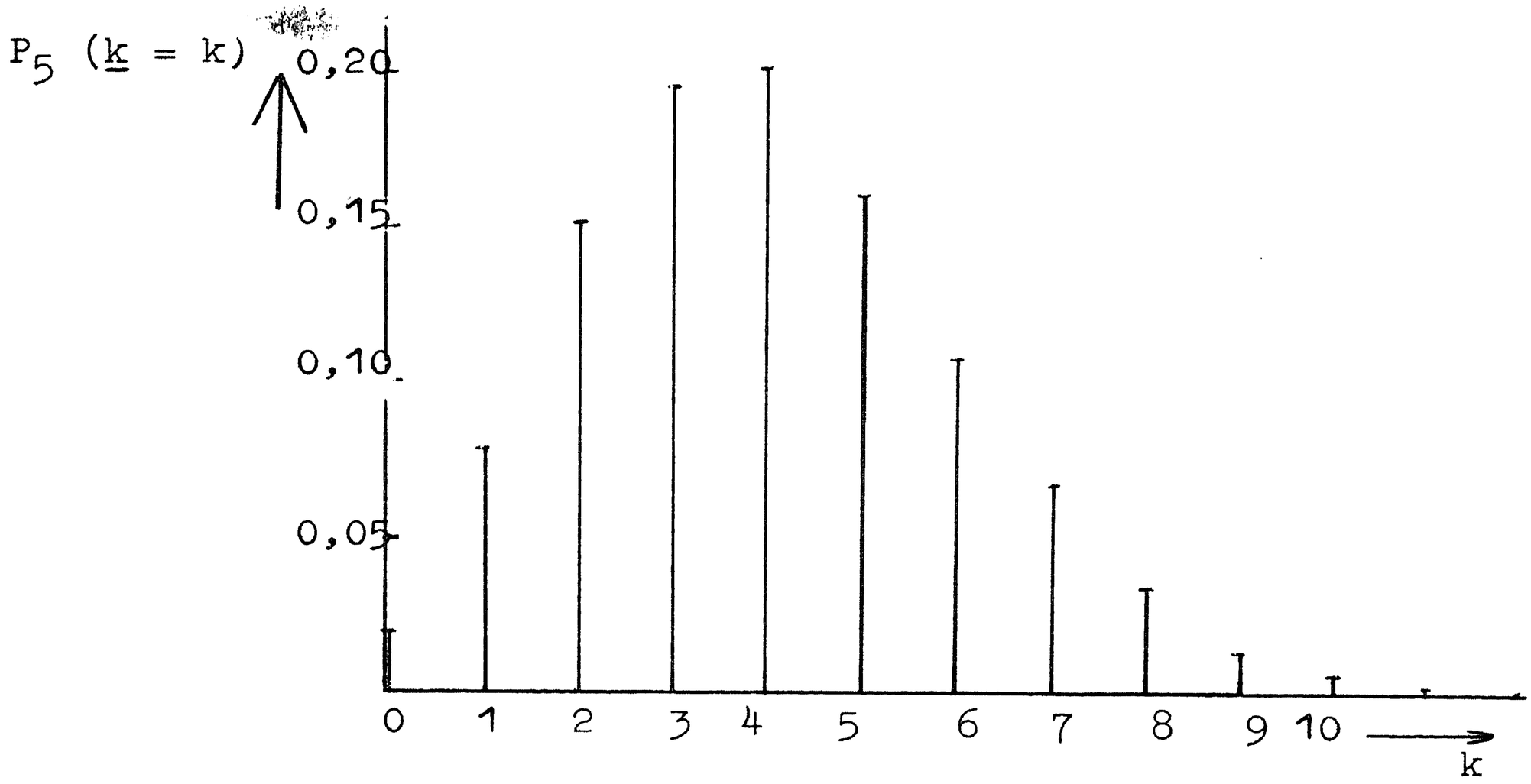
$$P_T\ \underline{k} = k = \frac{(\lambda T)^k}{k!} e^{-\lambda T},$$

waarin λ het gemiddelde aantal aankomsten per tijdseenheid is.

Om de door deze formule gegeven kansverdeling, de zgn. Poissonverdeling, te demonstreren denken we ons in dat aan de

informatiedienst van een stedelijke telefooncentrale gemiddeld 0,8 vragen per minuut gesteld worden. Blijkens de zojuist gegeven formule is de kansverdeling van het aantal vragen dat per vijf minuten gesteld zal worden, daar hier $\lambda = 0{,}8$ en $T = 5$:

$$P_5\ (\underline{k} = k) = \frac{4^k}{k!}\ e^{-4}$$



Figuur 2. De kansen bij de Poisson-aankomstenverdeling

$$P_5\ (\underline{k} = k) = \frac{4^k}{k!}\ e^{-4}$$

Bij de beschrijving van het aankomstproces van een tram of een bus zijn de wachtende passagiers niet zozeer geïnteresseerd in het aantal aankomsten per tijdseenheid (bv. per uur), als wel in de tijd die zal verstrijken tussen twee opeenvolgende aankomsten, het zgn. tussenaankomstinterval. De kansverdeling van de tussenaankomstintervallen volgt in het geval van een Poisson-aankomstproces uit de formule van de Poissonverdeling door de kans te berekenen dat gedurende T tijdseenheden geen enkel vraagelement arriveert.

$$P_T\ (\underline{k} = 0) = e^{-\lambda T} \cdot \frac{(\lambda T)^0}{0!} = e^{-\lambda T}$$

De gebeurtenis dat geen enkel vraagelement gedurende T tijdseenheden arriveert komt overeen met realisatie van een tussenaankomstinterval langer dan T tijdseenheden. De kansen op het voorkomen van deze gebeurtenissen zijn dus ook gelijk. Als we de lengte van het tussenaankomstinterval met $\underline{t}$ noteren, is dus te doorzien dat:

$$P(\underline{t} \geq T) = P_T(\underline{k} = 0) = e^{-\lambda T} .$$

Dus

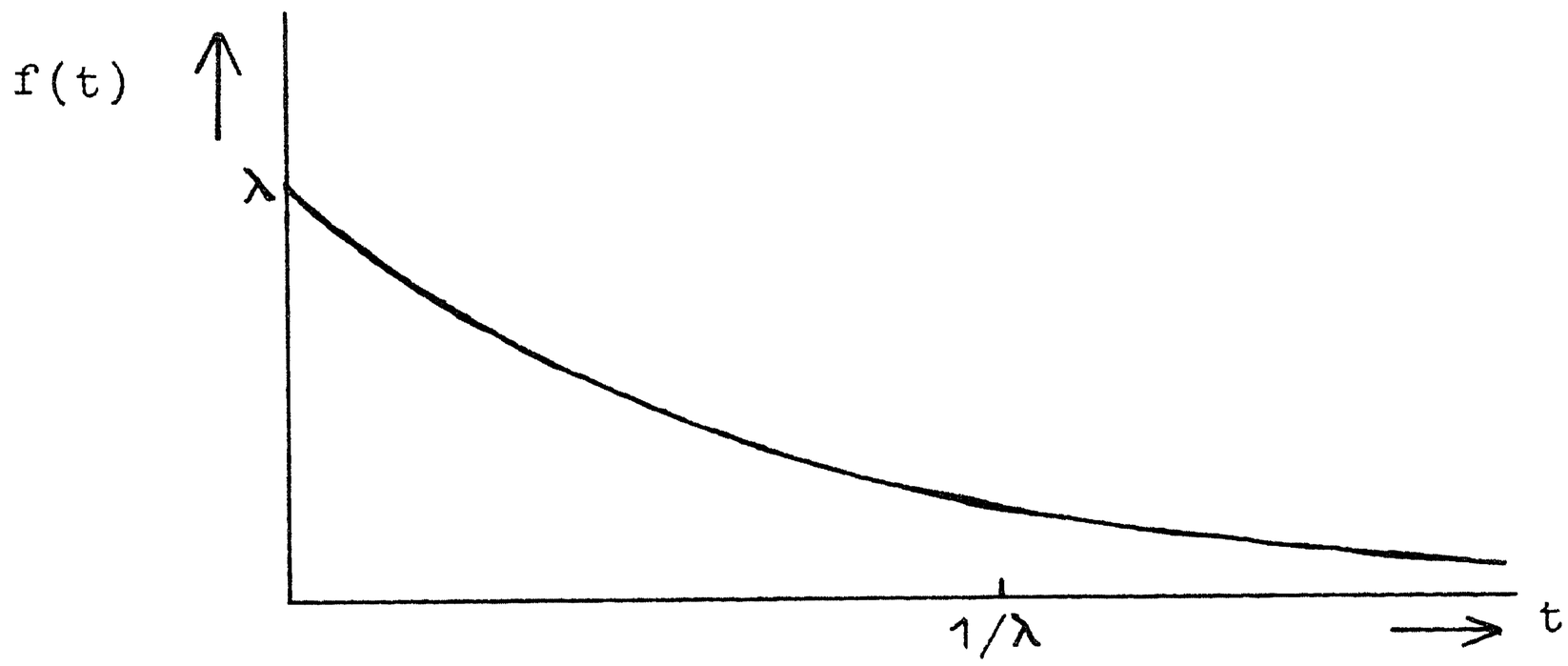
$$F(T) = P(\underline{t} \leq T) = 1-e^{-\lambda T} .$$

Differentieren van deze verdelingsfunctie geeft de verdelingsdichtheid

$$f(t) = \lambda e^{-\lambda t} .$$

Dit is de zgn. exponentiële verdeling.

De Poissonverdeling voor de aantallen aankomsten per tijdseenheid en de exponentiële verdeling voor de tussenaankomstintervallen treden dus altijd in combinatie op.



Figuur 3. De verdelingsdichtheid der exponentiële verdeling.

Uit figuur 3 blijkt dat de kleinste tussenaankomstintervallen de grootste kans hebben voor te zullen komen. Men kan bewijzen dat de gemiddelde lengte van een tussenaankomstinterval $\frac{1}{\lambda}$ tijdseenheden is. Intuïtief is dit ook gemakkelijk in verband te brengen met de definitie van λ als gemiddeld aantal aankomsten per tijdseenheid.

Het Poisson aankomstproces is gei trod ceerd aan de hand van een aantal kenmerken, waarvan de aa s e de onafhankelijkh id aangaf van de aantallen aankomsten gedurende disjuncte tijdsintervallen. Voor de exponentieel verdeeld tussenaankomstintervallen, waarmee we, zoals juist bewezen is, het Poisson-aankomstproces ook kunnen beschrijven, betekent dit kenmerk dat het verleden op geen enkel moment enige informatie kan geven over de lengte van het komende tussenaankomstinterval. Beschouwen we bijvoorbeeld een passagier die bij een tramhalte wacht, waar de trams met exponentieel verdeelde tussenaankomstintervallen plegen te arriveren. Wanneer hij hoort dat er sinds de laatste tram al veel tijd verstreken is mag deze passagier nu niet de conclusie trekken dat de eerstvolgende tram eerder zal komen dan wanneer hij zojuist een tram gemist had. De tijd die verstreken is sinds het vertrek van de vorige tram geeft immers als feit uit het verleden geen informatie over de tijd die nog zal verstrijken eer de volgende tram arriveert.

Heeft men een organisatie gevormd om de aankomsten te reguleren, maar blijken de aankomsten toch volgens een Poissonproces te verlopen, dan betekent het irrelevant zijn van alle informatie uit het verleden dat de organisatie geheel faalt, of althans zozeer faalt dat hij geen enkel nuttig effect meer heeft. Deze laatste restrictie omdat gebleken is dat het kan voorkomen dat ook zonder totale desintegratie der organisatie het actuele aankomstpatroon nauwelijks van een Poissonproces te onderscheiden is. Dit gebeurt in geval de afwijkingen van een regelmatig aankomstenschema groot worden in vergelijking met de beraamde tussenaankomstintervallen. Als voorbeeld kan dienen de aankomst van een tram bij een halte in een grote stad, waar zeer vele voorvallen in het drukke verkeer de dienstregeling kunnen doorkruisen.

Naast de rol van limiet-geval bij de desintegratie der

organisatie, levert het Poissonproces dikwijls ook een goede benadering voor de vraag naar een dienst waarvan zeer veel mensen potentiële gebruikers zijn, terwijl de kans dat een bepaalde persoon, in een bepaald, niet te lang interval beschouwd, gebruiker zal zijn heel klein is, zoals gelukkig de aankomst als spoedgeval in een ziekenhuis. Ook de vraag naar telefoonverbindingen komt voor deze categorie in aanmerking.

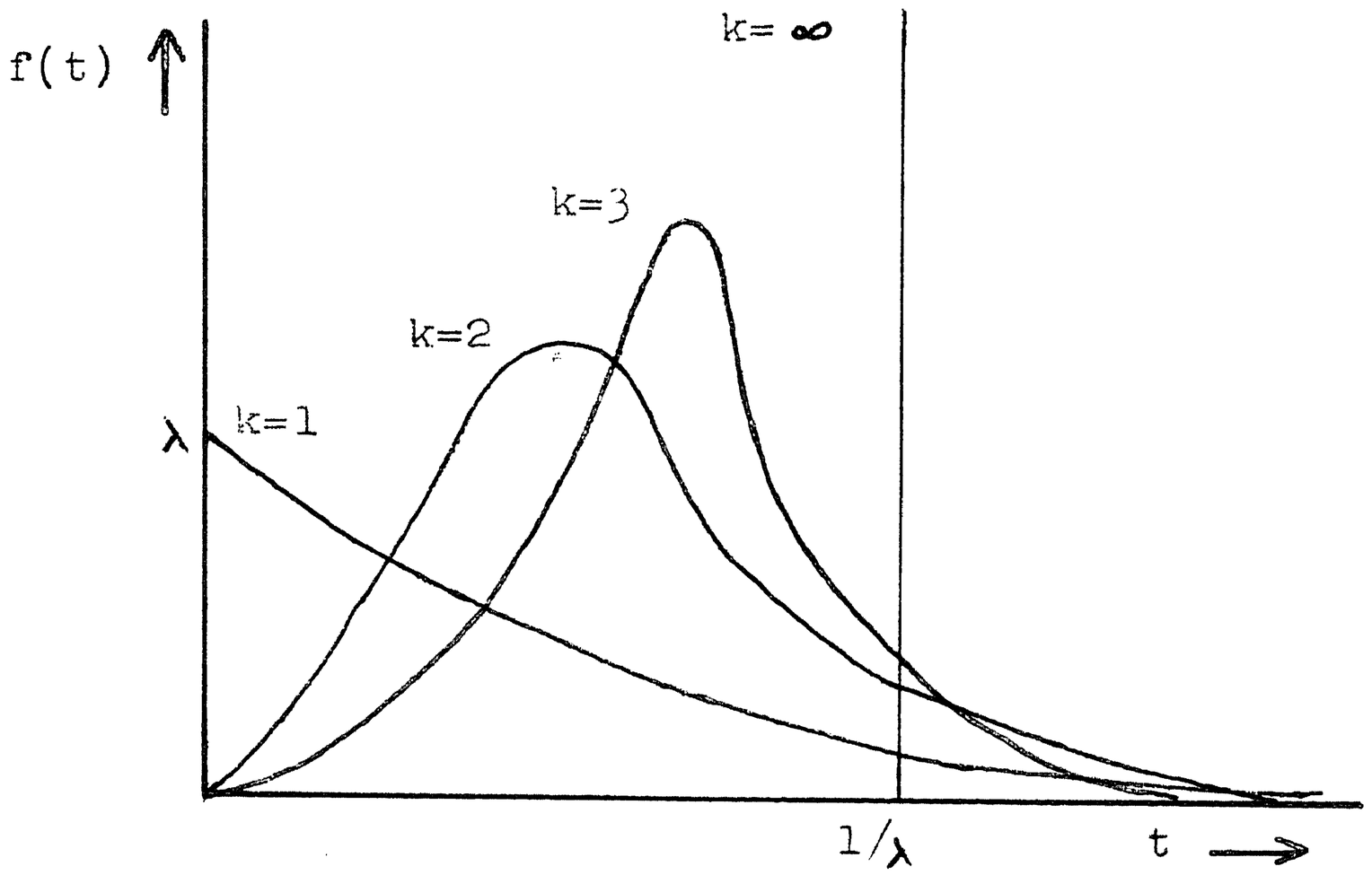
Ook een aankomstproces resulterend uit een combinatie van zeer vele op zichzelf goed functionerende organisaties kan op een waarnemer, die van deze organisaties niet op de hoogte is of wil zijn, de indruk maken van een Poissonproces. Zo komen de schepen op de Nieuwe Waterweg met grote nauwkeurigheid volgens een Poissonproces aan. Als de waarnemer wel op de hoogte is van het bestaan van deze organisaties zal hij, om reden van eenvoud vaak gebruik maken van het Poissonproces, omdat de informatie over de organisaties voor het beschrijven van het aankomstproces overbodig blijkt.

A.I. 2 Tegenover de willekeur in de aankomst staat de complete beheersing van dit proces, resulterend in constante tussenaankomstintervallen, een geval waar zeer wel mee te rekenen valt.

A.I. 3 Erlangverdelingen

Ook zeer vele overgangsgevallen tussen deze twee uitersten zijn wiskundig scherp te vatten. Zij vormen namelijk de groep van de zgn. Erlangverdelingen, van welke de verdelingsdichtheid naast λ een tweede parameter k vertoont:

$$f(t) = \frac{(\lambda k)^k}{(k-1)!} t^{k-1} e^{-k\lambda t}$$



Figuur 4. De verdelingsdichtheden der Erlangverdelingen voor k=1,2,3,..., .

Voor k=1 geeft de Erlangverdeling de verdelingsdichtheid van de exponentiële verdeling, $f(t) = \lambda e^{-\lambda t}$. We weten reeds dat $\frac{1}{\lambda}$ de gemiddelde lengte van een tussenaankomstinterval is. Deze waarde $\frac{1}{\lambda}$ blijkt voor alle k als gemiddeld tussenaankomstinterval te gelden. Zoals in figuur 4 te zien is wordt voor k=2,3,... de waarschijnlijkste waarde van t geleidelijk groter, en de spreiding van de verdeling geleidelijk kleiner, om voor k=∞ tenslotte naar geval A.I.2. te convergeren, dwz. naar een constante waarde $\frac{1}{\lambda}$ voor de tussenaankomstintervallen.
De spreiding van de tussenaankomstintervallen is dan nul.

A.I. 4 Aankomstproces afhankelijk van toestand van systeem

Er bestaan ook gevallen waar de vraagelementen bij aankomst bij het systeem hun toetreden tot dit systeem af zullen laten hangen van de lengte van de aanwezige rij. Men spreekt hier van verliesproblemen. Een extreem voorbeeld is een parkeerplaats, waar zich in het geheel geen rij kan vormen.

In andere gevallen als bij winkels, herenkappers, of bij vliegtuigen die vanwege een opstopping van het ene naar het andere vliegveld gedirigeerd worden, zullen zich bij een groeiende rij steeds minder vraagelementen metterdaad bij de rij aansluiten. Om dit in rekening te brengen kan een correctie op de gemiddelde aankomstfrequentie λ toegepast worden.

A. II Onderling afhankelijke aankomsten.

A. II. 1. Eindige populatie.

Een speciale behandeling krijgen gevallen waar de populatie der potentiële klanten niet zo groot is dat zij zonder de betrouwbaarheid der uitkomsten teniet te doen oneindig groot geacht kan worden. Dit geldt bijvoorbeeld voor een machinepark dat door een aantal reparateurs in bedrijf moet worden gehouden. De opgave is hier het juiste aantal reparateurs in dienst te nemen. Een defecte machine die op reparatie wacht kost geld, maar een op werk wachtende reparateur evenzeer.
De eindige populatie is er nu de reden van dat de frequentie van het defect raken der machines op ieder moment afhankelijk is van het aantal reeds defecte machines. Ook hier, maar op een andere wijze dan in A.I. 4., is het aankomstproces dus afhankelijk van het totaal aantal vraagelementen dat bediend wordt ofwel staat te wachten, dus van de toestand van het systeem.
Bij deze eindige populaties spreekt men van machineproblemen.

A. II. 2 Aankomst en bediening in groepen.

Soms is het nodig om in onze modellen rekening te houden met de omstandigheid dat klanten vaak in groepen bij de bedieningsinstallatie aankomen en in groepjes bediend worden.
Dit gebeurt immers in een restaurant, in een lift, bij douaneformaliteiten na aankomst van een vliegtuig etc.

A.II. 3. Markov - afhankelijkheid.

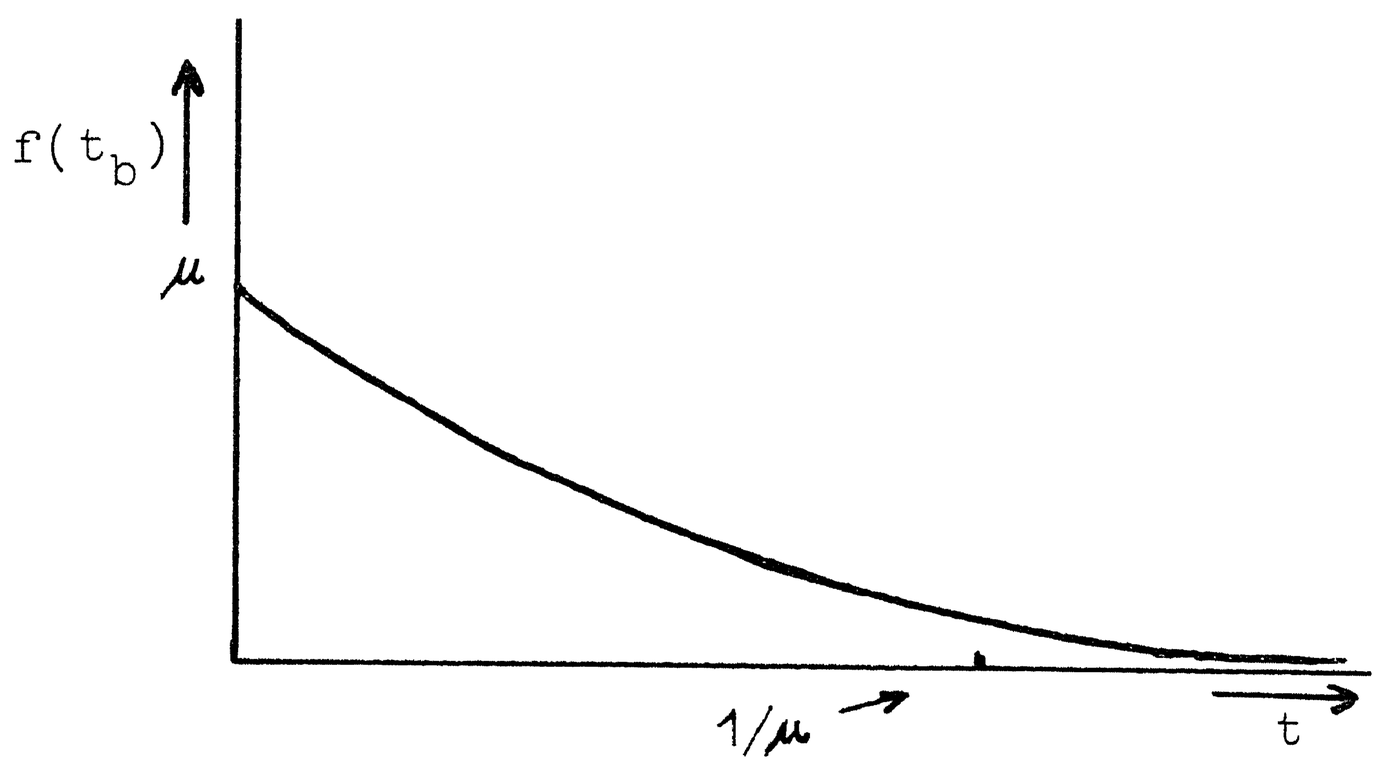
Door Prof. Runnenburg, verbonden aan het Mathematisch Centrum, wordt onderzocht welke mogelijkheden er zijn om Markov - afhankelijkheid tussen opeenvolgende tussenaankomst-intervallen in te voeren.

B. De benodigde bedieningstijd.

B.1 Exponentiële verdeling.

De tijd nodig voor het verlenen van de dienst, waar de verschillende vraagelementen uit een rij om verzoeken, kan naar gelang van de gevraagde dienst zeer uiteenlopen. Een zeer veel gebruikte veronderstelling is dat de benodigde bedieningstijden onderling onafhankelijk zijn en exponentieel verdeeld. Deze verdeling is reeds bekend uit de beschrijving der mogelijke tussenaankomstintervallen. De verdelingsdichtheid wordt gegeven door

$$f(t_b) = \mu e^{-\mu t_b}$$



Figuur 5. De verdelingsdichtheid der exponentiële verdeling.

Uit figuur 5 blijkt wederom, mede door de ligging van het gemiddelde $\frac{1}{\mu}$, dat de kleinste bedieningstijden de grootste kans hebben voor te zullen komen, terwijl de langere bedieningstijden toch nog zo veelvuldig optreden dat zij geenszins verwaarloosd mogen worden.

De exponentiële verdeling benadert vele gevallen in de praktijk, zoals de tijd dat een lijn bezet is door één telefoongesprek, de reparatieduur van een machine, en de opnameduur (boven een bepaald minimum) van patiënten in een ziekenhuis (ieder bed een loket) etc.

2. Erlang - verdelingen

Evenals bij het aankomstproces staan de Erlangverdelingen ook hier ten dienste om die processen te beschrijven, waarin de fluctuaties in de bedieningstijd kleiner zijn dan bij de volkomen "chaos" van exponentieel verdeelde bedieningstijden. De extreme mogelijkheid van constante bedieningstijden is hiervan dus weer een bijzonder geval.

C. Het bedieningsmechanisme.

Nu het probleem waarvoor we ons gesteld zien, namelijk het voldoen aan een stochastisch fluctuerende vraag naar een dienst van stochastische duur, in allerlei praktijksituaties herkend is en vervolgens wiskundig geformuleerd, zal nagegaan worden welke vrijheden men heeft bij het opstellen der bedieningsinstallatie die aan deze situatie het hoofd moet bieden. Nadat deze vrijheden in deze paragraaf kort geschetst zullen worden, zal in volgende paragrafen worden nagegaan hoe men criteria kan opstellen om uit deze mogelijkheden een keus te doen, en wat voor berekeningen tot de beslissing over de vorm van het bedieningsapparaat leiden.

C.1 Aantal loketten.

Het bedieningsmechanisme kan uit één loket bestaan. Hiervoor zal zich één rij formeren. Het woord rij moet niet letterlijk genomen worden. Ook restaurantbezoekers die verspreid over een restaurant zittend op een kellner (een wandelend loket) wachten, vormen een rij.

Bestaat het bedieningsmechanisme uit meer dan één loket, dan kunnen er op verschillende wijzen rijen ontstaan. Voor elk loket kan een aparte rij komen, zoals bij de kassa's van zelfbedieningswinkels. Vaak formeert zich echter één gemeenschappelijke rij voor alle loketten, zoals bij een kapper.

C.2 Bedieningsvolgorde.

In elk van deze rijen is van groot belang de regel welke de volgorde bepaalt waarin bediend wordt. Meestal geldt: "Die het eerst komt, het eerst maalt". Geheel tegengesteld is de mogelijkheid dat juist het laatst aangekomen vraagelement het eerst geholpen wordt, zoals in pakhuizen het geval kan zijn. Bij sommige automatische telefoonaansluitingen komt de keuze welke van verschillende lopende gespreksaanvragen doorverbonden zal worden willekeurig tot stand. De bedieningsvolgorde heet dan aselect.

In § 7 zal worden ingegaan op het geval dat men prioriteiten toekent, d.w.z. normen vaststelt op grond waarvan sommige vraageenheden voorrang krijgen. Dit kunnen richtlijnen zijn die men de organisatie van het systeem om welke reden dan ook heeft willen opleggen (ijlzaken in postkantoor), ofwel maatregelen die juist genomen zijn omdat zij de totale wachttijd in het systeem verminderen (reparatiewerkplaatsen).
In de tweede wereldoorlog werkte men in Engeland zeer hard aan prioriteitsregels voor het zo veilig mogelijk doen landen

van vliegtuigen die in zwermen terugkwamen van bombardementsvluchten, vaak aan het eind van hun brandstof en beschadigd.

C. 3. Bedieningssnelheid.

Bij veranderingen in het bedieningsmechanisme (mechanisatie etc.) blijkt dikwijls dat men andere verdelingsfuncties zal moeten gebruiken om het bedieningsproces te beschrijven. Zo zal wanneer door nieuwe apparatuur de wensen van een bepaald deel der vraagelementen sneller vervuld worden, niet alleen de gemiddelde bedieningstijd maar ook de vorm van de verdelingsfunctie veranderen.

Wordt de dienst door mensen verleend, dan kan het zin hebben in het model in te voeren dat de bedieningssnelheid groter wordt naarmate de rij wachtenden groeit.

C. 4. In §6 zullen in detail enkele meer gecompliceerde aanpassingen der organisatie besproken worden.

5. Conclusies uit de wiskundige modellen.

Het is nuttig een onderscheid te maken tussen de toestand van het systeem $\underline{n}$, en de nu in te voeren stochastische variabele $\underline{m}$ die de rijlengte weergeeft. Het verschil tussen $\underline{n}$ en $\underline{m}$ bestaat uit het element dat juist bediend wordt, zo aanwezig. De kans op de aanwezigheid van een dergelijk element is $1-p_o$, als p_o de kans is dat er géén vraagelement in het systeem is.

Men kan bewijzen dat bij een Boissonaankomstenverdeling met gemiddelde λ en een gemiddelde potentiële bedieningscapaciteit van μ vraagelementen per tijdseenheid geldt, wanneer $\mu > \lambda$:

$$1-p_o = \frac{\lambda}{\mu} .$$

Dit quotiënt λ/μ wordt de <u>bezettingsgraad</u> genoemd, en wordt aangeduid met ρ :

$$\rho \stackrel{def}{=} \frac{\lambda}{\mu} .$$

Omdat het gemiddelde aantal vraagelementen dat in bediening is dus $\frac{\lambda}{\mu}.1+(1-\frac{\lambda}{\mu}).0 = \frac{\lambda}{\mu}$ blijkt te zijn, is:

$$\mathcal{E}_{\underline{n}} = \mathcal{E}_{\underline{m}} + \frac{\lambda}{\mu} . \qquad (1)$$

We zullen de resultaten beschouwen van een model waarin als veronderstellingen gelden dat de <u>aankomsten Poisson-verdeeld zijn</u> en de <u>benodigde bedieningstijd</u> de <u>Erlang-verdeling</u> met parameter k heeft. We weten dat dit inhoudt dat als

$\underline{t}_b$ = de benodigde bedieningstijd,

de gemiddelde benodigde bedieningstijd gegeven wordt door

$$\mathcal{E}_{\underline{t}_b} = \frac{1}{\mu} . \qquad (2)$$

Tevens onderscheiden we

$\underline{t}_w$ = de wachttijd van een vraagelement.

$\underline{t}_s$ = de totale tijd die een element in het systeem doorbrengt.

Uiteraard geldt $\underline{t}_s = \underline{t}_w + \underline{t}_b$

en dus $\mathcal{E}\underline{t}_s = \mathcal{E}\underline{t}_w + \mathcal{E}\underline{t}_b$

wegens (2) $$\mathcal{E}\underline{t}_s = \mathcal{E}\underline{t}_w + \frac{1}{\mu} \quad . \tag{3}$$

Tussen het gemiddelde aantal elementen in het systeem en de gemiddelde tijd die een element in het systeem doorbrengt, bestaat een betrekking die met behulp van de definitie van λ als gemiddeld aantal per tijdseenheid arriverende elementen redelijk te doorzien is, maar waarvan het bewijs minder eenvoudig is

$$\mathcal{E}\underline{n} = \lambda\, \mathcal{E}\underline{t}_s \quad . \tag{4}$$

Met behulp van (3) en (4) kunnen we $\mathcal{E}\underline{n}$ en $\mathcal{E}\underline{t}_w$ in elkaar uitdrukken

$$\mathcal{E}\underline{n} = \lambda\, \mathcal{E}\underline{t}_w + \frac{\lambda}{\mu} \quad . \tag{5}$$

Aanvaarden we nu als een resultaat van de wachttijd-theorie dat voor de betreffende veronderstellingen geldt

$$\mathcal{E}\underline{m} = \frac{k+1}{2k} \cdot \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{k+1}{2k} \cdot \frac{\rho^2}{1-\rho} \tag{6}$$

dan volgt uit (1)

$$\mathcal{E}\underline{n} = \frac{k+1}{2k} \cdot \frac{\lambda^2}{\mu(\mu-\lambda)} + \frac{\lambda}{\mu} = \frac{k+1}{2k} \cdot \frac{\rho^2}{1-\rho} + \rho \tag{7}$$

en uit (5)

$$\mathcal{E}\underline{t}_w = \frac{k+1}{2k} \qquad \frac{\lambda}{\mu(\mu-\lambda)} \quad .$$

Voorbeeld 1.

Beschouw een wachttijdsysteem met Poissonaankomstproces en bedieningstijden die exponentieel verdeeld zijn, d.w.z. Erlang-verdeeld met parameter k=1.
De gemiddelde toestand van het systeem volgt uit (7) door substitutie van k=1:

$$\mathcal{E}_{\underline{n}} = \frac{\rho}{1-\rho} .$$

Deze functie is in figuur 6 uitgezet.

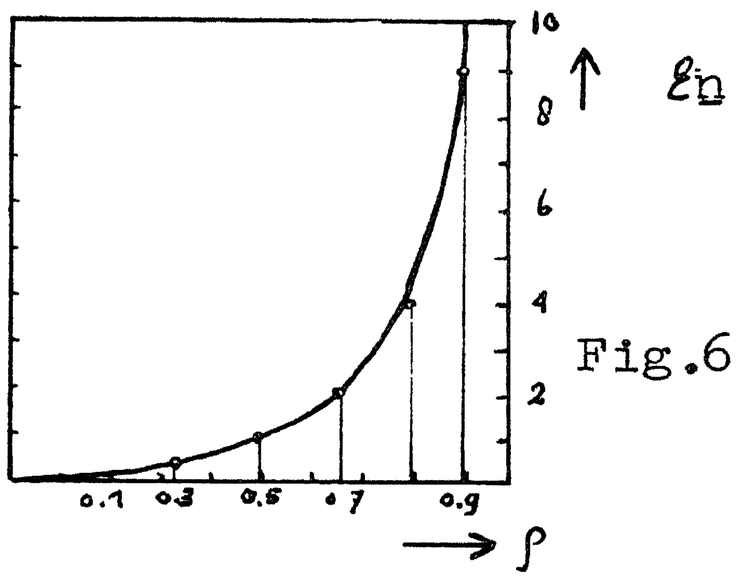


Fig.6 Het gemiddelde aantal elementen in het systeem bij verschillende bezettingsgraden.

Figuur 6 toont hoe sterk de stagnatie toeneemt als de bezettingsgraad boven $^2/_3$ komt, om voor waarden in de buurt van $\rho = 1$ zelfs boven alle grenzen te groeien. Hoe graag men de potentiële capaciteit van het bedieningssysteem ook volledig zou willen benutten, boven de 80% kan men zonder enorme opstoppingen niet gaan. Men bedenke namelijk dat $\mathcal{E}(\underline{n})$ slechts een gemiddelde is, van tijd tot tijd zal de gevormde rij veel langer zijn. Dit alles is vooral een gevolg van de omstandigheid dat tijdens de leegloop bedieningscapaciteit voor altijd verloren gaat. In §4 is reeds aangekondigd dat in het volgende zal worden nagegaan hoe de beheersbare factoren in het wachttijdsysteem op elkaar moeten worden afgestemd om een bottle-neck situatie op optimale wijze het hoofd te bieden.

In detail zal thans worden ingegaan op de verschillende instrumenten waarmee opstoppingen te vermijden zijn.

1) Allereerst moet worden getracht het aankomstproces (vgl. §4A) onder controle te krijgen. Hierin ligt immers de wortel van het kwaad, samen met de zich eveneens vaak aan onze controle onttrekkende benodigde bedieningstijd, welke als tweede punt behandeld zal worden.

De verdeling van de wachttijd en die van de rijlengte zijn in het geval van exponentiële tussenaankomstintervallen bekend, zowel voor één als voor meer loketten. In het geval van één loket kan men de invloed van nivellering der tussenaankomstintervallen analytisch volgen door Erlang-verdelingen in het model te substitueren met steeds grotere parameter k. Voor de waarde $k=\infty$, die constante bedieningstijd aangeeft, blijkt de gemiddelde wachttijd, $\mathcal{E}t_W$, zeer sterk gereduceerd te zijn t.o.v. een exponentieel verdeelde bedieningstijd.

In het geval van meerdere loketten is men er nog niet in geslaagd de berekeningen uit te voeren en gebruikt men electronische rekenmachines om simulatiemethoden toe te passen. De aldus verkregen resultaten zijn bijzonder nuttig, doch hun precisie staat ver achter bij analytisch verkregen resultaten tenzij men aan een onderzoek veel tijd van de machine (en dus zeer veel geld) wil spenderen.

Doet men dit niet dan dreigt men zich een gechargeerd gunstig beeld te vormen van de te verwachten situatie. Het is dan namelijk zeer wel mogelijk dat een met kleine kans optredende samenloop van omstandigheden, die juist de grote opstoppingen zal kunnen veroorzaken, zijn invloed nog niet heeft doen gelden.

2) De <u>benodigde bedieningstijd</u> moet tevens zoveel mogelijk <u>onder controle</u> gebracht worden.

Vergelijken we namelijk de waarden van $\mathcal{E}_{\underline{m}}$ volgende uit (6) voor k=1 (exponentiële bedieningstijd), met die voor $k=\infty$ (constante bedieningstijd), dan blijkt de laatste situatie een halvering van de verwachting van de rijlengte $\mathcal{E}_{\underline{m}}$ t.o.v. het exponentiële geval met zich mee te brengen, evenals voor $\mathcal{E}_{\underline{t}_w}$, de verwachting van de wachttijd.

<u>Voorbeeld 2</u>.

Een alleenwerkende kapper doet precies 25 minuten over het knippen van één zijner klanten. Deze arriveren volgens Poisson-proces, gemiddeld één klant per 40 minuten. Het antwoord op de vraag hoe lang een klant gemiddeld op het knippen moet wachten wordt verkregen door substitutie van $k=\infty$ in vergelijking (8). Hierdoor verkrijgt men de verwachting voor de wachttijd in het geval van constante bedieningstijd:

$$\left[\mathcal{E}\,\underline{t}_w\right]_{const} = \frac{\lambda}{2\mu(\mu-\lambda)} .$$

Aangezien $\lambda = 0{,}025$ klanten per minuut en $\mu = 0{,}04$ klanten per minuut is $\rho = 5/8$ en

$$\left[\mathcal{E}\,\underline{t}_w\right]_{const} = \frac{0{,}025}{0{,}08 \times 0{,}015} = 20{,}8 \text{ minuten} .$$

Zouden de wensen van de klanten uiteenlopen, terwijl daarbij de gemiddelde bedieningstijd niet verandert, dan zou dit de gemiddelde wachttijd aanmerkelijk doen oplopen tot een maximale waarde voor het geval de bedieningstijd exponentieel verdeeld is met een verwachting van 25 minuten. Nu moet immers k=0 gesubstitueerd worden in vgl.(8), hetgeen geeft:

$$\left[\mathcal{E}\,\underline{t}_w\right]_{exp} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{0{,}025}{0{,}04 \cdot 0{,}015} = 41{,}6$$

minuten.

We z'en dat deze waarde prec'es tweemaal zo groot is als $[\mathcal{E}_{\underline{t}_w}]_{cons}$.

Naast de acties om de aankomsten en de wensen der vraag-elementen te beïnvloeden bestaan er aanpassingen van de bedieningsinstallatie waardoor opstoppingen tegengegaan kunnen worden.

Hiertoe behoren:

3) Opvoeren der bedieningssn:lheid.
4) Verdelen van de vraag over meerdere parallel werkende loketten.

Propositie (3) reduceert zowel, de wachttijd $\underline{t}_w$ als de benodigde bedieningstijd t_b en is b.v. te bereiken door verdergaande mechanisering of automatisering.

Propositie (4) reduceert alleen $\underline{t}_w$.

Van geval tot geval zal men de doeltreffendheid van beide methoden moeten afwegen tegen de kosten. Zo bleek bij het opvoeren van de capaciteit van havenoutillage voor het lossen van ertsboten de eerste methode de voorkeur te verdienen, totdat op een zeker punt dit verder opvoeren van de lossingssnelheid inefficient werd. Vanaf dat moment was verlenging van de kademuren en uitbreiding van het aantal kranen met verhoogde bedieningssnelheid de goedkoopste manier om de bezettingsgraad en daarmee opstoppingen te reduceren.

Gaan we nu nog even terug naar het voorbeeld van de kapper. Deze merkt namelijk dat zijn zaak drukker gaat lopen. Hij trekt hieruit de consequentie een bediende aan te zullen nemen ($n_1=1 \rightarrow n_2=2$) zodra zijn klanten gemiddeld langer dan 25 minuten dreigen te moeten wachten. Bij welke aankomstfrequentie is dit moment gekomen?

Oplossing:
Hiertoe moet de λ' gevonden worden die voldoet aan de vergelijking:

$$\mathcal{E}\, t_w = 25 = \frac{\lambda'}{0{,}08\ (0{,}04 - \lambda')}$$

Hieruit blijkt $\lambda' = {}^{2}/_{75}$

Dus zodra het gemiddelde interval tussen twee aankomsten beneden ${}^{75}/_{2} = 37\frac{1}{2}$ minuut komt, moet naar een bediende uitgekeken worden.

De keuze van de 25 minuten is hier intuïtief geweest. De kapper voelt dat er dan klanten weg zullen gaan lopen aan wie hij, met behulp van een bediende zou kunnen verdienen.

Gecompliceerdere beslissingen van deze aard kunnen zinvol voorbereid worden door probleemstelling in stukken te hakken.

De ondernemer zal namelijk, na eerst de bedieningstijd zoveel mogelijk constant gemaakt te hebben, zijn organisatie moeten inrichten naar die ρ, waarvoor de <u>som</u> van

(a) de verwachting van de kosten ten gevolge van formering van een wachtrij=

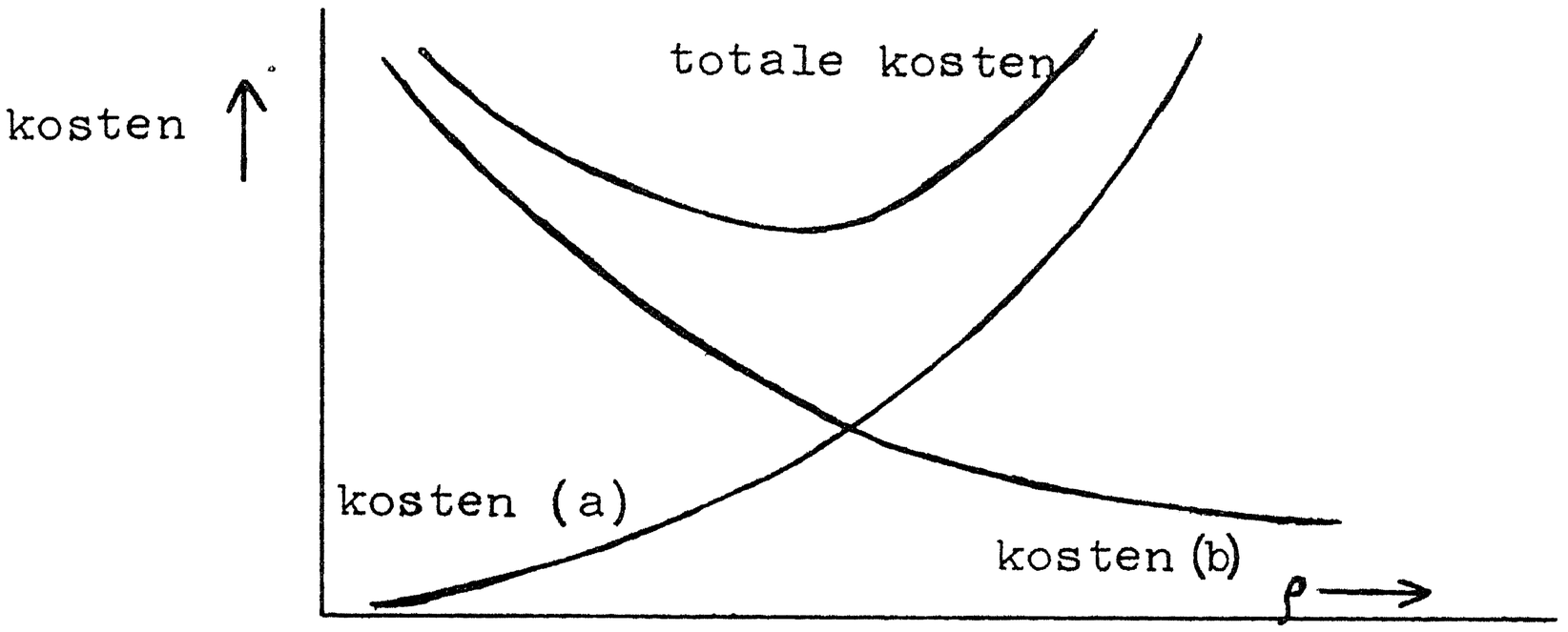
$$= \sum_{n=0}^{\infty} \{ \text{kosten toestand } n \} \cdot p_n(\rho)$$

en

(b) de exploitatiekosten bij een bepaalde ρ,

<u>minimaal</u> is

Dit minimum kan bijvoorbeeld grafisch bepaald worden, van welke procedure figuur 7 een indruk geeft.



Figuur 7. De verwachting van de kosten ten gevolge van formering van een wachtrij, de exploitatiekosten en hun som bij verschillende bezettingsgraden ρ .

Ad kosten (a)

De ondernemer zal allereerst een maat moeten vinden voor de door hem aan een bepaalde toestand ondervonden nadelen. Dit kan een kostprijsberekening zijn, zoals bv. de berekening van de kosten verbonden aan het inproductief staan van een machine gedurende een uur, maar ook quantificering van het goodwillverlies door weglopende of wegblijvende klanten (beducht voor de rij).

Verder is p_n, de kans op toestand n in een stationair systeem, een functie van ρ, vandaar de notatie $p_n(\rho)$. Zo kan men in voorbeeld 1 van § 5 bewijzen dat $p_n = \rho^n(1-\rho)$.

Als ρ toeneemt worden de grotere waarden van n waarschijnlijker. Deze immers duiden opstoppingen aan, die zoals we zagen veelvuldiger voorkomen naarmate de bezettingsgraad ρ stijgt. Vandaar dat de kosten (a) bij grotere ρ zullen stijgen.

Ad kosten (b)
Anderzijds zijn ook de exploitatiekosten een functie van ρ: Een kleinere ρ betekent bij een vaste vraag een bedienings-appa aat met grotere potent'ële apa iteit, dus ook hogere exploitatiekosten. Men bedenke dat in de kosten (b) de keus verwerkt zal zitten welke de ondernemer heeft moeten maken tussen de verschillende manieren om kleinere waarden van ρ te realiseren. Dit is dus de keus tussen het opvoeren der bedieningssnelheid en het verhogen van het aantal parallel werkende loketten.

Op de volgende wijze wordt dit alles in een concreet voorbeeld verwerkt.

Voorbeeld 3.

In een zeer groot machinepark raken, op een manier die als een Poissonproces beschouwd kan worden, gemiddeld drie machines per uur defect. De tijd die nodig zal zijn voor reparatie is exponentieel verdeeld. Men heeft berekend dat ieder inproductief uur van een machine de ondernemer f5.- kost. Er is keus tussen twee soorten monteurs, de ene snel werkend maar duur, de andere langzaam werkend maar goedkoop.
Een langzame monteur vraagt f.3.- per uur, gedurende welke tijd hij gemiddeld vier machines zal kunnen repareren. De snelle monteur vraagt f. 5.- per uur, maar hij kan gemiddeld zes machines per uur repareren. Zoals gezegd, zijn dit gemiddelden van een exponentiële verdeling.

De vraag is nu of men een langzame of een snelle monteur zal moeten aannemen, òf dat misschien twee langzame monteurs de goedkoopste oplossing zal brengen.

We berekenen hiertoe voor ieder der drie gevallen de verwachte totale kosten per uur.

Om de kosten van de inproductieve tijd der machines te berekenen herinneren we eraan dat de defecte machines een rij vormen voor de monteur (het loket). De verwachting van de toestand van dit systeem, $\mathcal{E}_{\underline{n}}$, geeft aan hoeveel machines gemiddeld aan het arbeidsproces onttrokken zullen zijn. Door k=1 te substitueren in formule (7) voor $\mathcal{E}_{\underline{n}}$ houden we rekening met de exponentieel verdeelde bedieningstijd en vinden voor de verwachting van het aantal per uur in het gehele machinepark niet-productieve machine-uren:

$$\mathcal{E}_{\underline{n}} = \frac{\lambda^2}{\mu(\mu-\lambda)} + \frac{\lambda}{\mu} = \frac{\lambda}{\mu-\lambda} \quad .$$

I Allereerst beschouwen we <u>één langzame goedkope monteur</u>. Uit de gegevens blijkt dat in zijn geval $\lambda=3$ en $\mu_I = 4$ machines per uur, zodat

$$[\mathcal{E}_{\underline{n}}]_I = \frac{\lambda}{\mu_I-\lambda} = 3 \text{ machine-uren per uur.}$$

De verwachte kosten ten gevolge van het inproductief staan der machines bedragen f.15.- per uur. Met de loonkosten van f.3.- per uur komen de totale kosten van de langzame goedkope monteur dus op f.18.- per uur. De bezettingsgraad

$$\rho_I = \frac{\lambda}{\mu_I} = \frac{3}{4} \quad .$$

II De <u>snelle dure monteur</u> kost aan loon f.5.- per uur.

Wegens $\lambda = 3$ en $\mu_{II} = 6$ machines per uur wordt

$$[\mathcal{E}_n]_{II} = \frac{\lambda}{\mu_{II}-\lambda} = 1 \text{ machine-uur per uur.}$$

Een dergelijk uur kost f.5.-, waardoor de totale kosten van een snelle dure monteur f.10.- per uur zullen bedragen.

Doordat de bezettingsgraad $\rho_{II} = \frac{\lambda}{\mu_{II}} = \frac{3}{6} = \frac{1}{2}$ aanzienlijk lager ligt dan $\rho_I = \frac{3}{4}$ blijkt deze tweede propositie zeer veel goedkoper.

III <u>Twee langzame goedkope monteurs</u> kosten de ondernemer per uur f. 6.- aan loon. Zij vormen twee loketten die de defecte machines bedienen. Nog steeds is $\lambda = 3$, maar de potentiële bedieningscapaciteit is nu gemiddeld $2\mu_I = 8$ machines per uur geworden. De bezettingsgraad bedraagt dus

$$\rho_{III} = \frac{\lambda}{2\mu_I} = \frac{3}{8}.$$

Op iets gecompliceerdere wijze dan in de hierboven behandelde gevallen volgt nu uit de resultaten van het wiskundig model voor twee loketten, dat het gemiddelde aantal machines dat telkens buiten werking is

$$\mathfrak{E}\ \underline{n} = 0{,}873 \text{ machines}$$

bedraagt.

Dit geeft 0,873 x f. 5,- = f. 4,37 per uur aan kosten door inproductiviteit der machines, hetgeen de totale kosten per uur bij dit alternatief op f. 10,37 brengt.

We zullen dus de <u>snelle dure monteur</u> aannemen.

Stel dat de snelle monteur een hoger loon gaat vragen terwijl de langzame monteurs f.3.-- per uur blijven verdienen. Het zal nu goedkoper worden op de langzame monteurs over te schakelen zodra de snelle monteur meer gaat verdienen dan het verschil tussen de totale kosten der langzame monteurs en de aan hemzelf inhaerente kosten door inproductiviteit der machines, in dit geval:

f.10,37 - f.5.- = f.5,37.

Gebruik makend van de zojuist beschreven resultaten der wachttijdtheorie, kan voor alle mogelijke combinaties van loonkosten der monteurs berekend worden welke beslissing bij het te werk stellen der monteurs optimaal zal zijn,

gegeven de kosten dat een machine een uur inproductief staat. De keuze zal hierbij gaan tussen 2 langzame monteurs, 1 snelle monteur of 1 langzame monteur. Er zal een figuur geconstrueerd worden waarin de optimale beslissing afgelezen kan worden voor iedere combinatie van

$l_1 \stackrel{def}{=}$ de loonkosten van een langzame monteur en

$l_2 \stackrel{def}{=}$ de loonkosten van een snelle monteur.

De eenheden waarin de zojuist gedefinieerde en de nog in te voeren kosten uitgedrukt worden, zullen niet nader worden gepreciseerd. Men kan bijvoorbeeld aan guldens denken.
Zijn

$c_1 \stackrel{def}{=}$ de kosten per uur door inproductiviteit der machines bij twee langzame monteurs,

dan bedragen de totale kosten bij twee langzame monteurs

$T_1 \stackrel{def}{=} 2l_1 + c_1$.

Uit berekening III op blz. 224 blijkt dat $c_1 = 4{,}37$ wanneer de kosten van een inproductief uur op 5 gesteld worden. Dus

$$T_1 = 2l_1 + 4{,}37$$

Definiëren we analoog

$c_2 \stackrel{def}{=}$ de kosten door inproductiviteit der machines bij 1 snelle monteur,

en

$$T_2 \stackrel{def}{=} l_2 + c_2 \ ,$$

dan worden de totale kosten bij 1 snelle monteur blijkens het resultaat $c_2 = 5$ (sub II van blz. 223)

$$T_2 = l_2 + 5.$$

Definiëren we tenslotte

$c_3 \stackrel{def}{=}$ de kosten door inproductiviteit der machines bij 1 langzame monteur

en

$$T_3 \stackrel{def}{=} l_1 + c_3$$

dan worden, rekening houdend met het resultaat c_3=15 sub I van blz. 223, de totale kosten bij 1 langzame monteur:

$$T_3 = l_1 + 15 .$$

In het vlak opgespannen door de l_1-as en de l_2-as correspondeert met ieder punt met coördinaten $(l_1; l_2)$ een combinatie van loonkosten. We zullen alleen dat deel van het vlak beschouwen waarvoor

$$l_1 \geq 0 \quad ; \quad l_2 \geq 0 .$$

De twee langzame monteurs zullen de goedkoopste mogelijkheid vormen bij die combinatie van loonkosten $(l_1 ; l_2)$ waarvoor geldt:

$$T_1 < T_2$$
$$\text{en } T_1 < T_3$$

ofwel

$$l_2 > 2l_1 - 0{,}63$$
$$\text{en } l_1 < 10{,}63 .$$

Deze relaties houden in dat de keuze van twee langzame monteurs optimaal zal zijn bij die punten die zowel boven de lijn $l_2 = 2l_1 - 0{,}63$ als tussen de lijn $l_1 = 10{,}63$ en de l_2-as liggen. Het domein van deze punten is horizontaal gearceerd.

Eén snelle monteur biedt de goedkoopste oplossing wanneer

$$T_2 < T_1$$
$$\text{en } T_2 < T_3$$

ofwel

$$l_2 < 2l_1 - 0{,}63$$
$$\text{en } l_1 > l_2 - 10 .$$

Aan deze eisen voldoen de punten $(l_1 ; l_2)$ in het niet gearceerde gebied.

In het resterende deel van het beschouwde vlak voldoen de coördinaten der punten aan

$$l_2 > l_1 + 10$$

$$\text{en } l_1 > 10{,}63$$

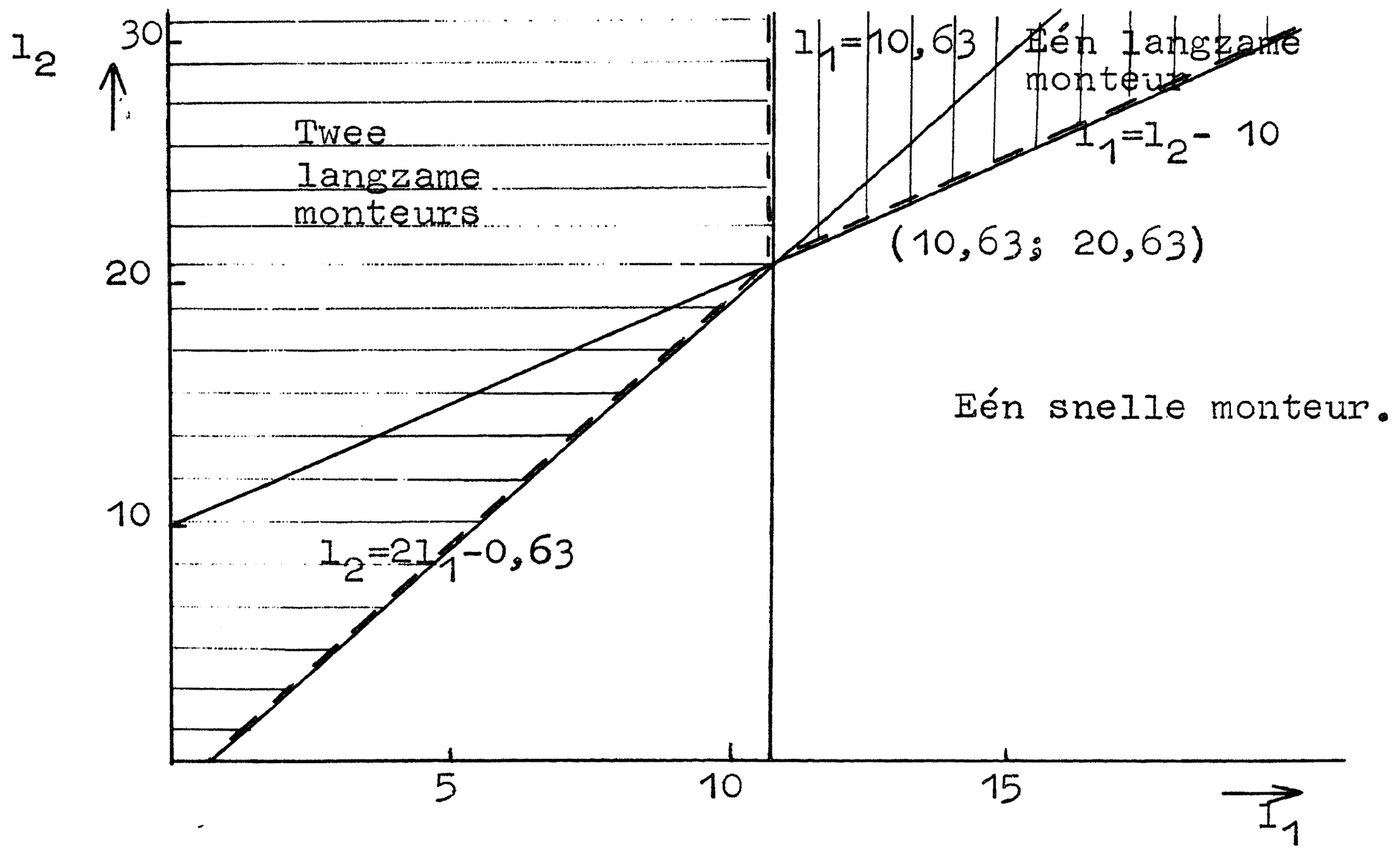
gecorrespondeerd met

$$T_3 < T_2$$

$$\text{en } T_3 < T_1 \ .$$

Hier vormt dus één langzame monteur de goedkoopste oplossing. Het betreffende gebied is verticaal gearceerd.

Denkt men zich de kosten niet uitgedrukt in guldens maar ziet men de getallen inderdaad als verhoudingscijfers, dan is duidelijk dat deze mogelijkheid niet buiten beschouwing gelaten mocht worden.



figuur 8. De optimale keuze van de monteurs bij verschillende combinaties der loonkosten.

6. Andere gevolgen der wachttijdtheorie voor de organisatie-opzet

6.1

Vele fusies van bedrijven worden aangegaan vanuit de verwachting in de toekomst met het grotere systeem tot een efficiëntere bedrijfsvoering te zullen komen. Met een voorbeeld zal in het volgende worden getoond hoeveel efficiënter bij vergroting juist die systemen gaan werken waarin het toeval een rol speelt, de systemen dus met een stochastisch karakter.

We vergelijken daartoe twee systemen waarin volkomen identieke machines door gelijkelijk capabele monteurs worden gerepareerd. In W. Feller, "Probability theory and its applications" vinden wij hiervan een voorbeeld, waarin van de machines zowel de looptijd (lengte van de periode tot de eerstvolgende storing) als de reparatietijd exponentieel verdeeld zijn, met respectievelijke parameters λ en μ. De verhouding tussen deze twee parameters is in het daar doorgerekende geval 0,1.
Van de twee vergeleken systemen omvat het ene zes machines en één monteur, het andere, grotere systeem twintig machines en drie monteurs.

Om de doelmatigheid van beide systemen te vergelijken zullen we ons bedienen van twee verliescoëfficiënten voor de machines:

$$q_n \overset{def}{=} \frac{\text{verwachting van het aantal machines, wachtend of in reparatie}}{\text{totaal aantal machines}}$$

$$q_m \overset{def}{=} \frac{\text{verwachting van het aantal machines in de wachtrij}}{\text{totaal aantal machines}}$$

en een verlies-coëfficiënt voor de monteurs:

$$q_r \overset{def}{=} \frac{\text{verwachting van het aantal monteurs zonder werk}}{\text{totaal aantal monteurs}}$$

Het ligt voor de hand dat het systeem beter werkt naarmate de verliesquotiënten kleiner zijn.

Bij vergelijking van de resultaten der berekeningen, geordend in de volgende tabel, blijkt dat het tweede systeem aanzienlijk efficiënter werkt, niettegenstaande het meer machines per monteur omvat.

	Systeem I	Systeem II
Aantal machines	6	20
Aantal monteurs	1	3
Aantal machines per monteur	6	6 2/3
q_n=verliescoëff. mach,wachtend of in rep.	0,1408	0,1063
q_m=verliescoëff. mach. in wachtrij	0,0547	0,0169
q_r=verliescoëff. monteurs	0,4845	0,4042

Tabellen in een in het Zweeds verschenen artikel van C. Palm (Industritidningen Norden, 1947) geven voor verschillende waarden van de voorkomende parameters de meest economische verhouding van het aantal machines per monteur.

In L.G. Peck en R.N. Hazelwood, Finite queueing tables, Wiley New York 1958, zijn tevens dergelijke quotiënten die een maat voor de efficiency vormen, voor allerlei mogelijke systemen getabelleerd.

6.2

Sinds lang is het streven naar werkverdeling in het productieproces zeer algemeen aanvaard. Een lopende band produceert in een hoog tempo, o.a. doordat eenieder die deelheeft aan het productieproces in zijn eigen taak gespecialiseerd is. In elk geval zijn minder investeringen aan gereedschappen, machines en vakopleiding vereist dan wanneer ieder in zijn eentje het gehele product zou maken. Het vlot functionneren van deze serieschakeling van de verschillende onderdelen van het productieproces staat of

valt evenwel met de wijze van aankomst van de te bewerken producten. Deze zal namelijk strikt volgens schema moeten verlopen, met name met constante tussenpozen. Ook de tijd nodig om de verschillende bewerkingen te verrichten moet constant zijn, en overal ongeveer gelijk. Uiteraard moeten deze bedieningstijden kleiner zijn dan de intervallen tussen de aankomsten, wil men steeds toenemende stagnatie vermijden.
Gedemonstreerd zal nu worden, aan de hand van een voorbeeld ontleend aan Flagle, Operations Research and Systems Engineering, dat deze in het deterministische geval zo efficiënt gebleken serieschakeling de meest ongelukkige organisatieopzet blijkt te zijn wanneer de tussenaankomstintervallen en de bedieningstijden niet meer constant zijn. Dit in het bijzonder als de aankomst en iedere phase van de bediening in het geval van uiterste desintegratie van de organisatie een Poissonproces benaderen.

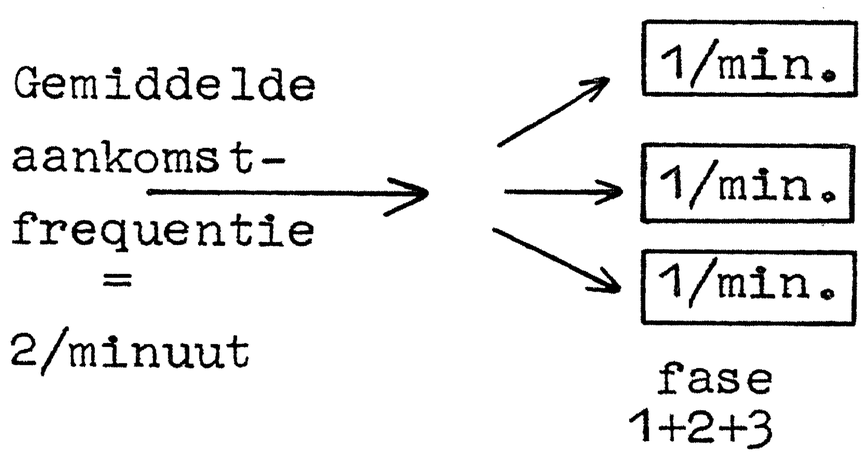
We zullen vraagelementen beschouwen die volgens een Poissonproces aankomen met een gemiddelde van $\lambda = 2$ per minuut en aan wie een dienst verleend moet worden die een exponentieel verdeelde tijd vergt, zeg gemiddeld 1 minuut. We kunnen nu parallel drie loketten installeren die ieder gemiddeld één klant per minuut helpen. We hebben dan een drie-lokettensysteem, zoals schematisch is weergegeven in figuur 9[a]. De bezettingsgraad wordt

$$\rho = \frac{2}{3} \quad .$$

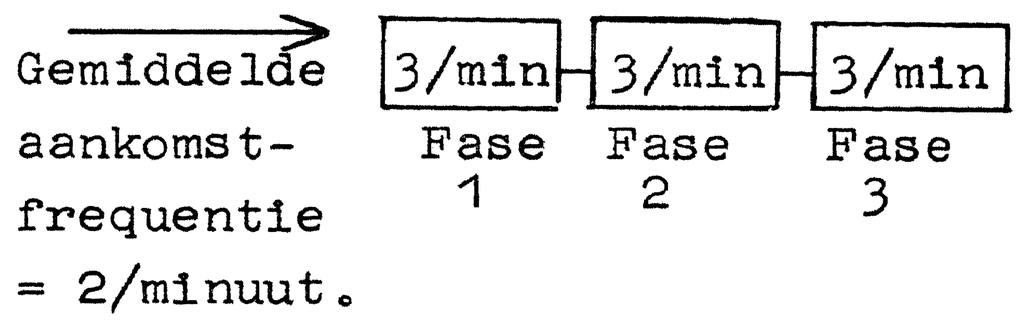
Immers

$$\rho \stackrel{\text{def}}{=} \frac{\text{aankomstfrequentie}}{\text{potentiële bedieningscapaciteit}}$$

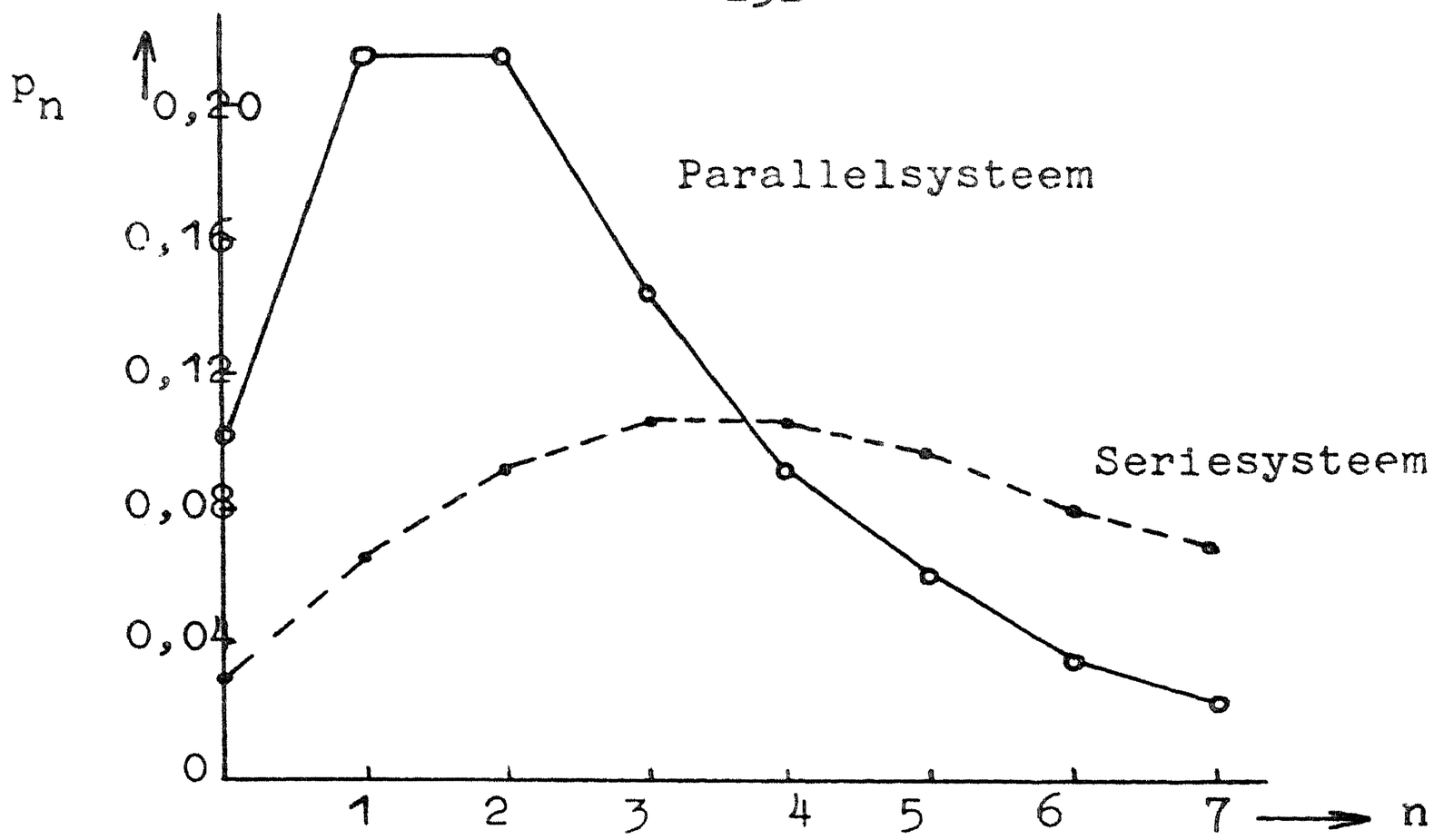
Voor dit systeem is p_n , de kans op toestand n , op te lossen. Het resultaat is uitgezet in figuur 10, in de punten verbonden door de getrokken lijn.
Bij waarden van n groter dan 3, ontstaat één gemeenschappelijke wachtrij.
We kunnen evenwel ook werkverdeling toepassen, en drie loketten in serie achter elkaar schakelen, zodat elk loket een derde van het werk doet en daarvoor een exponentiëel verdeeld tijdsinterval van gemiddeld 1/3 min. nodig heeft. Ieder van de drie fasen kan dus in de tijdseenheid (1 minuut) gemiddeld $\mu = 3$ klanten bedienen. Figuur 9^b geeft deze serieschakeling van de loketten schematisch weer. Deze organisatie brengt echter met zich mee dat zich voor elk van deze fasen een wachtrij kan vormen, hetgeen reeds doet vermoeden dat de kans op opstoppingen vergeleken met het vorige geval aanmerkelijk groter zal zijn. Dit blijkt ook uit de uitkomst van de (overigens gecompliceerde) berekening van de kans op toestand n .
Deze waarden corresponderen met die punten in figuur 10 die verbonden zijn door de gestippelde lijn.



Figuur 9^a. Parallelschakeling der loketten.



Figuur 9^b. Serieschakeling der loketten.



Figuur 10. De kansen op toestand n bij parallelschakeling en serieschakeling der loketten.

Bij het parallelsysteem blijken de toestanden met kleine n dus veel waarschijnlijker dan bij het seriesysteem, hetgeen wil zeggen dat bij de laatste opstoppingen veelvuldiger optreden.

De kans als klant niet te hoeven wachten is bij het drie-lokettensysteem:

$$p_0 + p_1 + p_2 = 0{,}555,$$

tegen

$$p_0 = 0{,}037$$

bij het seriesysteem.

We zien dus dat de kans op opstoppingen bij het seriesysteem nu dermate toeneemt dat de nadelen die hieruit voortvloeien de voordelen van de werkverdeling al snel in de schaduw zullen stellen.

Het geval van zowel aselecte input als output is natuurlijk uit oogpunt van opstoppingen wel het somberste wat men kan hebben. In de praktijk zal een systeem gedeeltelijk gedetermineerd zijn doordat de intervallen tussen de aankomsten en de bedieningstijden met meer of minder succes constant gehouden worden.

Het seriesysteem zal er dan naarmate dit beter lukt steeds beter afkomen. Men moet evenwel de invloed van kleine afwijkingen van een strak schema niet onderschatten. Gebleken is namelijk dat het dienstverlenend systeem dermate gevoelig is voor afwijkingen in de vraagsector dat bij kleine verstoringen al spoedig grote opstoppingen het gevolg zijn. Deze systemen zijn echter zeer moeilijk analytisch door te rekenen. Men past ook hier simulatiemethoden toe met behulp van electronische rekenmachines.

7 Prioriteiten

Wellicht zal uit het voorgaande zijn gebleken dat de wachttijdtheorie een poging is om na te gaan hoe wachten van mensen en goederen is te voorkomen, of althans is te reduceren. Het best kunnen opstoppingen aangepakt worden in de wortel van het kwaad, de stochastische tussenaankomstintervallen en de stochastische bedieningstijden. Kunnen deze niet verder geuniformiseerd worden, dan blijft als uitweg over het systeem enigszins te ontlasten (d.w.z. de bezettingsgraad ρ te doen dalen). Er zijn criteria geschetst wannéér dit te doen en hoe. De wachttijdtheorie heeft evenwel nog één pijl op zijn boog, zoals in § 4 reeds is aangekondigd. Men kan nagaan of het wel verstandig is de klanten in volgorde van aankomst te bedienen. De conclusie is dat men inderdaad de totale wachttijd in het systeem kan doen afnemen door hiervan af te wijken en op geschikte wijze prioriteiten te verlenen, d.w.z.: aan sommige klanten op grond van bepaalde normen voorrang geven.
Een zeer belangrijk resultaat van de theorie is bv. quantitatief inzicht in de regel dat in het geval twee klanten op bediening wachten de gemiddelde wachttijd geminimaliseerd wordt wanneer, onafhankelijk van de volgorde van aankomst, degeen met de kortste benodigde bedieningstijd het eerst behandeld wordt.

Het criterium op grond waarvan hier prioriteit moet worden verleend is dus: "Hoe kortere bedieningstijd, hoe hogere prioriteit".

Allereerst zal men moeten besluiten of men de werkzaamheden aan een vraagelement al dan niet onderbreekt op het moment van arriveren van een hoger geclasseerd element. Is dit het geval, dan wordt aan de bediening van dit laatste element onmiddellijk begonnen (zgn. pre-emptive service). Zo kent men in reparatiewerkplaatsen "stopwerk" dat direct onderbroken moet worden indien "spoedwerk" verricht moet worden. Handelt men evenwel iedere eenmaal begonnen dienst af, dan wordt het vraagelement met de hogere prioriteit pas geholpen als een loket vrijkomt (mits in de tussentijd niet een element met nog hogere prioriteit is gearriveerd). Men zal de kosten verbonden aan het onderbreken van een dienst moeten afwegen tegen de geldelijke voordelen die het nog sneller bedienen van hogere prioriteiten (wier wachttijd heel kostbaar kan zijn) met zich meebrengen.

Het kan gebeuren dat de vraagelementen, die volgens een Poissonverdeling binnenkomen, duidelijk in twee of meer groepen te onderscheiden zijn. We beperken ons tot het geval van twee groepen en we veronderstellen dat de elementen van beide groepen een exponentieel verdeelde bedieningstijd hebben en onafhankelijk van elkaar aankomen.
Een fractie p, resp. q van de vraagelementen behoort tot groep I resp. groep II en heeft als gemiddelde bedieningstijd μ_1, resp. μ_2. ($p+q=1$, $\mu_1 < \mu_2$). We zullen voorrang geven aan de elementen uit groep I.

Ook al zijn er nu twee groepen, het heeft toch nog zin om te spreken van de gemiddelde wachttijd $\mathcal{E}\underline{t}_w^*$ van een klant, immers

$$\mathcal{E}\,\underline{t}_w^* = p\;\;\mathcal{E}\,\underline{t}_{w_1} + q\;\;\mathcal{E}\,\underline{t}_{w_2} \qquad (1)$$

waarin $\mathcal{E}\underline{t}_{w_1}$, resp. $\mathcal{E}\underline{t}_{w_2}$ de gemiddelde wachttijd is van een klant uit groep I, resp. groep II. Een interessant resultaat van de theorie is dat de verhouding van $\mathcal{E}\underline{t}_{w_1}$ en $\mathcal{E}\underline{t}_{w_2}$ niet van μ_1 en μ_2 afhangt, noch van p; er geldt nl.

$$\frac{\mathcal{E}\underline{t}_{w_1}}{\mathcal{E}\underline{t}_{w_2}} = 1 - \rho \quad . \tag{2}$$

Verder kan men laten zien dat $\mathcal{E}t^*_w$ kleiner is dan de gemiddelde wachttijd bij het niet toekennen van prioriteiten, en dat de besparing groter is, naarmate ρ dichter bij 1 ligt.

Ook wanneer geen duidelijke splitsing in groepen van vraagelementen met een verschillende benodigde bedieningstijd in het oog valt, kan men met succes prioriteiten verlenen. Stel hiertoe dat er volgens een Poissonproces vraagelementen arriveren, die een dienst verlangen die een exponentieel verdeelde tijd vergt, met gemiddelde μ. We máken nu twee groepen: In groep I komen de klanten wier benodigde bedieningstijd $\leq c/\mu$ geschat wordt (c is een nader te bepalen constante) , en in groep II de overige elementen met geschatte bedieningstijd $> c/\mu$.

Geeft men weer voorrang aan de elementen uit groep I, dan kan men de constante c zó bepalen dat de gemiddelde wachttijd minimaal wordt. Deze optimale waarde van c blijkt altijd groter dan 1 te zijn: de grens tussen de twee groepen ligt dus boven de gemiddelde bedieningstijd.
We kunnen het aanbrengen van een splitsing in groepen toepassen op voorbeeld 1 van § 5, waar de aankomsten een Poissonproces volgen met gemiddelde $\lambda = 0{,}8$ aankomsten per minuut.

De bedieningstijden zijn exponentieel verdeeld rond een gemiddelde van $\mu = 1$ minuut per klant. De bezettingsgraad is dus $\rho = 0,8$.

Worden geen prioriteiten toegekend, dan kan $\mathcal{E}\,\underline{t}_w$, de verwachting van de wachttijd, worden berekend uit de formule onderaan blz. 215 in §5, met k=1 gesubstitueerd vanwege de exponentieele bedieningstijd. Hierbij moet tevens μ door $1/\mu$ vervangen, omdat μ in deze paragraaf als de gemiddelde benodigde bedieningstijd gedefinieerd is, en niet zoals in §5 als het gemiddelde aantal klanten dat per tijdseenheid bediend kan worden:

$$\mathcal{E}\,\underline{t}_w = \frac{2k}{k+1} \cdot \frac{\lambda}{1/\mu\,(1/\mu - \lambda)} = 4 \text{ minuten}$$

Voor de constante c, die de meest geschikte grens $c\mu$ tussen de twee groepen markeert, vinden we op grond van de gegevens, met behulp van de theorie 1,718.

De fractie klanten uit groep I, wier bedieningstijd $\leq 1,718\,\mu$ is, blijkt dan 0,82 te zijn. Tevens volgt uit de theorie dat hun gemiddelde wachttijd $\mathcal{E}\,\underline{t}_{w_1} = 1,36$ minuut bedraagt.

De fractie klanten uit groep II, met bedieningstijd $> 1,718\mu$, is dus 1 - 0,82 = 0,18. Zij hebben een aanzienlijk grotere gemiddelde wachttijd, immers volgens formule 2 uit deze paragraaf is

$$\mathcal{E}\,\underline{t}_{w_2} = \frac{\mathcal{E}\,\underline{t}_{w_1}}{1-\rho} \Rightarrow \mathcal{E}\,\underline{t}_{w_2} = 6,80 \text{ minuten.}$$

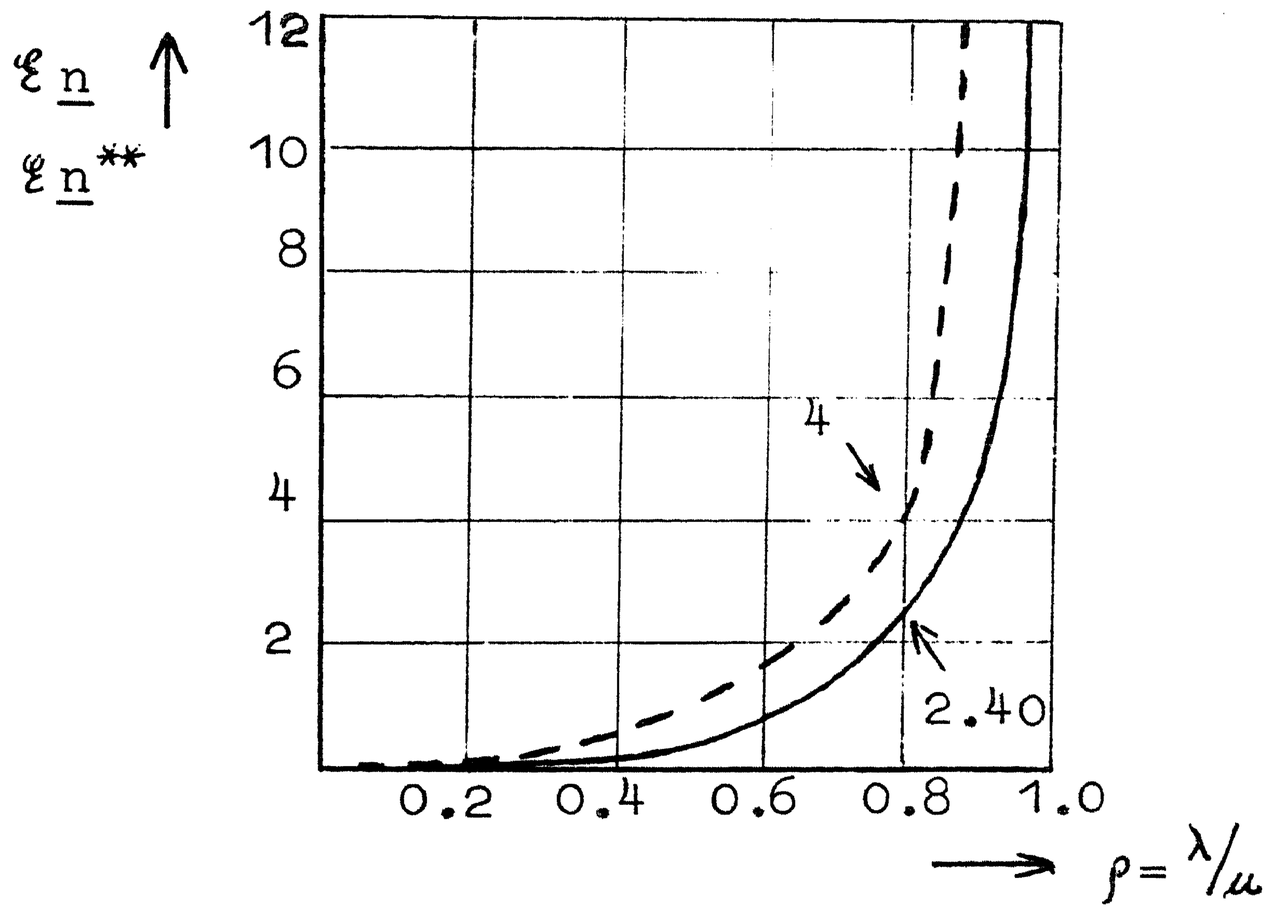
De gemiddelde wachttijd van een willekeurige klant $\mathcal{E}\,\underline{t}_w^*$, om welks reductie het gaat, wordt nu blijkens formule 1 uit deze paragraaf: $\mathcal{E}\,\underline{t}_w^* = 0,82.1,36 + 0,18.6,80 = 2,34$ minuten, hetgeen een aanzienlijke verbetering betekent t.o.v. $\mathcal{E}\,\underline{t}_w = 4$ minuten, de gemiddelde wachttijd zonder het toekennen van prioriteiten.

De boven beschreven splitsing in twee groepen kan men voor iedere groep afzonderlijk weer uitvoeren. Een verdere daling van de gemiddelde wachttijd van een willekeurige klant is het gevolg.

Zet men dit proces ad infinitum voort dan wordt uiteindelijk een splitsing in ∞ veel groepen verkregen. Kent men vervolgens prioriteiten toe naar afnemende bedieningstijd, dan wil dit zeggen, dat bij vrijkomen van een loket dàt element uit de rij aan de beurt zal zijn, waarvan de bediening de minste tijd vergt. Het "Die het eerst komt, het eerst maalt", dat bij een splitsing in eindig veel prioriteitsklassen binnen deze klassen nog geldt, is dus helemaal doorbroken. Past men deze rijdiscipline toe dan worden de verwachtingen van de toestand van het systeem uit de figuur van voorbeeld 1 in § 5 bij verschillende ρ (de gestippelde lijn in fig.11) gereduceerd tot de waarden van de getrokken lijn.
Voor $\rho = 0{,}8$ bedraagt $\mathfrak{E}_{\underline{n}}^{**} \stackrel{def}{=}$ de verwachting van de toestand van het systeem voor oneindig veel prioriteitsklassen:

$$\mathfrak{E}_{\underline{n}}^{**} = 2{,}40,$$

zoals uit figuur 11 af te lezen is.



Figuur 11. De verwachting van de toestand van het systeem zonder prioriteiten (gestippelde lijn) en bij oneindig veel prioriteitsklassen (getrokken lijn), bij verschillende bezettingsgraden, ontleend aan T.E. Phipps, zie Operations Research 4, 1956, p.76.

Dit resultaat kunnen we met behulp van formule (5) uit §5 omrekenen naar de corresponderende gemiddelde wachttijd $\mathcal{E}\underline{t}_w^{**}$

Immers

$$\mathcal{E}\underline{n}^{**} = \lambda \quad \mathcal{E}\underline{t}_w^{**} + \rho$$

geeft

$$\mathcal{E}\underline{t}_w^{**} = \quad 2 \text{ minuten}$$

hetgeen een verbetering betekent ten opzichte van

$$\mathcal{E}\underline{t}_w^{*} = 2{,}34 \text{ minuten},$$

de waarde van de verwachting van de wachttijd van een willekeurige klant bij twee prioriteitsklassen.

In de praktijk zal men de benodigde bedieningstijd (bv. de tijd nodig voor een bepaalde reparatie aan een machine) vaak niet zò scherp kunnen schatten dat men de laatste methode met succes zal kunnen uitvoeren. Men zal meestal volstaan met een beperkt aantal prioriteitsklassen. Zo kan men bij een splitsing in twee groepen voor de twijfelgevallen een menggroep maken, die de prioriteit tussen de eerste en tweede groep krijgt.

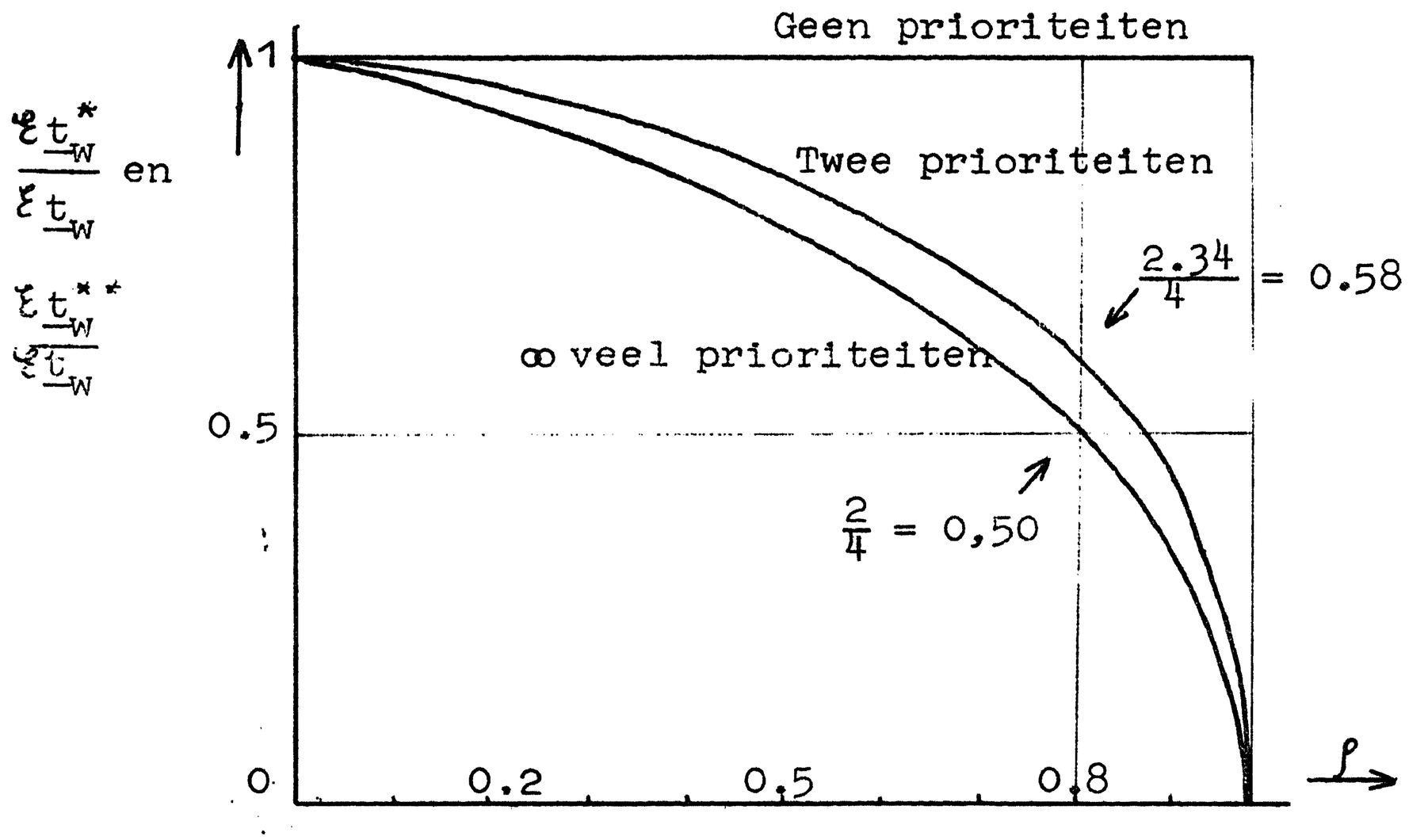
Dat men practisch altijd zal moeten afzien van een splitsing in ∞ veel groepen is eigenlijk niet bijzonder spijtig omdat het grootste deel van de mogelijke vermindering van de gemiddelde wachttijd al blijkt op te treden bij een splitsing in twee groepen. In de volgende figuur is dit te zien voor systemen waarvan de intervallen tussen opvolgende aankomsten én de bedieningstijden beide exponentieel verdeeld zijn. In de behandelde voorbeelden was $\rho = 0{,}8$. De daar gevonden resultaten zijn in figuur 12 apart aangegeven.

Hier varieren de ρ-waarden tussen 0 en 1, zie figuur 12.

De quotienten

$$\frac{\mathcal{E}\,t_w^{*}}{\mathcal{E}\,t_w} \qquad \text{en} \qquad \frac{\mathcal{E}\,t_w^{**}}{\mathcal{E}\,t_w}$$

geven een maat voor de reductie van de wachttijd in de gevallen dat respectievelijk twee en ∞ veel prioriteiten worden toegekend.



Figuur 12. Het effect van verschillende soorten prioriteitsverlening, ontleend aan F.W. Steutel; zie Statistica Neerlandica 13 (1959) nr 4, waar tevens een literatuurverwijzing te vinden is.

In deze § werd speciaal de wachttijd van een willekeurige klant geminimaliseerd. Zijn de kosten van oponthoud voor alle prioriteitsgroepen dezelfde en zijn deze evenredig met de wachttijd, dan worden aldus ook de kosten geminimaliseerd. Gaat men differentiatie aanbrengen in de kosten verbonden aan oponthoud van de verschillende groepen, dan kan hiermee rekening worden gehouden zonder dat dit de wachttijdberekeningen bijzonder compliceert.