Prestatie-Analyse van Telecommunicatiesystemen

door Frits C. Schoute

Inhoudsopgave

Lijst van figuren

Lijst van tabellen

Voorwoord

Telecommunicatie maakt in onze samenleving een sterke groei door. Niet alleen is er een toename van het telefonieverkeer, nog sterker groeit de hoeveelheid data die via communicatienetwerken uitgewisseld wordt. Het vakgebied van telecommunicatie is zeer in beweging, nu nieuwe standaards afgesproken worden en nieuwe technieken worden toegepast. Steeds meer mensen raken diepgaand betrokken bij telecommunicatie.

Dit betekent dat er een sterke behoefte is aan opleidingen en lesmateriaal op het gebied van telecommuncatie. Er zijn al enkele Nedelandstalige werken verschenen die het functioneren van moderne telecommunicatie beschrijven. Naast deze kwalitatieve aanpak is er ook behoefte aan een kwantitatieve analyse van telecommunicatiesystemen. Het is het doel van dit boek om in deze laatste behoefte te voorzien.

Voor gebruikers, beheerders en producenten van telecommunicatiesystemen is het van belang om te kunnen specificeren hoeveel verkeer een systeem kan verwerken en onder welke condities. In dit boek worden de basis begrippen van telecommunicatie verkeer uitgelegd en met uitgewerkte voorbeelden geïllustreerd. Het boek is ontstaan uit leerstof voor een post-HBO cursus. Het is bedoeld voor HBO en universitaire studenten en voor diegenen die in de praktijk betrokken zijn bij de kwantitatieve aspecten van telecommunicatiesystemen.

Het is wenselijk dat diegene die deze stof gaat bestuderen enige kennis van waarschijnlijkheidsrekening heeft, hoewel bij eerdere cursussen gebleken is dat dit niet strikt noodzakelijk is. Basisbegrippen uit de waarschijnlijkheidsrekening, zoals verwachtingswaarde, variantie en kansverdelingen, worden kort herhaald. Uitgebreider wordt ingegaan op Markov-processen.

Markov-modellen komen steeds weer terug bij de presentatie van de elementaire resultaten van wachtsystemen, verliessystemen en bij betrouwbaarheidsberekeningen. Dit wordt uitvoerig behandeld en steeds worden er uitgewerkte voorbeelden gegeven van toepassingen op vraagstukken uit de prestatie-analyse van telecommunicatiesystemen.

Als een bijproduct geeft dit boekje een aantal aardige toepassingen van eenvoudige kansrekening. Het hoofddoel is echter de basiskennis te presenteren die gebruikt kan worden om zodanig telecommunicatiesystemen te specificeren dat ze “op maat” zijn. Zowel voor de producent als de gebruiker leidt dit tot optimale prestatie van de geïnvesteerde middelen.

Dit boek kwam tot stand dankzij de cursus Digitale Telecommunicatie waar ik het vak Prestatie-analyse mocht verzorgen. Dank aan de organisatoren van de cursus en aan de vele

cursisten. Hun reacties hebben bijgedragen tot de kwaliteit van dit boek. In het bijzonder noem ik André van Wieringen, die het gehele manuscript heeft doorgewerkt en met tal van verbeteringen is gekomen.

Voor de realisatie van de electronische versie van dit boek ben ik veel dank verschuldigd aan Geert Awater. Nu het boek is uitverkocht, kan via het World Wide Web aan de vraag naar het boek voldaan worden.

Hoofdstuk 1

Inleiding

1.1 Prestatie-analyse van telecommunicatiesystemen

Reeds aan het begin van deze eeuw werden kwantitatieve aspecten van telecommunicatiesystemen bestudeerd. De meest bekende naam uit die tijd is die van A.K. Erlang, een Deense wiskundige die, onder meer, met waarschijnlijkheidstheorie een formule gaf voor de blokkeringskans voor een telefoonoproep op een bundel van N lijnen. Deze theorie wordt aangeduid met Verkeerstheorie (Engels: Teletraffic Theory).

In de laatste decennia is ook vanuit de computer science een grote uitbreiding ontstaan van de toepassing van waarschijnlijkheidstheorie op verlies- en vooral wachtproblemen. In het Engels wordt dit (Computer) Performance Analysis genoemd, ofwel, in het Nederlands, Prestatie-analyse (van computersystemen). Vanwege de huidige tendens van integratie van Verkeerstheorie en Prestatie-analyse hebben wij gekozen voor de titel: Prestatie-analyse van Telecommunicatiesystemen.

1.2 Motivatie

In andere boeken wordt vooral geschreven over hoe telecommunicatiesystemen werken, een in de eerste plaats kwalitatieve beschrijving. Hier willen we ingaan op de kwantitatieve kant van zulke systemen. We zullen leren vragen te beantwoorden als: hoe lang is de gemiddelde wachttijd van een bericht, hoe groot is de blokkeringskans voor een telefoonoproep, hoeveel neemt de betrouwbaarheid van een systeem toe als het dubbel uitgevoerd is, enzovoorts.

Als we uitspraken doen over bijvoorbeeld wachttijden dan is dit op zijn best het geven van kentallen zoals het gemiddelde en overschrijdingskansen van bepaalde wachttijden. Omdat vaak de parameters van het probleem slechts bij benadering bekend zijn, is het niet altijd even zinvol om de gemiddelde wachttijd in zes cijfers achter de komma te voorspellen. Wel zinvol is het om uit te rekenen hoe de kentallen afhangen van de parameter waarden en van de verschillende opties in het systeem ontwerp. Hiervoor wordt de theoretisch basis aangedragen. En de talrijke figuren en uitgewerkte voorbeelden illustreren hoe die theorie zinvol kan worden toegepast.

1.3 Indeling van dit boek

Omdat de prestaties van de systemen die we beschouwen beïnvloed worden door het toevalsgedrag van externe gebeurtenissen, hebben we in eerste instantie te maken met waarschijnlijkheidsrekening. De achtergrond van de studenten in de waarschijnlijkheidsrekening en de onderliggende wiskunde is zeer verschillend. We zullen daarom eerst in hoofdstuk 2 een herhaling geven van die basisbegrippen uit de waarschijnlijkheidsrekening die voor ons van belang zijn. In hoofdstuk 3 worden een aantal voor ons belangrijke kansverdelingen behandeld.

Met behulp van Markov-processen kunnen verlieskansen, wachttijden en betrouwbaarheidsmaten op een universele wijze geanalyseerd worden. Daarom gaan we in hoofdstuk 4 in op de berekening van de kansverdeling van de toestand van een Markov-proces.

Als het gaat om wachttijden van bits die als pakket over een lijn verzonden moeten worden of om berichten binnen een processor die verwerkt moeten worden, hebben we steeds te maken met wachtsystemen. In hoofdstuk 5 worden de basisresultaten uit de wachttijdtheorie gegeven en steeds met uitgewerkte voorbeelden toegepast op telecommunicatie problemen.

Het klassieke gedeelte van prestatie-analyse van telecommunicatiesystemen is de berekening van verlieskansen voor telefonieverkeer. Ook buiten de telefonie heeft deze kansberekening zijn toepassing. In hoofdstuk 6 gaan we hierop in.

Tenslotte laten we in hoofdstuk 7 zien dat we met het inmiddels geoefende begrip van Markov-processen ook betrouwbaarheid van systemen kunnen berekenen.

Hoofdstuk 2

Begrippen uit de kansrekening

2.1 Stochastische variabelen

Het gedrag van de systemen die we beschouwen wordt beïnvloed door toevalligheden zoals de aankomst van een datapakket van een andere centrale, of bijvoorbeeld een oproep van een telefoonabonnee. We beschrijven dit met toevalsgrootheden, ofwel stochastische variabelen. We beschouwen discrete en continue stochastische variabelen.

Een continue stochastische variabele neemt waarden aan op de reële rechte $(-\infty, +\infty)$ of de positieve reële rechte $(0, +\infty)$. Een voorbeeld is het verschil in aankomsttijd tussen twee opeenvolgende datapakketten in een centrale. Dit kan gemodelleerd worden als een positief reëel getal waarvan de waarde door het toeval bepaald is.

Een discrete stochastische variabele neemt waarden aan in een aftelbare verzameling. Bijvoorbeeld als we een octet (8 bits) verzenden over een transmissiekanaal met ruis dan is het aantal foute ontvangen bits een toevalsgrootheid die de waarden 0, 1, tot en met 8 kan aannemen. De verzameling $\{0, 1, \ldots 8\}$ is een eindige (en aftelbare) verzameling.

Als we praten over het aantal wachtenden in een wachtrij dan wordt vaak, voor de mathematische eenvoud, de wachtruimte van onbeperkte omvang gemodelleerd. Dit betekent dat de stochastische variabele die het aantal wachtenden op een bepaald tijdstip voorstelt, waarden aanneemt in de aftelbare verzameling $\{0, 1, 2, \ldots\}$.

2.2 Verdelingsfunctie

Om iets meer te kunnen zeggen over een stochastische variabele moet de kans worden gespecificeerd dat bepaalde waarden aangenomen kunnen worden. Een manier om dit te specificeren is de verdelingsfunctie (van die stochastische variabele). Elke stochastische variabele (discreet of continu) heeft een verdelingsfunctie. Een verdelingsfunctie $F(\cdot)$ is gedefinieerd door:

$$F(x) = \Pr\{X \leq x\} \tag{2.1}$$

Hierin betekent $\Pr\{A\}$: de kans (Probability) op de gebeurtenis A. De waarde van $F(\cdot)$ in het punt x geeft de kans dat de toevalsgrootheid (=stochastische variabele) X kleiner dan of gelijk aan x is.

Voorbeeld 2.1

Terwille van de grafiek van dit voorbeeld beschouwen we een transmissiekanaal met een behoorlijke kans op bitfouten. De toevalsgrootheid X stelt het aantal fouten in één octet voor. Voor de kansverdeling van X geldt de volgende tabel:

# fouten	kans	cumulatief
0	0.430	0.430
1	0.383	0.813
2	0.149	0.962
3	0.033	0.995
4	0.005	1.000
5	0.000	1.000
6	0.000	1.000
7	0.000	1.000
8	0.000	1.000

Gevraagd: De grafiek van de verdelingsfunctie van de discrete stochastische variabele X.

De gevraagde grafiek is getekend in figuur 2.1. Wat opvalt zijn de discontinuïteiten in de verdelingsfunctie. De waarde van de verdelingsfunctie in het punt 0.9 is $F(0.9) = \Pr\{X \leq 0.9\} = \Pr\{X = 0\} = 0.430$. Diezelfde uitkomst krijgen we in het punt 0.99, in 0.999, enzovoort. Echter in het punt 1.0 geldt ineens $F(1.0) = \Pr\{X \leq 1.0\} = \Pr\{X = 0 \text{ of } X = 1\}$ $= 0.430 + 0.383 = 0.813$. Vandaar de discontinuïteit en vandaar de knobbel getekend links aan het horizontale lijnstuk dat aangeeft dat die functie waarde hoort bij de onderliggende x waarde.

Het argument van een verdelingsfunctie loopt van $-\infty$ tot $+\infty$ en de waarde van een verdelingsfunctie gaat dan van 0 tot 1. We schrijven dit als:

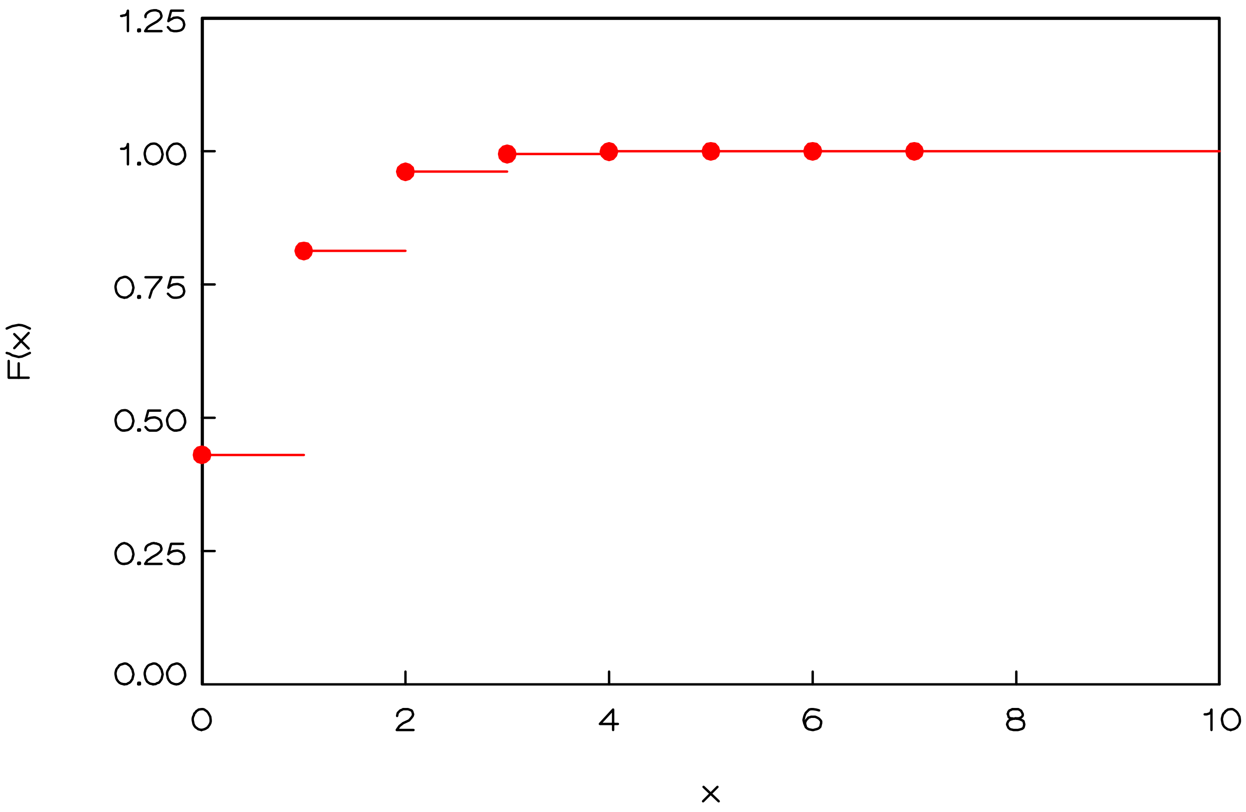
$$F(-\infty) = 0 \tag{2.2}$$

$$F(\infty) = 1 \tag{2.3}$$

In voorbeeld 2.1 worden die limietwaarden ook echt bereikt; nl.

$$F(x) = 0 \quad \text{voor} \quad x < 0 \tag{2.4}$$

$$F(x) = 1 \quad \text{voor} \quad x \geq 0 \tag{2.5}$$



Figuur 2.1: Verdelingsfunctie van aantal bitfouten in een octet bij voorbeeld 2.1.

Als het gaat om de verdelingsfunctie van een continue stochastische variabele dan is ook het verloop van de verdelingsfunctie continu.

Voorbeeld 2.2

Gegeven is een andere stochastische variabele X, die de tijd voorstelt die verloopt tussen twee opeenvolgende aankomsten van een berichtenstroom, met de verdelingsfunctie:

$$F(x) = \begin{cases} 1 - e^{-x/2} & \text{voor } x > 0 \\ 0 & \text{voor } x \leq 0 \end{cases} \tag{2.6}$$

In figuur 2.2 is een grafiek getekend van deze verdelingsfunctie voor $x > 0$.

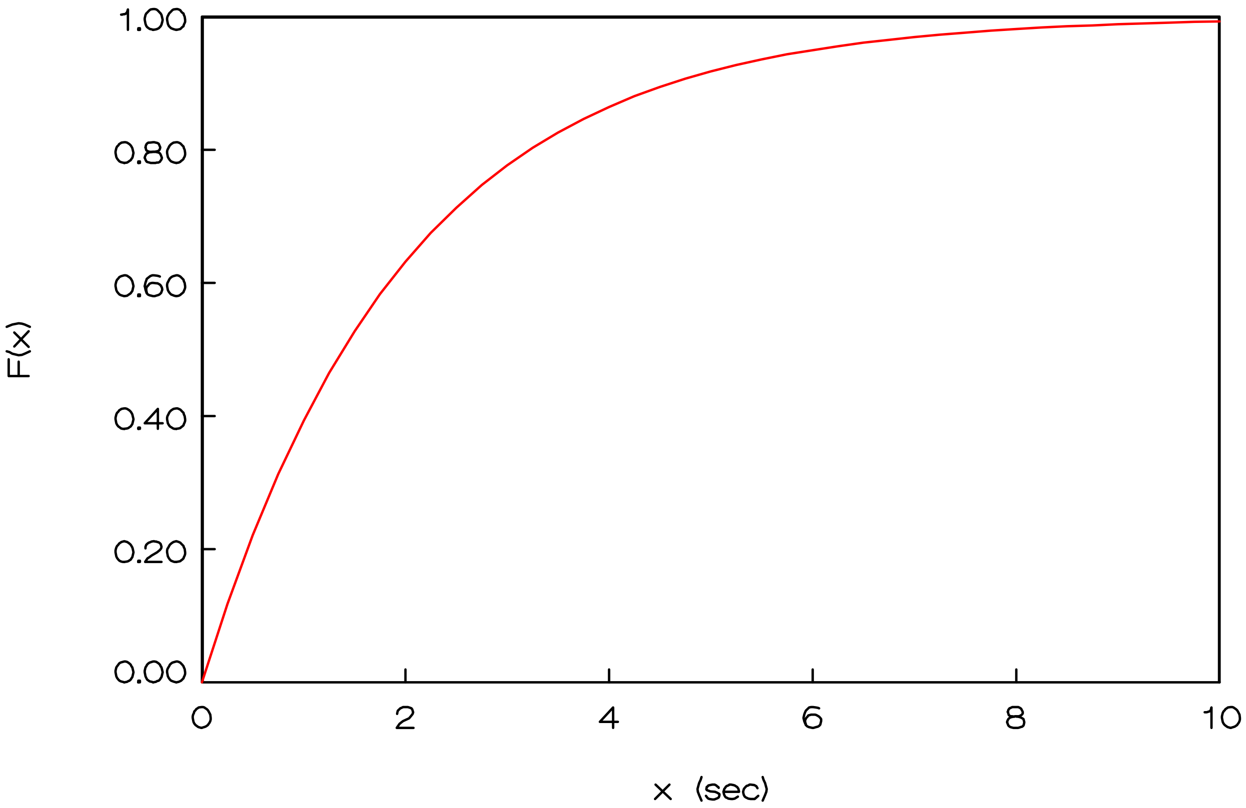
Gevraagd:

a) De kans dat de tussenaankomsttijd X kleiner is dan of gelijk is aan 1 tijdseenheid.

b) De kans dat de tussenaankomsttijd X groter is dan 2 tijdseenheden.

c) Bereikt de verdelingsfunctie van X zijn onder en bovengrens?

Het antwoord op a) volgt direct uit de definitie van verdelingsfunctie. Namelijk, $\Pr\{X \leq 1\} = F(1) = 0.393$.



Figuur 2.2: Verdelingsfunctie van de tussenaankomsttijd van twee opeenvolgende berichten bij voorbeeld 2.2.

Bij b) willen we de kans weten op het complement van de gebeurtenis $X \leq 2$. Deze kans is 1 minus de kans op $X \leq 2$. Dus, $\Pr\{X > 2\} = 1 - \Pr\{X \leq 2\} = 1 - F(2) = e^{-1} = 0.368$.

De onderlimiet 0 van de verdelingsfunctie wordt bereikt voor $x \leq 0$, omdat we hier met kans 1 te maken hebben met een positieve toevalsgrootheid. Maar voor deze stochastische variabele is geen bovengrens aan te geven die nooit overschreden wordt. De verdelingsfunctie bereikt dan ook nooit de waarde 1. Er geldt wel:

$$\lim_{x \to \infty} F(x) = 1 \tag{2.7}$$

De verdelingsfunctie geeft de kans dat de stochastische variabele een gegeven waarde niet overschrijdt. De kans dat de stochastische variabele in een bepaald interval $(a, b]$ valt is ook eenvoudig uit te drukken in termen van de verdelingsfunctie $F(\cdot)$. Het gaat om de kans $\Pr\{a < X \leq b\}$. Deze is gelijk aan de kans op de gebeurtenis $X \leq b$ verminderd met de kans op de gebeurtenis $X \leq a$:

$$\Pr\{a < X \leq b\} = F(b) - F(a) \tag{2.8}$$

Deze formule geldt zowel voor continue als discrete stochastische variabelen.

2.3 Kansdichtheid

Alleen voor een continue stochastische variabele is een kansdichtheid gedefinieerd en wel als de afgeleide van de verdelingsfunctie. Het heeft de interpretatie van kans per eenheid van tijd (in het voorbeeld dat we bekeken), of per eenheid van lengte, al naar gelang de eenheid waarin de stochastische variabele is uitgedrukt. Voor een verdelingsfunctie $F(\cdot)$ wordt de bijbehorende kansdichtheid met een kleine letter, $f(\cdot)$, aangegeven.

Per definitie is dus:

$$f(x) = \frac{dF(x)}{dx} \tag{2.9}$$

Omdat we weten dat $F(-\infty) = 0$, kan de verdelingsfunctie ook weer terug gekregen worden uit de kansdichtheid:

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{2.10}$$

De definitie van de kansdichtheid kan ook geschreven worden als:

$$f(t) = \lim_{\Delta t \to \infty} \frac{F(t + \Delta t) - F(t)}{\Delta t} \tag{2.11}$$

Hieruit volgt dat de kans dat de stochastische variabele een waarde aanneemt in het infinitesimale interval $(t, t + dt]$ gelijk is aan:

$$\Pr\{t < X \leq t + dt\} = f(t)dt \tag{2.12}$$

Voorbeeld 2.3

De tijd die nodig is om een bericht van de magnetische schijf van een electronic mail systeem te halen (opvraagtijd), wil men modelleren als een stochastische variabele. De volgende gegevens zijn gemeten:

minimum	20 ms
maximum	80 ms

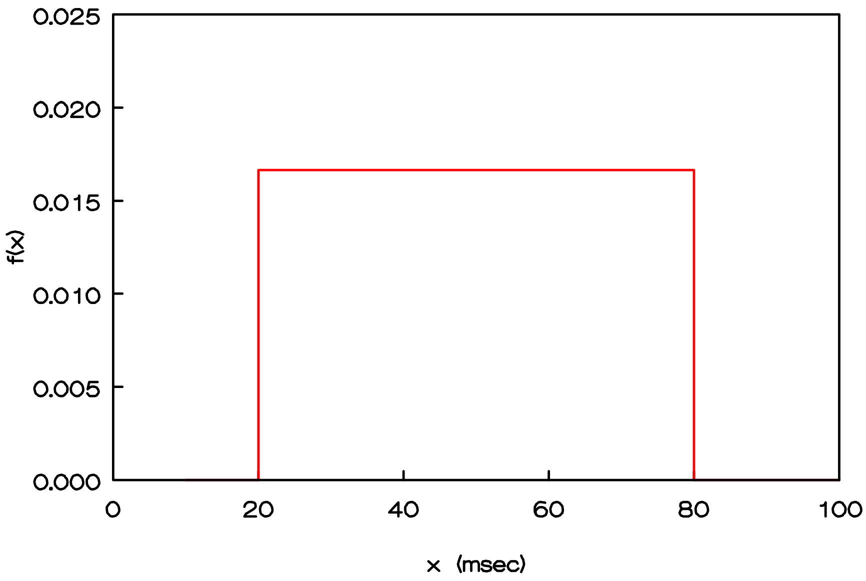
alle tussenliggende tijden zijn even waarschijnlijk.

Gevraagd: wat is de verdelingsfunctie en de kansdichtheid van deze stochastische variabele? (Van daaruit kunnen weer berekend worden: het gemiddelde, de variantie etc.).

Het gegeven dat alle tijden even waarschijnlijk zijn laat zich vertalen in een continue stochastische variabele met een kansdichtheid die konstant is op het interval $(20, 80]$ (De tijdschaal van ms wordt impliciet verondersteld).

De grafiek van de kansdichtheid is getekend in figuur 2.3

De hoogte van de rechthoek in bovenstaande grafiek volgt uit het gegeven dat de oppervlakte onder de curve gelijk aan 1 moet zijn. Dit volgt weer uit (2.1) en (2.10). Uit (2.10) volgt ook dat de waarde van de verdelingsfunctie in het punt x gelijk is aan de oppervlakte van



Figuur 2.3: Kansdichtheid van opvraagtijd in voorbeeld 2.3.

de rechthoek van figuur 2.4 links van x. De grafiek van de verdelingsfunctie is gegeven in figuur 2.4.

Voorbeeld 2.4

In voorbeeld 2.2 werd de verdelingsfunctie van tussenaankomsttijd van twee opeenvolgende berichten gegeven. Dit was een continue stochastische variabele en heeft dus ook een kansdichtheid.

Gevraagd: Hoe ziet de kansdichtheid van deze stochastische variabele eruit?

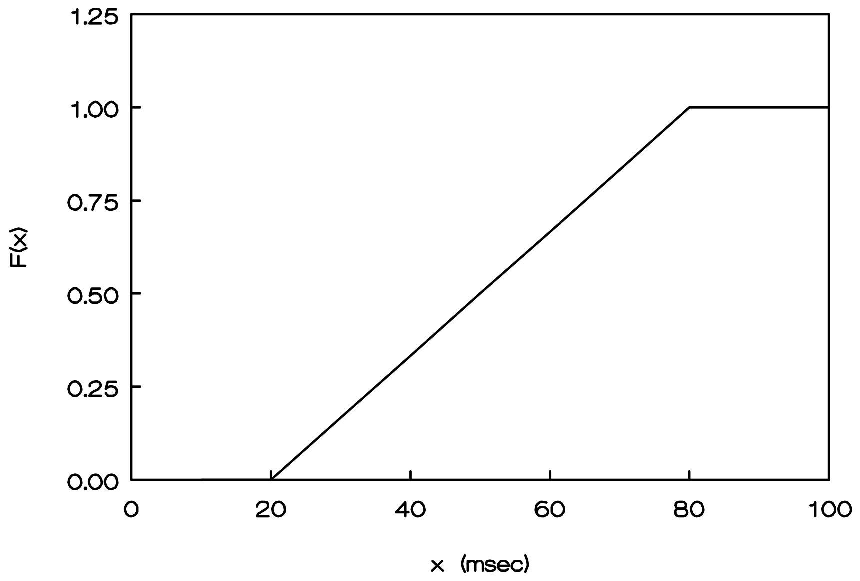
Differentiëren van $F(\cdot)$ zoals gedefinieerd in voorbeeld 2.2 levert op:

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & \text{voor } x > 0 \\ 0 & \text{voor } x < 0 \end{cases} \tag{2.13}$$

De kansdichtheid voor $x > 0$ is getekend in figuur 2.5.

2.4 Histogram

Bij kansverdeling van een discrete stochastische variabele hoort een histogram. Hierin geeft de hoogte van het balkje p_i, dat getekend is bij een bepaalde mogelijke uitkomst x_i van de



Figuur 2.4: Verdelingsfunctie van opvraagtijd in voorbeeld 2.3.

stochastische variabele, de kans aan dat die uitkomst zal optreden. Deze kans is gelijk aan de sprong die de verdelingsfunctie maakt in dat punt.

$$p_i = \Pr\{X = x_i\} = F(x_i) - F(x_{i-1}) \tag{2.14}$$

Hierin is $\{x_0, x_1, \ldots\}$ de verzameling van mogelijke uitkomsten.

Het histogram zoals hierboven beschreven is een manier om een kansverdeling van een discrete stochastische variabele te definiëren.

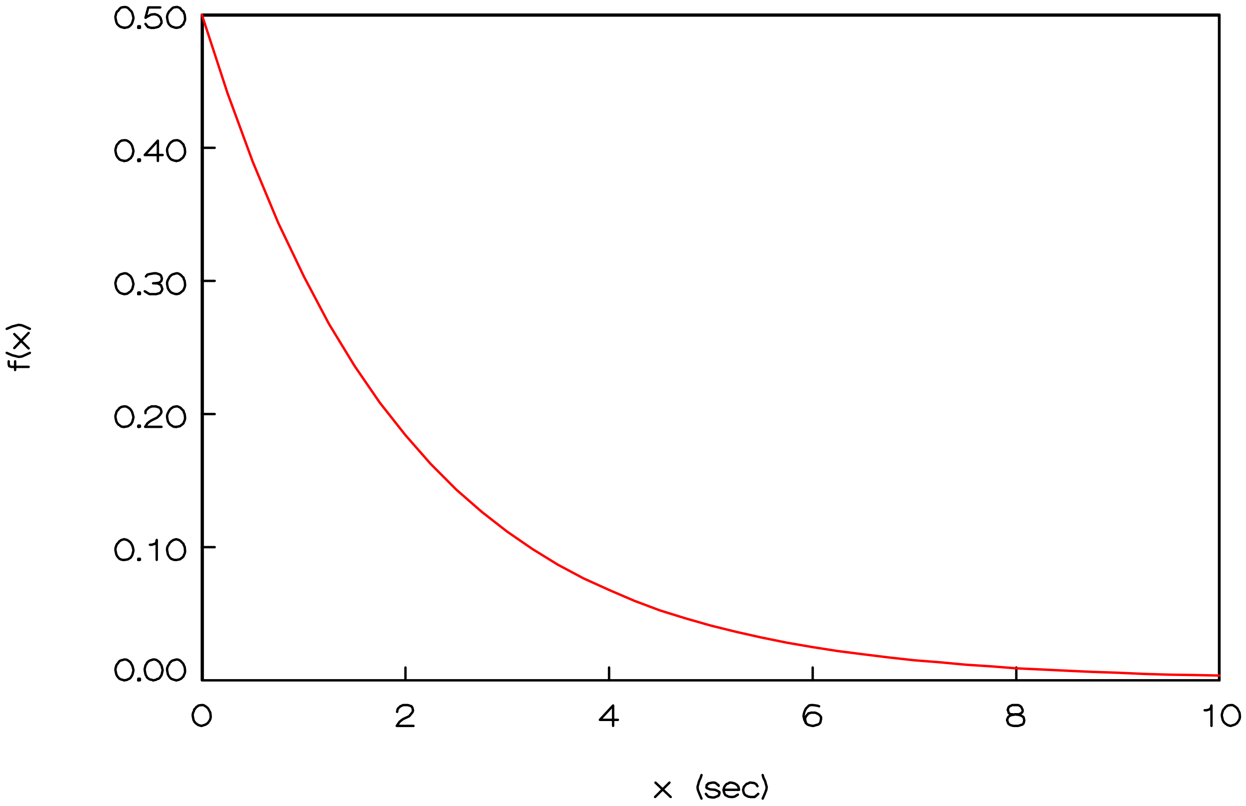
Voorbeeld 2.5

De toevalsgrootheid X uit voorbeeld 2.1 die het aantal bitfouten in een octet geeft is een discrete stochastische variabele.

Gevraagd: Hoe ziet het histogram van de kansverdeling van deze stochastische variabele X eruit?

In de tabel in voorbeeld 2.1 die de kansverdeling van X geeft staan zowel de waarden van de verdelingsfunctie (meest rechtse kolom) als de sprongen van de verdelingsfunctie $F(x_i) - F(x_{i-1}) = p_i$. Het histogram is dan als in figuur 2.6:

In het geval dat men waarnemingen doet van een discrete stochastische variabele wordt ook een histogram gevormd door in elk balkje te tellen hoe vaak de desbetreffende waarde is voorgekomen. Wanneer deze balkjes nu genormeerd worden door te delen door het totale aantal waarnemingen, dan gaat dit histogram-uit-waarnemingen in de limiet (als het aantal waarnemingen $\to \infty$ over in het histogram dat de kansverdeling definieert.



Figuur 2.5: Kansdichtheid van de tussenaankomsttijd van twee opeenvolgende berichten in voorbeeld 2.4.

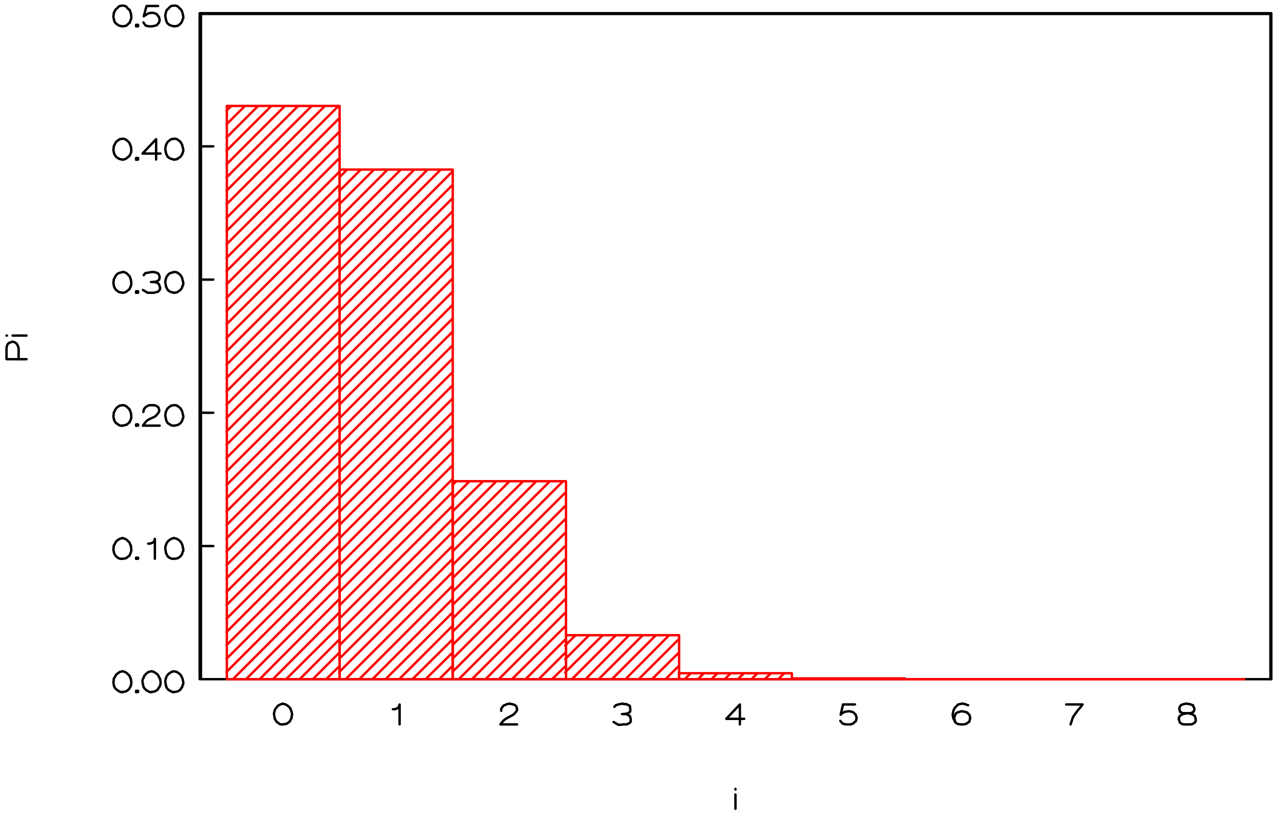
Ook bij waarnemingen van een continue stochastische variabele wordt vaak een histogram getekend door voor een reeks van intervallen, met breedte b_i, te tellen hoe vaak een waarneming in dat interval komt. Hier is een limietstelling wat moeilijker te preciseren; we kunnen zeggen dat als het aantal waarnemingen $\to \infty$ en $b_i \to 0$, dan gaat het histogram in vorm steeds meer lijken op de kansdichtheid van die stochastische variabele.

2.5 Kentallen van een kansverdeling

2.5.1 Gemiddelde

De meest primaire karakterisering van een stochastische variabele en zijn kansverdeling is de verwachtingswaarde. Dit is een gewogen gemiddelde van alle mogelijke uitkomsten. Synoniem met de term verwachtingswaarde wordt ook gebruikt: gemiddelde, eerste moment en de symbolen m_1 en $\mathrm{E}[X]$.

Voor een discrete stochastische variabele X met mogelijke uitkomsten $\{x_0, x_1, \ldots, x_{n-1}\}$



Figuur 2.6: Histogram van het aantal bitfouten in een octet in voorbeeld 2.5.

wordt de verwachtingswaarde gegeven door

$$m_1 = \mathrm{E}[X] = \sum_{i=0}^{n-1} x_i p_i \tag{2.15}$$

Voor een continue stochastische variabele krijgen we een soortgelijke gewogen som van mogelijke waarden:

$$m_1 = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) dx \tag{2.16}$$

Als de continue stochastische variabele alleen waarden > 0 kan aannemen dan kan het een enkele keer handig zijn de verwachtingswaarde met de volgende, niet intuïtieve, formule uit te rekenen:

$$m_1 = \mathrm{E}[X] = \int_{0}^{\infty} [1 - F(x)] dx \qquad \text{indien altijd } X > 0 \tag{2.17}$$

Voorbeeld 2.6

Gevraagd: Het gemiddelde voor de stochastische variabele van voorbeeld 2.1 en voor die van voorbeeld 2.2.

Het gemiddelde aantal bitfouten in een octet is een gewogen gemiddelde van de uitkomsten

0,1,2,...,8 waarbij de weegfactor 1 een uitkomst gelijk is aan de kans op die uitkomst. Overeenkomstig formule (2.15) met de kansen uit voorbeeld 2.1, vinden we voor het gemiddelde aantal bitfouten (met weglating van termen die nul zijn):

$$m_1 = 1 \times 0.383 + 2 \times 0.149 + 3 \times 0.033 + 4 \times 0.005 = 0.800 \tag{2.18}$$

De tijd tussen twee aankomsten is een continue stochastisch variabele. De verdelingsfunctie van deze stochastische variabele werd gegeven in voorbeeld 2.2:

$$F(x) = \begin{cases} 1 - e^{-x/2} & \text{voor } x > 0 \\ 0 & \text{voor } x \leq 0 \end{cases} \tag{2.19}$$

en de kansdichtheid in voorbeeld 2.4:

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2} & \text{voor } x > 0 \\ 0 & \text{voor } x < 0 \end{cases} \tag{2.20}$$

Berekening van het gemiddelde volgens formule (2.16) komt losjes gesproken neer op het vermenigvuldiging van de uitkomst x met de kans $f(x)dx$; de sommatie van deze infinitesimale termen is niets anders dan de integraal in (2.16). Om deze integraal uit te rekenen zouden we gebruik moeten maken van partiële integratie. Deze techniek komen we verder in deze studie van prestatie-analyse van telecommunicatiesystemen niet meer tegen, vandaar dat we liever gebruik maken van formule (2.17). (Deze formule is uit (2.16) af te leiden door middel van partiële integratie).

Substitutie van de kansdichtheid $F(\cdot)$ in (2.17) levert:

$$m_1 = \int_0^\infty e^{-x/2} dx = -2e^{-x/2} \Big|_0^\infty = 2 \tag{2.21}$$

Dus de gemiddelde tijd tussen twee aankomsten in voorbeeld 2.2 is 2 tijdseenheden.

2.5.2 Tweede moment

Na het 1^e moment komt uiteraard het 2^e moment en zo kunnen we nog een tijdje doorgaan. In feite is een kansverdeling volledig bepaald door al zijn momenten. Voor onze toepassingen is het voldoende om maar tot het tweede moment te gaan. We geven eerst formeel de definitie:

$$m_2 = \mathrm{E}\left[X^2\right] = \sum_{i=0}^{n-1} x_i^2 p_i \tag{2.22}$$

voor een discrete variabele, en

$$m_2 = \mathrm{E}\left[X^2\right] = \int_{-\infty}^{-\infty} x^2 f(x) dx \tag{2.23}$$

voor een continue stochastische variabele. Het tweede moment is dus een gewogen gemiddelde van de kwadraten van de mogelijke waarden die aangenomen kunnen worden.

Voorbeeld 2.7

Gevraagd: het tweede moment van de kansverdeling van voorbeeld 2.1. Voorbeeld 2.2 slaan we verder over, omdat anders de integraalrekening meer op de voorgrond komt dan de basisbegrippen uit de kansrekening die we nodig hebben bij prestatie-analyse.

Toepassing van formule (2.22) op het voorbeeld van de bitfouten in een octet geeft, net als bij het gemiddelde, een gewogen som, maar dan een gewogen som van kwadraten:

$$m_2 = 1^2 \times 0.383 + 2^2 \times 0.149 + 3^2 \times 0.033 + 4^2 \times 0.005 = 1.356 \tag{2.24}$$

Dit getal zegt nu nog niet veel. Straks kunnen we het gebruiken bij de berekening van variantie en standaarddeviatie in de volgende voorbeelden.

2.5.3 Variantie

Een directe toepassing van de misschien nog wat abstracte grootheid m_2, vinden we, wanneer we van een stochastische variabele de variantie willen weten. De variantie kan namelijk direct uitgedrukt worden in het eerste en tweede moment. De variantie, $\mathrm{Var}[X]$, geeft aan hoe sterk een stochastische variabele, X, van zijn gemiddelde afwijkt:

$$\mathrm{Var}[X] = \sum_{i=0}^{n-1} (x_i - \mathrm{E}[X])^2 p_i \tag{2.25}$$

voor een discrete stochastische variabele en

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f(x) dx \tag{2.26}$$

voor een continue stochastische variabele. In beide gevallen geldt:

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 \tag{2.27}$$

ofwel

$$\mathrm{Var}[X] = m_2 - m_1^2 \tag{2.28}$$

Voorbeeld 2.8

Stel we willen een maat hebben voor de kwadratische afwijking van het gemiddelde van het aantal bitfouten in een octet. Deze maat (de variantie) speelt onder andere een rol als we een kansverdeling willen benaderen met een Normale verdeling.

Voor de discrete stochastische variabele van het bekende voorbeeld 2.1 berekenen we de variantie met formule (2.27). We kunnen hierin de al eerder berekende waarden van het eerste en tweede moment substitueren. Dit geeft:

$$\mathrm{Var}[X] = 1.356 - 0.800^2 = 0.716 \tag{2.29}$$

2.5.4 Standaarddeviatie

De variantie is een gewogen som van kwadraten en is dus nooit negatief. Een andere maat voor de afwijking van het gemiddelde is de wortel uit de variantie en heet standaarddeviatie (σ). Voor discrete en continue stochastische variabelen geldt dus de definitie:

$$\sigma = \mathrm{StDev}\left[X\right] = \sqrt{\mathrm{Var}\left[X\right]} \tag{2.30}$$

2.5.5 Berekening van standaarddeviatie met calculator

Op de meeste wetenschappelijke calculators zit een toets om de standaarddeviatie van een verzameling van getallen te bepalen. Voor dezelfde verzameling getallen geven twee verschillende calculators soms twee verschillende berekende waarden van de standaarddeviatie op!

De reden is dat een verzameling getallen $\{x_0, x_1, \ldots, x_{n-1}\}$ op twee manieren kan worden geïnterpreteerd:

a) De verzameling definieert een kansverdeling waarbij x_i met kans $1/n$ optreedt (in het geval dat er k waarden van de verzameling aan elkaar gelijk zijn, dan is dat zo op te vatten dat die waarde met kans k/n optreedt).

b) De verzameling vormt een reeks monsters (samples) van een kansverdeling waarvan het gemiddelde en de standaarddeviatie nog moet worden geschat.

In beide gevallen wordt het gemiddelde gewoon gegeven door:

$$m_1 = \frac{1}{n} \sum_{i=0}^{n-1} x_i \tag{2.31}$$

Voor calculators die volgens methode a) werken, correspondeert de formule voor σ^2 (= variantie) met het bepalen van het verschil van het tweede moment en het kwadraat van het eerste moment. Dit geeft de formule

$$\sigma^2 = \frac{1}{n} \sum_{i=0}^{n-1} x_i^2 - m_1^2 \tag{2.32}$$

Voor calculators die volgens methode a) werken wordt formule (2.32) nog met een factor $n/(n-1)$ vermenigvuldigd. De reden is dat wanneer gemiddelde en standaarddeviatie moeten worden geschat, het geschatte gemiddelde zo goed mogelijk tussen de monsterwaarden in ligt, wat weer leidt tot een onderschatting van de variantie als de factor $n/(n-1)$ er niet zou zijn. Dit kan verduidelijkt worden in het extreme geval dat n slechts gelijk is aan 2 monsterwaarden. In dat geval ligt het geschatte gemiddelde precies tussen de twee monsterwaarden in. Om dan tot een unbiased schatter te komen van de variantie, zo hebben statistici bewezen, wordt bovenstaande formule nog met een factor 2/(2-1) vermenigvuldigd. In het algemene geval geeft dit de formule

$$\sigma^2 = \frac{1}{n-1}\sum_{i=0}^{n-1} x_i^2 - \frac{n}{n-1}m_1^2 \tag{2.33}$$

Voorbeeld 2.9

De tijd nodig voor het aanschakelen van een toon-ontvanger in een PABX wordt uitgebreid gemeten. Er blijkt dat deze tijd één op de drie keer 20 ms bedraagt en twee op de drie keer 15 ms Voor wachttijd berekeningen moet van deze bedieningsduur (aanschakeltijd) de variantie uitgerekend worden.

Gevraagd: Welke methode, a) of b), geeft direct de gewenste variantie?

De aanschakeltijd is op te vatten als een toevals-grootheid waarvan de kansverdeling bekend is. Namelijk met kans 2/3 is de uitkomst 15 en met kans 1/3 is de uitkomst 20. Daarom moet methode 1) toegepast worden. Mijn calculator geeft onder de toets σ_x, na kwadrateren, de uitkomst 5.556. De ingevoerde data-waarden waren $\{15, 15, 20\}$.

Ter verificatie: als bij deze vraag ten onrechte methode 2) wordt toegepast (op mijn calculator met de toets S_x) dan is de uitkomst 8.333.

2.5.6 Coëfficiënt van variatie

Als één na laatste in deze serie over kentallen van een kansverdeling introduceren we nog de definitie van coëfficiënt van variatie. Dit is de standaarddeviatie genormeerd met het gemiddelde en wordt aangegeven met de letter C:

$$C = \frac{\sigma}{m_1} \tag{2.34}$$

Onder andere bij formules voor gemiddelde wachttijd in een wachtrijsysteem is dit een makkelijke karakteristieke grootheid van bijvoorbeeld de bedieningsduurverdeling.

2.5.7 Percentielen

Voor een stochastische variabele X, met verdelingsfunctie $F(\cdot)$, is het percentiel t_p de uitkomst die met $p\%$ kans niet overschreden wordt. Voor p kan elk getal van 0 tot 100 ingevuld worden.

In formulevorm is de definitie van t_p (het p-de percentiel):

$$\Pr\{X \le t_p\} = p\% \tag{2.35}$$

Dit wil bijvoorbeeld zeggen dat X met 99% kans kleiner is dan, of gelijk aan t_{99}. Met 1% kans is X groter dan t_{99}.

Formule (2.35) kan ook geschreven worden als:

$$F(t_p) = p\% \tag{2.36}$$

Let wel, p is gegeven en t_p moet gevonden worden. Het gaat dus om de inverse functie van $F(\cdot)$.

Voorbeeld 2.10

Gegeven de verdelingsfunctie van de tijd tussen twee opeenvolgende aankomsten van voorbeeld 2.2:

$$F(x) = 1 - e^{-x/2} \qquad \text{voor } x > 0 \tag{2.37}$$

Men wil weten:

a) welke tussenaankomsttijd wordt slechts met 1% kans overschreden

b) hoe groot is de kans dat de gemiddelde tussenaankomsttijd (= 2 tijdseenheden) niet overschreden wordt

c) waar ligt het 50-ste percentiel van deze kansverdeling?

Bij a) gaat het om het 99-ste percentiel. Immers t_{99} wordt met 99% kans niet en met 1% kans wel overschreden. Overeenkomstig formule (2.36), bepalen we t_{99} uit:

$$1 - e^{t_{99}/2} = 0.99 \tag{2.38}$$

met een klein beetje algebra volgt hieruit:

$$t_{99} = -2\ln(0.01) = 9.210 \tag{2.39}$$

Bij vraag b) hebben we in eerste instantie alleen het begrip verdelingsfunctie nodig. De gevraagde kans is:

$$F(2) = 1 - e^{-1} = 0.632 \approx 63\% \tag{2.40}$$

Maar vergelijken we dit met formule (2.36) dan zien we dat bij deze kansverdeling het 63-ste percentiel bijna gelijk is aan het gemiddelde.

Bij symmetrische verdeling is juist het 50-ste percentiel (dit is de mediaan van de verdeling) gelijk aan het gemiddelde. Bij deze verdeling is dat niet het geval.

Het antwoord op c) vinden we op dezelfde wijze als bij a), en hier komt uit:

$$t_{50} = -2\ln(0.5) = 1.386 \tag{2.41}$$

Het bepalen van percentielen komt neer op het bepalen van de inverse functie van de verdelingsfunctie. Bij de verdelingsfunctie van een discrete stochastische variabele stuit dit wiskundig gezien op problemen. In dat geval moeten we t_p definiëren als de kleinste waarde τ waarvoor geldt: $\Pr\{X \leq \tau\} \geq p\%$. In de grafiek van zo'n verdelingsfunctie (zie figuur 2.1) komt dat neer op het bepalen van de plaats van de "knobbel" die ter hoogte zit van $p\%$ of, als die "knobbel" er niet is, de eerste die boven $p\%$ is.

Hoofdstuk 3

Speciale verdelingen

Een aantal verdelingen komen “van nature” voor bij de prestatie analyse van telecommunicatiesystemen. De belangrijkste hiervan zullen we in dit hoofdstuk opsommen. We geven van die verdelingen toepassingen en eigenschappen zoals verwachtingswaarde (gemiddelde) en variantie.

3.1 Geometrische verdeling

Beschouw een experiment dat met kans p “succes” oplevert en met kans $1-p$ “faalt”. De toevalsgrootheid X definiëren we als het aantal keren dat we het experiment moeten doen totdat het eerste “succes” optreedt. We schrijven “succes” tussen aanhalingstekens omdat deze benaming, die het meest gebruikelijk is, ook kan slaan op het optreden van een bitfout, waarbij het experiment het ontvangen van een bit uit een bitstroom is. Als de “succes”-kansen van de opeenvolgende experimenten onafhankelijk van elkaar zijn, dan heeft X een geometrische verdeling.

Opdat het eerste “succes” optreedt bij het i-de experiment, moeten we eerst $i-1$ keer “falen” meemaken, de kans hierop is $(1-p)^{i-1}$, en dan één maal “succes”, de kans daarop is p. De i-de kans van de geometrische verdeling wordt nu gegeven door:

$$p_i = (1-p)^{i-1}p \qquad i = 1, 2, \ldots \tag{3.1}$$

De verwachtingswaarde (eerste moment) wordt volgens de definitie van het vorige hoofdstuk gegeven door:

$$m_1 = \sum_{i=1}^{\infty} i(1-p)^{i-1}p \tag{3.2}$$

Om dit te berekenen (en voor later) is de volgende hulpformule en afleiding handig:

$$\sum_{i=0}^{\infty} i(1-p)^{i-1} \quad = \quad -\frac{d}{dp}\sum_{i=1}^{\infty}(1-p)^i \tag{3.3}$$

$$= \frac{d}{dp}\frac{p-1}{p} \tag{3.4}$$

$$= \frac{1}{p^2} \tag{3.5}$$

Dan is voor de geometrische verdeling het eerste moment:

$$m_1 = \frac{1}{p} \tag{3.6}$$

Zo kunnen we, met iets meer moeite, ook de variantie van de geometrische verdeling berekenen. De afleiding laten we hier achterwege; het resultaat is:

$$\sigma^2 = \frac{1-p}{p^2} \tag{3.7}$$

Voorbeeld 3.1

Een mogelijke methode om met een computer een willekeurig aankomstproces (bijvoorbeeld van telefoon oproepen) te simuleren is om de tijdas in kleine stukjes te verdelen en per tijdsleuf (zo'n stukje) te loten of er een aankomst is in die tijdsleuf. Stel dat de tijdas is opgedeeld in tijdsleuven van 200 ms. De grootheid p geeft de kans op een aankomst voor elke tijdsleuf. De loting voor elke tijdsleuf is onafhankelijk van de lotingen in voorgaande tijdsleuven.

Gevraagd:

a) Hoe groot moet p gekozen worden opdat het aankomstproces gemiddeld 1 aankomst per 2 seconden genereert?

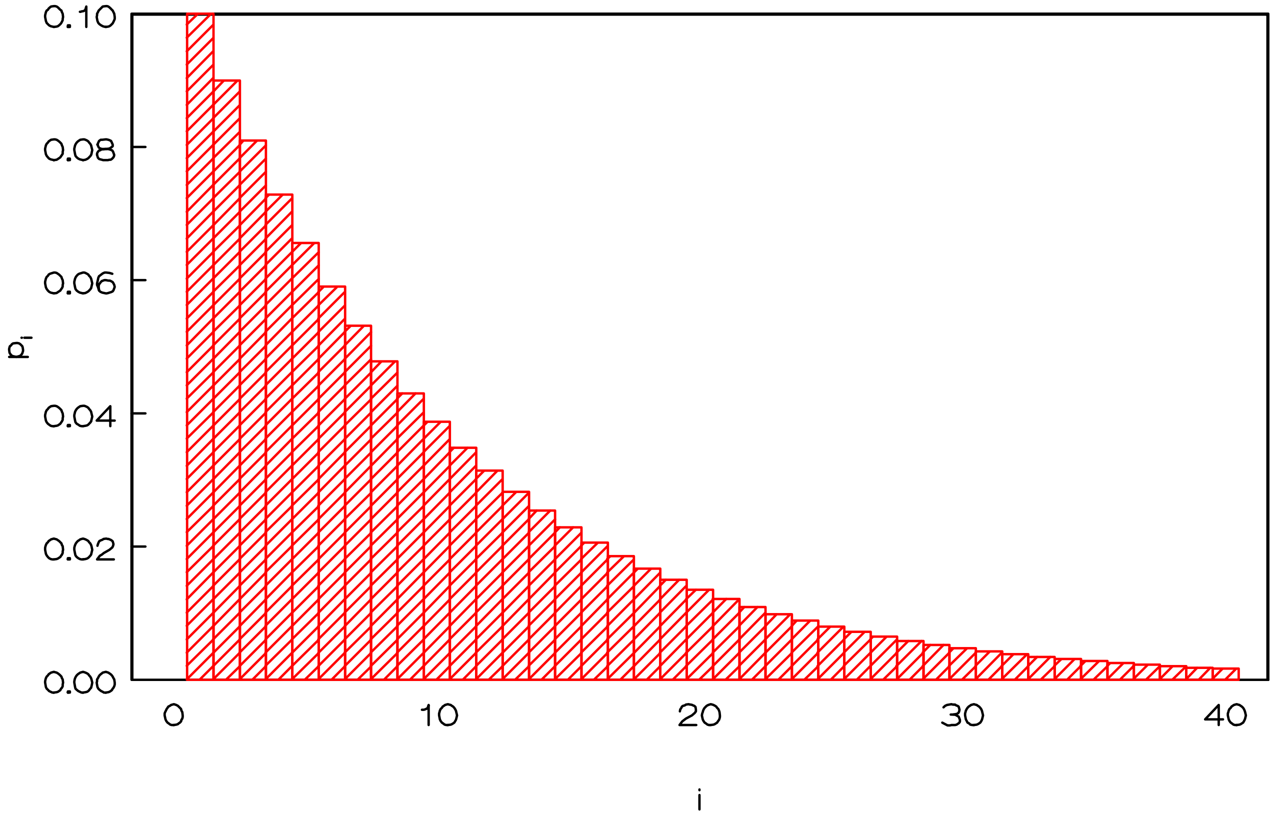
b) Hoe ziet de kansverdeling van het nummer van de tijdsleuf met de eerste aankomst eruit voor de in a) gevonden p?

c) De simulatie wordt gestart op $t = 0$ s. Hoe groot is de kans dat er op $t = 2$ seconde nog geen aankomst heeft plaatsgevonden?

d) Hoe hangt de coëfficiënt van variatie van de tussenaankomsttijd af van de waarde van p?

In een interval van 2 seconden zitten 10 tijdsleuven. Eén aankomst per 2 seconden betekent dus één aankomst per 10 tijdsleuven. Elke loting per tijdsleuf is op te vatten als een experiment met kans p op "succes" (aankomst). Het aantal tijdsleuven dat verloopt tot en met de eerste aankomst is gemiddeld $1/p$. Elke volgende aankomst duurt weer gemiddeld $1/p$ tijdsleuven. Het antwoord op a) is dan dat met $p = 0.1$ er gemiddeld één aankomst per 10 tijdsleuven is.

Ad b): Het nummer van de tijdsleuf met de eerste aankomst geeft aan het aantal keren dat de loting gedaan moet worden totdat er een "succes" is. Deze grootheid is een discrete



Figuur 3.1: Histogram aantal lotingen tot eerste aankomst (Geometrische verdeling).

stochastische variabele met een geometrische verdeling. Het histogram ziet er uit als in figuur 3.1 weergegeven.

Ad c): Op $t = 2$ seconden zijn er 10 tijdsleuven "achter de rug". De kans dat er in die 10 tijdsleuven geen aankomst was is gelijk aan de kans dat de eerste aankomst in tijdsleuf 11 of later valt. Deze kans is gelijk aan de waarschijnlijkheidsmassa in de staafjes rechts van 10:

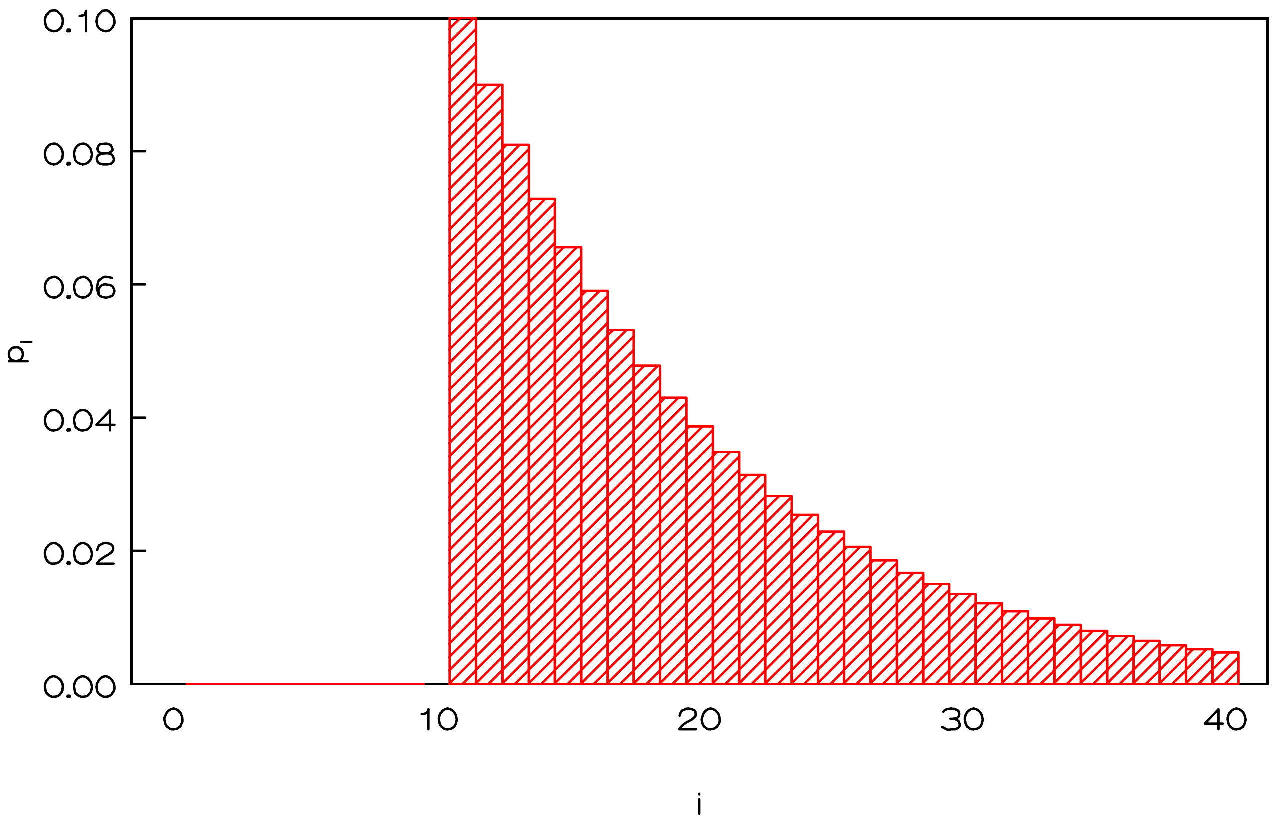
$$\begin{aligned} \Pr\{X > 10\} &= 1 - \Pr\{X \leq 10\} && (3.8)\\ &= 1 - \sum_{i=1}^{10} 0.9^{i-1} 0.1 && (3.9)\\ &= 1 - \frac{1 - 0.9^{10}}{1 - 0.9} 0.1 && (3.10)\\ &= 0.9^{10} && (3.11)\\ &= 0.349 && (3.12) \end{aligned}$$

Bij d) gaat het om de coëfficiënt van variatie van een geometrisch verdeelde stochastische variabele. We moeten hier m_1 uit (3.6) en σ uit (3.7) invullen in de definitie (2.34). Dit geeft:

$$C = \frac{\sqrt{1-p}}{p} \frac{1}{1/p} = \sqrt{1-p} \tag{3.13}$$

Voor $p = 0.1$ is de coëfficiënt van variatie 0.949; hoe kleiner p wordt, des te dichter komt C bij 1. Alleen voor kansen p groter dan 0.1 neemt C duidelijk af; eerst langzaam en dan steeds steiler gaat C naar 0 als p de waarde 1 nadert.

De geometrische verdeling heeft, als enige discrete verdeling, de geheugenloze eigenschap. Dit wil zeggen, dat wanneer er al k experimenten gedaan zijn én er is nog geen "succes" opgetreden, dan is de kansverdeling van het aantal experimenten dat nog gedaan moet worden (na die eerste k) totdat "succes" optreedt weer dezelfde als de oorspronkelijke kansverdeling.



Figuur 3.2: Geheugenloze eigenschap geometrische verdeling. Kansverdeling van eerste aankomst, gegeven geen aankomst in eerste tien sleuven.

Iets wiskundiger geformuleerd: de kansverdeling van $X - k$ gegeven $X > k$, is gelijk aan de kansverdeling van X. Dit kan als volgt worden ingezien. De geometrische kansverdeling wordt gekenmerkt door de constante verhouding (ratio $= 1-p$) tussen opeenvolgende kansen in het histogram. Wanneer nu gegeven is: $X > k$, dan betekent dit dat de eerste k kansen in het histogram 0 geworden zijn. De resterende kansen worden genormeerd (om de som van de kansen weer gelijk aan 1 te maken), hetgeen de ratio van die kansen onaangetast laat. De conditionele kansverdeling (dit betekent de kansverdeling gegeven $X > k$) is daardoor een verschoven versie van de oorspronkelijke kansverdeling.

3.2 Binomiale verdeling

Wanneer we n onafhankelijke experimenten doen en elk experiment heeft kans p op "succes" dan heeft de stochastische variabele X, die het totaal aantal "successen" voorstelt, een binomiale verdeling met parameters n en p.

De kansen van de binomiale verdeling worden gegeven door:

$$p_i = \binom{n}{i}(1-p)^{n-i}p^i \qquad 0 \leq i \leq n \tag{3.14}$$

De factor p_i staat voor de i "successen" die moeten optreden, de factor $(1-p)^{n-i}$ voor de voor de $n-i$ keer "falen" en de binomiaal coëfficiënt "n over i" staat voor het aantal mogelijkheden om i objecten uit een verzameling van n te kiezen (zonder onderscheid te maken in de volgorde).

De binomiaal coëfficiënt wordt als volgt berekend:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{1} \cdot \frac{n-1}{2} \cdot \cdots \cdot \frac{n-k+1}{k} \tag{3.15}$$

Voor de binomiale verdeling worden het gemiddelde en de variantie gegeven door:

$$m_1 = np \tag{3.16}$$

$$\sigma^2 = np(1-p) \tag{3.17}$$

Voorbeeld 3.2

We gaan verder met het model van voorbeeld 3.1, waarbij in elke tijdsleuf van 200 ms één aankomst kon plaatsvinden met kans 0.1.

Gevraagd: De kans dat, in een interval van 20 seconden, 5 of minder aankomsten plaatsvinden.

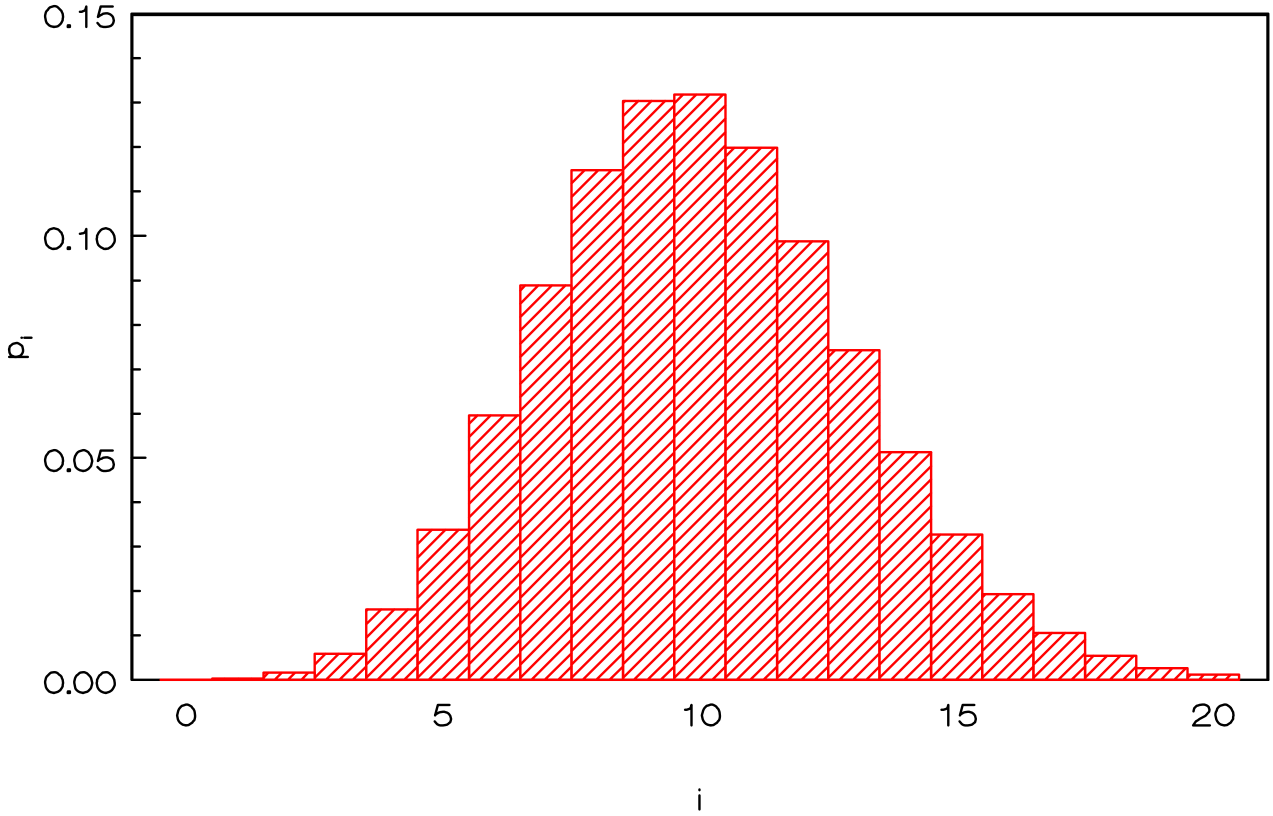
Het aantal aankomsten in 20 seconden (=100 tijdsleuven) komt overeen met het aantal "successen" in 100 experimenten. Deze toevalsgrootheid is binomiaal verdeeld met parameters $n = 100$ en $p = 0.1$. De verwachtingswaarde van het aantal aankomsten is 10. De kans op 5 of minder aankomsten is:

$$= p_0 + p_1 + p_2 + p_3 + p_4 + p_5 \tag{3.18}$$

$$= \binom{100}{0}0.9^{100} + \binom{100}{1}0.9^{99}0.1 + \cdots + \binom{100}{5}0.9^{95}0.000001 \tag{3.19}$$

$$= 0.000027 + 0.000295 + 0.001623 + 0.005892 + 0.015875 + 0.033866 \tag{3.20}$$

$$= 0.058 \tag{3.21}$$



Figuur 3.3: Histogram aantal successen in 100 experimenten met $p = 0.1$ (binomiale verdeling).

3.3 Poisson-verdeling

De Poisson-verdeling wordt vaak verward met Poisson-proces. Het Poisson-proces is een stochastisch proces; hierop komen we in de laatste paragraaf nog terug. De Poisson-verdeling is een kansverdeling van een discrete stochastische variabele X. De twee begrippen hebben wel veel met elkaar te maken: als een aankomstproces een Poisson-proces is, dan is de kansverdeling van het aantal aankomsten in een gegeven tijdsinterval een Poisson- verdeling.

Een typisch voorbeeld van zo'n aankomstproces zijn de oproepen die binnenkomen bij een telefooncentrale. Als dat aankomstproces een Poisson-proces is met gemiddeld λ aankomsten per seconde en we definiëren X als het aantal aankomsten in een interval van t seconden, dan heeft X een Poisson-verdeling met parameter $\alpha = \lambda t$.

De Poisson-verdeling met parameter α wordt gegeven door:

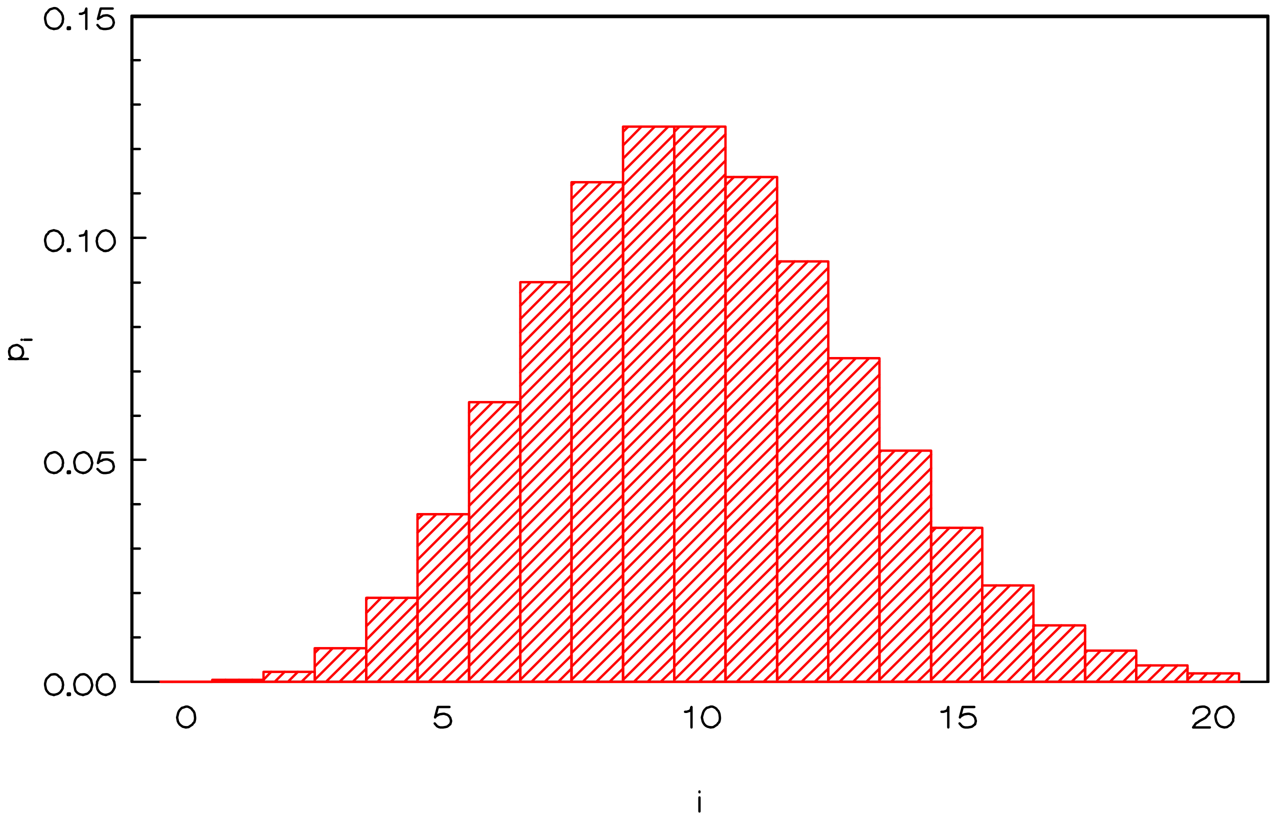
$$p_i = e^{-\alpha}\frac{\alpha^i}{i!} \tag{3.22}$$

Voor de Poisson-verdeling worden het gemiddelde en de variantie gegeven door:

$$m_1 = \alpha \tag{3.23}$$

$$\sigma^2 = \alpha \tag{3.24}$$

Figuur 3.4 toont een voorbeeld van een Poissonverdeling.



Figuur 3.4: Histogram Poisson-verdeling met $\alpha = 10$.

Voorbeeld 3.3

Modelleer het aankomstproces van oproepen bij een telefooncentrale als een Poisson-proces. Om een vergelijking te maken met het simulatie model dat in de twee voorgaande voorbeelden besproken werd, stellen we dezelfde vraag.

Gevraagd: Hoe groot is de kans op 5 of minder aankomsten in 20 seconden, bij gemiddeld 1 aankomst per 2 seconden?"

Het aantal aankomsten in een gegeven tijdsinterval, is een Poisson- verdeelde discrete stochastische variabele. De parameter van de Poisson-verdeling (het gemiddelde) is in dit voorbeeld $\alpha = 10$. De kans op 5 of minder aankomsten is:

$$= p_0 + p_1 + p_2 + p_3 + p_4 + p_5 \tag{3.25}$$

$$= e^{-10} + e^{-10}10 + e^{-10}\frac{10^2}{2!} + e^{-10}\frac{10^3}{3!} + e^{-10}\frac{10^4}{4!} + e^{-10}\frac{10^5}{5!} \tag{3.26}$$

$$= 0.000045 + 0.000454 + 0.002270 + 0.007567 + 0.018917 + 0.037833 \tag{3.27}$$

$$= 0.067 \tag{3.28}$$

De Poisson-verdeling is meer willekeurig dan de binomiale verdeling (grotere coëfficiënt van variatie) en heeft daardoor meer kansmassa in de extremen. Echter voor kansen tot op ongeveer 1% nauwkeurig is, bij deze parameters, de Poisson-verdeling een goede benadering voor de binomiale verdeling. Hierop komen we in de volgende paragraaf terug.

3.4 Normale Verdeling (Gaussische Verdeling)

Wanneer we een aantal onafhankelijke stochastische variabelen sommeren, dan gaat de kansverdeling van de som steeds meer lijken op een normale verdeling. Dit is de basis van het belang van de normale verdeling; immers vaak zal een waarneming die we doen bestaan uit de som van een groot aantal toevalsgrootheden. Voordat we de formule geven van de kansdichtheid van een normaal verdeelde continue stochastische variabele, nog eerst iets over het sommeren van stochastische variabelen en over de centrale limiet stelling.

Stel we hebben twee stochastische variabelen X_1 en X_2. Dan is $X_1 + X_2$ uiteraard ook een stochastische variabele waarvan het gemiddelde gelijk is aan de som van de gemiddelden (van X_1 respectievelijk X_2) en de variantie gelijk aan de som van de varianties. Dus:

stochast	gemiddelde	variantie
X_1	$m_{1,1}$	σ_1^2
X_2	$m_{1,2}$	σ_2^2
$X_1 + X_2$	$m_{1,1} + m_{1,2}$	$\sigma_1^2 + \sigma_2^2$

De centrale limietstelling luidt als volgt: Stel $X_1, X_2, \ldots$, vormen een reeks van onafhankelijke, gelijk-verdeelde stochastische variabelen. Definieer $S_n = X_1 + X_2 + \cdots + X_n$. Dan nadert de verdeling van S_n naar een normale verdeling als $n \to \infty$.

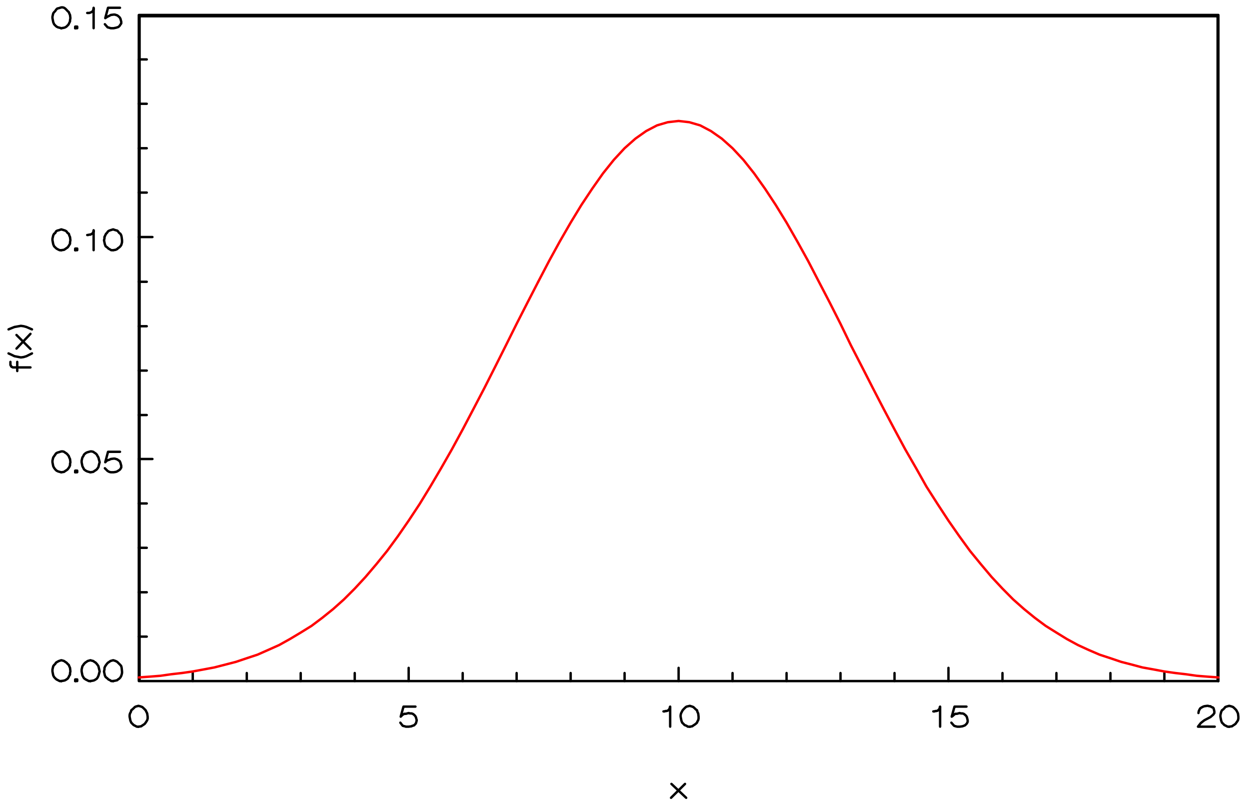
Direct gekoppeld met de centrale limietstelling is het feit dat de Poisson-verdeling steeds meer op een normale verdeling gaat lijken als de parameter van de Poisson-verdeling groter wordt. Het vergelijken van de figuren 3.4 en 3.5 suggereert dit al. Merk wel op dat de Poisson-verdeling een discrete verdeling is en de normale verdeling een continue verdeling. Dit "steeds meer lijken op" moet gepreciseerd worden in termen van de verdelingsfuncties. De Poisson-verdeling is op zijn beurt weer een limietgeval van de binomiale verdeling. Samenvattend:

$$\text{binomiaal} \to \text{Poisson} \to \text{normaal}$$

De eerste benadering kan men redelijkerwijs gebruiken (absolute fout in de orde van één procent of minder) als voor de binomiale verdeling $n > 10$ en $p < 0.1$ is. Voor de tweede benadering moet gelden $\alpha > 10$.

De kansdichtheid van een normaal verdeelde stochastische variabele met gemiddelde μ en variantie σ^2 is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3.29}$$



Figuur 3.5: Kansdichtheid normale verdeling met $\mu = 10$ en $\sigma^2 = 10$.

De verdelingsfunctie van een normaal verdeelde stochastische variabele is niet in elementaire functies uit te drukken. Vandaar dat men in dat geval met tabellen of benaderende algorithmes moet werken. We kunnen toe met één tabel, omdat een willekeurig normaal verdeelde stochastische variabele te herleiden is tot een standaard normaal verdeelde stochastische variabele. Bij de standaard normale verdeling is het gemiddelde 0 en de variantie 1. Voor de standaard normale verdeling is de kansdichtheid:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{3.30}$$

De bijbehorende verdelingsfuctie wordt de error function, Erf$(\cdot)$, genoemd:

$$\text{Erf}(x) = \int_{-\infty}^{x} f(t)dt \tag{3.31}$$

Stel, we hebben een normaal verdeelde stochastische variabele X, met gemiddelde μ en variantie σ^2 (dit geven we aan met $X \sim N(\mu, \sigma^2)$), en we willen weten hoe groot $\Pr\{X \leq x\}$ is. Om deze vraag te herleiden naar de standaard normale verdeling definiëren we de stochastische variabele Y, die direct gekoppeld is aan X, volgens

$$Y = \frac{X - \mu}{\sigma} \tag{3.32}$$

De translatie over μ en het normeren met σ heeft tot gevolg dat Y standaard normaal verdeeld is ($Y \sim N(0,1)$). De verdelingsfunctie van X kan nu uitgedrukt worden in $\mathrm{Erf}(\cdot)$, n.l.:

$$\Pr\{X \leq x\} = \Pr\{Y \leq \tfrac{x-\mu}{\sigma}\} = \mathrm{Erf}\left(\frac{x-\mu}{\sigma}\right) \tag{3.33}$$

In de volgende tabel (die in twee stukken gedeeld is) worden waarden van de error function function gegeven. Door de symmetrie van de kansdichtheid van de standaard normale verdeling ten opzichte van $x = 0$, kan makkelijk de ene tabel uit de andere afgeleid worden. Er geldt namelijk

$$\mathrm{Erf}(-x) = 1 - \mathrm{Erf}(x) \tag{3.34}$$

Voorbeeld 3.4

Bij een nieuwe PABX worden de ISDN toestellen en terminals gevoed vanuit de peripheral circuit boards. De voeding levert alleen vermogen aan een toestel als dat actief is. Er is opgegeven dat in het drukke uur, een toestel met kans 0.1 actief is. Het aantal toestellen aan de PABX is 200. Om te bepalen of het vermogen van de voeding voldoende is, wil men weten hoe groot de kans is dat, in het drukke uur, meer dan 30 toestellen tegelijkertijd actief zijn.

Ter vereenvoudiging van het probleem, veronderstellen we dat de activiteitsperioden van de verschillende toestellen onafhankelijk van elkaar zijn. Het aantal actieve toestellen op een willekeurig moment van het drukke uur, X, is dan een binomiaal verdeelde stochastische variabele (200 "experimenten"; "succes" slaat op het actief zijn van een toestel). Voor het gemiddelde en de variantie van X geldt:

$$m_1 = 20 \tag{3.35}$$

$$\sigma^2 = 200(1-0.1)0.1 = 18 \tag{3.36}$$

Een rechtstreekse sommatie van de binomiale kansen p_{30} tot en met p_{200} stuit op moeilijkheden zoals het berekenen van de binomiaal coëfficiënt:

$$\binom{200}{30} = 409681705022127773530866523638950880 \tag{3.37}$$

Ook met een benadering van de binomiale verdeling met een Poisson-verdeling krijgen we numerieke problemen.

Om de gevraagde kans te berekenen is hier een benadering met de normale verdeling het meest praktische. Stel W is normaal verdeeld met gemiddelde 20 en variantie 18. Definieer $Y = (W-20)/\sqrt{18}$:

$$\begin{aligned}
\Pr\{W > 30\} &= 1 - \Pr\{W \leq 30\} && (3.38)\\
&= 1 - \Pr\{Y \leq 2.36\} && (3.39)\\
&= \mathrm{Erf}(-2.36) && (3.40)\\
&= 0.00914 \qquad \text{(volgens tabel 3.2)} && (3.41)
\end{aligned}$$

3.5 Negatief exponentiële verdeling

Het limiet geval van de (discrete) geometrische verdeling is de (continue) negatief exponentiële verdeling. Ook hier kan de verdeling geïnterpreteerd worden als: hoe lang duurt het totdat een "succes" optreedt. Een typisch voorbeeld is de tijd die zal verstrijken tot de eerstvolgende telefoon oproep, wanneer er gemiddeld λ oproepen per tijdseenheid zijn.

Laat T een continue stochastische variabele zijn met de volgende kansdichtheid

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{voor } 0 \leq t \\ 0 & \text{voor } t < 0 \end{cases} \tag{3.42}$$

en dus de volgende verdelingsfunctie

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & \text{voor } 0 \leq t \\ 0 & \text{voor } t < 0 \end{cases} \tag{3.43}$$

dan is T negatief exponentieel verdeeld met gemiddelde en variantie:

$$m_1 = \frac{1}{\lambda} \tag{3.44}$$

$$\sigma^2 = \frac{1}{\lambda^2} \tag{3.45}$$

Voorbeeld 3.5

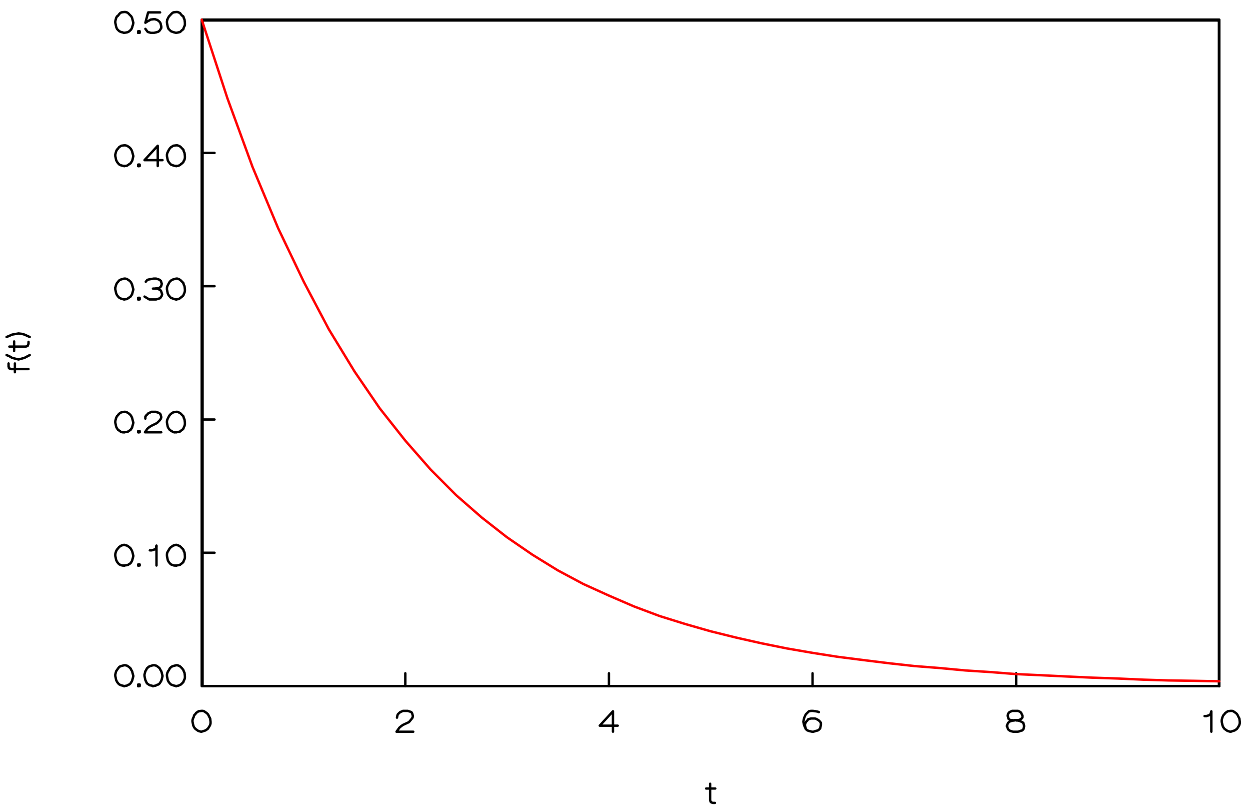
In dit voorbeeld vergelijken we de simulatie methode van voorbeeld 3.1 met een aankomstproces waarbij de tijd tot de eerste aankomst (na $t = 0$) en de tijden tussen twee opeenvolgende aankomsten negatief exponentieel verdeeld zijn met gemiddelde 2 (dus $\lambda = 0.5$).

Als tegenhangers van vragen c) en d) uit voorbeel 3.1 kunnen we ons afvragen hoe groot de kans is dat er op $t = 2$ nog geen aankomst heeft plaatsgevonden en hoe groot de coëfficiënt van variatie is van een negatief exponentieel verdeelde stochastische variabele.

Het antwoord op de laatste vraag is direct af te lezen uit (3.44) en (3.45). Immers $C = m_1/\sigma = 1$. Dit is in overeenstemming met de bovenstaande stelling dat de negatief exponentiële verdeling een limietgeval is van de geometrische verdeling. Wordt het interval $(0, 2)$ opgedeeld in steeds meer tijdsleuven en vermindert de kans op een aankomst in een tijdsleuf verhoudingsgewijs dan gaat ook voor de geometrische verdeling de coëfficiënt van variatie naar 1.

De kans op geen aankomst gedurende de eerste twee seconden is het complement van de kans dat de eerste aankomst plaatsvindt op een tijdstip $T \leq 2$:

$$\begin{aligned} \Pr\{T > 2\} &= 1 - \Pr\{T \leq 2\} && (3.46) \\ &= 1 - F(2) && (3.47) \\ &= e^{-1} && (3.48) \\ &= 0.368 && (3.49) \end{aligned}$$



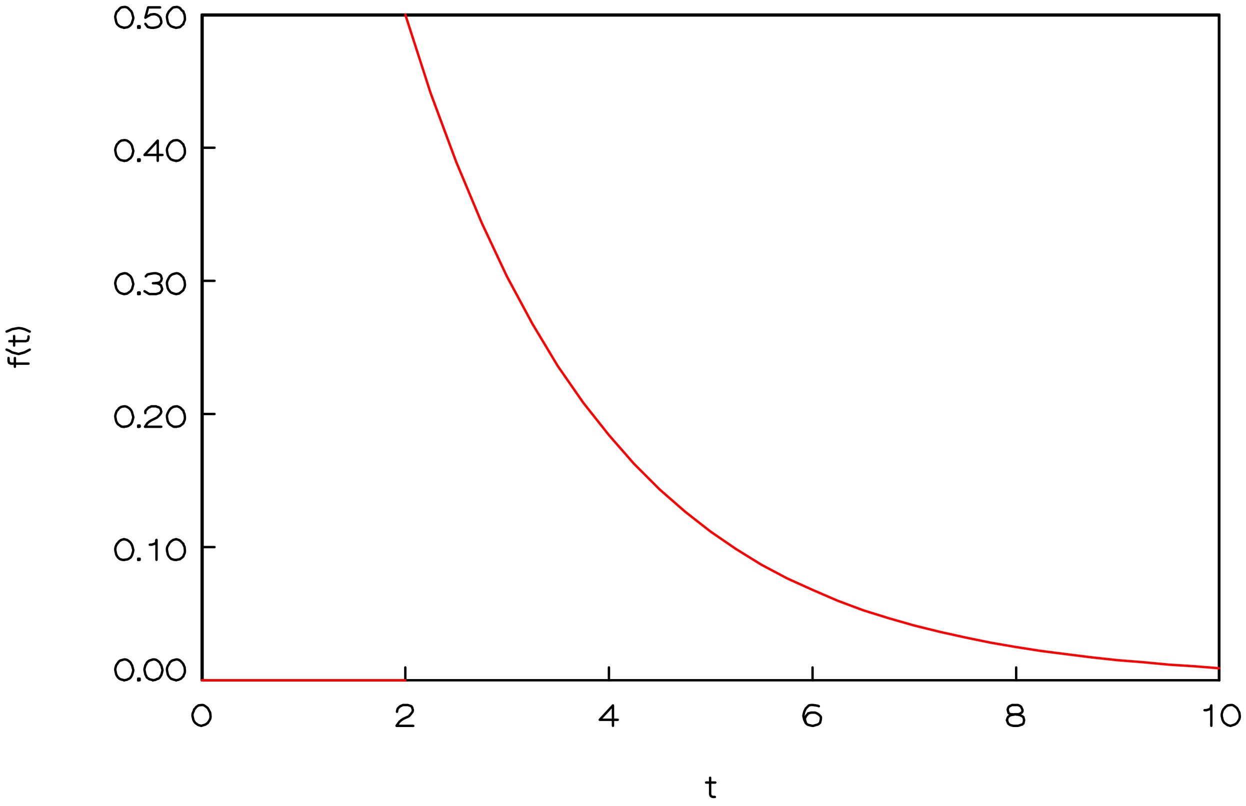
Figuur 3.6: Kansdichtheid negatief exponentiële verdeling met $m_1 = 2$ (dus $\lambda = 0.5$).

Net als de geometrische verdeling, heeft ook de negatief exponentiële verdeling de geheugenloze eigenschap. Dit is ook de reden dat bij prestatie-analyse en wachttijdtheorie meestal verondersteld wordt dat de tijd tussen aankomsten of de bedieningsduur negatief exponentieel verdeeld is. Namelijk in dat geval hoeft in de analyse de tijd die verstreken is sinds de laatste aankomst (respectievelijk de bedieningsduur die reeds verstreken is) niet meegenomen te worden in de analyse, omdat de kansverdeling van de tijd die nog te gaan is, onafhankelijk is van de reeds verstreken tijd.

3.6 Poisson-proces

Het Poisson-proces is niet een kansverdeling maar een toevalsproces waarbij nog als extra dimensie een tijdparameter hoort. Het kan gezien worden als een proces waarbij met stochastische tussenpozen een gebeurtenis (bijvoorbeeld een aankomst) plaatsvindt. Een reden om het Poisson-proces toch hier te bespreken is dat bij dit proces die stochastische tussenpozen negatief exponentieel verdeeld zijn.

Figuur 3.8 laat, als functie van de tijdparameter, het geaccumuleerde aantal gebeurtenissen



Figuur 3.7: Geheugenloze eigenschap negatief exponentiële verdeling. Kansdichtheid eerste aankomst gegeven geen aankomsten voor $t = 2$.

(aankomsten) zien voor een willekeurige realisatie van een Poisson-proces. De driehoekjes onderaan geven de tijdstippen aan waarop de aankomsten plaatsvonden.

Er zijn drie gelijkwaardige karakteriseringen te geven voor een Poisson-proces met intensiteit λ:

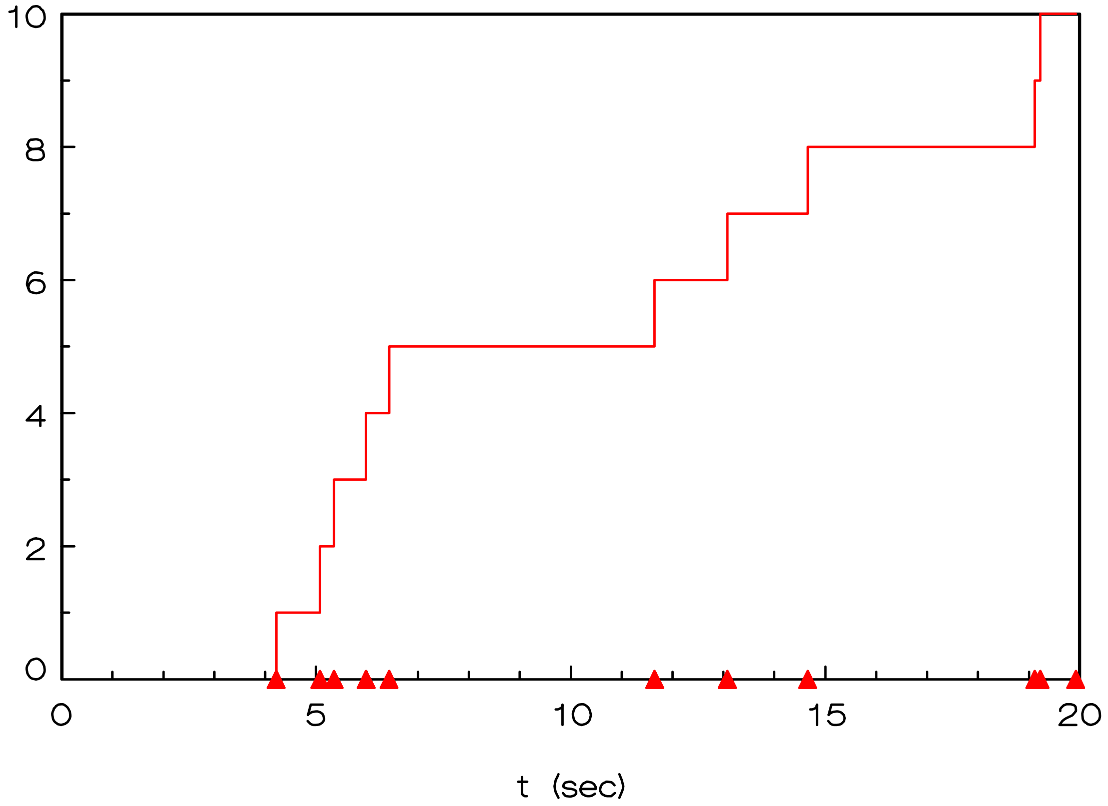
a) De tijdsintervallen tussen de aankomsten zijn onafhankelijke stochastische variabelen die negatief exponentieel verdeeld zijn met gemiddelde $1/\lambda$.

b) Voor een willekeurig tijdsinterval van lengte Δt geldt voor de limiet $\Delta t \to 0$ dat:

$$\Pr\{2 \text{ aankomsten in } \Delta t\} \to 0 \tag{3.50}$$
$$\Pr\{1 \text{ aankomst in } \Delta t\} \to \lambda \Delta t \tag{3.51}$$
$$\Pr\{0 \text{ aankomsten in } \Delta t\} \to 1 - \lambda \Delta t \tag{3.52}$$

en deze gebeurtenissen zijn onafhankelijk van aankomsten in een ander interval.

c) Voor elke positieve t, heeft het aantal aankomsten in het tijdsinterval $(0, t)$ een Poisson-verdeling met gemiddelde λt.



Figuur 3.8: Het geaccumuleerde aantal aankomsten van een Poisson-proces met $\alpha = 0.5$, als functie van de tijd. Verkregen uit een simulatie met negatief exponentieel verdeelde tussenaankomsttijden.

De gelijkwaardigheid van de karakteriseringen wil zeggen dat elke van de drie beweringen de andere impliceert:

$$\text{a)} \Longleftrightarrow \text{b)} \Longleftrightarrow \text{c)} \tag{3.53}$$

We zouden bewering a) kunnen kiezen als definitie van het Poisson-proces. Bij de voorbeelden van dit hoofdstuk zagen we al dat een aankomst proces gesimuleerd kan worden door bijvoorbeeld elke 200 ms te loten of er een aankomst is. De bewering bij b) is dat hoe kleiner de lengte van zo'n tijdsleuf is, des te meer gaat die simulatie op een Poisson-proces lijken. Dan moet wel de kans op een aankomst in zo'n tijdsleuf proportioneel gehouden worden met de lengte van de tijdsleuf: $\lambda\Delta t$. Bij een echt Poisson-proces is er in elk interval Δt een zekere kans op twee of meer aankomsten, echter die kans is evenredig met $(\Delta t)^2$. Voor Δt voldoende klein is dan de gebeurtenis van twee of meer aankomsten te verwaarlozen.

In voorbeeld 3.1 presenteerden we een methode om het Poisson-proces bij benadering te simuleren door elke tijdsleuf te loten of er een aankomst plaatsvindt (er zijn efficiëntere methoden om dit te simuleren). In voorbeeld 3.2 zagen we dat het aantal aankomsten in een interval binomiaal verdeeld is. Als nu de lengte van elke tijdsleuf korter gemaakt wordt, het aantal tijdsleuven in het interval verhoudingsgewijs vergroot wordt en de kans op een aankomst evenredig kleiner wordt, dan gaat die binomiale verdeling steeds meer op een Poisson-verdeling lijken. Met deze constructie komen we bij bewering c) uit.

x	Erf(x)	x	Erf(x)	x	Erf(x)	x	Erf(x)
0.00	0.50000	1.00	0.84134	2.00	0.97725	3.00	0.99865
0.02	0.50798	1.02	0.84614	2.02	0.97831	3.02	0.99874
0.04	0.51595	1.04	0.85083	2.04	0.97932	3.04	0.99882
0.06	0.52392	1.06	0.85543	2.06	0.98030	3.06	0.99889
0.08	0.53188	1.08	0.85993	2.08	0.98124	3.08	0.99896
0.10	0.53983	1.10	0.86433	2.10	0.98214	3.10	0.99903
0.12	0.54776	1.12	0.86864	2.12	0.98300	3.12	0.99910
0.14	0.55567	1.14	0.87286	2.14	0.98382	3.14	0.99916
0.16	0.56356	1.16	0.87698	2.16	0.98461	3.16	0.99921
0.18	0.57142	1.18	0.88100	2.18	0.98537	3.18	0.99926
0.20	0.57926	1.20	0.88493	2.20	0.98610	3.20	0.99931
0.22	0.58706	1.22	0.88877	2.22	0.98679	3.22	0.99936
0.24	0.59483	1.24	0.89251	2.24	0.98745	3.24	0.99940
0.26	0.60257	1.26	0.89617	2.26	0.98809	3.26	0.99944
0.28	0.61026	1.28	0.89973	2.28	0.98870	3.28	0.99948
0.30	0.61791	1.30	0.90320	2.30	0.98928	3.30	0.99952
0.32	0.62552	1.32	0.90658	2.32	0.98983	3.32	0.99955
0.34	0.63307	1.34	0.90988	2.34	0.99036	3.34	0.99958
0.36	0.64058	1.36	0.91309	2.36	0.99086	3.36	0.99961
0.38	0.64803	1.38	0.91621	2.38	0.99134	3.38	0.99964
0.40	0.65542	1.40	0.91924	2.40	0.99180	3.40	0.99966
0.42	0.66276	1.42	0.92220	2.42	0.99224	3.42	0.99969
0.44	0.67003	1.44	0.92507	2.44	0.99266	3.44	0.99971
0.46	0.67724	1.46	0.92785	2.46	0.99305	3.46	0.99973
0.48	0.68439	1.48	0.93056	2.48	0.99343	3.48	0.99975
0.50	0.69146	1.50	0.93319	2.50	0.99379	3.50	0.99977
0.52	0.69847	1.52	0.93574	2.52	0.99413	3.52	0.99978
0.54	0.70540	1.54	0.93822	2.54	0.99446	3.54	0.99980
0.56	0.71226	1.56	0.94062	2.56	0.99477	3.56	0.99981
0.58	0.71904	1.58	0.94295	2.58	0.99506	3.58	0.99983
0.60	0.72575	1.60	0.94520	2.60	0.99534	3.60	0.99984
0.62	0.73237	1.62	0.94738	2.62	0.99560	3.62	0.99985
0.64	0.73891	1.64	0.94950	2.64	0.99585	3.64	0.99986
0.66	0.74537	1.66	0.95154	2.66	0.99609	3.66	0.99987
0.68	0.75175	1.68	0.95352	2.68	0.99632	3.68	0.99988
0.70	0.75804	1.70	0.95543	2.70	0.99653	3.70	0.99989
0.72	0.76424	1.72	0.95728	2.72	0.99674	3.72	0.99990
0.74	0.77035	1.74	0.95907	2.74	0.99693	3.74	0.99991
0.76	0.77637	1.76	0.96080	2.76	0.99711	3.76	0.99992
0.78	0.78230	1.78	0.96246	2.78	0.99728	3.78	0.99992
0.80	0.78814	1.80	0.96407	2.80	0.99744	3.80	0.99993
0.82	0.79389	1.82	0.96562	2.82	0.99760	3.82	0.99993
0.84	0.79955	1.84	0.96712	2.84	0.99774	3.84	0.99994
0.86	0.80511	1.86	0.96856	2.86	0.99788	3.86	0.99994
0.88	0.81057	1.88	0.96995	2.88	0.99801	3.88	0.99995
0.90	0.81594	1.90	0.97128	2.90	0.99813	3.90	0.99995
0.92	0.82121	1.92	0.97257	2.92	0.99825	3.92	0.99996
0.94	0.82639	1.94	0.97381	2.94	0.99836	3.94	0.99996
0.96	0.83147	1.96	0.97500	2.96	0.99846	3.96	0.99996
0.98	0.83646	1.98	0.97615	2.98	0.99856	3.98	0.99997

Tabel 3.1: Waarden van $\mathrm{Erf}(x) = \Pr\{Y \leq x\}$, met $Y \sim N(0,1)$, voor $x \geq 0$.

x	Erf(x)	x	Erf(x)	x	Erf(x)	x	Erf(x)
0.00	0.50000	-1.00	0.15866	-2.00	0.02275	-3.00	0.00135
-0.02	0.49202	-1.02	0.15386	-2.02	0.02169	-3.02	0.00126
-0.04	0.48405	-1.04	0.14917	-2.04	0.02068	-3.04	0.00118
-0.06	0.47608	-1.06	0.14457	-2.06	0.01970	-3.06	0.00111
-0.08	0.46812	-1.08	0.14007	-2.08	0.01876	-3.08	0.00104
-0.10	0.46017	-1.10	0.13567	-2.10	0.01786	-3.10	0.00097
-0.12	0.45224	-1.12	0.13136	-2.12	0.01700	-3.12	0.00090
-0.14	0.44433	-1.14	0.12714	-2.14	0.01618	-3.14	0.00084
-0.16	0.43644	-1.16	0.12302	-2.16	0.01539	-3.16	0.00079
-0.18	0.42858	-1.18	0.11900	-2.18	0.01463	-3.18	0.00074
-0.20	0.42074	-1.20	0.11507	-2.20	0.01390	-3.20	0.00069
-0.22	0.41294	-1.22	0.11123	-2.22	0.01321	-3.22	0.00064
-0.24	0.40517	-1.24	0.10749	-2.24	0.01255	-3.24	0.00060
-0.26	0.39743	-1.26	0.10383	-2.26	0.01191	-3.26	0.00056
-0.28	0.38974	-1.28	0.10027	-2.28	0.01130	-3.28	0.00052
-0.30	0.38209	-1.30	0.09680	-2.30	0.01072	-3.30	0.00048
-0.32	0.37448	-1.32	0.09342	-2.32	0.01017	-3.32	0.00045
-0.34	0.36693	-1.34	0.09012	-2.34	0.00964	-3.34	0.00042
-0.36	0.35942	-1.36	0.08691	-2.36	0.00914	-3.36	0.00039
-0.38	0.35197	-1.38	0.08379	-2.38	0.00866	-3.38	0.00036
-0.40	0.34458	-1.40	0.08076	-2.40	0.00820	-3.40	0.00034
-0.42	0.33724	-1.42	0.07780	-2.42	0.00776	-3.42	0.00031
-0.44	0.32997	-1.44	0.07493	-2.44	0.00734	-3.44	0.00029
-0.46	0.32276	-1.46	0.07215	-2.46	0.00695	-3.46	0.00027
-0.48	0.31561	-1.48	0.06944	-2.48	0.00657	-3.48	0.00025
-0.50	0.30854	-1.50	0.06681	-2.50	0.00621	-3.50	0.00023
-0.52	0.30153	-1.52	0.06426	-2.52	0.00587	-3.52	0.00022
-0.54	0.29460	-1.54	0.06178	-2.54	0.00554	-3.54	0.00020
-0.56	0.28774	-1.56	0.05938	-2.56	0.00523	-3.56	0.00019
-0.58	0.28096	-1.58	0.05705	-2.58	0.00494	-3.58	0.00017
-0.60	0.27425	-1.60	0.05480	-2.60	0.00466	-3.60	0.00016
-0.62	0.26763	-1.62	0.05262	-2.62	0.00440	-3.62	0.00015
-0.64	0.26109	-1.64	0.05050	-2.64	0.00415	-3.64	0.00014
-0.66	0.25463	-1.66	0.04846	-2.66	0.00391	-3.66	0.00013
-0.68	0.24825	-1.68	0.04648	-2.68	0.00368	-3.68	0.00012
-0.70	0.24196	-1.70	0.04457	-2.70	0.00347	-3.70	0.00011
-0.72	0.23576	-1.72	0.04272	-2.72	0.00326	-3.72	0.00010
-0.74	0.22965	-1.74	0.04093	-2.74	0.00307	-3.74	0.00009
-0.76	0.22363	-1.76	0.03920	-2.76	0.00289	-3.76	0.00008
-0.78	0.21770	-1.78	0.03754	-2.78	0.00272	-3.78	0.00008
-0.80	0.21186	-1.80	0.03593	-2.80	0.00256	-3.80	0.00007
-0.82	0.20611	-1.82	0.03438	-2.82	0.00240	-3.82	0.00007
-0.84	0.20045	-1.84	0.03288	-2.84	0.00226	-3.84	0.00006
-0.86	0.19489	-1.86	0.03144	-2.86	0.00212	-3.86	0.00006
-0.88	0.18943	-1.88	0.03005	-2.88	0.00199	-3.88	0.00005
-0.90	0.18406	-1.90	0.02872	-2.90	0.00187	-3.90	0.00005
-0.92	0.17879	-1.92	0.02743	-2.92	0.00175	-3.92	0.00004
-0.94	0.17361	-1.94	0.02619	-2.94	0.00164	-3.94	0.00004
-0.96	0.16853	-1.96	0.02500	-2.96	0.00154	-3.96	0.00004
-0.98	0.16354	-1.98	0.02385	-2.98	0.00144	-3.98	0.00003

Tabel 3.2: Waarden van $\mathrm{Erf}(x) = \Pr\{Y \leq x\}$, met $Y \sim N(0,1)$, voor $x \leq 0$.

Hoofdstuk 4

Markov-Processen

Markov-processen vormen een speciaal soort van stochastische processen en zijn voor ons van belang omdat een groot aantal basis modellen van wachtrij- en verliessystemen, alsmede eenvoudige betrouwbaarheidsmodellen, als een Markov-proces beschreven kunnen worden.

4.1 Stochastisch proces

Een systeem waarvan de toestand zich ontwikkelt in de tijd, mede bepaald door toevallige gebeurtenissen, is een voorbeeld van een stochastisch proces.

Een stochastisch proces kan gedefinieerd worden als een familie van stochastische variabelen $X(t)$, waarbij t de tijdparameter is. Voor elk paar t_1, t_2 zijn $X(t_1)$ en $X(t_2)$ stochastische variabelen die onderling statistisch afhankelijk zijn.

4.2 Markov-eigenschap

Een stochastisch proces $X(t)$ heeft de Markov-eigenschap indien het verloop van het proces voor $\tau > t$ volledig bepaald wordt door $X(t)$. Dat wil zeggen dat geen extra informatie over $X(\tau)$ verkregen wordt door naast $X(t)$ ook nog $X(s)$ te kennen wanneer $s < t$.

De waarde van $X(t)$ is de toestand van het Markov-proces op tijdstip t. Als we t identificeren met "heden" dan wil de Markov-eigenschap zeggen dat alle informatie van het systeem over het gedrag in de "toekomst" gegeven wordt door de huidige toestand en dat kennis van het proces uit het "verleden" hier niets aan toevoegt.

De Markov-eigenschap impliceert dat de verblijftijd in een gegeven toestand een "geheugenloze" verdeling moet hebben.

Een Markov-proces is nu, per definitie, een stochastisch proces dat aan de Markov-eigenschap voldoet. Een Markov-proces wordt ook wel Markov-keten genoemd indien de toestandsruimte discreet is, d.w.z. $X(t)$ kan alleen waarden aannemen in een aftelbare verzameling. In de gevallen die wij bestuderen is de toestandsruimte steeds de discrete verzameling $\{0, 1, 2, \ldots\}$ of een eindige deelverzameling daarvan.

Afhankelijk van de waarden die de tijdparameter kan aannemen spreken we van een continue-tijd Markov-keten of een discrete-tijd Markov-keten. Bij een continue-tijd Markov-keten is t een continue variabele, meestal met waarden in het interval $[0, \infty)$. Bij een discrete-tijd Markov-proces is t een discrete variabele, meestal met waarden uit de verzameling $\{0, 1, 2, \ldots\}$.

Conceptueel is een discrete-tijd Markov-proces nog het meest eenvoudige. Voordat we daarmee beginnen, eerst nog een korte herhaling van matrixrekening, omdat daarmee zo compact de waarschijnlijkheidsanalyse van Markov-ketens beschreven kan worden.

4.3 Matrixrekening

Vectoren en matrices zullen we aangeven met hoofdletters en de elementen waaruit zij zijn opgebouwd worden aangegeven met de corresponderende kleine letter.

Vectoren worden gebruikt om de kansverdeling van een Markov-keten over zijn toestandsruimte weer te geven. De overgangskansen staan in een matrix: in rij i kolom j staat de kans dat het proces van toestand i naar toestand j gaat. Om de gewenste resultaten (de kansverdeling op het volgende tijdstip) te krijgen, moeten we vector-matrixvermenigvuldiging doen, en niet matrix-vectorvermenigvuldiging. Dit betekent dat hier de vectoren liggende (één rij, n kolommen) vectoren zijn.

Nota bene: de nummering van de toestandsruimte van het Markov-proces loopt hier van 1 tot en met n. Voor de uitleg van matrixrekening is dit verreweg het meest voor de hand liggend. Echter als de toestand van het Markov-proces bijvoorbeeld het aantal wachtende taken voorstelt, dan is het natuurlijker om de nummering van 0 tot n te laten lopen. In zo'n geval moet de nummering (de indices van de vector- en matrixelementen) even aangepast worden. Bij wachtsystemen met onbeperkte wachtruimte loopt de nummering van 0 tot ∞ (in dat geval passen we geen matrixrekening toe).

Laat $V(t)$ een (liggende) n-dimensionale vector zijn:

$$V(t) = \begin{bmatrix} v_1(t) & v_2(t) & \cdots & v_n(t) \end{bmatrix} \tag{4.1}$$

en P een $n \times n$ matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{4.2}$$

Stel, we willen uitgaande van $V(0)$ en P het vector-matrixprodukt $V(1)$ berekenen:

$$V(1) = V(0)P \tag{4.3}$$

Dan is $V(1)$ een gewogen som van de rijen van P, de gewichtsfactoren zijn de elementen van $V(0)$. Het j-de element van $V(1)$ hangt alleen af van elementen uit de j-de kolom van P:

$$v_j(1) = \sum_{k=1}^{n} v_k(0)p_{kj} \tag{4.4}$$

Voorbeeld 4.1

Gegeven de vector $V(0)$ en matrix P:

$$V(0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{4.5}$$

$$P = \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.8 & 0.0 & 0.2 \\ 0.0 & 0.8 & 0.2 \end{bmatrix} \tag{4.6}$$

Gevraagd: Het vector matrixproduct $V(0)P$.

We hadden al gelezen dat de vector-matrixvermenigvuldiging op te vatten is als een gewogen som van de rijen van P. Door de speciale keuze van $V(0)$ (alle elementen, behalve de eerste, gelijk aan 0) krijgt de eerste rij van P al het gewicht. Omdat de gewichtsfactor 1 is, is het effect dat gewoon de eerste rij van P geselecteerd wordt:

$$V(0)P = \begin{bmatrix} 0.8 & 0.2 & 0.0 \end{bmatrix} \tag{4.7}$$

Matrix-matrixvermenigvuldiging kan gezien worden als n keer vector-matrixvermenigvuldiging. Stel, we willen, uitgaande van de $n \times n$ matrices P en Q, hun produkt, de $n \times n$ matrix R, berekenen:

$$R = PQ \tag{4.8}$$

dan wordt de i-de rij van R bepaald door de i-de rij van P:

$$r_{ij} = \sum_{k=1}^{n} p_{ik}q_{kj} \tag{4.9}$$

Opmerking: in het algemeen is PQ niet gelijk aan QP.

Voorbeeld 4.2

Gegeven de matrix P van voorbeeld 4.1.

Gevraagd: Wat is P^2?

Het kwadraat van een (vierkante) matrix is niet anders dan het produkt van die matrix met zichzelf. De elementen van P^2 kunnen recht-door-zee berekend worden met (4.9). Iets vlotter gaat het door in te zien dat elke rij van P^2 een gewogen som is (gewichts factoren 0.8 en 0.2) van twee rijen van P. Het resultaat is:

$$\begin{bmatrix} 0.80 & 0.16 & 0.04 \\ 0.64 & 0.32 & 0.04 \\ 0.64 & 0.16 & 0.20 \end{bmatrix} \tag{4.10}$$

Matrix-matrixvermenigvuldiging van niet-vierkante matrices is ook mogelijk. De enige restrictie is dat het aantal kolommen van de linker matrix gelijk is aan het aantal rijen van de rechter matrix. In het algemene geval wordt een $m \times n$ matrix met een $n \times q$ matrix vermenigvuldigd; het product is een $m \times q$ matrix. Een bijzonder geval is $q = 1$. Dan is de rechter matrix een gewone (staande) vector en hebben we te maken met gewone matrix-vectorvermenigvuldiging.

4.4 Eindige Markov-keten met discrete tijdparameter

Bij een eindige Markov-keten bestaat de toestandsruimte uit, zeg, n mogelijke toestanden, die we hier, overeenkomstig de notatie in de vorige paragraaf, van 1 tot n nummeren (soms ligt het ook voor de hand om de nummering bij 0 te beginnen). De tijdparameter doorloopt de verzameling $\{0, 1, 2, \ldots\}$.

De evolutie van een eindige Markov-keten met discrete tijdparameter wordt bepaald door zijn overgangsmatrix P. Hierin geeft het element op de i-de rij, j-de kolom de kans aan dat de toestand op tijdstip t gelijk aan j zal zijn, gegeven dat de toestand op tijdstip $t-1$ gelijk aan i was:

$$p_{ij} = \Pr\{X(t) = j \mid X(t-1) = i\} \tag{4.11}$$

Laat $V(t-1)$ de kansverdeling van $X(t-1)$ zijn:

$$v_i(t-1) = \Pr\{X(t-1) = i\} \tag{4.12}$$

De kans om op tijdstip t in toestand j te zijn wordt berekend door uit te gaan van de kansverdeling op tijdstip $t-1$ en voor elke toestand i de kans op een overgang van i naar j in aanmerking te nemen.

$$v_j(t) = \sum_{i=1}^{n} v_i(t-1) p_{ij} \tag{4.13}$$

In matrixnotatie kan dit kortweg geschreven worden als:

$$V(t) = V(t-1)P \tag{4.14}$$

We kunnen, uitgaande van een gegeven kansverdeling $V(0)$, t maal de kansverdeling rechts vermenigvuldigen met P, om zo bij $V(t)$ uit te komen:

$$V(t) = V(0)P^t \tag{4.15}$$

Voorbeeld 4.3

We beschouwen de inhoud van een berichtenbuffer als een discrete tijd Markov-proces. $X(t)$ is het aantal berichten in de buffer aan het eind van het t-de interval. Het t-de interval duurt van tijdstip $t-1$ tot tijdstip t. In dit voorbeeld kan de berichtenbuffer maximaal 2 berichten bevatten. De binnenkomst van een bericht valt precies samen met een interval en berichten die arriveren als de buffer vol is gaan verloren. Alleen als er geen aankomst is en als de buffer niet leeg is dan wordt in een interval één bericht verwerkt. De kans op een aankomst in een interval is 0.2.

Gevraagd: De matrix van overgangskansen en de kansverdeling van het aantal berichten in de buffer op $t = 3$, gegeven dat de buffer leeg is op $t = 0$.

Uit bovenstaande beschrijving volgt dat als de buffer leeg dan kan er met kans 0.2 één bericht bij komen of er gebeurt niets. Als er één bericht in de buffer is dan komt er met kans 0.2 nog één bij of, met kans 0.8, gaat er één weg. Als de buffer vol is dan verdwijnt er één bericht met kans 0.8 of de buffer blijft vol. Deze drie zinnen zijn samengevat in de drie regels van de overgangsmatrix P:

$$P = \begin{bmatrix} 0.8 & 0.2 & 0.0 \\ 0.8 & 0.0 & 0.2 \\ 0.0 & 0.8 & 0.2 \end{bmatrix} \tag{4.16}$$

Het gegeven dat de buffer leeg op $t = 0$ laat zich vertalen in de begin verdeling waarbij alle kansmassa in het eerste element zit (met kans 1 zijn er nul berichten in de buffer):

$$V(0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{4.17}$$

De kansverdeling op $t = 3$ is volgens (4.15) gelijk aan $V(3) = V(0)P^3$. Omdat we al eerder de producten $V(0)P$ en P^2 hebben uitgerekend, ontbinden we $V(0)P^3$ in die twee factoren, zodat hier nog maar één vector-matrixproduct berekend hoeft te worden.

$$V(3) = \begin{bmatrix} 0.8 & 0.2 & 0.0 \end{bmatrix} \begin{bmatrix} 0.80 & 0.16 & 0.04 \\ 0.64 & 0.32 & 0.04 \\ 0.64 & 0.16 & 0.20 \end{bmatrix} = \begin{bmatrix} 0.768 & 0.192 & 0.040 \end{bmatrix} \tag{4.18}$$

Voor de overgangsmatrices die redelijkerwijs voorkomen geldt, dat op den duur (als $t \to \infty$) de kansverdeling van $V(t)$ naar een limiet V gaat die niet meer afhangt van de begin verdeling $V(0)$. Markov-ketens waarvoor zo'n limiet V bestaat worden regulier genoemd.

Niet-reguliere Markov-ketens hebben te maken met pathologische gevallen waarin er met kans 1 een cyclus van toestanden doorlopen wordt of waarin de toestandsruimte kan worden opgedeeld in twee of meer deelverzamelingen zodanig dat er geen overgangen tussen die deelverzamelingen voorkomen. Ook kan het zijn dat de evenwichtsverdeling niet bestaat omdat bij een oneindige Markov-keten de waarschijnlijksheidmassa zich steeds verder naar rechts verplaatst (de verwachtingswaarde van $X(t)$ wordt steeds groter als $t \to \infty$); dit komt bijvoorbeeld voor bij een wachtrijsysteem waar de aankomst- intensiteit groter is dan de verwerkingsintensiteit.

Voor reguliere Markov-ketens is de evenwichtsverdeling V gedefinieerd als

$$V = \lim_{t \to \infty} V(t) \tag{4.19}$$

Bij het berekenen van achtereen volgens $V(t)$, $V(t+1)$, $V(t+2), \ldots$ ziet men dat de verandering van deze kansverdeling steeds geringer wordt. $V(t)$ gaat in de limiet naar de kansverdeling V waarvoor geldt dat nog eens vermenigvuldigen met P in het geheel geen verandering van de kansverdeling tot gevolg heeft. Dus de evenwichtsverdeling V voldoet aan het volgende stelsel lineaire vergelijkingen:

$$V = VP \tag{4.20}$$

Men is geïnteresseerd in het berekenen van de evenwichtsverdeling omdat dit het lange termijn gedrag van een systeem kwantificeert. Bijvoorbeeld de evenwichtskans dat de buffer vol is, is op de lange termijn gelijk aan de fractie van de tijd dat de buffer vol is. Hieruit volgt dan weer hoe groot de fractie van de berichten die verloren gaat is.

Eén mogelijkheid om de evenwichtsverdeling te berekenen is via formule (4.20). Als we I definiëren als de eenheidsmatrix (op de diagonaal 1-en en verder overal 0), en O als de nul-vector (overal 0) dan is (4.20) ook te schrijven als:

$$O = V(P - I) \tag{4.21}$$

Dit zijn n vergelijkingen met n onbekenden maar het zijn niet n onafhankelijke vergelijkingen. Dit blijkt al uit het linker lid, waarin alleen nullen staan. Dit betekent dat als V een oplossing is dan is een willekeurig veelvoud van V ook een oplossing. Om de oplossing eenduidig vast te leggen moeten we gebruik maken van het gegeven dat de som van de kansen gelijk aan 1 is. Dat geeft als extra vergelijking de normeringsvergelijking:

$$\sum_{i=1}^{n} v_i = 1 \tag{4.22}$$

Als men de beschikking heeft over een routine voor matrixinversie, zoals die bijvoorbeeld als standaardfunctie in het spreadsheet programma Lotus 123 vanaf versie 2 zit, dan kan hiervan als volgt gebruik gemaakt worden om de evenwichtsverdeling te berekenen:

a) Construeer de matrix M door in $(P - I)$ de laatste kolom te vervangen door een kolom van 1-en (een willekeurige andere kolom is ook mogelijk, dit verandert de positie van de 1 in de hieronder gedefinieerde vector E).

b) Definieer de liggende vector $E = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$

c) dan is de evenwichtsverdeling: $V = EM^{-1}$

In de matrix $(P - I)$ zijn de rijsommen steeds gelijk aan 0; er gaat dus geen informatie verloren door één kolom weg te laten. Het effect van een kolom van 1-en is dat vermenigvuldiging van V met die kolom gelijk is aan de som van de elementen van V. Stel dat de laatste kolom van $(P-I)$ vervangen is door 1-en, dan is het matrix-vectorproduct VM gelijk aan $E = \left[\begin{array}{cccccc} 0 & 0 & \cdots & 0 & 0 & 1 \end{array}\right]$ De laatste 1 komt van de normeringsvergelijking (4.22). Het bovenstaande recept gebruikt dat de oplossing van het stelsel vergelijkingen $VM = E$, gegeven wordt door $V = EM^{-1}$. Hierin is M^{-1} de inverse van de matrix M, dat wil zeggen die matrix waarvoor geldt: $M^{-1}M = I$.

Voorbeeld 4.4

Voor het model van de berichtenbuffer van voorbeeld 4.3 willen we de fractie van de tijd dat de buffer vol is weten.

We construeren de matrix M, door in de overgangsmatrix P de diagonaal-elementen met 1 te verminderen en de rechter kolom te vervangen door 1-en:

$$M = \left[\begin{array}{rrr} -0.2 & 0.2 & 1 \\ 0.8 & -1 & 1 \\ 0 & 0.8 & 1 \end{array}\right] \tag{4.23}$$

met behulp van bijvoorbeeld Lotus 123 vinden we:

$$M^{-1} = \left[\begin{array}{rrr} -2.143 & 0.714 & 1.429 \\ -0.952 & -0.238 & 1.190 \\ 0.762 & 0.190 & 0.048 \end{array}\right] \tag{4.24}$$

De evenwichtsverdeling van de Markov-keten, V, wordt nu gegeven door:

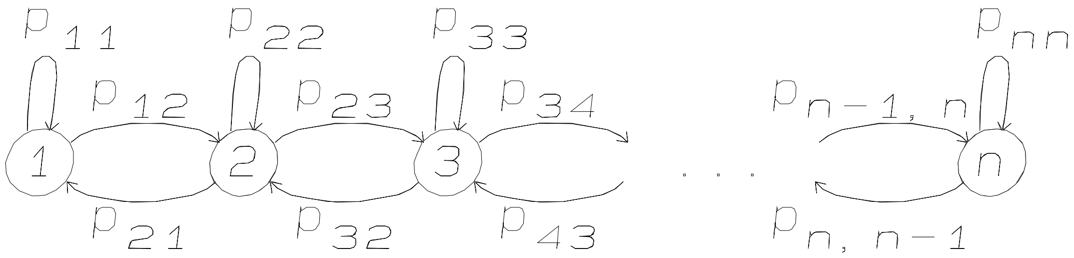
$$V = EM^{-1} = \left[\begin{array}{ccc} 0.762 & 0.190 & 0.048 \end{array}\right] \tag{4.25}$$

De evenwichtsverdeling is te interpreteren als de fractie van de tijd dat het proces in de verschillende toestanden is. Dus de buffer is vol voor een fractie 0.048 van de tijd.

Bij het definiëren van een Markov-keten is het altijd nuttig om het bijbehorende toestandsdiagram te tekenen. Voor een eindige Markov-keten met discrete tijdparameter ziet dit er als volgt uit:

- Voor elke toestand een cirkel met in de cirkel het nummer van die toestand
- Voor elke $p_{ij} > 0$ een pijl van cirkel i naar cirkel j, met naast de pijl de waarde van p_{ij}

Het berekenen van de evenwichtsverdeling kan nog eenvoudiger dan volgens (4.21) en (4.22) gebeuren in het geval dat vanuit toestand i alleen maar overgangen mogelijk zijn naar toestanden $i-1$, $i+1$ en i zelf. Het bijbehorende "n dimensionale" toestandsdiagram is weergegeven in figuur 4.1.



Figuur 4.1: Toestandsdiagram van een discrete-tijd Markov-keten. Dit is niet de meest algemene vorm, omdat hier alleen overgangen plaatsvinden naar toestanden met een nummer dat één hoger of één lager is.

In dit geval volgt uit evenwichtsoverwegingen dat de kans op een overgang van i naar $i+1$ even groot moet zijn als de kans op een overgang van $i+1$ naar i. Er vindt een overgang plaats van i naar $i+1$ als de toestand gelijk is aan i (kans v_i) en als gegeven toestand i de bestemming $i+1$ gekozen wordt (kans $p_{i,i+1}$). Met dezelfde redenering vinden we de kans op een overgang van $i+1$ naar i. Het gelijkstellen van deze kansen levert:

$$v_i p_{i,i+1} = v_{i+1} p_{i+1,i} \tag{4.26}$$
$$\Longleftrightarrow$$
$$v_{i+1} = v_i \frac{p_{i,i+1}}{p_{i+1,i}} \qquad i = 1, 2, \ldots, n-1 \tag{4.27}$$

Dit leidt tot het volgende algoritme om de evenwichtsverdeling V te berekenen.

Ga uit van een ongenormeerde verdeling W, gedefinieerd door:

$$w_1 = 1 \tag{4.28}$$
$$w_{i+1} = w_i \frac{p_{i,i+1}}{p_{i+1,i}} \qquad i = 1, 2, \ldots, n-1 \tag{4.29}$$

Bepaal w_{som}:

$$w_{\text{som}} = \sum_{i=1}^{n} w_i \tag{4.30}$$

Dan wordt de evenwichtsverdeling V gegeven door:

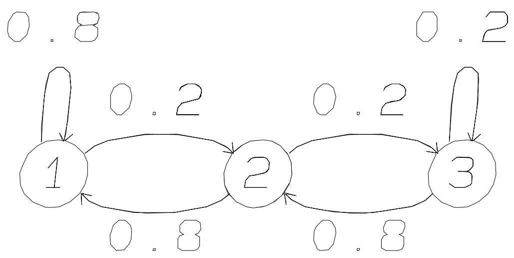
$$v_i = \frac{w_i}{w_{\text{som}}} \qquad i = 1, 2, \ldots, n \tag{4.31}$$

Voorbeeld 4.5

Gegeven een discrete-tijd Markov-keten met het toestandsdiagram als gegeven in figuur4.2:

Gevraagd: De evenwichtsverdeling van deze Markov-keten.

Merk op dat dit de Markov-keten is van het voorbeeld van de berichtenbuffer. We kunnen dus het gevraagde antwoord aflezen uit de berekening van voorbeeld 4.4. Echter, we willen



Figuur 4.2: Toestandsdiagram bij voorbeeld 4.5.

hier de eenvoudige structuur van het toestandsdiagram gebruiken, om zonder matrixinversie de evenwichtsverdeling te vinden.

We passen de formules (4.28)–(4.31) toe en vinden zo:

$$\begin{aligned} w_1 &= 1 && (4.32)\\ w_2 &= \frac{0.2}{0.8} w_1 = \frac{1}{4} && (4.33)\\ w_3 &= \frac{0.2}{0.8} w_2 = \frac{1}{16} && (4.34)\\ w_{\mathrm{som}} &= \frac{21}{16} && (4.35) \end{aligned}$$

en de gevraagde evenwichtsverdeling:

$$\begin{aligned} v_1 &= \frac{16}{21} = 0.762 && (4.36)\\ v_2 &= \frac{4}{21} = 0.190 && (4.37)\\ v_3 &= \frac{1}{21} = 0.048 && (4.38) \end{aligned}$$

4.5 Markov-keten met continue tijdparameter

Het verschil met de in de vorige paragraaf behandelde Markov-keten is in de tijdparameter. Hier is de tijdparameter een continue grootheid. Het is een element van de positieve reële rechte $[0, \infty)$.

Bij Markov-ketens (continue of discrete tijdparameter) is de toestandsruimte altijd discreet. Mogelijke toestandruimtes van Markov-ketens zijn bijvoorbeeld:

- $\{1, 2, \ldots, n\}$
- $\{A, B, C, D, E, F, G\}$

- $\{0, 1, \ldots, n\}$
- $\{0, 1, 2, \ldots\}$

Deze laatste (onbegrensde) toestandsruimte komen we tegen bij het beschrijven van een wachtrij systeem met een onbegrensd aantal wachtplaatsen. De toestand van het wachtrij systeem op een tijdstip t wordt gegeven door het aantal klanten in het systeem op dat tijdstip.

In de rest van dit hoofdstuk zullen we steeds uitgaan van een toestandsruimte die begint bij "0". Dus één van de laatste twee genoemde mogelijke toestandsruimtes.

Omdat hier de tijdparameter continu is, houden we veranderingen bij van de Markov-keten in infinitesimale tijdsintervalletjes. Het is daarom niet onredelijk om de aandacht te beperken tot die Markov-processen waarbij de toestandsovergangen alleen plaatsvinden naar buur-toestanden. Markov-processen waarbij alleen overgangen plaatsvinden naar buur-toestanden worden geboorte-sterfte-processen genoemd.

Voor een geboorte-sterfte-proces zijn de overgangskansen van de toestand op tijdstip $t - \Delta t$ naar een toestand op tijdstip t hieronder gegeven. Beschouw Δt als een klein getal, later beschouwen we de limiet $\Delta t \to 0$. Termen van de orde $(\Delta t)^2$ en hoger zijn daarom hieronder samengevat als $o(\Delta t)$, wat wil zeggen dat deze termen ten opzichte van Δt verwaarloosd kunnen worden als $t \to 0$.

$$\Pr\{X(t) = i \mid X(t - \Delta t) = i - 1\} = \lambda_{i-1}\Delta t + o(\Delta t) \tag{4.39}$$
$$\Pr\{X(t) = i \mid X(t - \Delta t) = i\} = 1 - \lambda_i\Delta t - \mu_i\Delta t + o(\Delta t) \tag{4.40}$$
$$\Pr\{X(t) = i \mid X(t - \Delta t) = i + 1\} = \mu_{i+1}\Delta t + o(\Delta t) \tag{4.41}$$

Hierin is λ_i de geboorte-intensiteit in toestand i en μ_i de sterfte-intensiteit in toestand i. Zo'n intensiteit stelt, na vermenigvuldiging met Δt, de kans voor op een overgang naar een toestand met één meer of één minder.

Als voorheen definiëren we

$$v_i(t) = \Pr\{X(t) = i\} \tag{4.42}$$
$$V(t) = \begin{bmatrix} v_1(t) & v_2(t) & \cdots \end{bmatrix} \tag{4.43}$$

De kans dat het proces op tijdstip t in toestand i is wordt, op termen van de orde $o(\Delta t)$ na, volledig bepaald door de kans dat het op tijdstip $t - \Delta t$ in toestand $i - 1$, i of $i + 1$ was:

$$v_i(t) = v_{i-1}(t - \Delta t)\lambda_{i-1}\Delta t \tag{4.44}$$
$$+v_i(t - \Delta t)(1 - \lambda_i\Delta t - \mu_i\Delta t) \tag{4.45}$$
$$+v_{i+1}(t - \Delta t)\mu_{i+1}\Delta t + o(\Delta t) \tag{4.46}$$

Uit bovenstaande vergelijking kunnen we de afgeleide bepalen van de kans om in toestand i te zijn, door de term $v_i(t - \Delta t)$ naar het linkerlid te brengen, te delen door Δt, en dan de limiet $t \to 0$ te nemen:

$$\frac{d}{dt}v_i(t) = v_{i-1}(t)\lambda_{i-1} + v_i(t)(-\lambda_i - \mu_i) + v_{i+1}(t)\mu_{i+1} \qquad i = 0, 1, 2, \ldots \tag{4.47}$$

Merk op dat als de index $i = 0$ is, dan valt de eerste term van het rechterlid weg. Ook is in die toestand de sterfte-intensiteit $\mu_0 = 0$. We krijgen dan voor $i = 0$:

$$\frac{d}{dt} v_0(t) = v_0(t)(-\lambda_0) + v_1(t)\mu_1 \tag{4.48}$$

In het geval van een continue-tijd Markov-keten met een eindige toestandsruimte $\{0, 1, \ldots, n\}$, geldt dat $\lambda_n = 0$ en krijgen we een soortgelijke aanpassing van de differentiaalvergelijking (4.47).

De differentiaalvergelijkingen (4.47) kunnen weer compact opgeschreven worden in matrix-notatie. We definiëren daartoe de volgende tri-diagonale matrix Q. Deze matrix wordt de infinitesimaalgenerator van de Markov-keten genoemd.

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ -\mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & -\mu_2 & -\lambda_2 - \mu_2 & \lambda_1 & 0 & \cdots \\ 0 & 0 & \ddots & & \ddots & \ddots \\ \vdots & \vdots & \ddots & & & \end{bmatrix} \tag{4.49}$$

De differentiaalvergelijking voor $V(t)$, de kansverdeling op tijdstip t, wordt dan een gewone eerste orde lineaire differentiaalvergelijking.

$$V(t) = V(t)Q \tag{4.50}$$

In veel gevallen is men geïnteresseerd in de evenwichtsverdeling van de continue-tijd Markov-keten. De berekening daarvan wordt weer even eenvoudig als dat was voor een discrete-tijd Markov-keten met alleen overgangen naar buurtoestanden. Dat behandelen we aan het eind van dit hoofdstuk.

Eerst gaan we nog in op het tijdsafhankelijke gedrag van de kansverdeling (de oplossing van de differentiaalvergelijking (4.50)). Dit vraagt iets meer geavanceerde begrippen uit de matrixrekening. Voorbeeld 4.6 laat een toepassing zien waar het nodig is om de vector $V(t)$ als functie van t te berekenen.

De oplossing van (4.50) wordt gegeven door:

$$V(t) = V(0)e^{Qt} \tag{4.51}$$

In het geval dat $V(t)$ en Q gewoon scalaire grootheden zijn (gewone getallen, en niet een vector en een matrix), is het niet zo moeilijk om te verifiëren dat $V(t)$ aan (4.50) voldoet. In het vector matrix geval is (4.52) evenzo de oplossing van (4.50), maar dan moeten we nog wel even aangeven hoe e^{Qt} gedefinieerd is. Deze definitie gaat met behulp van de reeksontwikkeling van de e-macht:

$$e^{Qt} = I + Qt + \frac{1}{2}Q^2t^2 + \frac{1}{3!}Q^3t^3 + \frac{1}{4!}Q^4t^4 + \cdots \tag{4.52}$$

I stelt de eenheidsmatrix voor (1-en op de diagonaal en verder 0-en). Het berekenen van machten van matrices hebben we al gezien aan het begin van dit hoofdstuk. Het vermenigvuldigen van een matrix met een scalar (bijvoorbeeld t of $1/2$) betekent het vermenigvuldigen van elk matrixelement met die scalar. Het optellen van matrices is niets anders dan het optellen van de overeenkomstige elementen.

Het is natuurlijk onbegonnen werk, om alle machten van Q uit te rekenen. Om toch verder te kunnen, maken we gebruik van het feit dat Q te diagonaliseren is:

$$Q = U^{-1}DU \tag{4.53}$$

Hierin is U de matrix gevormd door de linker eigenvectoren van Q. D is een diagonaalmatrix, met op de diagonaal de eigenwaarden van Q. Een willekeurige macht van Q is nu te schrijven als:

$$\begin{aligned} Q^n &= \underbrace{\left(U^{-1}DU\right)\left(U^{-1}DU\right)\cdots\left(U^{-1}DU\right)}_{n} && (4.54)\\ &= U^{-1}D^nU && (4.55)\end{aligned}$$

De n-de macht van een diagonaalmatrix D is direct op te schrijven; dit is de diagonaalmatrix met op de diagonaal de n-de machten van de diagonaalelementen van D.

De reeksontwikkeling (4.53) is nu te schrijven als:

$$\begin{aligned} e^{Qt} &= U^{-1}\left(I + Dt + \frac{1}{2}D^2t^2 + \frac{1}{3!}D^3t^3 + \frac{1}{4!}D^4t^4 + \cdots\right)U && (4.56)\\ &= U^{-1}e^{Dt}U && (4.57)\end{aligned}$$

De berekening van e^{Dt} vereist geen verdere matrixoperaties, want deze matrix is overal 0, met uitzondering van de diagonaalelementen die gevormd worden door $e^{\text{diagonaalelement van } D}$.

We hebben hiermee volledig het tijdsafhankelijke gedrag van de kansverdeling gespecificeerd. Een voorbeeld met een eindige Markov-keten zal de bovenstaande berekeningen verduidelijken.

Voorbeeld 4.6

De bezetting van twee telefoonlijnen wordt gemodelleerd als een continue tijd Markov-keten. De toestandsruimte (het aantal bezette lijnen) is $\{0, 1, 2\}$. De overgangsintensiteiten zijn $\lambda_0 = \lambda_1 = 1$ en $\mu_1 = 1$ en $\mu_2 = 2$. De tijdseenheid is de gemiddelde houdtijd van een gesprek. Op $t = 0$ zijn beide lijnen bezet: $V(0) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$.

Gevraagd: Hoe groot is de kans dat op $t = 0.5$ beide lijnen nog steeds, of weer opnieuw, bezet zijn? Het antwoord geeft een indicatie voor het instellen van de vertraging voor de automatische oproepherhaler.

De infinitesimaalgenerator Q voor dit voorbeeld is:

$$Q = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 2 & -2 \end{bmatrix} \tag{4.58}$$

Q is te schrijven als $Q = U^{-1}DU$. De lezer kan verifiëren dat de rijen van U de linker (eenheids-) eigenvectoren, en de diagonaalelementen van D de eigenwaarden van van Q zijn.

$$Q = U^{-1}DU = \tag{4.59}$$
$$\begin{bmatrix} 0.6000 & 0.2472 & 0.6472 \\ 0.6000 & -0.6472 & -0.2472 \\ 0.6000 & 0.8000 & -0.8000 \end{bmatrix} \begin{bmatrix} 0.00 & 0.0000 & 0.0000 \\ 0.00 & -3.6180 & 0.0000 \\ 0.00 & 0.0000 & -1.3820 \end{bmatrix} \begin{bmatrix} 0.6667 & 0.6667 & 0.3333 \\ 0.3090 & -0.8090 & 0.5000 \\ 0.8090 & -0.3090 & -0.5000 \end{bmatrix}$$

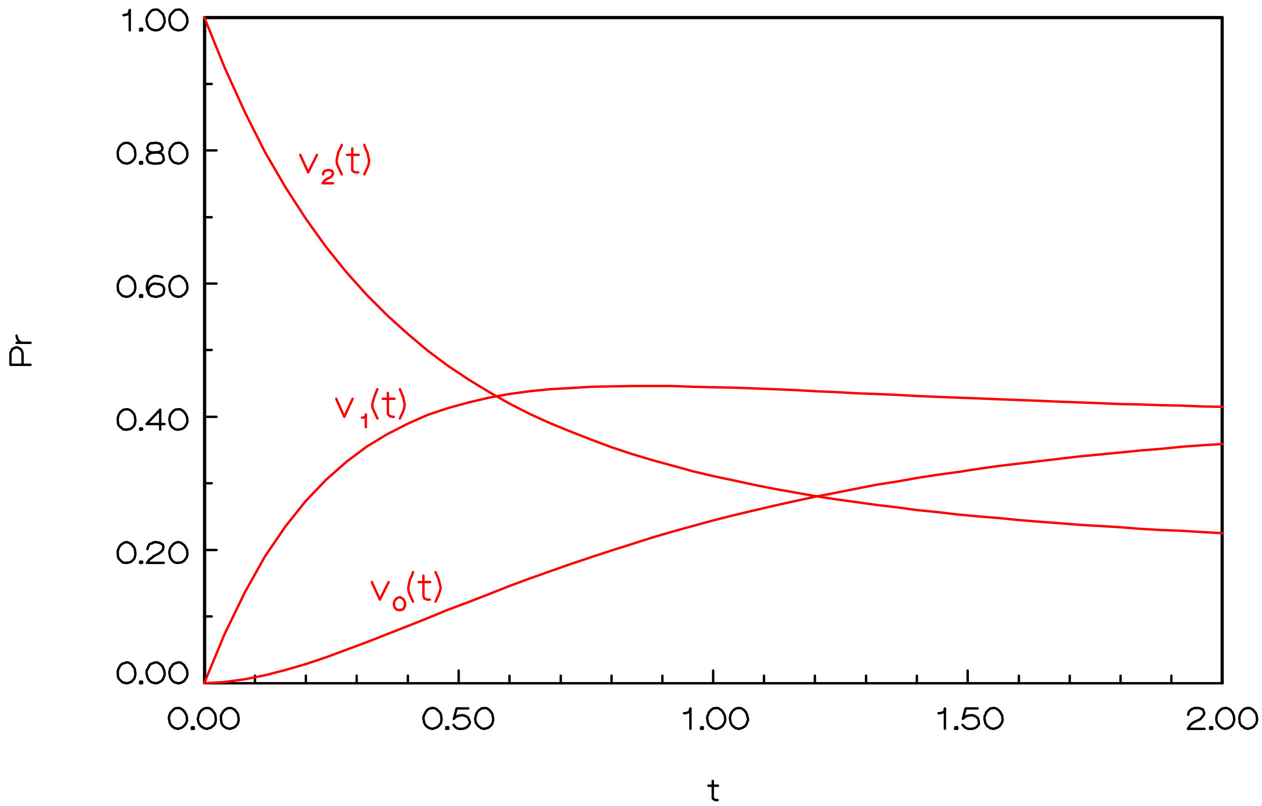
Volgens (4.52) en (4.57) is het verloop van de kansverdeling als functie van t:

$$V(t) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} U^{-1}e^{Dt}U \tag{4.60}$$

Het uitvoeren van bovenstaande matrixvermenigvuldigingen geeft een vector waarvan de elementen een som zijn van een constante term en twee e-machten:

$$v_0(t) = 0.4 + 0.2472e^{-3.618t} - 0.6472e^{-1.382t} \tag{4.61}$$
$$v_1(t) = 0.4 - 0.6472e^{-3.618t} + 0.2472e^{-1.382t} \tag{4.62}$$
$$v_2(t) = 0.2 + 0.4000e^{-3.618t} + 0.4000e^{-1.382t} \tag{4.63}$$

Een grafiek van de drie elementen van $V(t)$ is getekend in figuur 4.3.



Figuur 4.3: Kansverdeling van het aantal bezette lijnen als functie van t, gegeven dat op $t = 0$ beide lijnen bezet zijn.

Het antwoord op de vraag is:

$$v_2(0.5) = 0.2 + 0.4e^{-1.809} + 0.4e^{-0.691} \quad (4.64)$$
$$= 0.466 \quad (4.65)$$

Dat wil zeggen dat wanneer na een halve gemiddelde houdtijd weer opnieuw gekeken wordt of een lijn vrij is (nadat op $t = 0$ beide lijnen bezet waren) dan is de kans op succes 53%.

De matrixnotatie maakt het eenvoudig om de oplossing van het stelsel differentiaalvergelijkingen op te schrijven. Het voert te ver om daar hier verder op in te gaan. De matrixnotatie komt ook van pas als men de beschikking heeft over routines voor het oplossen van stelsels van lineaire vergelijkingen.

De evenwichtsverdeling

$$V = \lim_{t\to\infty} V(t) \quad (4.66)$$

wordt gekenmerkt door het niet meer veranderen (in de tijd gezien) van de kansverdeling, met andere woorden, de waarde van de afgeleide is 0. Dit geeft de volgende verzameling van vergelijkingen voor de evenwichtsverdeling:

$$0 = v_{i-1}\lambda_{i-1} + v_i(-\lambda_i - \mu_i) + v_{i+1}\mu_{i+1} \qquad i = 0, 1, 2, \ldots \quad (4.67)$$

Hierbij is de sterfte-intensiteit in toestand 0 per definitie gelijk aan 0: $\mu_0 = 0$. Bij een eindige Markov-keten is ook in toestand n de geboorte-intensiteit gelijk aan 0: $\lambda_n = 0$. Als $i = 0$ is in (4.67) dan moet voor v_{i-1} de waarde 0 ingevuld worden. De ongedefinieerde waarde van λ_{i-1} speelt dan geen rol meer.

Voor het oplossen van de evenwichtsvergelijking moet ook nog gebruik gemaakt worden van de normeringsvergelijking

$$\sum_{i=0}^{\infty} v_i = 1 \quad (4.68)$$

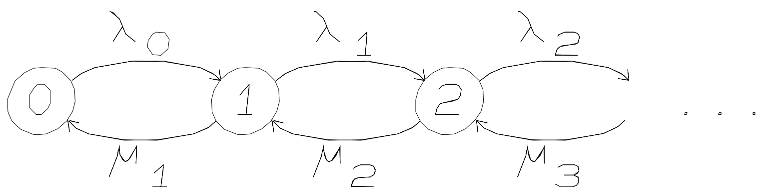
Evenals het stelsel differentiaalvergelijkingen (4.47), $i = 1, 2, \ldots$, kan het stelsel evenwichtsvergelijkingen kan in matrixvorm geschreven worden:

$$O = VQ \quad (4.69)$$

waarin O de nulvector $\begin{bmatrix} 0 & 0 & \cdots \end{bmatrix}$ is en Q de infinitesimaalgenerator zoals gedefinieerd in (4.49).

Omdat een geboorte-sterfte-proces alleen overgangen tussen buur-toestanden kent, heeft het toestandsdiagram van het Markov-proces een één-dimensionale vorm, en kan de evenwichtsverdeling makkelijk sequentieel uitgerekend worden, zoals verderop zal worden gegeven.

Het toestandsdiagram is weer een praktisch hulpmiddel voor het beschrijven van een geboorte-sterfte-proces. De algemene vorm van het toestandsdiagram is getekend in figuur 4.4.



Figuur 4.4: Toestandsdiagram van een geboorte-sterfte-proces. De waarden langs de pijlen zijn overgangsintensiteiten.

Het is van belang te realiseren dat de waarden, die langs de pijlen staan van een toestandsdiagram van een continue-tijd Markov-keten, een iets andere interpretatie hebben dan in het geval van een discrete-tijd Markov-keten. In het laatste geval stonden naast de pijlen overgangskansen. In het geval van een continue-tijd Markov-keten zijn die waarden overgangsintensiteiten. Een overgangsintensiteit is: kans per tijdseenheid. Pas na vermenigvuldiging met Δt, heeft bijvoorbeeld $\lambda_i \Delta t$ de interpretatie van de kans om van toestand i naar $i+1$ te gaan in een interval van lengte Δt. Daarom staan in een toestandsdiagram van een continue-tijd Markov-keten ook geen pijlen van een toestand naar zichzelf omdat hier geen sprake is van een overgangsintensiteit; wel van een overgangskans die bijvoorbeeld een waarde heeft van $1 - \lambda_i \Delta t - \mu_i \Delta t$, maar dan kan er geen factor t buiten haakjes gehaald worden.

Voor het bepalen van de evenwichtsverdeling van een continue-tijd Markov-keten (als het een geboorte-sterfte-proces is) kunnen we hetzelfde algoritme gebruiken als voor discrete-tijd Markov-ketens met een één-dimensionale toestandsruimte. Veel toepassingen van continue-tijd Markov-ketens hebben betrekking op systemen met een oneindige toestandsruimte, bijvoorbeeld wachtrijsystemen met onbegrensde wachtrij-capaciteit. Daarom formuleren we het algoritme voor een oneindige continue-tijd Markov-keten:

Ga uit van een ongenormeerde verdeling W, gedefinieerd door:

$$w_0 = 1 \tag{4.70}$$

$$w_{i+1} = w_i \frac{\lambda_i}{\mu_{i+1}} \qquad i = 1, 2, \ldots, n-1 \tag{4.71}$$

Bepaal w_{som}:

$$w_{\text{som}} = \sum_{i=0}^{\infty} w_i \tag{4.72}$$

Dan wordt de evenwichtsverdeling V gegeven door:

$$v_i = \frac{w_i}{w_{\text{som}}} \qquad i = 0, 1, 2, \ldots \tag{4.73}$$

4.6 Markov-modellen

De kracht van de Markov-modellen is dat de toestand van het systeem alle informatie bevat die we in onze analyse willen beschouwen. We hebben in de voorbeelden gezien (en meer voorbeelden volgen nog) dat met een toestandsruimte die een deelverzameling is van de natuurlijke getallen zinvolle, niet-triviale prestatie-analyse gedaan kan worden.

In het meer algemene Markov-model kan de toestand van het systeem een vector zijn met geheel-tallige waarden, of reële waarden. Dit kan nog meer algemeen gemaakt worden, maar dan verliest het de kracht van de eenvoud.

Het feit dat alle informatie zit in de toestand van het systeem, betekent met name, dat de tijd die al verstreken is (gedurende het verblijf in een toestand) niet bepalend is voor het toekomstig gedrag van het Markov-proces. Daarom heeft de kansverdeling van de verblijftijd van het proces in één toestand noodzakelijkerwijs de geheugenloze eigenschap. Bij een discrete-tijd Markov-keten is de verblijftijd geometrisch verdeeld en bij een continue-tijd Markov-keten heeft de verblijftijd de negatief exponentiële verdeling.

In de volgende hoofdstukken worden de Markov-modellen toegepast op wachtsystemen, verliessystemen en betrouwbaarheidsrekening.

Hoofdstuk 5

Wachtsystemen

Een eerste toepassing van onze kennis van stochastische variabelen en Markov-processen is de analyse van de $M/M/1$ en de $M/M/n$ wachtrij. Deze notatie van wachtrijsystemen komt van de statisticus D.G. Kendall. Eerst in het kort de betekenis van de symbolen in Kendall's notatie.

5.1 Kendall's notatie

Een wachtrij- of verliessysteem kan gespecificeerd worden met 3, 4 of 5 symbolen, gescheiden door een schuine streep, bijvoorbeeld:

$M/M/1$

$M/D/n/k/k$

Het eerste symbool slaat op het aankomstproces en het tweede symbool op het bedieningsproces. Wij zullen op deze posities de letters M of G of D tegenkomen.

De M is van Markovian. We hebben in het vorige hoofdstuk gezien dat, bij een continue-tijd Markov-proces, de verblijftijd in een toestand negatief exponentieel verdeeld is. Vertaald in termen van een wachtrij systeem (waarbij de toestand het aantal klanten in het systeem is), betekent dit dat de tijd tussen aankomsten van klanten negatief exponentieel verdeeld is (eerste symbool = M), respectievelijk, dat de bedieningsduur negatief exponentieel verdeeld is (tweede symbool = M). Omdat negatief verdeelde tussenaankomsttijden een karakterisering van een Poisson-proces is, betekent de eerste M ook dat het aankomstproces een Poisson-proces is, als tenminste de klantenpopulatie oneindig groot is. Deze laatste conditie wordt drie alinea's verderop verduidelijkt.

De G staat voor General en wordt gebruikt als de verdeling van de tussenaankomsttijd respectievelijk bedieningsduur niet een bekende verdeling is. De D van Deterministic wordt gebruikt als de tussenaankomsttijd respectievelijk bedieningsduur constant is. Dit is te zien als een verdeling waarbij alle kansmassa in een punt geconcentreerd is.

Het derde symbool slaat op het aantal bedieners ofwel loketten. Dus de $M/M/1$ queue, is de bekende single server queue waarbij het aankomstproces een Poisson-proces is en de bedieningsduur negatief exponentieel verdeeld.

Indien vermeld, geeft het vierde symbool het aantal plaatsen in het systeem en het vijfde symbool het totale aantal klanten. Indien niet vermeld dan worden deze parameters impliciet ∞ verondersteld. Met het aantal plaatsen wordt de som bedoeld van het aantal bedieningsplaatsen (= het aantal loketten) en het aantal plaatsen waar klanten kunnen wachten. We hebben te maken met een verliessysteem indien het aantal plaatsen in het systeem kleiner is dan het totale aantal klanten.

In het voorbeeld van de $M/D/n/k/k$ wachtrij zijn de tussenaankomsttijden negatief exponentieel verdeeld, de bedieningsduur is constant en er is voor elke klant één plaats: ofwel éénvan de n bedieners ofwel éénvan de $k-n$ wachtplaatsen. Bij een eindige populatie van klanten is de parameter van de negatief exponentiële verdeling van de tussenaankomsttijden afhankelijk van het aantal "vrije" klanten (d.w.z. klanten die niet bij een bediener én niet op een wachtplaats zijn). Daarom is het aankomstproces geen Poisson-proces; bij een Poisson-proces is de aankomstintensiteit constant en vermindert niet nadat er bijvoorbeeld een aankomst heeft plaatsgevonden.

5.2 Stochastische variabelen wachtsystemen.

Bij de analyse van een wachtrij systeem zijn we geïnteresseerd in hoeveel klanten er zijn en hoelang ze er zijn; dit kan betrekking hebben op alleen de wachtrij, alleen de loketten of het hele systeem (= wachtplaatsen + loketten). Dit levert de volgende 6 stochastische variabelen:

$$\begin{array}{ll} N_q & \text{het aantal klanten in het systeem} \\ N_w & \text{het aantal wachtende klanten} \\ N_s & \text{het aantal klanten onder bediening} \\ T_q & \text{de verblijftijd van een klant in het systeem} \\ T_w & \text{de wachttijd van een klant} \\ T_s & \text{de bedieningsduur van een klant} \end{array} \tag{5.1}$$

We schrijven N_q en niet $N_q(t)$ omdat we het steeds hebben over de evenwichtsverdeling. Tussen bovenstaande toevalsgrootheden zijn een paar voor de hand liggende relaties:

$$N_q = N_w + N_s \tag{5.2}$$

$$T_q = T_w + T_s \tag{5.3}$$

En ook nog een relatie die iets minder voor de hand ligt, maar wel zeer bruikbaar is; hieraan is de volgende paragraaf gewijd.

5.3 De formule van Little

Deze formule geeft de relatie tussen het gemiddelde aantal klanten in het systeem, en de gemiddelde verblijftijd van een klant in het systeem:

$$\mathrm{E}[N_q] = \lambda \mathrm{E}[T_q] \tag{5.4}$$

Of met woorden: als er gemiddeld λ klanten per seconde aankomen en elke klant blijft gemiddeld $\mathrm{E}[T_q]$ seconden dan zal men gemiddeld $\lambda\mathrm{E}[T_q]$ klanten aantreffen. Merk op de dimensies van linker en rechter lid:

$$\text{aantal klanten} = \text{klanten per seconde} \times \text{seconde per klant} \tag{5.5}$$

De formule van Little is zeer algemeen geldig; hangt bijvoorbeeld niet af van de bedieningsvolgorde (First Come First Served, Last In First Out) en geldt ook bij systemen met prioriteiten. Een algemeen bewijs voert hier te ver, maar voor het geval van FCFS kan de formule van Little als volgt aangetoond worden. Een klant verblijft gemiddeld $\mathrm{E}[T_q]$ seconden in het systeem. In die tijd komen er gemiddeld $\lambda\mathrm{E}[T_q]$ klanten aan. Bij vertrek laat de beschouwde klant dus gemiddeld $\lambda\mathrm{E}[T_q]$ klanten achter. Bij aankomst trof de klant onder beschouwing gemiddeld genomen $\mathrm{E}[N_q]$ klanten aan. Voor de evenwichtsverdeling moet het gemiddeld aantal aangetroffen klanten, $\mathrm{E}[N_q]$, gelijk zijn aan het gemiddeld aantal achtergelaten klanten, $\lambda\mathrm{E}[T_q]$.

De formule van Little geldt ook voor de twee andere paren van stochastische variabelen:

$$\mathrm{E}[N_w] = \lambda\mathrm{E}[T_w] \tag{5.6}$$
$$\mathrm{E}[N_s] = \lambda\mathrm{E}[T_s] \tag{5.7}$$

Voorbeeld 5.1

Een processor is 80% van de tijd bezig met het verwerken van data-pakketten. Gemiddeld zijn er 3.2 pakketten aan het wachten op verwerking.

Gevraagd: Hoe groot is de gemiddelde wachttijd van een pakket (uitgedrukt in de gemiddelde verwerkingstijd)?

N_s kan de waarden 1 of 0 aannemen, afhankelijk van of er een pakket verwerkt wordt of niet. Uit het gegeven $\Pr\{N_s = 1\} = 0.8$ volgt dat $\mathrm{E}[N_s] = 0.8$. Little's resultaat voor $\mathrm{E}[N_s]$ geeft dan:

$$\lambda = \frac{0.8}{\mathrm{E}[T_s]} \tag{5.8}$$

Substitutie van λ in Little's resultaat voor $\mathrm{E}[N_w]$ geeft, met het gegeven $\mathrm{E}[N_w] = 3.2$, de gemiddelde wachttijd

$$\mathrm{E}[T_w] = 3.2\frac{\mathrm{E}[T_s]}{0.8} = 4\mathrm{E}[T_s] \tag{5.9}$$

5.4 M/M/1 wachtsysteem

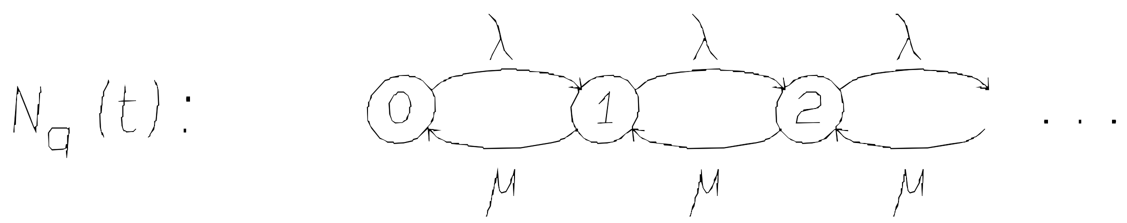
Bij het $M/M/1$ wachtsysteem komen klanten aan, voor één loket, volgens een Poisson-proces met intensiteit λ. Dit betekent dat de kans op een aankomst in Δt gelijk is aan $\lambda \Delta t$ en dat de tijd tussen twee aankomsten negatief exponentieel verdeeld is met gemiddelde $1/\lambda$. De bedieningsduur is ook negatief exponentieel verdeeld; het gemiddelde van deze verdeling $1/\mu$. Hier is μ eveneens op te vatten als een intensiteit: zolang het aantal klanten in het systeem > 0 is, zal in een intervalletje van Δt met kans $\mu \Delta t$ een klant vertrekken.

De gemiddelde bedieningsduur van een klant is de inverse van de verwerkingscapaciteit van het loket:

$$\mathrm{E}\left[T_s\right] = \frac{1}{\mu} \tag{5.10}$$

De bezettingsgraad van een wachtsysteem ρ is de verhouding van de aankomstintensiteit λ en de verwerkingscapaciteit μ. Voor een $M/M/1$ wachtsysteem is de bezettingsgraad van het loket gelijk aan het produkt van het aantal klanten (dat aankomt) per seconde en het aantal seconden (bediening) per klant:

$$\rho = \frac{\lambda}{\mu} \tag{5.11}$$



Figuur 5.1: Toestandsdiagram van het aantal klanten in het $M/M/1$ wachtsysteem.

Bij het berekenen van de evenwichtsverdeling volgens (4.71)–(4.73) zien we aan de hand van het toestandsdiagram in figuur 5.1 dat de verhouding tussen twee opeenvolgende kansen, $w_{i+1} : w_i$, gelijk is aan ρ. Dit geeft:

$$\begin{aligned} w_i &= \rho^i \qquad i = 0, 1, 2, \ldots && (5.12) \\ w_{\mathrm{som}} &= \sum_{i=0}^{\infty} \rho^i = \frac{1}{1-\rho} && (5.13) \end{aligned}$$

De evenwichtsverdeling, die de kansen geeft om op een willekeurig moment i klanten in het systeem aan te treffen, wordt gegeven door:

$$v_i = (1-\rho)\rho^i \qquad i = 0, 1, 2, \ldots \tag{5.14}$$

Dit lijkt op de geometrische verdeling, met $(1-\rho)$ i.p.v. p, maar hier begint de index bij 0 (vergelijk (2.1)). De verwachtingswaarde van de verdeling (5.14) geeft het gemiddelde aantal klanten in het systeem:

$$\mathrm{E}\left[N_q\right] = \sum_{i=0}^{\infty} i v_i = \frac{\rho}{1-\rho} \tag{5.15}$$

Met de formule van Little weten we dat het gemiddelde aantal klanten gelijk is aan λ keer de gemiddelde verblijftijd in het systeem. Daaruit volgt voor de gemiddelde verblijftijd van een klant in het systeem:

$$\begin{aligned} \mathrm{E}[T_q] &= \frac{1}{\lambda}\mathrm{E}[N_q] && (5.16) \\ &= \frac{1}{\mu-\lambda} && (5.17) \end{aligned}$$

De vraag die misschien wel rijst is: “als λ groter is dan μ, is de verblijftijd dan negatief?” Het antwoord is dat we bij de te bestuderen wachtsystemen er stilzwijgend van uit gaan dat de evenwichtsverdeling van het Markov-proces bestaat. Voor wachtsystemen betekent dit dat $\rho < 1$ (dus bij één loket dat $\lambda < \mu$) moet zijn.

Een bijzondere eigenschap van de $M/M/1$ wachtrij is dat de verdeling van de verblijftijd negatief exponentieel is. Dit betekent dat we expliciet de kans dat de verblijftijd kleiner is dan, of gelijk aan t kunnen opschrijven. De kans wordt gegeven door de negatief exponentiële verdelingsfunctie van T_q:

$$\Pr\{T_q \leq t\} = 1 - e^{-(\mu-\lambda)t} \qquad t \leq 0 \tag{5.18}$$

Percentielen van de verblijftijdverdeling kunnen direct uit de verdelingsfunctie (5.18) berekend worden. We leerden in sectie 2.5.7 (hoofdstuk 2) dat bijvoorbeeld het 99-ste percentiel, t_{99}, per definitie de eigenschap heeft dat de toevalsgrootheid T_q met 99% kans kleiner is dan, of gelijk aan, t_{99}. Het 99-ste percentiel is daarom de oplossing van de volgende vergelijking:

$$\Pr\{T_q \leq t_{99}\} = 1 - e^{-(\mu-\lambda)t_{99}} = 0.99 \tag{5.19}$$

hieruit volgt dat:

$$t_{99} = \frac{\ln(100)}{\mu-\lambda} \tag{5.20}$$

In figuur 5.2 staat ook het 63-ste percentiel getekend. Deze waarde is bijzonder omdat $63\% \approx 1 - e^{-1}$, waaruit direct is af te leiden dat

$$t_{63} = \frac{1}{\mu-\lambda} = \mathrm{E}[T_q] \tag{5.21}$$

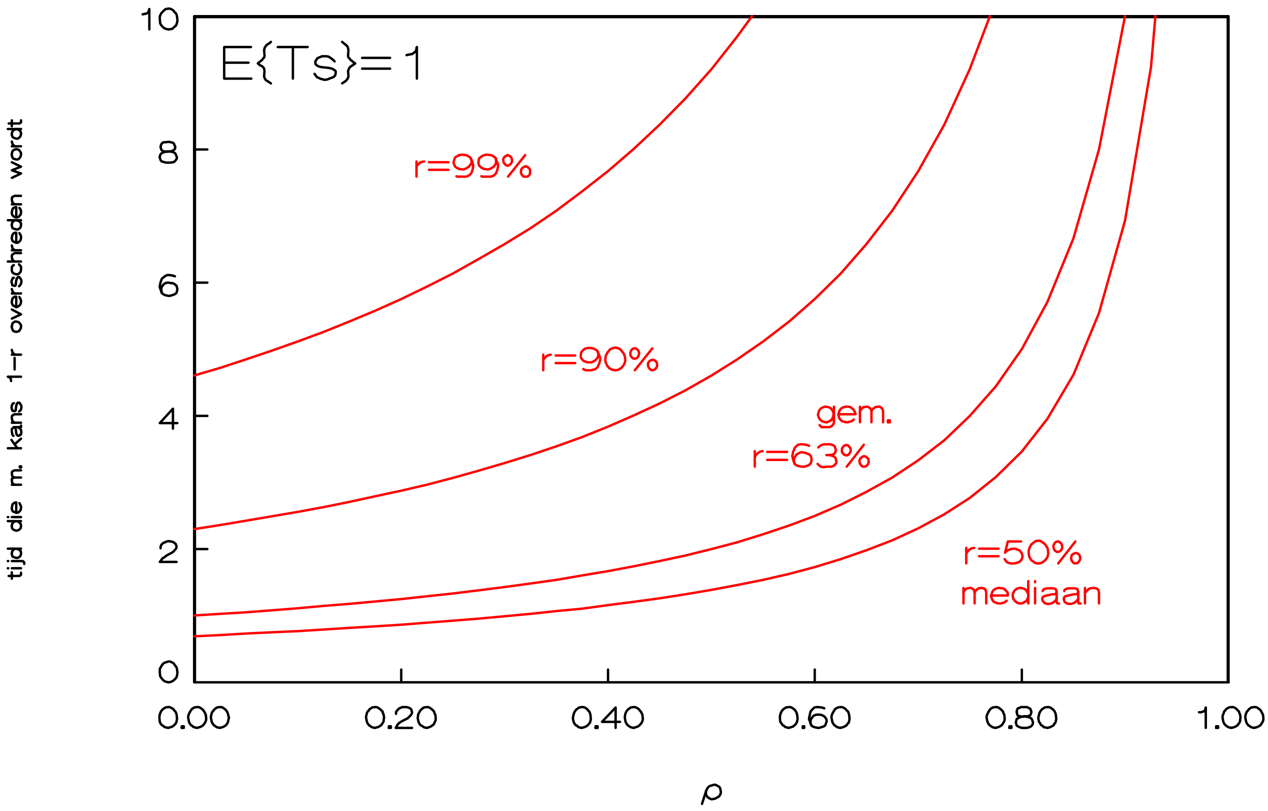
Dus t_{63} is de gemiddelde verblijftijd, en verder is t_{50} de mediaan van de verblijftijdverdeling.

Voorbeeld 5.2

Het eenvoudigste model van een processor die taken van verschillende lengte verwerkt is het $M/M/1$ model. Hierbij wordt zowel voor de tijd die verstrijkt tussen twee aankomsten als voor de bedieningsduur van een taak, de geheugenloze eigenschap verondersteld. We stellen verder dat de gemiddelde bedieningsduur 0.5 seconde is.

Men wil voor zo'n processor bepalen hoeveel taken er verwerkt kunnen worden bij een gegeven eis voor de responstijd en hoe sterk de responstijd toeneemt bij een 10% hogere bezettingsgraad. In voorbeeld 5.3 vergelijken we dit met een systeem met twee langzame processoren.

Gevraagd:



Figuur 5.2: Percentielen van de verblijftijdverdeling van de $M/M/1$ wachtrij. Verticaal is de verblijftijd uitgezet die met kans $1 - r$ overschreden wordt.

a) Hoe hangt de gemiddelde responstijd van het systeem (de gemiddelde verblijftijd van een klant) af van de aankomstintensiteit λ?

b) Hoeveel taken per seconde kunnen er verwerkt worden als de gemiddelde responstijd 2.5 seconden mag zijn?

c) Hoeveel neemt de responstijd toe als aankomstintensiteit van taken met 10% toeneemt?

Het gegeven $\mathrm{E}[T_s] = 0.5$ seconde impliceert: $\mu = 2\ \mathrm{s}^{-1}$ De relatie tussen $\mathrm{E}[T_q]$ en λ volgt uit (5.16) en is in dit geval:

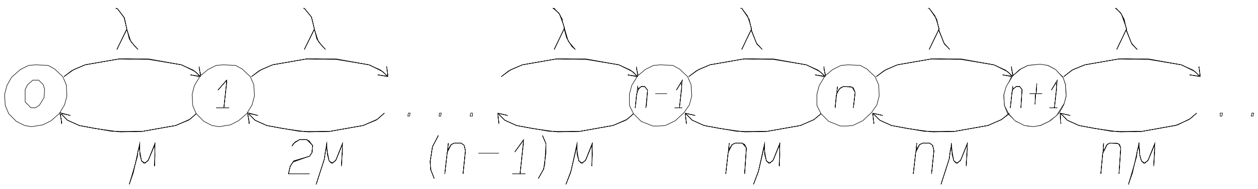
$$\mathrm{E}[T_q] = \frac{1}{2 - \lambda} \tag{5.22}$$

Met het gegeven $\mathrm{E}[T_q] = 2.5$ seconden, vinden we $\lambda = 1.6$. Het antwoord op b) is dat het systeem 1.6 taken per seconde kan verwerken.

Indien λ toeneemt tot $\lambda = 1.76$ dan wordt de gemiddelde responstijd $1/0.24 = 4.17$ seconde Ten opzichte van 2.5 seconden is dit een toename van 67%.

5.5 M/M/n wachtsysteem

Het Markov-model voor de $M/M/n$ wachtrij lijkt veel op het Markov-model voor de $M/M/1$ wachtrij; het verschil is dat nu bij k klanten in het systeem de kans op het vertrek van een klant in een tijdsintervalletje Δt gelijk is aan $k\mu\Delta t$ (voor $0 < k \leq n$) of gelijk is aan $n\mu\Delta t$ (voor $k > n$). Hierin is μ de bedieningsintensiteit van één loket. In het toestandsdiagram (figuur 5.3) zien we dan de "sterfte"-intensiteit oplopen van μ tot $n\mu$, om daarna constant te blijven.



Figuur 5.3: Toestandsdiagram van aantal klanten in een $M/M/n$ wachtrij.

De formules voor bijvoorbeeld gemiddelde verblijftijd zijn wel wat ingewikkelder en iets minder resultaten worden in expliciete vorm gegeven. Diegene die een misschien wel terechte weerzin voelt opkomen bij de formules in deze paragraaf realisere zich de eenvoud van het onderliggende Markov-model en van de numerieke procedure om de evenwichtsverdeling te berekenen. Uit de evenwichtsverdeling volgt direct het gemiddelde aantal klanten in het systeem en daaruit (met de formule van Little) de gemiddelde verblijftijd. De gemiddelde wachttijd is dan gemiddelde verblijftijd minus gemiddelde bedieningstijd.

Klanten worden bediend aan één van de n loketten en de gemiddelde bedieningsduur is daar:

$$\mathrm{E}[T_s] = \frac{1}{\mu} \tag{5.23}$$

De bezettingsgraad ρ is gedefinieerd als de verhouding van de aankomstintensiteit en de maximale verwerkingscapaciteit:

$$\rho = \frac{\lambda}{n\mu} \tag{5.24}$$

Uit het toestandsdiagram kunnen we voor de ongenormeerde evenwichtsverdeling W de volgende gelijkheden aflezen:

$$w_i = \frac{(n\rho)^i}{i!} \qquad i = 0, 1, 2, \ldots, n-1 \tag{5.25}$$

$$w_i = \frac{n^n}{n!}\rho^i \qquad i = n, n+1, n+2, \ldots \tag{5.26}$$

$$w_{\text{som}} = \sum_{i=0}^{n-1} \frac{(n\rho)^i}{i!} + \frac{n^n}{n!}\frac{\rho^n}{1-\rho} \tag{5.27}$$

De evenwichtsverdeling is dan:

$$v_i = \frac{w_i}{w_{\text{som}}} \qquad i = 0, 1, 2, \ldots \tag{5.28}$$

Als we de staart van de evenwichtsverdeling sommeren, te beginnen met $i = n$ (dit is de eerste toestand waarbij alle loketten bezet zijn) dan levert dit ons de kans dat alle loketten bezet zijn. Dit is ook de kans dat een klant bij aankomst geen vrij loket vindt. Het resultaat dat de gelijkheid van deze kansen stelt (en ook bewezen is) wordt PASTA genoemd. PASTA is het acroniem voor Poisson Arrivals See Time Averages. Voor de $M/M/n$ wachtrij is de kans dat een klant geen vrij loket vindt, v_w, mede van belang omdat het een compacte formule voor het gemiddelde aantal klanten in het systeem geeft. De kans v_w wordt ook genoteerd als $C(n,\rho)$ of als $E_{2,n}(\rho)$. De formule voor $E_{2,n}$ heet de Erlang-C-formule. In hoofdstuk 6 komt de Erlang-B-formule aan bod.

$$\begin{aligned} v_w &= \sum_{i=n}^{\infty} v_i && (5.29)\\ &= \sum_{i=n}^{\infty} \frac{(n\rho)^n}{n!} \frac{1}{w_{\text{som}}} && (5.30)\\ &= \frac{(n\rho)^n}{n!(1-\rho)} \frac{1}{w_{\text{som}}} && (5.31)\\ &= C(n,\rho) = E_{2,n}(\rho) && (5.32) \end{aligned}$$

De formule voor het gemiddelde aantal klanten in het systeem kan in termen van de Erlang-C-formule gegeven worden:

$$\mathrm{E}[N_q] = n\rho \left(1 + \frac{C(n,\rho)}{n(1-\rho)}\right) \tag{5.33}$$

Hieruit volgt de gemiddelde verblijftijd van een klant in het $M/M/n$ wachtsysteem:

$$\mathrm{E}[T_q] = \frac{1}{\mu} \left(1 + \frac{C(n,\rho)}{n(1-\rho)}\right) \tag{5.34}$$

Voor de $M/M/n$ wachtrij (inclusief $n = 1$) is de formule voor de gemiddelde wachttijd, $\mathrm{E}[T_w]$, niet opgeschreven. Deze kan direct verkregen worden door $\mathrm{E}[T_q]$ te verminderen met de gemiddelde bedieningsduur $\mathrm{E}[T_s]$.

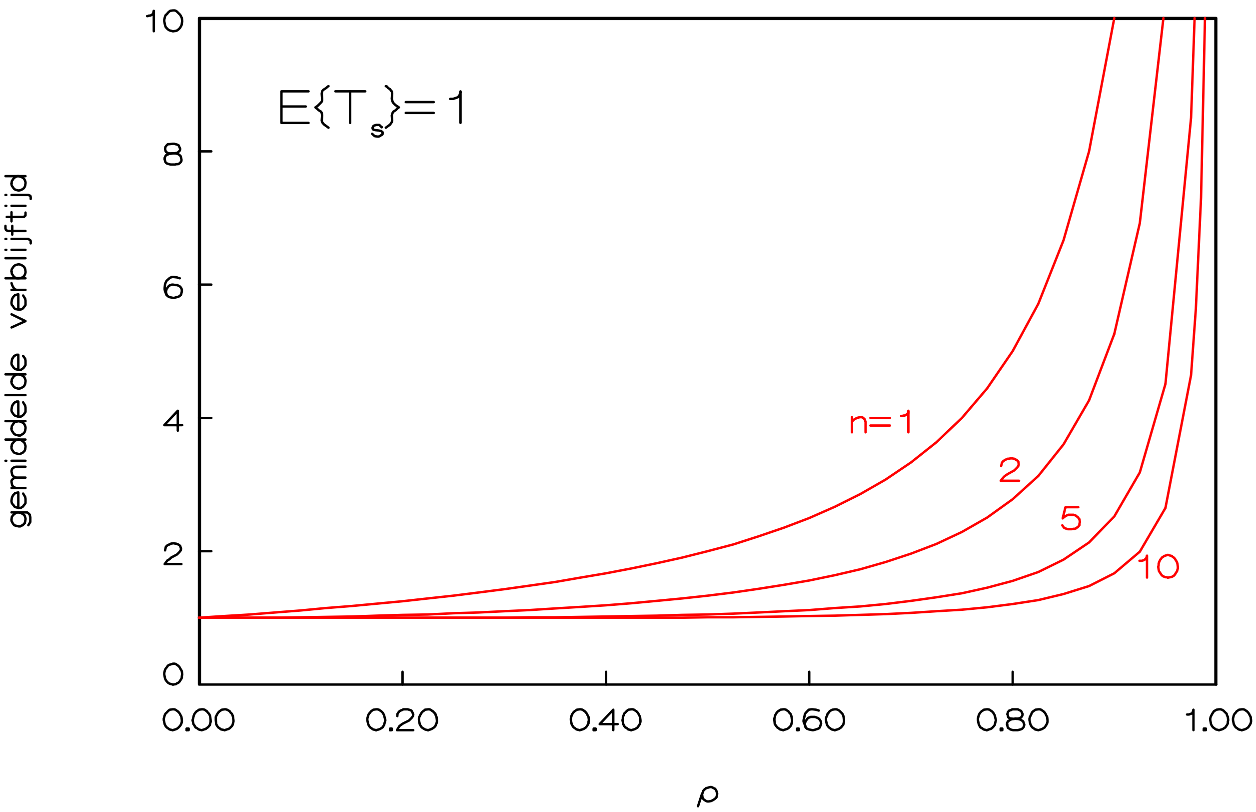
Voorbeeld 5.3

Stel dat we, in plaats van één processor met verwerkingstijd 0.5 seconde (zoals in voorbeeld 5.2), het systeem ook kunnen implementeren met twee goedkope processoren met een gemiddelde verwerkingstijd van 1 seconde per taak.

Om deze optie van twee langzame processoren te vergelijken met de optie van één snelle processor, stellen we weer dezelfde vragen.

Gevraagd:

a) Hoe hangt de gemiddelde responstijd van het systeem (de gemiddelde verblijftijd van een klant) af van de aankomstintensiteit λ?



Figuur 5.4: Gemiddelde verblijftijd van een klant in een $M/M/n$ wachtrij. De tijdseenheid is gelijk aan de gemiddelde bedieningsduur.

b) Hoeveel taken per seconde kunnen er verwerkt worden als de gemiddelde responstijd 2.5 seconden mag zijn?

c) Hoeveel neemt de responstijd toe als aankomstintensiteit van taken met 10% toeneemt?

De uitwerking van formule (5.34) voor de gemiddelde verblijftijd (wachttijd + verwerkingstijd) vraagt eerst om uitwerking van (5.25) en (5.32). Voor de $M/M/2$ wachtrij levert dit:

$$w_{\text{som}} = 1 + 2\rho + \frac{4\rho^2}{2(1-\rho)} \tag{5.35}$$

$$C(2,\rho) = \frac{2\rho^2}{1+\rho} \tag{5.36}$$

$$\mathrm{E}\,[T_q] = \frac{1}{\mu}\frac{1}{1-\rho^2} \tag{5.37}$$

Het gegeven $\mathrm{E}\,[T_s] = 1$ seconde impliceert: $\mu = 1$ sec^{-1}. Voor de $M/M/2$ wachtrij is $\rho = \lambda/2\mu$. De gemiddelde verblijftijd uitgedrukt in λ is dan:

$$\mathrm{E}\,[T_q] = \frac{4}{4-\lambda^2} \tag{5.38}$$

Met het gegeven $\mathrm{E}[T_q] = 2.5$ seconden, vinden we $\lambda = 1.549$. Het antwoord op b) is dat het systeem 1.549 taken per seconde kan verwerken.

Indien λ toeneemt tot $\lambda = 1.704$ dan wordt de gemiddelde responstijd $4/(4 - 2.904) = 3.65$ seconde. Ten opzichte van 2.5 seconden is dit een toename van 46%

5.6 M/D/1 wachtsysteem

Het $M/D/1$ wachtsysteem kenmerkt zich door een constante bedieningsduur bij het loket. Het aankomstproces is (nog steeds) een Poisson-proces. De bedieningsduur schrijven we weer als T_s, maar T_s is nu een gedegenereerde stochastische variabele die altijd dezelfde waarde aanneemt. Om de overeenkomst te handhaven met de $M/M/1$ wachtrij, is die vaste waarde $1/\mu$:

$$T_s = \frac{1}{\mu} \tag{5.39}$$

Het aankomstproces is een Poisson-proces met intensiteit λ (klanten per tijdseenheid). De bezettingsgraad van de $M/D/1$ wachtrij is het product van het aantal klanten per tijdseenheid en het aantal tijdseenheden bedieningsduur per klant:

$$\rho = \frac{\lambda}{\mu} \tag{5.40}$$

Als we de toestand van een $M/D/1$ wachtrij op een willekeurig moment zouden moeten beschrijven, dan moet, naast het aantal klanten in het systeem, ook vermeld worden hoever de bediening van de klant bij het loket gevorderd is, omdat nu de bedieningsduur niet een geheugenloze verdeling heeft. Het lijkt daarom dat de eenvoudige Markov-keten beschrijving zoals we die kennen van de $M/M/1$ wachtrij hier niet opgaat. Toch is er een eenvoudig model mogelijk, namelijk door te kijken naar het aantal klanten in het systeem op discrete tijdstippen. We kiezen deze tijdstippen op onderling gelijke afstand van precies één bedieningsduur:

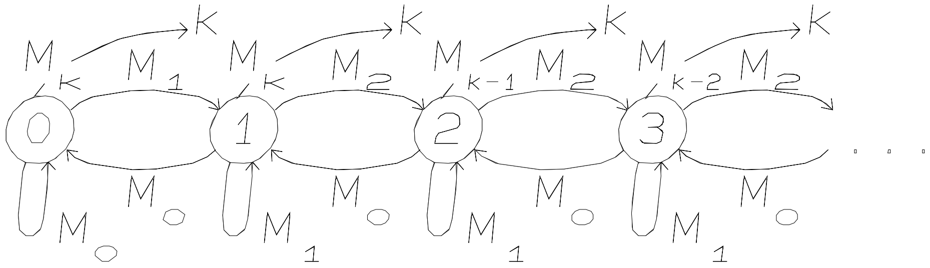
$$t = 0,\ \frac{1}{\mu},\ \frac{2}{\mu},\ \frac{3}{\mu}, \ldots \tag{5.41}$$

Als t de bovenstaande waarden doorloopt dan is het proces $N_q(t)$ een discrete tijd Markov-keten. Ondanks dat niet gegeven is hoeveel van de bedieningsduur verstreken is van de klant bij het loket (als er eentje is) is het wel zeker dat een eventuele bediening, die gaande is op een bepaald tijdstip, afgelopen zal zijn op het volgende discrete tijdstip. Ook is zeker dat er tussen twee tijdstippen van de rij (5.41) er nooit meer dan één bediening beëindigd zal worden.

Het toeval in de overgangen van de discrete-tijd Markov-keten $N_q(t)$ wordt bepaald door het aankomstproces. Het aantal aankomsten in een tijdsinterval van lengte $1/\mu$ is Poisson verdeeld met gemiddelde $\lambda/\mu = \rho$. De kans op k aankomsten M_k wordt dan gegeven door de bekende formule:

$$M_k = e^{-\rho}\frac{\rho^k}{k!} \tag{5.42}$$

Als er op een bepaald tijdstip i klanten in het systeem zijn ($i > 0$), dan zijn er met kans M_k het volgende discrete tijdstip $i - 1 + k$ klanten in het systeem. Dit geeft de overganskansen vanuit de toestand i ($i > 0$). Vanuit toestand 0 wordt met kans M_k een sprong gemaakt naar toestand k. Vanuit elke toestand gaat dus een hele waaier van overgangskansen naar rechts in het toestandsdiagram. In figuur 5.5 is dit aangegeven met een onderbroken pijl. Bijvoorbeeld vanuit toestand 3 zijn overgangen mogelijk naar toestand 2 met kans M_0, naar zichzelf met kans M_1, naar 4 met kans M_2, en zo verder, naar k met kans M_{k-2}. Dit laatste staat in het diagram als $M_{k-2} \longrightarrow k$.



Figuur 5.5: Toestandsdiagram discrete tijd Markov-keten bij $M/D/1$ model.

Voor het opschrijven van de evenwichtskansen van de discrete-tijd Markov-keten is het handig om Z_k te definiëren als de kans op k of meer aankomsten:

$$Z_k = \sum_{i=k}^{\infty} M_i = 1 - \sum_{i=0}^{k-1} M_i \tag{5.43}$$

De evenwichtskans om 0 klanten in het systeem te treffen is 1 minus de kans dat het loket bezet is:

$$v_0 = 1 - \rho \tag{5.44}$$

Het bepalen van de overige evenwichtskansen gaat met een soortgelijk evenwichtsargument als in het geval dat er alleen overgangen naar buurtoestanden mogelijk zijn (zie pagina 41). Denk een verticale lijn tussen toestand 0 en 1. De kans dat een willekeurige overgang er een is die van rechts naar links over die lijn gaat, is gelijk aan $M_0 v_1$. De kans op een overgang over die lijn van links naar rechts is $(M_1 + M_2 + \cdots)v_0 = Z_1 v_0$. Uit de evenwichtsveronderstelling volgt dan:

$$v_1 = \frac{Z_0}{M_0} v_0 \tag{5.45}$$

Plaats nu een denkbeeldige lijn tussen 1 en 2. De kans op een sprong naar links, over de lijn, is $M_0 v_2$. Een sprong naar rechts, over de lijn, kan komen uit toestand 0 (kans $M_2 + M_3 + \cdots$) of uit toestand 1 (kans $M_2 + M_3 + \cdots$). In geval van evenwicht is een sprong naar rechts even waarschijnlijk als een sprong naar links, dus:

$$v_2 = \frac{Z_2}{M_0}(v_0 + v_1) \tag{5.46}$$

Zo kan de redenering voortgezet worden. We schrijven hier de resultaten tot en met v_4 op:

$$v_3 = \frac{1}{M_0}(Z_2 v_2 + Z_3(v_0 + v_1)) \tag{5.47}$$

$$v_4 = \frac{1}{M_0}(Z_2 v_3 + Z_3 v_2 + Z_4(v_0 + v)1)) \tag{5.48}$$

De verdelingsfunctie van de wachttijd wordt gegeven door een redelijk complexe uitdrukking die we hier in zijn algemeenheid niet kunnen behandelen. Echter als we ons beperken tot die waarden van de verdelingsfunctie waar t een veelvoud van de bedieningsduur is, dan blijkt dat er slechts nog een simpele optelling van bovenstaande kansen van de evenwichtsverdeling nodig is.

De reden is dat de wachttijd van een nieuwe klant, die k of minder klanten aantreft in het systeem, altijd kleiner is dan k maal de bedieningsduur $1/\mu$. Omgekeerd volgt uit het feit dat de wachttijd voor een "test-klant" kleiner is dan k/μ dat het aantal klanten dat deze klant in het systeem treft kleiner of gelijk aan k moet zijn. Als we dit combineren met het PASTA (Poisson Arrivals See Time Averages) resultaat dan kunnen we compact in formule notatie schrijven:

$$\Pr\{T_w \leq \tfrac{k}{\mu}\} = \Pr\{N_q \leq k\} = \sum_{k=0}^{k} v_i \qquad k = 0, 1, 2, \ldots \tag{5.49}$$

Met een constante bedieningsduur is de verblijftijd van een klant in het systeem altijd precies $1/\mu$ langer dan de wachttijd. De verdelingsfunctie van de verblijftijd is dan:

$$\Pr\{T_q \leq \tfrac{k}{\mu}\} = \Pr\{N_q \leq k-1\} = \sum_{k=0}^{k-1} v_i \qquad k = 1, 2, 3, \ldots \tag{5.50}$$

De gemiddelde verblijftijd of wachttijd is niet direct af te leiden uit (5.49) respectievelijk (5.50) omdat de verdelingsfuncties alleen gegeven zijn voor veelvouden van de bedienings-duur, en bovendien zijn die formules alleen praktisch indien k niet te groot is. Voor de gemiddelde verblijftijd / wachttijd maken we gebruik van een algemeen resultaat dat in de volgende paragraaf gegeven wordt. In het geval van de $M/D/1$ wachtrij levert dit de volgende formules op:

$$\mathrm{E}[T_w] = \frac{\rho}{2(1-\rho)}\frac{1}{\mu} \tag{5.51}$$

$$\mathrm{E}[T_q] = \frac{2-\rho}{2(1-\rho)}\frac{1}{\mu} \tag{5.52}$$

Voorbeeld 5.4

Een processor verwerkt taken met een verwerkingstijd (bedieningsduur) van 100 ms per taak. De toegestane bezettingsgraad van de processor, $\rho_{\max}$, wordt bepaald door de eis dat, met slechts 10% kans, de responstijd langer dan 300 ms mag zijn.

Aanvankelijk had men $\rho_{\max}$ berekend volgens het $M/M/1$ model (dus uitgaande van een negatief exponentieel verdeelde bedieningsduur). Dit gaf een waarde van $\rho_{\max}$ die in de praktijk veel te laag bleek. Toen merkte men op dat de bedieningsduur geen toevalsgrootheid is, maar een deterministische waarde heeft. Het $M/D/1$ model is dus van toepassing. De aanname van een Poisson-aankomstproces (de eerste M) blijft wel gerechtvaardigd zolang de aankomsten gegenereerd worden door een groot aantal onafhankelijke bronnen.

Gevraagd: Wat is de waarde van $\rho_{\max}$,

a) volgens het $M/M/1$ model?

b) volgens het $M/D/1$ model?

We kiezen onze tijdseenheid gelijk aan de (gemiddelde) bedieningsduur $T_s = 100$ ms Dan is $\mu = 1$, en $\lambda = \rho$. De eis is:

$$\Pr\{T_q > 3\} = 0.1 \tag{5.53}$$

Voor het $M/M/1$ model halen we deze overschrijdingskans uit de verdelingsfunctie voor de verblijftijd (5.18):

$$\Pr\{T_q > 3\} = e^{-(1-\lambda)3} = 0.1 \tag{5.54}$$

Hieruit kunnen we λ oplossen; deze waarde van λ geeft dan (vanwege $\mu = 1$) ook de maximaal toegestane waarde voor ρ:

$$\rho_{\max} = \frac{3 - \ln(10)}{3} = 0.232 \tag{5.55}$$

Voor de $M/D/1$ wachtrij, wordt 1 minus de overschrijdingskans gegeven door formule (5.50). Hieruit volgt dat de kans op een verblijftijd groter dan 3 maal de bedieningsduur gelijk is aan:

$$\Pr\{T_q > 3\} = 1 - v_0 - v_1 - v_2 \tag{5.56}$$

Om de drie evenwichtskansen te berekenen passen we de formules (5.42)–(5.47) toe. Daarmee schrijven we achtereenvolgens op:

$$\begin{array}{rclcrcl}
M_0 & = & e^{-\rho} & \quad & Z_0 & = & 1 \\
M_1 & = & e^{-\rho}\rho & & Z_1 & = & 1 - e^{-\rho} \\
M_2 & = & e^{-\rho}\rho^2/2 & & Z_2 & = & 1 - e^{-\rho}(1+\rho) \\
v_0 & = & 1 - \rho & & & & \\
v_1 & = & (e^{\rho} - 1)(1 - \rho) & & & & \\
v_2 & = & (e^{\rho} - 1 - \rho)e^{\rho}(1 - \rho) & & & &
\end{array} \tag{5.57}$$

Bij het sommeren van v_0, v_1 en v_2, vallen sommige termen tegen elkaar weg; met nog wat algebra vinden we dan

$$v_0 + v_1 + v_2 = e^{\rho}(e^{\rho} - \rho)(1 - \rho) = 0.9 \tag{5.58}$$

Het rechterlid, 0.9, komt van de eis $\Pr\{T_q > 3\} = 0.1$. De maximaal toegestane bezettingsgraad, $\rho_{\max}$, is de oplossing van bovenstaande vergelijking. De oplossing van deze zogeheten transcendente vergelijking is niet in elementaire functies uit te drukken. We hebben daarom $\rho_{\max}$ bepaald met een binary search. Deze zoek methode is eenvoudig te programmeren en kan in een paar regels Pascal opgeschreven worden als:

```
left := 0; right := 1;
for i := 1 to 20 do
begin
  rho := (left + right)/2;
  if exp(rho) * (exp(rho) - rho) * (1 - rho) > 0.9
  then
    left := rho
  else
    right := rho
end;
```

`rho` convergeert dan naar de gevraagde oplossing:

$$\rho_{\max} = 0.587 \tag{5.59}$$

5.7 M/G/1 wachtsysteem

Voor dit wachtsysteem, waarbij de bedieningsduur een willekeurige kansverdeling kan hebben, beperken wij ons tot het presenteren van de Pollaczek-Khinchine-formule voor de gemiddelde wachttijd.

Van de bedieningsduurverdeling hoeft alleen het gemiddelde, $\mathrm{E}[T_s]$, en de coëfficiënt van variatie gegeven te zijn (ofwel het eerste en tweede moment van de verdeling, want het gemiddelde is m_1 en het kwadraat van de coëfficiënt van variatie, $C_s^2 = (m_2 - m_1^2)/m_1^2$).

De gemiddelde wachttijd is:

$$\mathrm{E}[T_w] = \frac{\rho}{2(1-\rho)}(1 + C_s^2)\mathrm{E}[T_s] \tag{5.60}$$

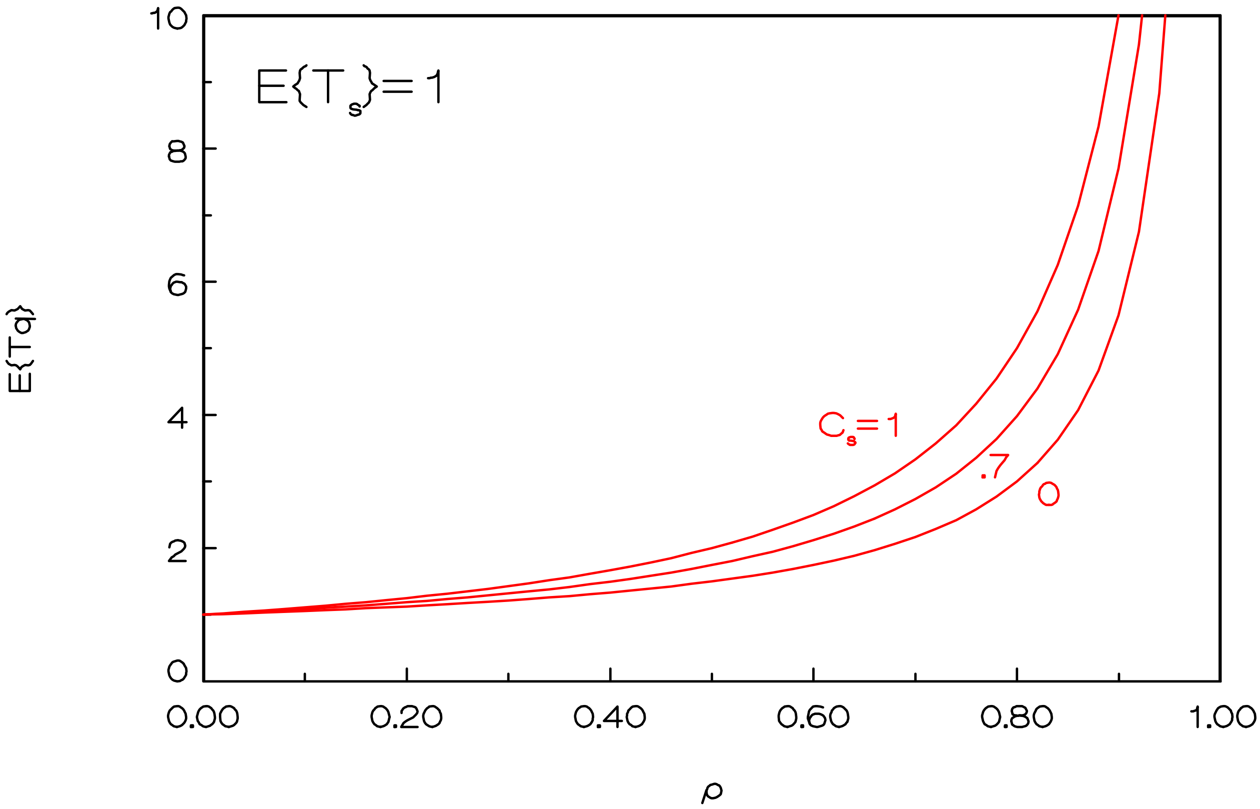
Hieruit blijkt hoe de gemiddelde wachttijd bepaald wordt door de coëfficiënt van variatie van de bedieningsduurverdeling. In het geval van een constante bedieningsduur is $C_s^2 = 0$; dit geeft ons weer formule (5.51). Bij een negatief exponentieel verdeelde stochastische variabele is $C_s^2 = 1$. Vergelijken we de $M/M/1$ en de $M/D/1$ wachtrij dan is de factor $(1 + C_s^2) = 2$ respectievelijk 1. Dat impliceert een verdubbeling van de gemiddelde wachttijd voor de $M/M/1$ wachtrij ten opzichte van $M/D/1$.

De gemiddelde verblijftijd volgt direct uit (5.60) door daar de gemiddelde bedieningsduur bij te tellen:

$$\mathrm{E}[T_q] = \left[\frac{\rho}{2(1-\rho)}(1 + C_s^2) + 1\right]\mathrm{E}[T_s] \tag{5.61}$$

Het is eenvoudig te verifiëren dat substitutie van 1 voor C_s^2 in bovenstaande formule, formule (5.16) oplevert. Nota bene $\mathrm{E}[T_s] = 1/\mu$.

In figuur 5.6 is een grafiek van (5.61) voor verschillende waarden van C_s getekend. De twee extremen komen overeen met de gemiddelde verblijftijd voor respectievelijk het $M/M/1$ wachtsysteem ($C_s = 1$) en het $M/D/1$ wachtsysteem ($C_s = 0$). Hier tussenin ligt de gemiddelde verblijftijd van een wachtsysteem waarbij de coëfficiënt van variatie van de bedieningsduur gelijk is aan 0.7, dus $C_s^2 = 0.49$.



Figuur 5.6: Gemiddelde verblijftijd voor het $M/G/1$ wachtsysteem.

Voorbeeld 5.5

Twee experts, Arie en Bernard, zijn het niet eens over hoe ze berichten zullen versturen over een 64 kbit/s link. De volgende berichten komen voor:

lengte	informatie-octetten	aantal aangeboden berichten/s
kort	16	200
lang	64	25

Elk bericht heeft ook nog 2 overhead octetten.

Arie wil de lange berichten opdelen in 4 korte berichten, dat geeft volgens hem een korte gemiddelde wachttijd omdat de berichten kort zijn en allemaal van dezelfde lengte. Bernard vind dat geen goed idee omdat er dan meer overhead is (8 octetten voor een lang bericht).

Ze besluiten de twee opties (A: opsplitsen, B: niet opsplitsen) te vergelijken op basis van de geïdealiseerde modellen $M/D/1$ en $M/G/1$.

Gevraagd: Wat is de gemiddelde wachttijd van berichten, voordat ze op de lijn gezet kunnen worden, volgens het $M/D/1$ respectievelijk het $M/G/1$ model?

optie A:
De berichten worden verzonden in pakketten van 18 octetten lang. De aankomstintensiteit van pakketten is

$$\lambda_A = 200 + 4 \times 25 = 300 \text{ s}^{-1} \tag{5.62}$$

De bedieningsduur is de tijd nodig om een pakket op de lijn te zetten. Dit is het aantal bits in een pakket gedeeld door de bitsnelheid van het kanaal. Dit is

$$\frac{1}{\mu_A} = \frac{(16+2)8}{64000} = 2.25 \text{ ms} \tag{5.63}$$

De bezettingsgraad van het kanaal is dan $\rho_A = \lambda_A/\mu_A$:

$$\rho_A = 0.675 \tag{5.64}$$

Volgens (5.51) is dan de gemiddelde wachttijd van een pakket:

$$\mathrm{E}\left[T_{w_A}\right] = \frac{0.675}{2(1-0.0675)} 2.25 = 2.34 \text{ ms} \tag{5.65}$$

optie B:
De berichten worden verzonden in pakketten met lengtes van 18 en 66 octetten. De aankomstintensiteit van pakketten is:

$$\lambda_B = 200 + 25 = 225 \text{ s}^{-1} \tag{5.66}$$

De bedieningsduur is nu op te vatten als een toevalsgrootheid die met kans 200/225 gelijk is aan de tijd nodig om een kort pakket op de lijn te zetten en met kans 25/225 gelijk aan de tijd nodig om een lang pakket op de lijn te zetten. De gemiddelde bedieningsduur is:

$$\frac{1}{\mu_B} = \frac{200}{225}\frac{(16+2)8}{64000} + \frac{25}{225}\frac{(64+2)8}{64000} = 2.91667 \text{ ms} \tag{5.67}$$

De bezettingsgraad van het kanaal is dan $\rho_B = \lambda_B/\mu_B$:

$$\rho_B = 0.65625 \tag{5.68}$$

Voor het berekenen van de gemiddelde wachttijd hebben we nog het kwadraat van de coëfficiënt van variatie, C_s^2, nodig. Met (2.27) en (2.34) berekenen we dit uit het eerste en tweede moment van de bedieningsduurverdeling:

$$m_2 = \left(\frac{200}{225}\frac{144^2}{64^2} + \frac{25}{225}\frac{528^2}{64^2}\right) \times 10^{-6} = 12.0625 \times 10^{-6} \text{ s}^2 \tag{5.69}$$

$$C_s^2 = \frac{12.0625 - 2.91667^2}{2.91667^2} 0.417959 \tag{5.70}$$

Volgens (5.60) is dan de gemiddelde wachttijd:

$$\mathrm{E}\left[T_w B\right] = \frac{0.65625}{2(1-0.65625)} 1.417959 \times 2.91667 \text{ ms} \tag{5.71}$$

$$= 3.95 \text{ ms} \tag{5.72}$$

In de praktijk is het vaak nog meer van belang om de overschrijdingskans van een bepaalde verblijftijd of wachttijd te kennen dan om te weten wat de gemiddelde verblijftijd of wachttijd is. Het kan bijvoorbeeld zijn dat de gemiddelde wachttijd voldoende klein is, maar dat men wil weten hoe groot de kans is dat een wachttijd van 10 maal de gemiddelde bedieningsduur overschreden wordt. Zowel gebruiker als producent van een systeem moet eisen kunnen stellen, of verifiëren, dat een bepaalde wachttijd, bijvoorbeeld, met slechts 1% kans overschreden wordt.

Voor het $M/M/1$ wachtsysteem kan de overschrijdingskans van de verblijftijd uit de verdelingsfunctie gehaald worden en bij het $M/D/1$ systeem volgen overschrijdingskansen van een geheel aantal maal de bedieningsduur uit de evenwichtsverdeling van de Markov-keten. Wat kan er nu gezegd worden van een $M/G/1$ wachtsysteem?

Het is gelukt om een eenvoudige benaderende formule te vinden voor de percentielen van de wachttijdverdeling van de $M/G/1$ wachtrij. Laat t_x het x-de percentiel van de wachttijdverdeling zijn. Per definitie is de overschrijdingskans van t_x gelijk aan $1 - x/100$. De benaderende formule is bijvoorbeeld voor t_{99} en $\rho > 0.9$ tot een paar promille nauwkeurig. In de gevallen die we geverifieerd hebben (alle met $C_s < 1$) is de afwijking ten opzichte van het exacte resultaat in de orde van een paar procent of minder als ρ en x niet te klein zijn, zeg $\rho > 0.7$ en $x > 90$.

De benadering voor het x-de percentiel t_x (de wachttijd die met kans $p = 1 - x/100$ overschreden wordt) is:

$$t_x \approx \frac{1}{2} \ln\left(\frac{1}{1 - x/100}\right)\left(\frac{1 + C_s^2}{1-\rho} - \frac{1}{2}\right) \mathrm{E}\left[T_s\right] \tag{5.73}$$

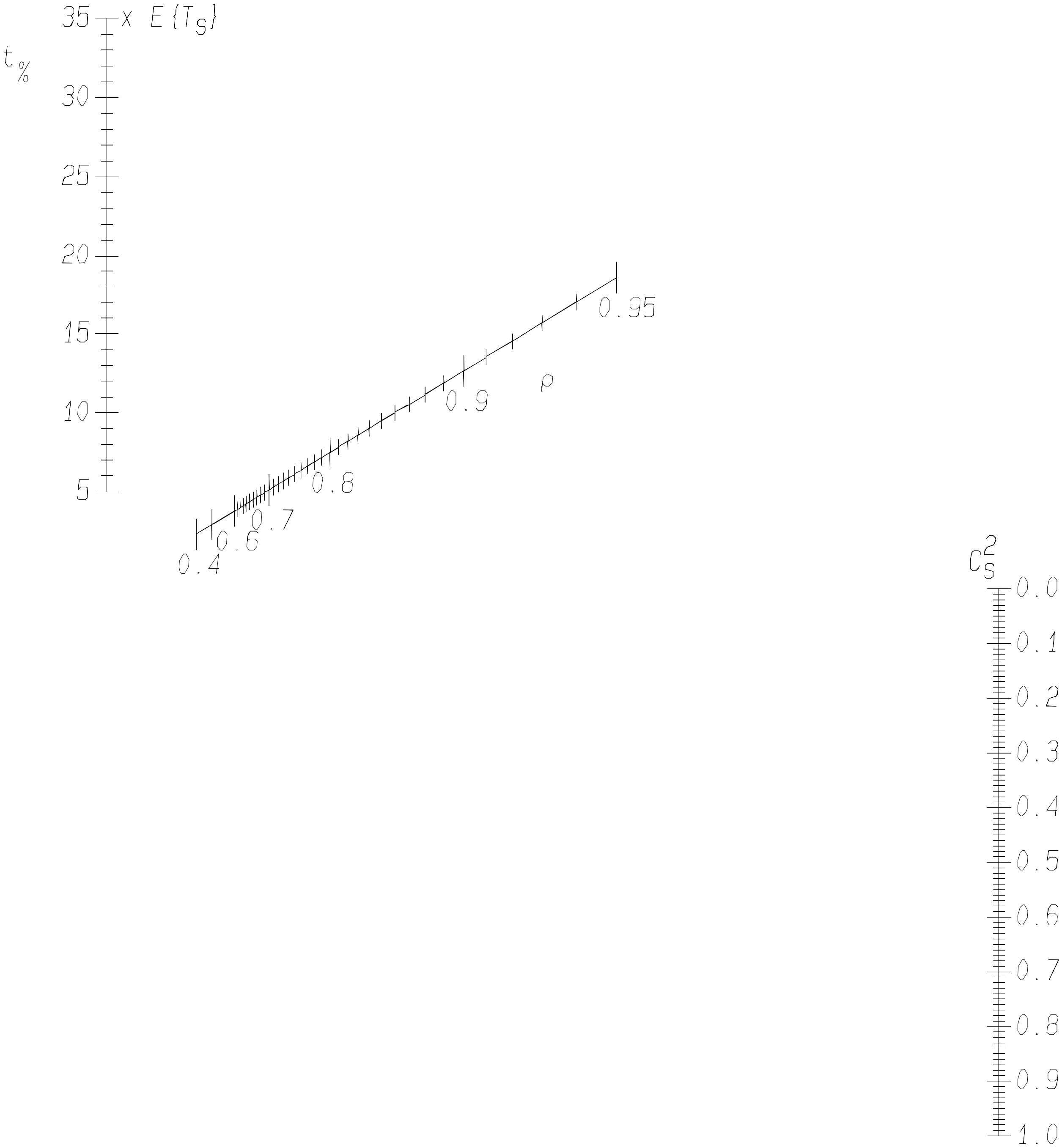
$$= \frac{1}{2} \ln(p)\left(\frac{1}{2} - \frac{1 + C_s^2}{1-\rho}\right) \mathrm{E}\left[T_s\right] \tag{5.74}$$

In figuur 5.7 is een nomogram gegeven waarmee eenvoudig vastgesteld kan worden hoe groot de toegestane bezettingsgraad is voor een loket als gegeven is dat een bepaalde wachttijd met 1% kans overschreden kan worden. In voorbeeld 5.6 wordt het gebruik van het nomogram nader toegelicht.

Voorbeeld 5.6

Gegeven een $M/G/1$ wachtsysteem met de eis dat de wachttijd met slechts 1% kans langer mag duren dan 10 maal de gemiddelde bedieningsduur.

Gevraagd:



Figuur 5.7: Het 99-ste percentiel van de wachttijdverdeling, $t_\%$, versus bezettingsgraad ρ, versus kwadraat van de coëfficiënt van variatie van de bedieningsduur C_s^2.

a) Als $C_s^2 = 0.5$, wat is dan de maximaal toegestane belasting $\rho_{\max}$?

b) Indien de belasting $\rho = \rho_{\max}$, waar ligt dan het 99-ste percentiel van de wachttijdverdeling voor een $M/D/1$ en voor een $M/M/1$ wachtsysteem?

De antwoorden kunnen uit het nomogram worden afgelezen. Bij a) is t_{99} gegeven ($10\mathrm{E}[T_s]$) en $C_s^2 = 0.5$. Trek een lijn van 10 in de linker kolom naar 0.5 in de rechter kolom. Deze snijdt de schuine balk in het punt

$$\rho_{\max} = 0.69 \tag{5.75}$$

Bij b) is ρ gegeven en $C_s^2 = 0$ voor $M/D/1$ en $C_s^2 = 1$ voor $M/M/1$. De lijn vanaf 0 in de rechter kolom, door 0.69 in de middenbalk komt uit bij:

$$t_{99} = 6.3\mathrm{E}[T_s] \qquad \text{voor } M/D/1 \tag{5.76}$$

De lijn vanaf 1 in de rechter kolom, door 0.69 in de middenbalk komt uit bij

$$t_{99} = 13.7\mathrm{E}[T_s] \qquad \text{voor } M/M/1 \tag{5.77}$$

5.8 M/G/1 Wachtsysteem met 2 prioriteiten

In deze paragraaf zullen we nog een aspect behandelen dat onder andere in computersystemen, met name bij de processor van een telecommunicatie systeem, veelvuldig toegepast wordt. Het gaat hier om het geven van prioriteit aan bepaalde taken. Het beheerssysteem (operating system) van een processor heeft meestal verschillende wachtrijen, voor elke prioriteit één. Zodra een taak beëindigd is, wordt eerst bij de hoogste prioriteit gekeken of daar nog een taak staat te wachten, daarna naar de lagere prioriteiten.

Men werkt met prioriteiten omdat sommige taken (klanten) meer vertraging kunnen velen dan andere. Verder is het zo dat het geven van prioriteit aan korte taken een gunstig effect heeft op de gemiddelde wachttijd.

Voor wachtsystemen met prioriteiten kunnen we hier niet zelf de resultaten afleiden met een eenvoudig Markov-model. We volstaan daarom met het geven van formules voor de gemiddelde wachttijd. We beperken ons tot een systeem met twee prioriteiten, omdat dat de notatie overzichtelijker maakt. Bovendien kunnen deze resultaten vaak ook toegepast worden in systemen met meer dan twee prioriteiten door respectievelijk alle hogere of alle lagere prioriteiten samen te voegen.

In een $M/G/1$ systeem met twee prioriteiten is het aankomstproces van klanten van elke prioriteit een Poisson-proces. We gebruiken de volgende notatie:

$$\lambda_i \qquad \text{aankomstintensiteit van klanten met prioriteit } i \tag{5.78}$$

Voor het nummeren van de prioriteiten gebruiken we de volgende conventie:

$$\begin{array}{ll} i = 1 & \text{voor de hoogste prioriteit} \\ i = 2 & \text{voor de laagste prioriteit} \end{array} \tag{5.79}$$

De bedieningsduur is voor elke prioriteit een stochastische variabele met zijn eigen kansverdeling. Zo kunnen we ook voor elke prioriteit de bezettingsgraad aangeven.

$$\begin{array}{ll} T_{s_i} & \text{bedieningsduur voor prioriteit } i \\ \rho_i = \lambda_i \mathrm{E}\left[T_{s_i}\right] & \text{bezettingsgraad door prioriteit } i \end{array} \tag{5.80}$$

We definiëren nog twee grootheden die in de uiteindelijke formules van pas komen. Deze zijn:

$$\mathrm{E}\left[T_a\right] = \frac{1}{2}\lambda_1 \mathrm{E}\left[T_{s_1}^2\right] \tag{5.81}$$

$$\mathrm{E}\left[T_b\right] = \frac{1}{2}\lambda_1 \mathrm{E}\left[T_{s_1}^2\right] + \frac{1}{2}\lambda_2 \mathrm{E}\left[T_{s_2}^2\right] \tag{5.82}$$

Hierin is $\mathrm{E}\left[T_a\right]$ de residutijd voor klanten van prioriteit 1. Dit is de tijd die een willekeurige aankomst van prioriteit 1 nog zou moeten wachten totdat de bediening die aan de gang is voltooid is, gewogen met de kans op wachten. Deze interpretatie van residutijd geldt in het geval dat er alleen klanten van prioriteit 1 zijn. $\mathrm{E}\left[T_b\right]$ is een som van twee residutijden.

Er zijn twee prioriteitsdisciplines te onderscheiden. De eerste wordt de Head Of Line (HOL) discipline genoemd. Hierbij wordt elke bediening zonder onderbreking voltooid. De tweede is de Preemptive Resume (PR) discipline. In dat geval wordt de bediening van een lagere prioriteit onderbroken bij aankomst van een hogere prioriteit en later weer hervat. Voor de gemiddelde wachttijd gelden de volgende formules:

Head Of Line discipline:

$$\mathrm{E}\left[T_{w_1}\right] = \frac{\mathrm{E}\left[T_b\right]}{1-\rho_1} \tag{5.83}$$

$$\mathrm{E}\left[T_{w_2}\right] = \frac{\mathrm{E}\left[T_b\right]}{(1-\rho_1)(1-\rho_1-\rho_2)} \tag{5.84}$$

Preemptive Resume:

$$\mathrm{E}\left[T_{w_1}\right] = \frac{\mathrm{E}\left[T_a\right]}{1-\rho_1} \tag{5.85}$$

$$\mathrm{E}\left[T_{w_2}\right] = \frac{\mathrm{E}\left[T_b\right]}{(1-\rho_1)(1-\rho_1-\rho_2)} + \frac{\rho_1}{1-\rho_1}\mathrm{E}\left[T_{s_2}\right] \tag{5.86}$$

Voorbeeld 5.7

Het voorbeeld 5.5 liet zien dat een kortere gemiddelde bedieningsduur leidt tot een kortere gemiddelde wachttijd (bij dezelfde of zelfs iets hogere belasting). Toch had Bernard (terecht) bezwaar tegen de conclusie dat dus het opsplitsen van lange pakketten wenselijk was. Er werd geen rekening gehouden met het feit dat een opgesplitst lang pakket vier maal een wachttijd zou kunnen ondervinden en de aanname van een Poisson-aankomstproces, bij optie A, was aanvechtbaar. Daarom stelt Bernard voor om te onderzoeken of het geven van voorrang aan korte pakketten niet tot een vergelijkbaar lage wachttijd leidt.

Gevraagd: Wat is de gemiddelde wachttijd voor berichten indien korte berichten voorrang krijgen?

Uit voorbeeld 5.5 lezen we de volgende gegevens:

$$\mathrm{E}\left[T_{s_1}\right] = \frac{18}{8000} = 2.25 \text{ ms} \qquad \lambda_1 = 200 \text{ s}^{-1} \tag{5.87}$$

$$\mathrm{E}\left[T_{s_2}\right] = \frac{66}{8000} = 8.25 \text{ ms} \qquad \lambda_2 = 25 \text{ s}^{-1} \tag{5.88}$$

Hieruit bepalen we de volgende waarden voor bezettingsgraad per prioriteitsklasse

$$\begin{array}{rcccl} \rho_1 & = & 200 \times 2.25 \times 10^{-3} & = & 0.45000 \\ \rho_2 & = & 25 \times 8.25 \times 10^{-3} & = & 0.20625 \end{array} \tag{5.89}$$

Voor de residu-tijden vinden we:

$$\mathrm{E}\left[T_a\right] = 0.5 \times 200 \times 2.252 \times 10^{-6} = 0.50625 \text{ ms} \tag{5.90}$$

$$\mathrm{E}\left[T_b\right] = 0.50625 + 0.5 \times 25 \times 8.25^2 \times 10^{-6} = 1.35703 \text{ ms} \tag{5.91}$$

Omdat de transmissie van een pakket (van lagere prioriteit) eerst afgemaakt zal worden, voordat een tussentijds aangekomen pakket van hogere prioriteit op de lijn gezet wordt, is hier de HOL prioriteitsdiscipline van toepassing. Invullen in de formules (5.84) levert dan:

$$\mathrm{E}\left[T_{w_1}\right] = \frac{1.35703}{0.55000} = 2.46733 \text{ ms} \tag{5.92}$$

$$\mathrm{E}\left[T_{w_2}\right] = \frac{1.35703}{0.55000(1 - 0.45000 - 0.20625)} \text{ ms} \tag{5.93}$$

De gemiddelde wachttijd wordt verkregen door bovenstaande gemiddelde wachttijden per discipline te wegen met gewichtsfactoren die overeenkomen met het relatieve voorkomen van de twee soorten pakketten. Dan is:

$$\mathrm{E}\left[T_w\right] = \frac{200}{225} 2.46733 + \frac{25}{225} 7.17769 = 2.99 \text{ ms} \tag{5.94}$$

5.9 Wachtsystemen met cyclische bediening

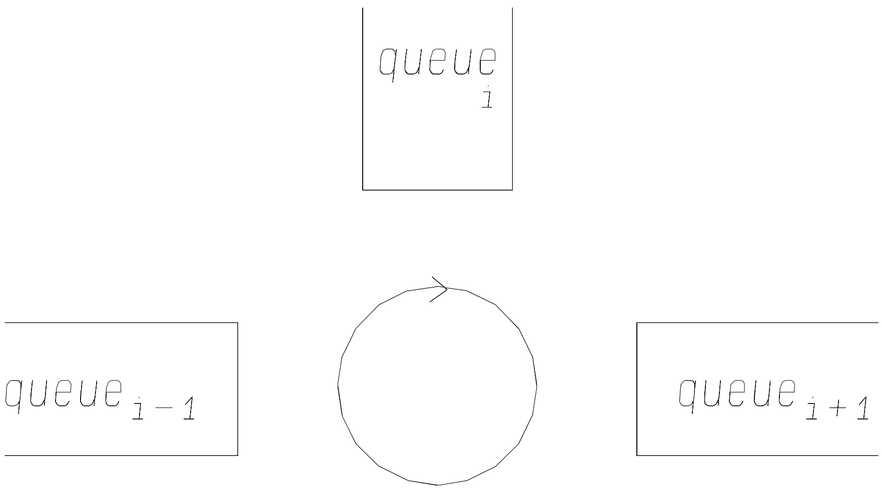
De toepassing van token-ring-protocollen in Local Area Networks heeft bijgedragen tot uitgebreide belangstelling voor wachtsystemen met cyclische bediening. Hierbij is er één server (het transmissie kanaal) en zijn er meer wachtrijen (de stations). De server loopt de wachtrijen cyclisch af om eventuele klanten te bedienen. (We vermijden het woord loket, omdat bij het spreken over “een loket dat de klanten afloopt,” de spraakverwarring compleet zal zijn).

Het model van cyclische bediening heeft ook toepassing in de analyse van een processor die verschillende stromen van taken verwerkt, zonder dat de ene stroom prioriteit heeft over de andere. Omdat we ook rekening houden met omschakeltijden kan het model ook gebruikt worden in een situatie waarbij de server na het beëindigen van een bediening niet direct aan een volgende klant kan beginnen maar eerst gedurende tenminste één omschakeltijd niet beschikbaar is.

Met een N-dimensionale toestandsruimte voor een systeem met N wachtrijen, wordt een exacte analyse van bijvoorbeeld de evenwichtsverdeling zeer complex. Alleen voor speciale gevallen, zoals $N = 2$ of een systeem waarbij alle N wachtrijen dezelfde parameters hebben, zijn er exacte oplossingen gevonden. Voor algemenere gevallen zij er diverse benaderende oplossingen.

In 1987 publiceerden Boxma en Meister in het tijdschrift Performance Analysis een goede benadering die practisch bruikbaar en vrij algemeen is. In deze sectie willen wij deze benadering presenteren en een toepassing geven. De benadering geeft een uitdrukking voor de gemiddelde wachttijd voor elke wachtrij.

Het model is geïllustreerd in figuur 5.8. De aankomsten van klanten bij elke wachtrij (queue) worden verondersteld een Poisson-proces te vormen. De aankomstintensiteit bij wachtrij i is λ_i. De bedieningsduur van een klant kan willekeurig verdeeld zijn, waarbij alleen het eerste moment β_i en het tweede moment $\beta_i^{(2)}$ gegeven hoeft te zijn. De index i geeft aan dat elke wachtrij zijn eigen bedieningsduurverdeling kan hebben.



Figuur 5.8: Nummero’s $i-1$, i en $i+1$ van N wachtrijen met cyclische bediening.

Met een omschakeltijd kan bijvoorbeeld voor een token-ring LAN de tijd gemodelleerd worden die het LAN nodig heeft om de permissie tot zenden (de token) door te geven van station

i naar station $i+1$. In het algemene model is de omschakeltijd een stochastische variabele. De tijd die verstrijkt na beëindiging van bediening bij queue i, totdat de volgende bediening bij queue $i+1$ begint is een stochastische variabele met gemiddelde s_i en variantie σ_i^2.

De wachtrijen zijn genummerd $1, 2, \ldots, N$. Na wachtrij i is wachtrij $i+1$ aan de beurt. De term $i+1$ moet ook zo opgevat worden dat na wachtrij N wachtrij 1 weer aan de beurt is; we hebben het immers over cyclische bediening.

We definiëren nog de volgende grootheden (uitgedrukt in aankomstintensiteit λ_i, gemiddelde bedieningsduur β_i en gemiddelde omschakeltijd s_i):

$$\begin{array}{rcll} \rho_i & = & \lambda_i\beta_i & \text{bezettingsgraad door queue } i \\ \rho & = & \sum_{i=1}^{N}\rho_i & \text{bezettingsgraad} \\ s & = & \sum_{i=1}^{N} s_i & \text{omschakeltijd per cyclus} \end{array} \tag{5.95}$$

De tijd die nodig is voor één volledige cyclus is een stochastische variabele die we aangeven met T_c. We kunnen beredeneren wat de verwachtingswaarde van T_c is. Namelijk, het gemiddelde aantal klanten dat per cyclus aankomt bij wachtrij i is $\mathrm{E}[T_c]\,\lambda_i$. De hoeveelheid werk die dit met zich meebrengt is $\mathrm{E}[T_c]\,\lambda_i\beta_i$. Zoveel tijd moet de server ook gemiddeld per bezoek aan wachtrij i besteden. Daarna volgt een omschakeltijd van gemiddeld s_i. Voor de hele cyclus sommeren we dit van $i=1$ tot en met $i=N$ om de gemiddelde cyclustijd te krijgen:

$$\mathrm{E}[T_c] = \mathrm{E}[T_c]\,\lambda_1\beta_1 + s_1 + \tag{5.96}$$

$$\mathrm{E}[T_c]\,\lambda_2\beta_2 + s_2 + \cdots \tag{5.97}$$

$$= \mathrm{E}[T_c]\,\rho + s \tag{5.98}$$

Hieruit kunnen we $\mathrm{E}[T_c]$ oplossen:

$$\mathrm{E}[T_c] = \frac{s}{1-\rho} \tag{5.99}$$

Zolang de evenwichtsverdeling bestaat, is dit resultaat algemeen geldig. Het is onafhankelijk van de wijze van bedienen: exhaustive (alle aanwezige klanten bij wachtrij i worden in één bezoek van de server bediend) of non-exhaustive (ten hoogste één klant per wachtrij wordt bediend per bezoek van de server).

De evenwichtsverdeling bestaat alleen indien voor elke wachtrij het aanbod van klanten kleiner is dan wat er per cyclus verwerkt wordt. De voorwaarde voor het bestaan van de evenwichtsverdeling kan inzichtelijk gemaakt worden voor het model dat de server doorgaat naar de volgende wachtrij nadat een klant van een wachtrij bediend is (non-exhaustive service). In dat geval moet voor elke wachtrij de verwachtingswaarde van het aantal klanten dat aankomt per cyclus kleiner dan 1 zijn: $\lambda_i\mathrm{E}[T_c] < 1$, voor elke waarde van i. Dit leidt direct tot de volgende evenwichtsvoorwaarde:

$$\max\{\lambda_i\}s < 1-\rho \tag{5.100}$$

De benaderingsformule van Boxma en Meister geldt voor non-exhaustive service. De formule is op het eerste gezicht wat indrukwekkend, maar, zoals we zullen zien zijn in de toepassingen sommige termen gelijk aan nul of zijn bijvoorbeeld alle termen van een sommatie aan elkaar gelijk.

De gemiddelde wachttijd bij queue i wordt gegeven door:

$$
\begin{aligned}
\mathrm{E}\left[T_{w_i}\right] &= \frac{1-\rho+\rho_i}{1-\rho-\lambda_i s} \frac{1-\rho}{(1-\rho)\rho+\sum_{j=1}^{N}\rho_j^2} \\
&\times \left[\frac{\rho}{2(1-\rho)}\sum_{j=1}^{N}\lambda_j\beta_j^{(2)} + \frac{\rho}{2s}\sum_{j=1}^{N}\sigma_j^2 + \frac{s}{2(1-\rho)}\sum_{j=1}^{N}\rho_j(1+\rho_j)\right] \qquad (5.101)
\end{aligned}
$$

Vaak zal men te maken hebben met modellen waarbij alle wachtrijen verondersteld worden dezelfde aankomstintensiteit en bedieningsduurverdeling te hebben. Als dan ook de omschakeltijd onafhankelijk is van i spreken we van een symmetrisch model. In het symmetrische geval is formule (5.101) niet meer een benadering, maar geeft exact de gemiddelde wachttijd.

In het asymmetrische geval is de formule een betere benadering naarmate alle wachtrijen verder van verzadiging zijn, d.w.z. naarmate voor alle i de waarden van het product $\lambda_i \mathrm{E}\left[T_c\right]$ verder van 1 ligt.

Voorbeeld 5.8

30 PC's moeten verbonden worden met een file-server voor het vastleggen van de transacties die met de PC gedaan worden. Men eist van de data-communicatie verbinding dat de gemiddelde responstijd (=wachttijd + transmissietijd) kleiner is dan 200 ms. De berichten die verzonden worden naar de file-server hebben een lengte van 1000 octetten. De tijd die verstrijkt tussen het genereren van twee opeenvolgende berichten vanuit één PC veronderstellen we negatief exponentieel verdeeld. Gemiddeld genereert één PC 1 bericht per seconde.

Arie stelt voor om de PC's via één 1 Mbit/s token-ring met de file-server te verbinden. Ondanks de omschakeltijd van 5 ms tussen “einde transmissie” van de ene PC en “begin transmissie” van de volgende, zou dit snel genoeg moeten zijn.

Bernard wil elke PC een vaste 64 kbit/s verbinding geven met de file-server. “Want,” zegt hij, “die token-ring is nog niet eens 16 keer sneller dan zo'n vaste verbinding (dan praat ik nog niet over de omschakeltijden) dus kan je daarmee nooit 30 PC's aan.”

Gevraagd: Wat is voor de twee verschillende opties de gemiddelde responstijd?

Voor de optie van Arie (token-ring) passen we formule (5.101) toe. Eerst zetten we de benodigde parameters op een rijtje:

$$
\begin{aligned}
N &= 30 \\
\lambda_i &= 1 \text{ s}^{-1} \\
\beta_i &= 8 \text{ ms} \\
\beta_i^{(2)} &= 64 \times 10^{-6} \text{ s}^2 \\
s_i &= 5 \text{ ms} \\
s &= 150 \text{ ms} \\
\sigma_i^2 &= 0 \\
\rho_i &= 0.008 \\
\rho &= 0.240
\end{aligned}
$$

Invullen geeft dan:

$$
\begin{aligned}
\mathrm{E}\left[T_{w_i}\right] &= \frac{0.768}{0.610}\frac{0.76}{0.18432}\left(\frac{0.24}{1.52}1.92 \times 10^{-3} + \frac{0.15}{1.52}0.24192\right) && (5.102)\\
&= 126 \text{ ms} && (5.103)
\end{aligned}
$$

Bij deze wachttijd moeten we nog optellen de transmissie-tijd

$$\mathrm{E}\left[T_{s_i}\right] = 8 \text{ ms} \tag{5.104}$$

om te komen tot een responstijd van

$$\mathrm{E}\left[T_{q_i}\right] = 134 \text{ ms} \tag{5.105}$$

Voor de optie van Bernard is het $M/D/1$ model van toepassing voor elke PC. De transmissietijd wordt verkregen door de berichtlengte, 8000 bits, te delen door de transmissiesnelheid 64000 bit/s. Dit geeft:

$$\mathrm{E}\left[T_{s_i}\right] = 125 \text{ ms} \tag{5.106}$$

De responstijd wordt gegeven door (5.52) met voor ρ ingevuld de waarde van $\lambda_i \mathrm{E}\left[T_{s_i}\right] = 0.125$, en voor $1/\mu$ de waarde van $\mathrm{E}\left[T_{s_i}\right]$. Dit geeft:

$$\mathrm{E}\left[T_{q_i}\right] = 134 \text{ ms} \tag{5.107}$$

Hoofdstuk 6

Markov Verliessystemen

In dit hoofdstuk wordt weer het model van een continue-tijd Markov-keten toegepast, ditmaal om, onder andere, de achtergrond van de in de (telefoon-)verkeerstheorie bekende Erlang-formule te belichten. In het tweede deel van dit hoofdstuk zien we hoe met hetzelfde Markov-model, door iets andere parameters in te vullen, rekening gehouden kan worden met een eindige populatie van klanten die een oproep kunnen plaatsen.

6.1 De Erlang-formule

Voor het dimensioneren van een telefonie bundel is de Erlang-formule het basisgereedschap. De formule geldt voor elk verliessysteem met een beperkt aantal loketten (lijnen). De enige veronderstelling die gemaakt moet worden is dat het aankomstproces een Poisson-proces is; de bedieningsduur verdeling (houdtijd van een lijn) kan willekeurig verdeeld zijn.

Er moet wel worden opgemerkt dat de huidige tendens van integratie van spraak en data, waarbij we virtuele verbindingen hebben, afbreuk doet aan de algemene toepasbaarheid van de Erlang-formule als basisgereedschap voor het dimensioneren van een telefonie bundel. Bij het mengen van virtuele verbindingen met verschillende karakteristieken wordt het Erlang-model complexer, omdat onder andere ook rekening gehouden moet worden met het al dan niet actief zijn van een virtuele verbinding.

Echter het Erlang-model is van toepassing op elk systeem met een Poisson-aankomstenproces en een beperkt aantal loketten, waarbij voor "loketten" van alles ingevuld kan worden. Het feit dat de formule niet afhangt van de kansverdeling van de bedieningsduur heeft in grote mate bijgedragen aan de toepasbaarheid en het belang van de formule.

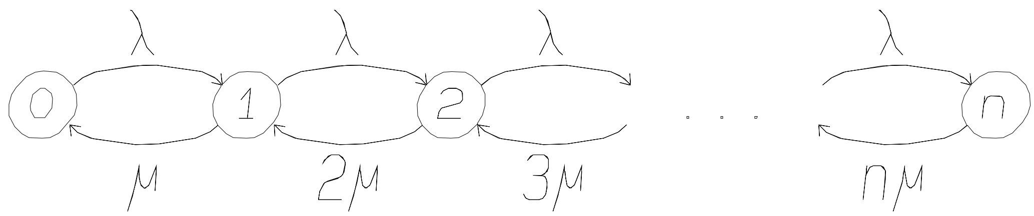
Zoals gezegd geldt de Erlang-formule voor een willekeurig $M/G/n/n$ systeem. De formule geeft de kans dat op een willekeurig moment (en vanwege PASTA voor een willekeurige aankomst) alle n loketten bezet zijn. De laatste n in $M/G/n/n$ slaat op het aantal plaatsen in het systeem: dit aantal is gelijk aan het aantal loketten. Klanten die geen vrij loket vinden (of in telefonie termen: oproepen die geen vrije lijn vinden) gaan verloren. Daarom heet dit een verliessysteem. De vijfde parameter is niet vermeld; dit betekent dat we te maken hebben met een oneindige populatie, wat weer inhoudt dat het aankomstproces een Poisson-proces is.

Voor de berekening van de blokkeringskans (de kans dat alle loketten bezet zijn) gaan we uit van het $M/M/n/n$ model. Dit model kan geanalyseerd worden met de inmiddels vertrouwde methode van het berekenen van de evenwichtsverdeling van een continue-tijd geboorte-sterfte-proces. Het is geen beperking dat we alleen kijken naar het $M/M/n/n$ systeem, omdat de evenwichtsverdeling van het $M/G/n/n$ verliessysteem niet afhankelijk is van de bedieningsduurverdeling. Dit is een opmerkelijk resultaat dat alleen geldt voor een puur verliessyteem. Dit resultaat kunnen we hier niet bewijzen.

Het toestandsdiagram van de continue-tijd Markov-keten is getekend in figuur 6.1. Hier zien we een constante aankomstintensiteit van λ aankomsten per tijdseenheid (gemiddeld) en een verwerkingsintensiteit die evenredig is met het aantal bezette lijnen. Een klant heeft een gemiddelde bedieningsduur van $1/\mu$ tijdseenheden. In telefonie termen wordt gezegd dat de gemiddelde houdtijd van een oproep $1/\mu = h$ tijdseenheden bedraagt. Het produkt van het aantal aankomsten per tijdseenheid en de gemiddelde houdtijd is de hoeveelheid aangeboden verkeer:

$$A = \frac{\lambda}{\mu} = \lambda h \tag{6.1}$$

Het verkeer wordt uitgedrukt in een dimensieloze eenheid die naar de wiskundige Erlang is genoemd. We zeggen dat het aangeboden verkeer A Erlang bedraagt.



Figuur 6.1: Toestandsdiagram van het $M/M/n/n$ verlies systeem (Erlang-model).

Met het bekende recept voor het berekenen van de evenwichts verdeling, lezen we voor de ongenormeerde kansen w_i uit het toestandsdiagram af dat:

$$w_0 = 1 \tag{6.2}$$

$$w_1 = A \tag{6.3}$$

$$w_2 = \frac{A^2}{2!} \tag{6.4}$$

$$w_i = \frac{A^i}{i!} \qquad i = 0, 1, 2, \ldots, n \tag{6.5}$$

De normeringsconstante is de som van de ongenormeerde kansen:

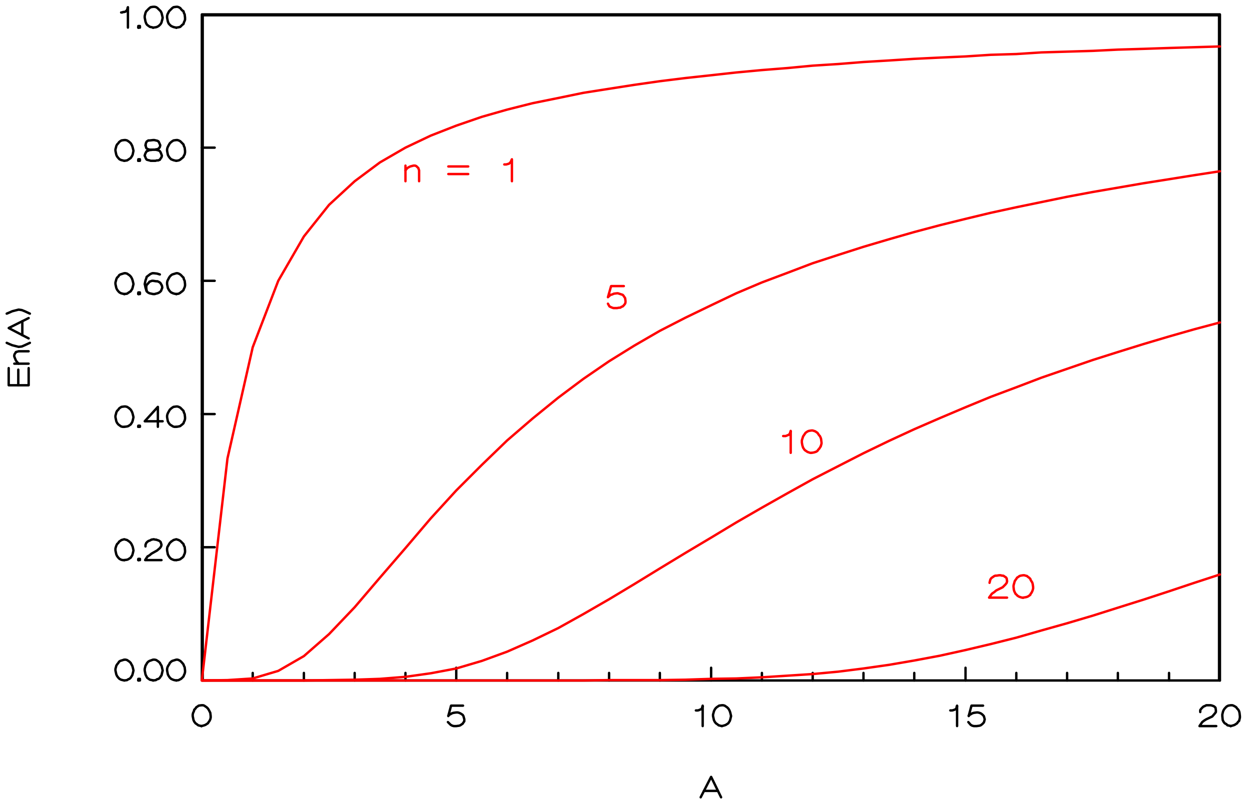
$$w_{\text{som}} = \sum_{i=0}^{n} \frac{A^i}{i!} \tag{6.6}$$

Waar het ons om gaat, is v_n, de kans dat alle lijnen bezet zijn. Vanwege de eigenschap: Poisson Arrivals See Time Averages (PASTA), is dit de blokkeringskans voor een oproep. Voor v_n is een hele reeks van alternatieve notaties in gebruik:

$$v_n = B(n, A) = E_{1,n}(A) = E_n(A) \tag{6.7}$$

De waarde van v_n wordt verkregen door w_n te delen door w_{som}. Dit geeft de beroemde Erlang-B-formule die, uitgeschreven, de volgende gedaante heeft:

$$E_n(A) = \frac{\frac{A^n}{n!}}{1 + \frac{A^1}{1!} + \cdots + \frac{A^n}{n!}} \tag{6.8}$$



Figuur 6.2: Blokkeringskans volgens het Erlang-model als functie van het aangeboden verkeer.

Voor het uitrekenen van $E_n(A)$ op een programmeerbare rekenmachine is er een handiger methode dan rechtstreeks via formule (6.8). Deze methode maakt gebruik van een eenvoudige, recursieve relatie voor $1/E_i(A)$. Definieer:

$$I_i(A) = \frac{1}{E_i(A)} \qquad i = 1, 2, \ldots \tag{6.9}$$

De recursie wordt begonnen door te constateren dat bij 0 lijnen de blokkeringskans gelijk aan 1 is, dus ook:

$$I_0(A) \;=\; 1 \tag{6.10}$$

$$I_i(A) \quad = \quad I_{i-1}(A)\frac{i}{A} + 1 \qquad i = 1, 2, \ldots \tag{6.11}$$

Voorbeeld 6.1

Bij een bedrijf met een PABX (Private Automatic Branch Exchange = Berijfstelefooncentrale) met vijf buitenlijnen blijkt dat in het drukke uur de blokkeringskans hoger is dan de gewenste waarde van 1%. Het aangeboden verkeer op de buitenlijnen in het drukke uur bedraagt 2.5 Erlang.

Gevraagd: Hoe groot is de blokkeringskans en hoeveel lijnen moeten erbij komen opdat de blokkeringskans kleiner is dan 1%?

De vraag is eenvoudig te beantwoorden met de recursieve relatie (6.11). Uitgaande van $A = 2.5$ Erlang en $I_0(2.5) = 1$ berekenen we achtereenvolgens

$$\begin{array}{rclcl} I_1(2.5) & = & 1\frac{1}{2.5} + 1 & = & 1.4 \\ I_2(2.5) & = & 1.4\frac{2}{2.5} + 1 & = & 2.12 \\ I_3(2.5) & = & 2.12\frac{3}{2.5} + 1 & = & 3.544 \\ I_4(2.5) & = & 3.544\frac{4}{2.5} + 1 & = & 6.6704 \\ I_5(2.5) & = & 6.6704\frac{5}{2.5} + 1 & = & 14.3408 \end{array} \tag{6.12}$$

Hieruit volgt direct de blokkeringskans van 2.5 Erlang op 5 lijnen:

$$E_5(2.5) = \frac{1}{14.3408} = 6.97\% \tag{6.13}$$

Om te komen tot een blokkeringskans van minder dan 1% moeten we doorgaan met de recursie totdat de inverse groter is dan 100:

$$\begin{array}{rclcl} I_6(2.5) & = & 14.3408\frac{6}{2.5} + 1 & = & 35.41792 \\ I_7(2.5) & = & 35.41792\frac{7}{2.5} + 1 & = & 100.170176 \end{array} \tag{6.14}$$

Dus 7 lijnen is voldoende en de blokkeringskans is dan:

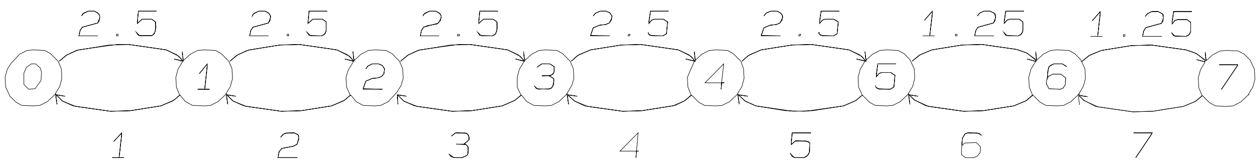
$$E_7(2.5) = \frac{1}{100.170176} = 1.00\% \tag{6.15}$$

Voorbeeld 6.2

Arie stelt voor om in plaats van twee (dure) buitenlijnen erbij te nemen, twee wachtplaatsen toe te voegen aan de PABX. Oproepen vanuit de PABX naar een buitenlijn, die alle buitenlijnen bezet vinden blijven dan wachten tot een buitenlijn vrij komt. Oproepen van buiten kunnen niet gebruik maken van de wachtplaatsen. We gaan, net als in voorbeeld 6.1, uit van 2.5 Erlang aangeboden verkeer, waarvan de helft gegenereerd wordt door oproepen vanuit de PABX (de andere helft door oproepen van buiten).

Gevraagd: Hoe groot is de blokkeringskans bij 2.5 Erlang aangeboden verkeer op 5 lijnen met 2 wachtplaatsen? Geef de blokkeringskans voor oproepen vanuit de PABX, die gebruik kunnen maken van de wachtplaatsen, en voor oproepen van buitenaf.

Het model lijkt op op het $M/M/5/7$ model, maar verschilt doordat alleen oproepen vanuit de PABX gebruik kunnen maken van de wachtplaatsen. Het toestands diagram van het aangepaste model is weergegeven in figuur 6.3:



Figuur 6.3: Het toestandsdiagram bij voorbeeld 6.2.

De toestand van het systeem is het aantal klanten dat een lijn belegd heeft plus eventuele wachtende klanten. De intensiteit waarmee lijnen vrij komen is gelijk aan het aantal bezette lijnen, dus nooit groter dan 5 per gemiddelde houdtijd (we kiezen onze tijdseenheid gelijk aan de gemiddelde houdtijd). Als het systeem in toestand 5 of 6 is, dan is de overgangsintensiteit naar een hogere toestand 1.25 per gemiddelde houdtijd, omdat die overgangen alleen plaatsvinden indien een oproep vanuit de PABX een wachtplaats bezet.

We berekenen eerst de evenwichts verdeling volgens (4.28)–(4.31)

$$\begin{array}{lllll}
w_0 & = & 1 & & \\
w_1 & = & 1\frac{2.5}{1}+1 & = & 2.5 \\
w_2 & = & 2.5\frac{2.5}{2}+1 & = & 3.125 \\
w_3 & = & 3.125\frac{2.5}{3}+1 & = & 2.604167 \\
w_4 & = & 2.604167\frac{2.5}{4}+1 & = & 1.627604 \\
w_5 & = & 1.627604\frac{2.5}{5}+1 & = & 0.813802 \\
w_6 & = & 0.813802\frac{1.25}{5}+1 & = & 0.203451 \\
w_7 & = & 0.203451\frac{1.25}{5}+1 & = & 0.050863 \\
w_{\text{som}} & = & 11.924887 & &
\end{array} \tag{6.16}$$

Voor oproepen vanuit de PABX is er alleen blokkering als het systeem in toestand 7 is. De kans hierop is (slechts):

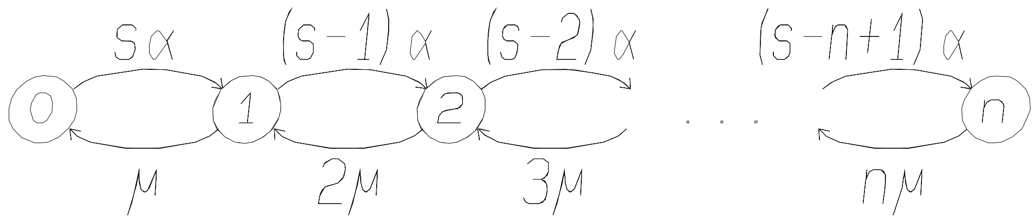
$$v_7 = \frac{0.050863}{11.924887} = 0.43\% \tag{6.17}$$

Voor oproepen van buiten is er blokkering als het systeem in toestand 5, 6 of 7 is. De kans hierop is:

$$v_5 + v_6 + v_7 = \frac{0.813802 + 0.203451 + 0.050863}{11.924887} = 8.96\% \tag{6.18}$$

6.2 De Engset-formule

Stel, we hebben een bundel van n lijnen en een klein aantal klanten. Elke klant kan één van de n lijnen beleggen. Indien één van de klanten een lijn belegd heeft, dan fungeert deze niet meer als een bron voor een nieuwe oproep, daarom neemt de aankomstintensiteit af als het aantal belegde lijnen toeneemt. In die zin onderscheidt het Engset-model zich van het Erlang-model. Bij Erlang is het aankomstproces een Poisson-proces met een constante aankomstintensiteit. Bij Engset is de aankomstintensiteit evenredig met het aantal vrije klanten.



Figuur 6.4: Toestandsdiagram van het $M/M/n/n/s$ verlies systeem (Engset-model).

Het Engset-model wordt eenvoudig als we een, inmiddels vertrouwd, toestandsdiagram tekenen voor het onderliggende Markov-proces dat het aantal bezette lijnen aangeeft. Zie figuur 6.4. Voor dit geboorte-sterfte-proces is de kans dat in een tijdsinterval Δt een lijn vrij komt, gegeven i bezette lijnen, $i\mu\Delta t$, net als voor het Erlang-model. Hierin is μweer de inverse van de gemiddelde houdtijd: $\mu = 1/h$. De kans dat in een tijdsinterval Δt een lijn bezet wordt, gegeven j vrije klanten, is $j\alpha\Delta t$. Hier is α de aankomstintensiteit per vrije klant, dat wil zeggen dat de tijd tot de eerstvolgende aankomst negatief exponentieel verdeeld verondersteld wordt met gemiddelde $1/(j\alpha)$. Indien het totaal aantal klanten gelijk aan s is en er zijn i lijnen bezet dan is het aantal vrije klanten $j = s - i$.

Het Engset-model past in de Kendall-classificatie en wordt genoteerd als het $M/M/n/n/s$ verliessysteem. De eerste M slaat op de negatief exponentieel verdeelde tussenaankomsttijden, de tweede M op de negatief exponentieel verdeelde houdtijd (bedieningsduur). De eerste n slaat op n lijnen (loketten) en de tweede n op het totaal aantal plaatsen in het systeem. Het is een puur verliessysteem omdat het aantal plaatsen in het systeem gelijk is aan het aantal lijnen. De laatste parameter slaat op het eindige totaal aantal klanten. De tussenaankomsttijden zijn negatief exponentieel verdeeld, maar de parameter van de negatief exponentiele verdeling verandert als de toestand van het proces verandert. Daarom noemen we het aankomstproces niet een Poisson proces.

Uit het toestandsdiagram kunnen de ongenormeerde evenwichtskansen afgelezen worden:

$$w_i = \binom{s}{i}\left(\frac{\alpha}{\mu}\right)^i \qquad i = 0, 1, \ldots, n \tag{6.19}$$

$$\alpha = \text{aankomstintensiteit per vrije klant} \tag{6.20}$$

De normeringsconstante is:

$$w_{\text{som}} = \sum_{i=0}^{n} \binom{s}{i} \left(\frac{\alpha}{\mu}\right)^i \tag{6.21}$$

De evenwichtsverdeling wordt weer gegeven door:

$$v_i = \frac{w_i}{w_{\text{som}}} \tag{6.22}$$

Om de blokkeringskans te berekenen zijn we in dit geval niet klaar met v_n. Vanwege het eindige aantal klanten is het aankomst proces geen Poisson-proces en we kunnen dus niet gebruik maken van de PASTA eigenschap. De blokkeringskans wordt verkregen door te kijken naar de verhouding van de kans op een blokkering in een willekeurig (maar klein) interval Δt en de kans op een aankomst in zo'n interval.

De kans op een aankomst in Δt is:

$$\sum_{i=0}^{n} \alpha(s-i) v_i \Delta t = \frac{\alpha s}{w_{\text{som}}} \sum_{i=0}^{n} \binom{s-1}{i} \left(\frac{\alpha}{\mu}\right)^i \Delta t \tag{6.23}$$

Het rechterlid werd verkregen door de kansen v_i in te vullen en gebruik te maken van de volgende, door uitschrijven te verifiëren, gelijkheid:

$$(s-i)\binom{s}{i} = s\binom{s-1}{i} \tag{6.24}$$

Een blokkering ontstaat alleen wanneer n lijnen bezet zijn en n van de $s-n$ vrije klanten een aankomst genereert. De kans dat dit gebeurt in een interval Δt is:

$$\alpha s(s-n) v_n \Delta t = \frac{\alpha s}{w_{\text{som}}} \binom{s-1}{n} \left(\frac{\alpha}{\mu}\right)^n \Delta t \tag{6.25}$$

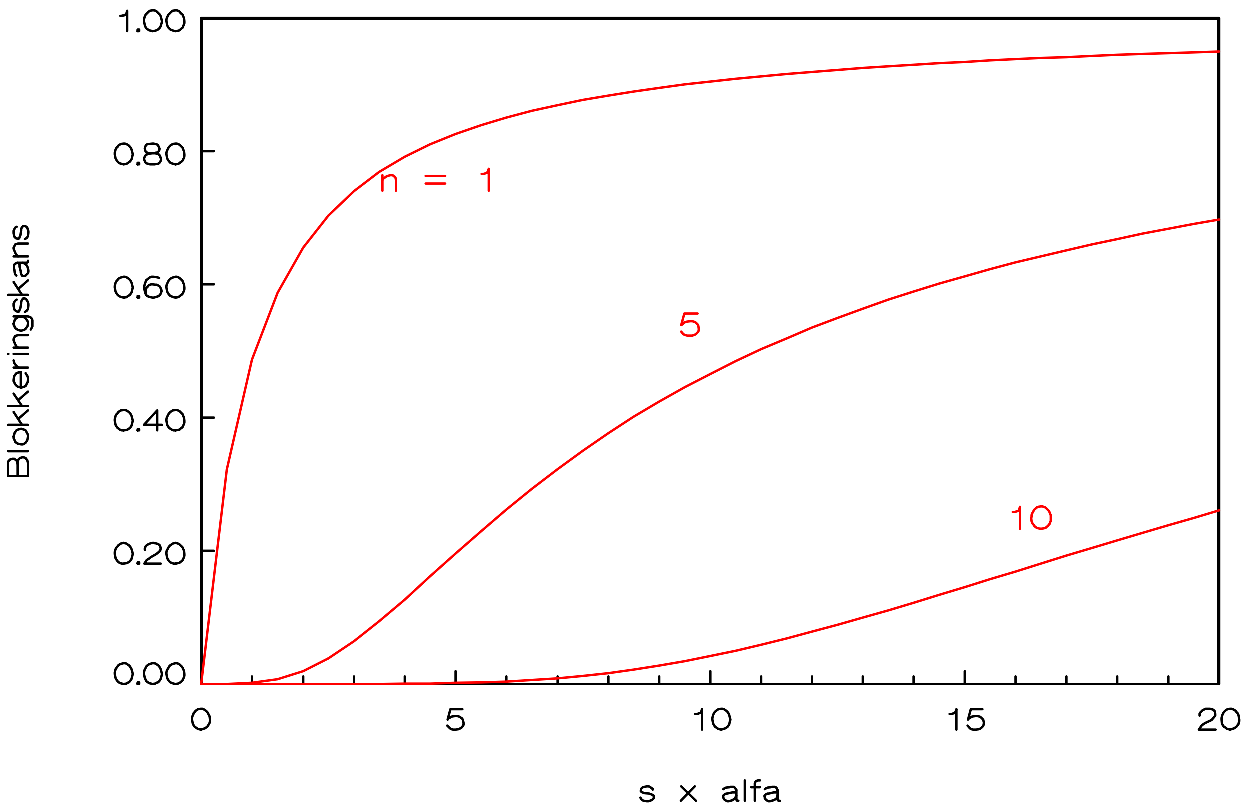
De Engset-blokkeringskans, P_{B}, wordt dan gegeven door het quotiënt van (6.25) en (6.23):

$$P_{\text{B}} = \frac{\binom{s-1}{n} \left(\frac{\alpha}{\mu}\right)^n}{\sum_{i=0}^{n} \binom{s-1}{i} \left(\frac{\alpha}{\mu}\right)^i} \tag{6.26}$$

Deze formule geldt dus voor een eindige populatie van s klanten die elk gedurende gemiddeld $1/\alpha$ vrij zijn en dan gedurende gemiddeld $1/\mu$ een lijn bezetten (als er tenminste geen blokkering optreedt; in dat geval begint weer een nieuwe vrije periode voor die klant).

Ook voor het Engset-model is recursieve berekening van de blokkeringskans mogelijk. We definiëren de $J_i(s,\alpha)$ als de inverse van de blokkeringskans voor s klanten, met oproepintensiteit α per vrije klant, op i lijnen.

$$P_{\text{B}} = \frac{1}{J_n(s,\alpha)} \tag{6.27}$$



Figuur 6.5: Blokkeringskans volgens het Engset-model. Ter vergelijking met figuur 6.2 staat langs de x-as de aankomstintensiteit wanneer alle $s = 20$ klanten vrij zijn.

Ter initialisatie gebruiken we weer dat bij 0 lijnen de blokkeringskans 1 is:

$$J_0(s,\alpha) = 1 \tag{6.28}$$

Zolang $i < s$, is de recursie voor $J_i(s,\alpha)$:

$$J_i(s,\alpha) = J_{i-1}(s,\alpha)\frac{i\mu}{(s-i)\alpha} + 1 \tag{6.29}$$

Voorbeeld 6.3

Bernard vraagt zich af of het Engset-model voor de PABX met 5 buitenlijnen en 25 abonnees niet tot een aanzienlijk lagere blokkeringskans leidt dan de 6.97% die in voorbeeld 6.1 berekend was. We definiëren een abonnee die niet een buitenlijn belegd heeft als een vrije klant, en gaan uit van een oproep instensiteit van $\alpha = 0.1$ (per gemiddelde houdtijd) per vrije klant.

Gevraagd: Hoe groot is de blokkeringskans P_B volgens het Engset-model?

De oproepintensiteit is uitgedrukt per gemiddelde houdtijd. Dat wil zeggen dat de gemiddelde houdtijd van een lijn 1 tijdseenheid is en dat de "bedieningsintensiteit" van een lijn

$\mu = 1$ is. We passen nu rechtstreeks de recursie (6.28)–(6.29) toe:

$$\begin{array}{lcccl} J_1(25,0.1) & = & 1\frac{1}{2.5}+1 & = & 1.416667 \\ J_2(25,0.1) & = & 1.416667\frac{2}{2.5}+1 & = & 2.231884 \\ J_3(25,0.1) & = & 2.231884\frac{3}{2.5}+1 & = & 4.043478 \\ J_4(25,0.1) & = & 4.043478\frac{4}{2.5}+1 & = & 8.701863 \\ J_5(25,0.1) & = & 8.701863\frac{5}{2.5}+1 & = & 22.754658 \end{array} \tag{6.30}$$

Volgens het Engset-model is dan de blokkeringskans:

$$P_{\mathrm{B}} = \frac{1}{22.754658} = 4.39\% \tag{6.31}$$

Hoofdstuk 7

Betrouwbaarheidsberekeningen

Betrouwbaarheidsberekeningen horen ook thuis in het rijtje: Markov-processen, Wachtsystemen, Verliessystemen, ... omdat het hier ook gaat om toevallige gebeurtenissen die de prestatie van bijvoorbeeld een telecommunicatiesysteem bepalen. De effecten van het parallel respectievelijk in serie zetten van deelsystemen kunnen in eerste instantie beschreven worden met een Markov-proces.

Slecht één hoofdstuk is te weinig om recht te doen aan een heel vak, in het Engels aangeduid met de term Reliability Theory, maar omdat een paar basisbegrippen van dit vak in de context van Markov-processen eenvoudig verduidelijkt kunnen worden, willen we hier toch even op in gaan.

7.1 Definities

Een systeem waarvan we de betrouwbaarheid of levensduur willen kwantificeren is opgebouwd uit een aantal units. Elke unit heeft een levensduur die beschreven wordt als een stochastische variabele. Laat T zo'n stochastische variabele zijn. T heeft dan een verdelingsfunctie $F(\cdot)$ en we nemen aan dat voor T ook de kansdichtheidsfunctie $f(\cdot) = F'(\cdot)$ bestaat. Het accentje naast F staat voor de eerste afgeleide.

Bij betrouwbaarheidsberekeningen is het gebruikelijk om te spreken over de reliability-functie $R(\cdot)$, deze is direct gerelateerd aan de verdelingsfunctie van T volgens:

$$R(t) = 1 - F(t) = \Pr\{T > t\} \tag{7.1}$$

Stel, we hebben n units allemaal met dezelfde reliability-functie $R(\cdot)$, die allemaal op tijdstip 0 "ingeschakeld" worden. Op tijdstip t zijn er nog $nR(t)$ in werking. De kans dat één van de oorspronkelijke n units uitvalt in het interval $(t, t + \Delta t)$ is gelijk aan $nf(t)\Delta t$ (plus termen van de orde Δt^2 en hoger). De kans op uitval per werkende unit per tijdseenheid is dan:

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)} \tag{7.2}$$

Bovenstaande grootheid $\lambda(t)$ is de failure rate, ofwel, het uitval-tempo op tijdstip t. Het is eenvoudig na te gaan dat, indien de levensduur verdeling negatief exponentieel is, de failure rate een constante is. In het geval dat er een bovengrens $t_{\max}$ is voor de levensduur, gaat de failure rate naar ∞ als $t \to t_{\max}$.

De Mean Time To Failure (MTTF) van een systeem definiëren we als de verwachtingswaarde van de bedrijfsduur van het systeem nadat het voor het eerst is ingeschakeld. Voor repareerbare systemen is dit de eerste bedrijfsduur. Het hangt af van de configuratie van het systeem hoe de bedrijfsduur van het systeem gerelateerd is aan de levensduur van de units waaruit het bestaat. Bijvoorbeeld, als alle units in serie staan, dan is de bedrijfsduur gelijk aan de levensduur van de kortst levende unit. Staan de units parallel, dan is de bedrijfsduur gelijk aan de levensduur van de langst levende unit.

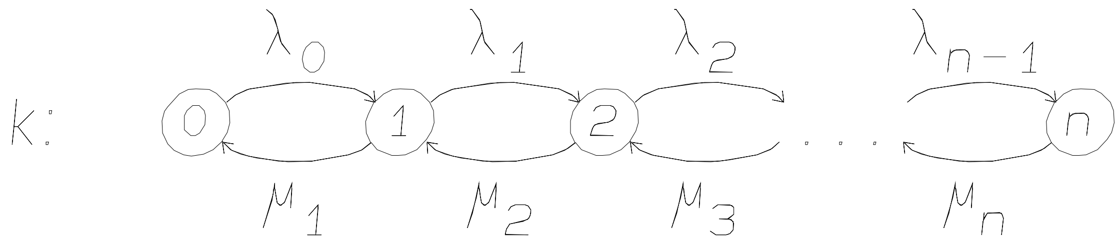
Voor repareerbare systemen definiëren we ook de Mean Time Between Failures (MTBF), zijnde de bedrijfsduur van het systeem nadat het gerepareerd is. In de volgende paragraaf zullen deze definities iets concreter worden gemaakt voor een model van een repareerbaar systeem met n parallelle units.

7.2 Repareerbare systemen

Wanneer we een systeem hebben dat uit een aantal units bestaat die elk een negatief exponentieel verdeelde levensduur hebben en ook een negatief exponentieel verdeelde reparatietijd, dan kan dat systeem beschreven worden als een Markov-proces. Hierbij is dan impliciet aangenomen dat het falen van één unit statistisch onafhankelijk is van het falen van de andere units. Vooral deze laatste aanname maakt dat de toepasbaarheid van het model enigszins beperkt is. Maar het feit dat met een eenvoudig geboorte-sterfte-proces snel inzicht verkregen kan worden in het effect van verschillende vormen van redundantie maakt toch een verdere bestudering de moeite waard.

We beschouwen een systeem met n units waarvan er slechts één hoeft te functioneren. De overige $n-1$ units zijn standby units, men zegt dan dat de redundantie $n-1$ is. We krijgen een geboorte-sterfte-proces door het aantal defecte units, k, te kiezen als de toestand van het systeem. Merk op dat de historische naamgeving geboorte-sterfte-proces weinig gepast is, omdat in die termen het falen van een unit correspondeert met een "geboorte" en het voltooien van een reparatie met een "sterfte" (misschien zouden we het moeten hebben over sterfte-geboorte-processen). Het toestandsdiagram voor het aantal defecte units staat getekend in figuur 7.1.

De levensduur van een unit is negatief exponentieel verdeeld met verwachtingswaarde $1/\lambda$. De levensduur gaat in wanneer de unit is ingeschakeld. Voor één ingeschakelde unit is de failure rate gelijk aan λ. Verder is m_k een door de architectuur van het systeem bepaalde grootheid die aangeeft hoeveel units zijn ingeschakeld, gegeven dat er k units defect zijn. De



Figuur 7.1: Toestandsdiagram voor het aantal defecte units van een repareerbaar systeem.

volgende terminologie wordt gebruikt:

$$\left.\begin{array}{ll} \text{cold standby} & m_k = 1 \\ \text{hot standby} & m_k = n - k \end{array}\right\} k = 0, 1, \ldots, n \tag{7.3}$$

De kans op een overgang van toestand k naar toestand $k+1$ in een interval Δt is $m_k \lambda \Delta t$. De overgangsintensiteit naar rechts vanuit een toestand k is:

$$\lambda_k = m_k \lambda \tag{7.4}$$

De reparatieduur van een unit werd ook negatief exponentieel verondersteld. De gemiddelde reparatie duur is $1/\mu$. De overgangsintensiteit naar links hangt af van de reparatie capaciteit die beschikbaar is voor het systeem. Twee uitersten zijn:

$$\begin{array}{lll} \mu_k = k\mu & k = 0, 1, \ldots, n & \text{capaciteit} \geq n \\ \mu_k = \mu & k = 1, 2, \ldots, n & \text{capaciteit} = 1 \end{array} \tag{7.5}$$

Met bovenstaande gegevens zijn we in staat om de evenwichtsverdeling van het aantal defecte units te berekenen. Omdat de redundantie van het hier bestudeerde systeem $n-1$ is, is het systeem nog operationeel zolang het aantal defecte units kleiner is dan n. Laat de vector V de evenwichtsverdeling weergeven, dan halen we hieruit voor de beschikbaarheid:

$$\text{beschikbaarheid} = 1 - v_n \tag{7.6}$$

De twee andere prestatiematen die we willen berekenen zijn de MTTF en de MTBF. Voor de MTTF gaat het om de eerste bedrijfsduur van het systeem; in dat geval is de begintoestand van het systeem $k = 0$ defecte units. Bij de MTBF is de begintoestand van het systeem $k = n - 1$, omdat de j-de bedrijfsduur van het systeem ingaat nadat alle units defect waren op het moment dat de eerst gerepareerde unit “uit de reparatiewerkplaats komt” en in bedrijf gesteld wordt.

Vanuit een willekeurige begintoestand willen we bepalen hoe lang het duurt totdat een failure optreedt, d.w.z. hoe lang het duurt totdat alle n units defect zijn. Daartoe definiëren we

$$f_k = \begin{array}{l} \text{verwachtingswaarde van de tijd die verstrijkt totdat} \\ \text{toestand } n \text{ wordt bereikt vanuit begintoestand } k \end{array} \tag{7.7}$$

In het randgeval dat de begintoestand $k = n$ is, verstrijkt er helemaal geen tijd totdat toestand n bereikt wordt, omdat het systeem al in toestand n is. Daarom kunnen we direct opschrijven:

$$f_n = 0 \tag{7.8}$$

Als de begintoestand $k = n - 1$ is dan zijn er twee overgangen vanuit die toestand mogelijk: met intensiteit $n - 1$ naar toestand n en met intensiteit $n - 1$ naar toestand $n - 2$. Dit betekent dat in een interval Δt de toestand $k = n - 1$ met kans $(\lambda_{n-1} + \mu_{n-1})\Delta t$ verlaten wordt. We weten ook dat de verblijftijd in een toestand negatief exponentieel verdeeld is. Hieruit volgt dat de gemiddelde verblijftijd in de toestand $k = n - 1$ gelijk is aan:

$$\frac{1}{\lambda_{n-1} + \mu_{n-1}} \tag{7.9}$$

Na bovenstaande tijd vindt er één van de twee al eerder genoemde overgangen plaats. In de tabel hieronder (7.10) staan die overgangen genoemd, met daarnaast de kans dat juist die overgang optreedt, gegeven dat de toestand $k = n - 1$ verlaten wordt, en daarnaast weer de tijd die het dán nog duurt totdat toestand $k = n$ bereikt wordt:

$$\begin{array}{ccc} \text{van } n-1 \text{ naar} & \text{kans} & \text{tijd tot } n \text{ bereikt wordt} \\ n & \dfrac{\lambda_{n-1}}{\lambda_{n-1} + \mu_{n-1}} & 0 \\ n-2 & \dfrac{\mu_{n-1}}{\lambda_{n-1} + \mu_{n-1}} & f_{n-2} \end{array} \tag{7.10}$$

We krijgen nu de waarde voor f_{n-1} door bij de waarde van (7.9) een gewogen som van de waarden uit de rechterkolom van (7.10) op te tellen. Dit geeft:

$$f_{n-1} = \frac{1}{\lambda_{n-1} + \mu_{n-1}} + \frac{\mu_{n-1}}{\lambda_{n-1} + \mu_{n-1}} f_{n-2} \tag{7.11}$$

Deze formule kan herschreven worden in de vorm:

$$-\mu_{n-1} f_{n-2} + (\lambda_{n-1} + \mu_{n-1}) f_{n-1} = 1 \tag{7.12}$$

Door uit te gaan van een willekeurige begintoestand k en dezelfde redenering te volgen, krijgen we de volgende reeks van relaties:

$$-\mu_k f_{k-1} + (\lambda_k + \mu_k) f_k - \lambda_k f_{k+1} = 1 \quad k = n-1, n-2, \ldots, 0 \tag{7.13}$$

In het geval $k = 0$ dient de term $\mu_0 f_{-1}$ gelijk aan nul gesteld te worden. De formules (7.8) en (7.13) vormen samen een stelsel van $n+1$ lineaire vergelijkingen met $n+1$ onbekenden, waaruit $f_0, f_1, \ldots, f_n$ opgelost kunnen worden. We kunnen ook dit stelsel in matrix-vectorvorm schrijven en dan rechtstreeks met daarvoor bestaande procedures oplossen. Daarmee hebben we dan ook de gezochte resultaten:

$$\text{MTTF} = f_0 \tag{7.14}$$

want dit is de gemiddelde tijd om vanuit begintoestand $k = 0$ de toestand $k = n$ te bereiken en

$$\text{MTBF} = f_{n-1} \tag{7.15}$$

want dit is de gemiddelde tijd om vanuit begintoestand $k = n - 1$ de toestand $k = n$ te bereiken.

Voorbeeld 7.1

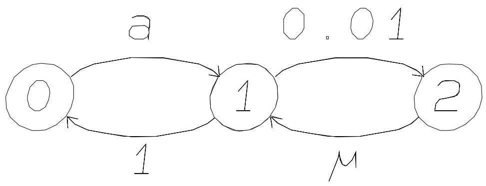
Bij een telecommunicatie systeem in de bush-bush moet een stroomvoorzieningssysteem geleverd worden dat uit betrouwbaarheidsoverwegingen dubbel uitgevoerd is. De gemiddelde reparatietijd voor één unit is één tijdseenheid. De gemiddelde bedrijfsduur van één unit is 100 tijdseenheden.

In het oorspronkelijke plan waren er faciliteiten om eventueel beide units tegelijk te kunnen repareren. Verder wilde men de tweede unit hot standby laten draaien om een korte spanningsonderbreking bij een enkelvoudig defect te voorkomen.

In het kader van de bezuinigingen moet ofwel de faciliteit van dubbele reparatie worden teruggebracht tot een enkelvoudige faciliteit, ofwel de tweede unit wordt cold standby. Om een enigszins gemotiveerde beslissing te kunnen nemen moet er eerst wat gerekend worden.

Gevraagd: Wat is, in geval van het oorspronkelijke plan en in geval van de verschillende bezuinigingsopties, de MTTF, de MTBF en de beschikbaarheid van het stroomvoorzieningssysteem?

In alle gevallen kunnen we het volgende toestandsdiagram tekenen voor het aantal defecte units.



Figuur 7.2: Toestandsdiagram bij voorbeeld 7.1.

Door de keuze van de tijdseenheid (gelijk aan de gemiddelde reparatieduur) is $\mu_1 = 1$. We korten μ_2 af als μ. De gemiddelde bedrijfsduur van één unit van 100 tijdseenheden houdt in dat $\lambda = 0.01$. De verschillende opties kunnen gekozen worden door voor a en μ de volgende waarden in te vullen:

hot standby	$a = 0.02$
cold standby	$a = 0.01$
dubbele faciliteit	$\mu = 2$
enkele faciliteit	$\mu = 1$

De berekening — met parameters a en μ — van de MTTF en MTBF vindt met formule (7.13) plaats. Hier is $n = 2$. Dit geeft twee lineaire vergelijkingen (één voor $k = 0$ en één voor $k = 1$):

$$af_0 - af_1 = 1 \tag{7.16}$$

$$-f_0 + 1.01 f_1 = 1 \tag{7.17}$$

Merk op dat hier al te zien is dat de MTTF en de MTBF niet afhangt van of er al dan niet een dubbele reparatie faciliteit is, ofwel van μ. Deze parameter is wel van invloed op de beschikbaarheid van het systeem. Bovenstaand stelsel van lineaire vergelijkingen kan in matrixnotatie geschreven worden als:

$$\begin{bmatrix} a & -a \\ -1 & 1.01 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{7.18}$$

De oplossing met matrixinversie is:

$$\begin{bmatrix} f_0 \\ f_1 \end{bmatrix} = \frac{100}{a} \begin{bmatrix} 1.01 & a \\ 1 & a \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{7.19}$$

Matrix-vector vermenigvuldiging geeft dan:

$$\mathrm{MTTF} = f_0 = \frac{101 + 100a}{a} \tag{7.20}$$

$$\mathrm{MTBF} = f_1 = \frac{100 + 100a}{a} \tag{7.21}$$

De beschikbaarheid van het systeem volgt direct uit de evenwichtsverdeling. Voor de ongenormeerde verdeling (zie (4.71)–(4.73)) vinden we:

$$w_0 = 1 \tag{7.22}$$

$$w_1 = a \tag{7.23}$$

$$w_2 = \frac{0.01a}{u} \tag{7.24}$$

De beschikbaarheid is de kans op 0 of 1 defecte unit:

$$1 - v_2 = \frac{1 + a}{1 + a + \frac{0.01a}{u}} \tag{7.25}$$

Het invullen van a en u voor de verschillende opties geeft:

	MTTF	MTBF	$1 - v_2$
hot standby en dubbele faciliteit	5150	5100	0.999902
cold standby en dubbele faciliteit	10200	10100	0.999950
hot standby en enkele faciliteit	5150	5100	0.999804
cold standby en enkele faciliteit	10200	10100	0.999901