

STICHTING
MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

A M S T E R D A M

STATISTISCHE AFDELING

Rapport S 265 (C 13)

Leergang Besliskunde

Hoofdstuk III

Stochastische grootheden, kansverdelingen en momenten

door

J. Kriens

(Bewerking van de hoofdstukken 7, 8 en 9 van rapport
S 200 (C8) door Prof. Dr J. Hemelrijk)

Februari 1960

1. Stochastische grootheden en eendimensionale kansverdelingen.

In deze en in de volgende paragrafen worden een aantal belangrijke begrippen en stellingen uit de kansrekening behandeld.

Definitie 1;1

Een grootheid $\underline{x}$, die verschillende waarden aan kan nemen, wordt een stochastische grootheid genoemd indien voor ieder gegeven reëel getal x de kans, dat $\underline{x}$ een waarde aanneemt, die hoogstens daaraan gelijk is, gedefinieerd is.

Voorbeelden

Grootheden, die in een wiskundig model door stochastische grootheden voorgesteld kunnen worden, zijn de volgende. Het aantal successen in een reeks onafhankelijke experimenten met ieder een kans p op succes. Het aantal worpen met een munt tot voor het eerst kruis verkregen wordt. Maar ook: de gewichtstoename van een proefdier tijdens een proefperiode; de levertijd van een uitgegeven order; het aantal fouten, dat in een week in een administratie gemaakt wordt en het aantal telefoongesprekken, dat gedurende een bepaald tijdsinterval over een bepaalde lijn gevoerd wordt.

Notatie

Stochastische grootheden geven wij aan met onderstreepte Latijnse letters: $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{x}_1$, $\underline{x}_2$, $\underline{x}_i$ enz., terwijl getalwaarden, die zij aan kunnen nemen, vaak beschouwd als algebraïsche variabelen, door dezelfde letters zonder onderstreping aangegeven worden.

(1;1) $P[\underline{x} \leq x] \stackrel{\text{def}}{=} \text{de kans dat } \underline{x} \text{ een waarde } \leq x \text{ aanneemt.}$

In plaats van onderstreepte letters worden vaak ook vet gedrukte letters of hoofdletters gebruikt.

Bijbehorende termen

De kansverdeling van een stochastische grootheid $\underline{x}$ is de verzameling van alle waarden x , die $\underline{x}$ aan kan nemen met de bijbehorende kansen $P[\underline{x} \leq x]$. De verdelingsfunctie van $\underline{x}$ is de functie

(1;2) $F(x) \stackrel{\text{def}}{=} P[\underline{x} \leq x],$

opgevat als functie van de algebraïsche variabele x .

Iedere verdelingsfunctie is een functie, die monotoon niet dalend is en uitsluitend waarden in het gesloten interval $[0,1]$ aanneemt, terwijl

$$(1;3) \quad F(-\infty) = 0, \quad F(+\infty) = 1.$$

Definitie 1;2

Discrete¹⁾ kansverdelingen zijn kansverdelingen, waarbij de beschouwde stochastische grootte een eindig of aftelbaar aantal waarden $x_1, x_2, \dots$ aan kan nemen, terwijl

$$(1;3) \quad p_i \stackrel{\text{def}}{=} P[X = x_i] > 0, \quad \sum_i p_i = 1$$

De bijbehorende verdelingsfuncties zijn trapfuncties:

$$(1;4) \quad F(x) = \sum_{x_i \leq x} p_i.$$

De belangrijkste discrete verdelingen zijn de volgende.

1. Het alternatief of de dichotomie. (voorbeeld 5;1 van paragraaf I.5):

$$(1;5) \quad P[X = 0] = p, \quad P[X = 1] = q = 1 - p$$

2. De binomiale verdeling (voorbeeld 5;6 van paragraaf I.5):

$$(1;6) \quad P[X = k] = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n; q = 1 - p)$$

Van deze verdeling bestaan uitvoerige tabellen, waarvan wij noemen: Tables of the Binomial Probability Distribution, National Bureau of Standards - Applied Mathematics Series 6 (1950), H.G. ROMIG, 50-100 Binomial Tables, John Wiley (1953).

Beide tabellen bevatten zowel de waarden van de afzonderlijke termen als van de verdelingsfunctie; p doorloopt de waarden 0,01 (0,01) 0,50 en n de waarden 2(1)49 respectievelijk 50(5)100. Het bepalen van kansen wanneer $n > 100$ is, zullen wij in hoofdstuk IV behandelen.

3. De hypergeometrische verdeling (voorbeeld 7 van paragraaf I.5):

1) Of: discontinue.

$$(1;7) \quad P[\underline{x} = k] = \frac{\binom{M}{k} \binom{N}{n-k}}{\binom{M+N}{n}} \quad (k \text{ geheel tussen } \max(n-N, 0) \text{ en } \min(M, n)).$$

Deze verdeling is niet zo uitvoerig getabelleerd als de binomiale. In een gegeven situatie moet men de waarden $P[\underline{x}=k]$ vaak zelf berekenen, hetgeen vergemakkelijkt wordt door het bestaan van tabellen met de binomiaalcoëfficiënten $\binom{m}{n}$; o.a. is er één opgenomen in TH.C. FRY, Probability and its Engineering Uses, D. van Nostrand Company, New York (1928), waarin m de waarden 1(1)100 doorloopt.

Verder kan deze verdeling in veel gevallen goed benaderd worden door de binomiale verdeling (vgl. de opmerking bij voorbeeld 5;7 in het eerste hoofdstuk).

4. De negatief binomiale verdeling (voorbeeld 8 van paragraaf I.5):

$$(1;8) \quad P[\underline{x} = x] = \binom{x-1}{k-1} p^k q^{x-k} \quad (x = k, k+1, \dots; q = 1-p).$$

Tussen de kansen (1;8) en (1;6) bestaat een eenvoudig verband: vermenigvuldigt men de kansen (1;6) met

$$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}.$$

en vervangt men k door x, dan ontstaan de kansen (1;8). Met behulp van deze relatie kunnen de kansen van de negatief binomiale verdeling altijd uit een tabel van de binomiale verdeling berekend worden.

5. De Poissonverdeling:

$$(1;9) \quad P[\underline{x} = k] = \frac{\lambda^k e^{-\lambda}}{k!} \quad (k = 0, 1, \dots; \lambda \text{ willekeurig positief}).$$

Deze verdeling is onder andere getabelleerd in: E.C. MOLINA, Poisson's Exponential Binomial Limit, D. van Nostrand Company, New York (1945); er is een tabel met afzonderlijke termen en een cumulatieve tabel; λ doorloopt de waarden 0,001(0,001)0,01(0,01)0,3(0,1)15,0(1)100.

Verder bevat het uitgebreide tabellenboek van E.S. PEARSON and H.O. HARTLEY, Biometrika Tables for Statisticians, Vol. I Cambridge (1954),

een tabel van afzonderlijke termen voor $\lambda=0,1(0,1)15$ (tabel 39) en een cumulatieve tabel voor $\lambda=0,0005(0,0005)0,005(0,005)0,05(0,05)1,0(0,1)5,0(0,25)10,0(0,5)20,0(1)67$ (tabel 7).

6. De (discrete) homogene verdeling:

$$(1;10) \quad P[X = k] = \frac{1}{n+1} \quad (k = 0, 1, \dots, n).$$

Ter illustratie zijn in figuur 1,1 de afzonderlijke kansen en de verdelingsfunctie getekend van een alternatief met kans $p = \frac{1}{3}$ en in figuur 1,2 de afzonderlijke kansen en de verdelingsfunctie van een Poissonverdeling met parameter $\lambda = 2,2$.

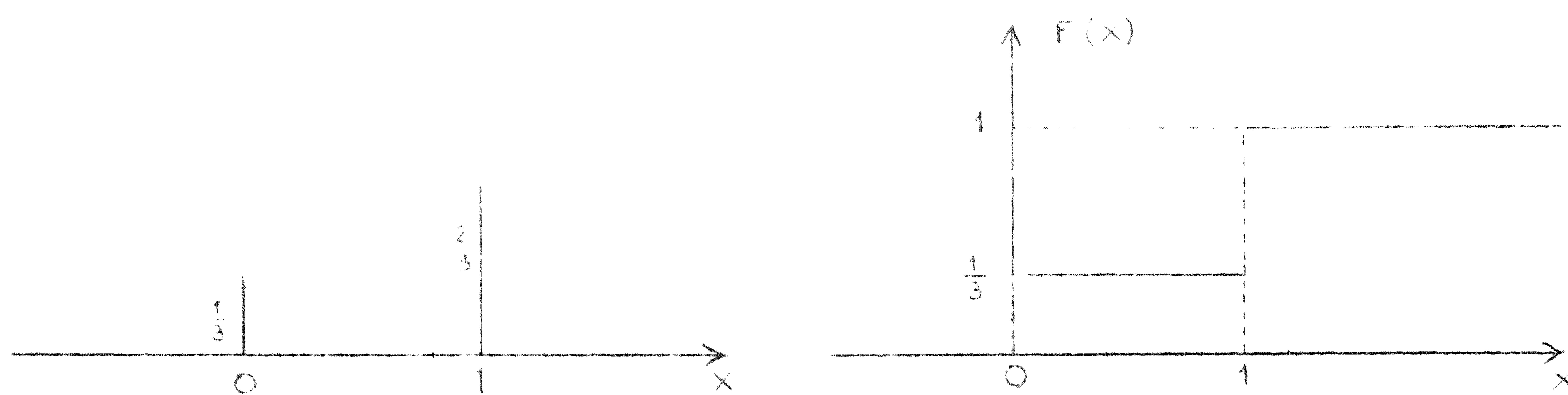


fig. 1,1

Afzonderlijke kansen en verdelingsfunctie van een alternatief met $p = \frac{1}{3}$

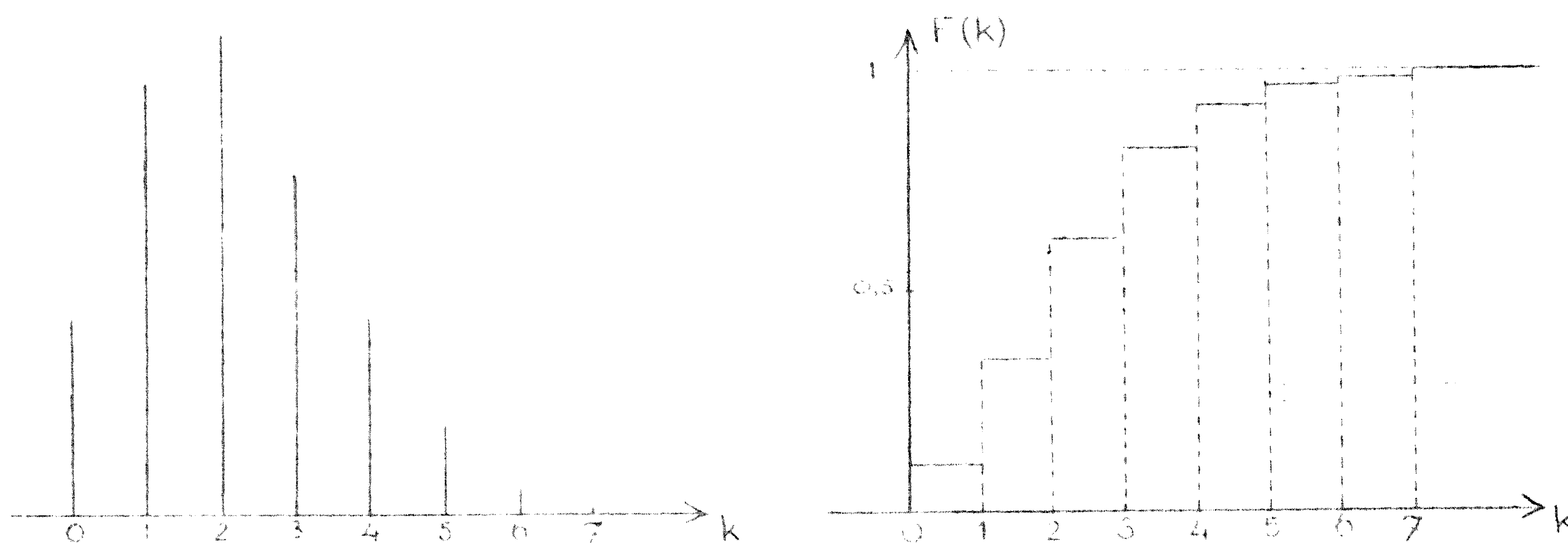


fig. 1,2

Afzonderlijke kansen en verdelingsfunctie van een Poissonverdeling met $\lambda = 2,2$

Definitie 1;3

Continue kansverdelingen zijn kansverdelingen, waarbij de stochastische grootte een continuüm van waarden (bijv. alle waarden van een eindig of oneindig interval) aan kan nemen, terwijl de verdelingsfunctie $F(x)$ voor iedere x continu is en voor iedere x , op hoogstens eindig veel na, continu differentiëerbaar. De afgeleide

$$(1;11) \quad f(x) \stackrel{\text{def}}{=} \frac{d}{dx} F(x)$$

wordt dan de verdelingsdichtheid genoemd. Derhalve geldt

$$(1;12) \quad F(x) = \int_{-\infty}^x f(v) dv \quad 1)$$

en, zoals op grond van de axioma's gemakkelijk te bewijzen is

$$(1;13) \quad P[x_1 \leq x \leq x_2] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx \quad (\text{vergelijk figuur 1,3}),$$

$$(1;14) \quad \int_{-\infty}^{\infty} f(x) dx = 1,$$

$$(1;15) \quad P[x = x] = 0 \quad \text{voor iedere } x.$$

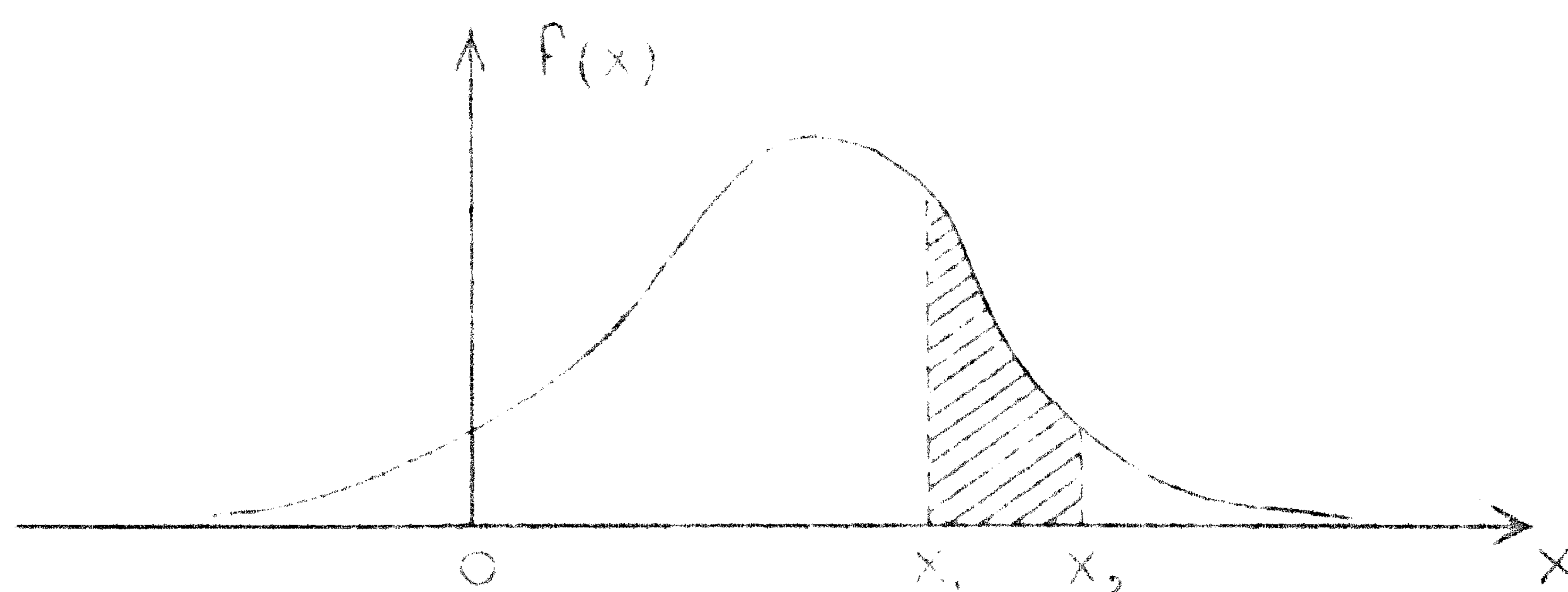


fig. 1,3

Meetkundige illustratie van $P[x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} f(x) dx.$

Deze kansverdelingen treden op als limiet van discrete kansverdelingen (zoals wij later zullen zien; zij worden daarom vaak voor benaderingsdoeleinden gebruikt) en als wiskundig model bij grootheden, die quantitatief op een bij benadering continue schaal afgelezen worden. Voorbeelden daarvan zijn: de lengte van een recruit,

- 1) Als de verdeling in werkelijkheid "begint" bij een punt $x=a$ kan men ook a als ondergrens van de integraal nemen; daar dan echter $F(x)=0$ voor $x \leq a$, dus ook $f(x)=0$ in dat gebied, is de ondergrens $-\infty$ steeds correct. Analoge overwegingen gelden voor de bovengrens, indien deze het "einde" van de verdeling overschrijdt.

de tijdsduur van een treinrit, de snelheid van een vliegtuig en de vraag per week naar een massa-artikel.

De belangrijkste continue verdelingen zijn hieronder opsomd.

7. De (continue) homogene verdeling:

$$(1;16) \quad f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{voor } \theta_1 \leq x \leq \theta_2, \\ 0 & \text{elders} \end{cases}$$

$$(1;17) \quad F(x) = \begin{cases} 0 & \text{voor } x < \theta_1, \\ \frac{x - \theta_1}{\theta_2 - \theta_1} & \text{voor } \theta_1 \leq x \leq \theta_2, \\ 1 & \text{voor } x > \theta_2. \end{cases}$$

De "eindpunten" θ_1 en θ_2 zijn hierin constanten en treden als parameters op.¹⁾

8. De normale verdeling:

$$(1;18) \quad f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad 2) \quad (-\infty < x < +\infty, -\infty < \mu < +\infty, \sigma > 0)$$

$$(1;19) \quad F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv.$$

Notatie

Bezit $\underline{x}$ een normale verdeling met parameters μ en σ , dan wordt dit aangegeven met: $\underline{x}$ is $N(\mu, \sigma)$ -verdeeld.

Is $\mu=0$ en $\sigma=1$, dus is $\underline{x}$ $N(0,1)$ -verdeeld, dan spreekt men van de gestandaardiseerde normale verdeling. Voor grootheden, die deze verdeling bezitten, gebruiken wij het symbool $\underline{u}$. De gestandaardiseerde normale verdeling is uitvoerig getabelleerd.

Tabellen van de verdelingsdichtheid en de verdelingsfunctie van $\underline{u}$ zijn onder andere opgenomen in de reeds genoemde verzameling van HARTLEY en PEARSON; x doorloopt hierin de waarden 0,00(0,01)

- 1) Parameters van kansverdelingen worden meestal door griekse letters aangegeven. In bepaalde gevallen, zoals bij de parameter p van een binomiale verdeling, zijn andere symbolen zozeer ingeburgerd, dat van deze regel afgeweken wordt.
- 2) Dat voor deze functie inderdaad de relatie (1;14) geldt vindt men bewezen in de appendix aan het eind van deze paragraaf (zie blz.10).

6,00 (tabel 1).

Een beknopte tabel is aan het eind van dit hoofdstuk toegevoegd.

9. De exponentiële verdeling:

$$(1;20) \quad f(x) = \lambda e^{-\lambda x} \quad (0 \leq x < \infty; \lambda \text{ willekeurig positief}).$$

$$(1;21) \quad F(x) = \int_0^x \lambda e^{-\lambda v} dv = 1 - e^{-\lambda x}.$$

Zowel de waarden van $f(x)$ als van $F(x)$ kunnen gemakkelijk gevonden worden met behulp van een tabel van natuurlijke logaritmen of van e-machten; een dergelijke tabel is bijvoorbeeld:

H.W. HOLTAPPEL, Tafels van e^x , P. Noordhoff, Groningen (1938); x loopt van 0,000 met sprongen van 0,001 op tot en met 9,999.

10. De χ^2 -verdeling:

$$(1;22) \quad f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (0 \leq x < \infty; n=1,2,\dots)$$

$$(1;23) \quad F(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^x v^{\frac{n}{2}-1} e^{-\frac{v}{2}} dv$$

In de formules (1;22) en (1;23) is Γ de γ -functie, die gedefiniëerd wordt door (vgl. ook II.6)

$$(1;24) \quad \Gamma(p) = \int_0^{\infty} e^{-x} x^{p-1} dx.$$

Naar de parameter n wordt de χ^2 -verdeling vaak genoemd: χ^2 -verdeling met n vrijheidsgraden.

Een tabel van de verdelingsfunctie voor $n = 1(1)30(2)70$ en $x = 0,001(0,001)0,01(0,01)0,1(0,1)2,0(0,2)10(0,5)20,0(1)40(2)120$ is opgenomen in de verzameling van HARTLEY en PEARSON (tabel 7).

11. De gamma-verdeling:

$$(1;25) \quad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \quad (0 \leq x < \infty; \alpha > 0, \lambda > 0)$$

$$(1;26) \quad F(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^x e^{-\lambda v} v^{\alpha-1} dv$$

De onder 9 en 10 genoemde verdelingen zijn speciale gevallen van de gamma-verdeling. De exponentiële verdeling verkrijgt men

door in (1;25) voor α de waarde 1 in te vullen ($\Gamma(1)=1$) en de χ^2 -verdeling met n vrijheidsgraden door $\lambda=\frac{1}{2}$ en $\alpha=\frac{n}{2}$ in te vullen. Een ander speciaal geval van de gamma-verdeling is de zogenaamde Erlang-verdeling, die ontstaat, wanneer men α alleen gehele positieve waarden laat doorlopen.

Onder de klasse van γ -verdelingen vallen vele verdelingen die in de praktijk vaak gebruikt kunnen worden. Voor een nadere beschouwing van dit belangrijke type verdelingen en een overzicht van de bestaande tabellen verwijzen wij naar GERDA KLERK-GROBBEN, De Gamma-verdeling; kort overzicht betreffende eigenschappen, schattings- en toetsingsmethoden, Rapport S 247 van de Statistische Afdeling van het Mathematisch Centrum (1959).

12. De beta-verdeling:

$$(1;27) \quad f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (0 \leq x \leq 1; \alpha > 0, \beta > 0)$$

$$(1;28) \quad F(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x v^{\alpha-1} (1-v)^{\beta-1} dv.$$

Waarden van $F(x)$ kunnen onder andere gevonden worden met behulp van tabel 16 en nomogram 17 uit de verzameling van HARTLEY en PEARSON; in de tabel doorloopt α de waarden 0,5(0,5)15,20,30,60, ∞ en β de waarden 0,5(0,5)5,6,7 $\frac{1}{2}$,10,12,15,20,30, ∞ .

13. De logarithmisch normale verdeling:

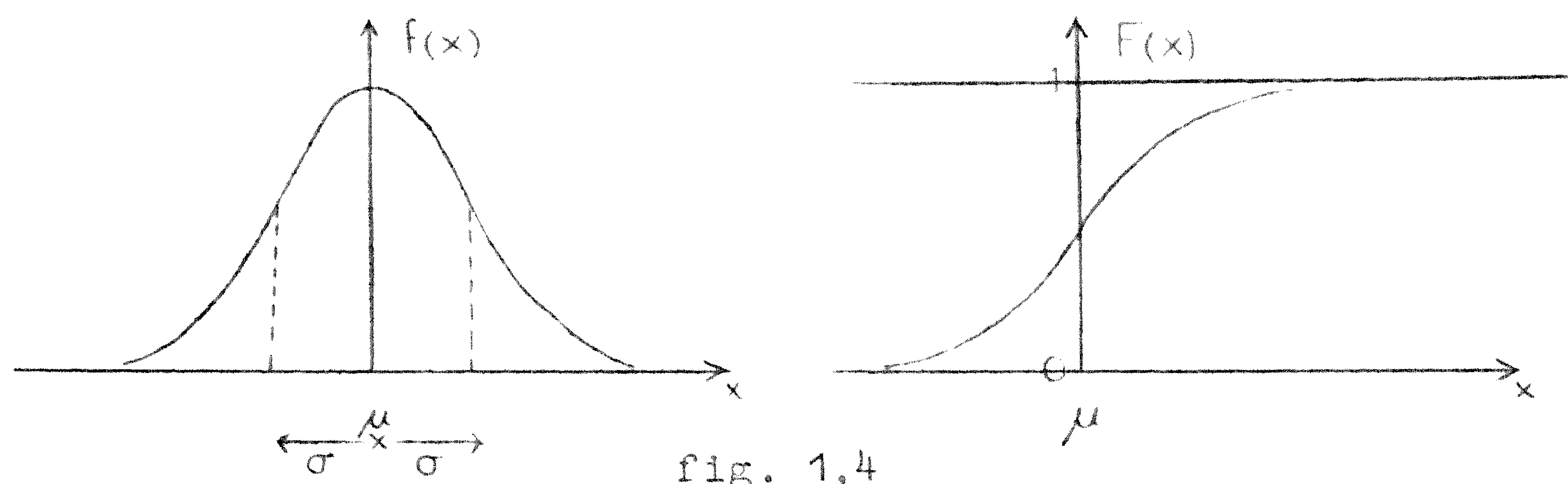
$$(1;29) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \frac{1}{x} \quad (0 \leq x < \infty; -\infty < \mu < +\infty; \sigma > 0)$$

$$(1;30) \quad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-\frac{(\log v - \mu)^2}{2\sigma^2}} \frac{1}{v} dv.$$

Deze verdeling kan uit de normale verdeling worden afgeleid door voor x te substitueren $\log y$; de grootheid x is dan zelf niet normaal verdeeld, maar de logaritme ervan wel. Deze transformatie kan ook gebruikt worden om met behulp van tabellen van de normale verdeling waarden van $f(x)$ en $F(x)$ te vinden.

Ter illustratie is in figuur 1,4 de verdelingsdichtheid en de verdelingsfunctie van een normale verdeling met parameters μ en σ getekend. De afstand van het punt $x=\mu$ tot de abscis van het buig-

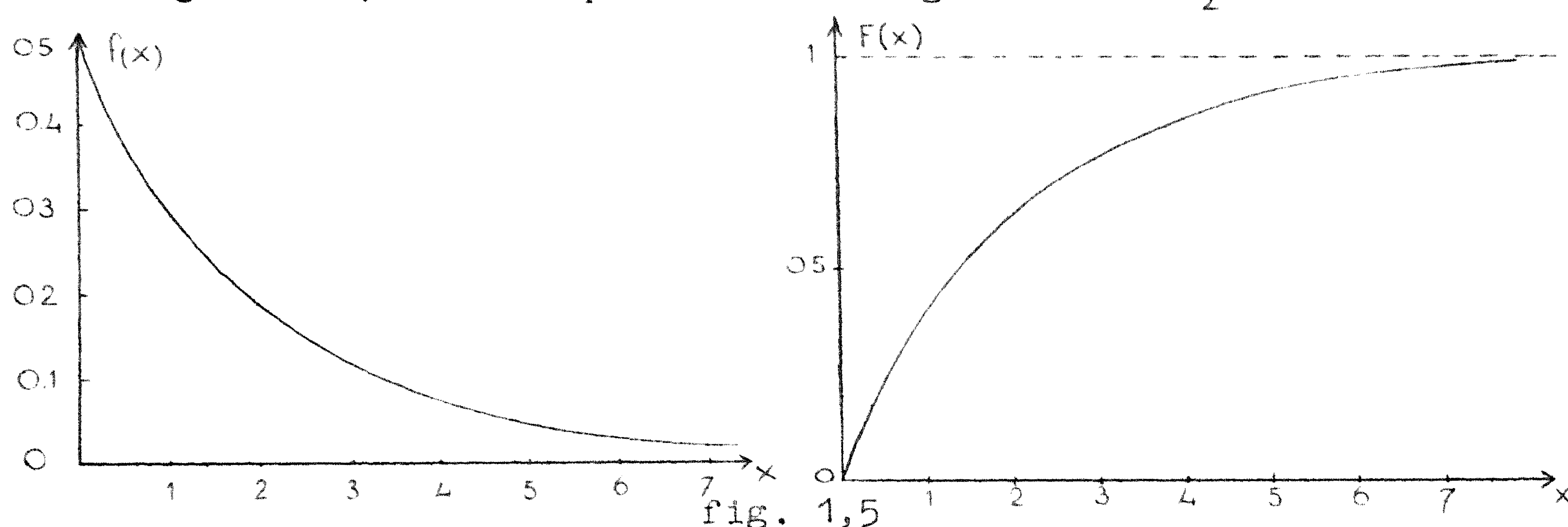
punt van $f(x)$ bedraagt σ , hetgeen men kan verifiëren door de



Verdelingsdichtheid en verdelingsfunctie van de normale verdeling

tweede afgeleide van $f(x)$ gelijk aan nul te stellen.

Een tweede illustratie geeft figuur 1,5 waarin de verdelingsdichtheid en de verdelingsfunctie van een exponentiële verdeling is getekend; voor de parameter λ is gekozen $\lambda = \frac{1}{2}$.



Verdelingsdichtheid en verdelingsfunctie van de exponentiële verdeling met parameterwaarde $\lambda = \frac{1}{2}$

Opmerking

Tenslotte willen wij nog verwijzen naar twee Nederlandse publicaties.

1. In een artikel van H J. PRINS, dat zal verschijnen in het tweede nummer van de jaargang 1960 van STATISTICA NEERLANDICA, wordt een overzicht gegeven van methoden om kansen behorende bij bepaalde verdelingen te berekenen met behulp van tabellen van andere verdelingen.

2. Onder redactie van de Vereniging voor Statistiek verschijnt bij Stenfert Kroese een losbladige verzameling "Statistische

Tabellen en Nomogrammen", die reeds een groot aantal van de hier genoemde verdelingen omvat.

Definitie 1;4

Gemengde kansverdelingen zijn kansverdelingen, die verkregen kunnen worden als gewogen gemiddelden van een discrete en een continue, d.w.z. waarvoor geldt:

$$(1;31) \quad F(x) = \lambda F_1(x) + (1-\lambda) F_2(x), \quad (0 < \lambda < 1)$$

waarin $F_1(x)$ een trapfunctie en $F_2(x)$ continu differentiëerbaar is.

Voorbeeld

De schadeuitkering, die in één jaar op een verzekeringspolis moet geschieden. Er is een positieve kans op uitkering 0 (geen verhaalbare schade), terwijl verder een (vrijwel) continue schaal van mogelijke uitkeringen bestaat.

Opmerking

De mogelijkheden zijn hiermede nog niet uitgeput. Men kan continue verdelingsfuncties beschouwen, die een niet continue afgeleide bezitten of die in oneindig veel punten niet differentiëerbaar zijn. Deze pathologische gevallen zijn voor de praktijk van weinig of geen belang en blijven hier buiten beschouwing.

Appendix

Wij bewijzen dat (1;18) voldoet aan (1;14), dus dat

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \frac{(x-u)^2}{\sigma^2}} dx = 1.$$

Geven wij het linkerlid aan door L en substitueren wij $u = \frac{x-u}{\sigma}$, dan komt er, daar $dx = \sigma du$ is,

$$L = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} u^2} du.$$

Derhalve is

$$L^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)} du dv$$

en overgang op poolcoördinaten r en φ via de transformatieformules $u=r \cos \varphi$ en $v=r \sin \varphi$ geeft daar $r^2=u^2+v^2$ en $du dv = r dr d\varphi$

$$L^2 = \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^{\infty} r e^{-\frac{1}{2} r^2} dr = \int_0^{\infty} e^{-\frac{1}{2} r^2} d\left(\frac{1}{2} r^2\right) = 1.$$

2. Twee- en meerdimensionale kansverdelingen.

Definitie 2;1

Twee stochastische grootheden $\underline{x}_1$ en $\underline{x}_2$ bezitten een simultane kansverdeling als voor ieder stel reële getallen x_1 en x_2 de (simultane) verdelingsfunctie

$$(2;1) \quad F(x_1, x_2) \stackrel{\text{def}}{=} P[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 \leq x_2]$$

gedefiniëerd is.

Voorbeelden

Het aantal worpen met een munt tot voor de tweede maal kruis verkregen wordt en het nummer van de worp, waarbij in die rij worpen de eerste keer kruis viel. De lichaamslengte en de borstomvang van een aselekt uit de Nederlandse mannelijke bevolking genomen man. De levertijd van een artikel en het aantal binnenkomende bestellingen per week voor dit artikel. Het aantal klanten dat zich per dag bij een loket meldt en de bedieningstijd per klant.

Definitie 2;2

Een discrete twee-dimensionale kansverdeling is een verdeling, waarbij in de twee-dimensionale ruimte R_2 met x_1 en x_2 als coördinaten een eindig of aftelbaar aantal punten $P_j = (x_{1,j}, x_{2,j})$ ($j=1,2,\dots$) aanwezig zijn, waarvoor geldt

$$(2;2) \quad P[P = P_j] = p_j > 0, \quad \sum_j p_j = 1,$$

waarin $\underline{x} = (\underline{x}_1, \underline{x}_2)$ een stochastisch punt in R_2 voorstelt.

De bijbehorende verdelingsfunctie is dan een terrasfunctie:

$$(2;3) \quad F(x_1, x_2) = \sum_{\substack{j \\ x_{1,j} \leq x_1 \wedge x_{2,j} \leq x_2}} p_j$$

Voorbeeld 2;1

Stellen $\underline{x}_1$ resp. $\underline{x}_2$ de nummers voor van de worpen, in een rij onafhankelijke worpen met een munt, waarbij voor het eerst resp. voor de tweede maal kruis verkregen wordt, dan is, zoals gemakkelijk te zien is, als de kans op kruis p is:

$$(2;4) \quad P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2] = p^2 q^{x_2-2} \quad (x_2 > x_1).$$

Deze kans is dus voor iedere $x_1 < x_2$ dezelfde. Dit doet ver-

moeden, dat bij gegeven x_2 alle mogelijke waarden van x_1 dezelfde kans zullen bezitten. Dat dit zo is kan als volgt bewezen worden (vgl. 1;8):

$$(2;5) \quad \begin{aligned} P[\underline{x}_1 = x_1 | \underline{x}_2 = x_2] &= \frac{P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2]}{P[\underline{x}_2 = x_2]} = \\ &= \frac{p^2 q^{x_2-2}}{(x_2-1)p^2 q^{x_2-2}} = \frac{1}{x_2-1} \end{aligned}$$

Onder de voorwaarde $\underline{x}_2 = x_2$ volgt $\underline{x}_1$ dus de discrete homogene verdeling over de getallen $1, 2, \dots, x_2-1$.

Definitie 2;3

Een continue twee-dimensionale kansverdeling is een kansverdeling, waarvoor de functie

$$(2;6) \quad f(x_1, x_2) \stackrel{\text{def}}{=} \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}$$

overal bestaat en continu is (discontinuïteit van f in eindig veel punten of op één of meer krommen is toegestaan; een exacte formulering hiervan zou te ver voeren). De functie f heet de (simultane) verdelingsdichtheid. De omgekeerde relatie van (2;6) is

$$(2;7) \quad F(x_1, x_2) = \iint_{\substack{v_1 \leq x_1 \\ v_2 \leq x_2}} f(v_1, v_2) dv_1 dv_2$$

Is G een gebied in de ruimte R_2 met coördinaten x_1 en x_2 en is $\underline{P} = (\underline{x}_1, \underline{x}_2)$, dan geldt

$$(2;8) \quad P[P \in G] = \iint_G f(x_1, x_2) dx_1 dx_2$$

Het bewijs van deze relatie is elementair als G een rechthoek is. Is G van een andere vorm, dan leidt overdekking van G met inkrimpende rechthoeken en limietovergang tot het doel. Dit maattheoretische bewijs en de eisen, die aan G opgelegd moeten worden, behandelen wij hier niet.

Voorbeeld 2;2

Twee grootheden $\underline{x}$ en $\underline{y}$ bezitten een tweedimensionale normale verdeling, als hun verdelingsdichtheid de volgende vorm heeft:

$$(2;9) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \frac{x-\mu_1}{\sigma_1} \frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right\}}$$

Hierin komen 5 parameters voor: $\mu_1, \mu_2, \sigma_1, \sigma_2$ en ρ .

Deze verdeling kan vaak als model gebruikt worden indien aan één aselekt uit een grote verzameling gekozen object twee afmetingen waargenomen worden, bijv. lengte en breedte van een meeuwen-ei.

Marginale verdelingen.

Bezitten $\underline{x}_1$ en $\underline{x}_2$ een twee-dimensionale discrete verdeling, dan geldt zowel voor $i=1$ als $i=2$

$$(2;10) \quad F_i(x_i) \stackrel{\text{def}}{=} P[x_i \leq x_i] = \sum_{x_{ij} \leq x_i} p_{ij}.$$

$F_i(x_i)$ heet de marginale verdelingsfunctie van $\underline{x}_i$.

De marginale kansen worden gegeven door

$$(2;11) \quad P[x_i = x_i] = \sum_{x_{ij} = x_i} p_{ij}.$$

Voorbeeld 2;3

De marginale verdeling van $\underline{x}_1$, behorende bij de door (2;4) gegeven verdeling, is volgens (2;11):

$$\begin{aligned} P[x_1 = x_1] &= \sum_{x_2=0}^{\infty} p^2 q^{x_1+x_2} = \sum_{x_2=0}^{\infty} p^2 q^{x_1+x_2+0} = \\ &= p q^{x_1-1} \sum_{i=1}^{\infty} p q^{i-1} = p q^{x_1-1}, \end{aligned}$$

daar $\sum_{i=1}^{\infty} p q^{i-1} = 1$. Deze marginale verdeling is dus dezelfde als

(1;8) met $k=1$. (Was dit ook direct in te zien? De lezer berekene de marginale verdeling van $\underline{x}_2$).

Bezitten $\underline{x}_1$ en $\underline{x}_2$ een twee-dimensionale continue verdeling, dan geldt volgens (2;8) voor de marginale verdelingsfunctie van $\underline{x}_1$:

$$(2;12) \quad F_1(x_1) = \int_{v_1=-\infty}^{x_1} \int_{v_2=-\infty}^{+\infty} f(v_1, v_2) dv_1 dv_2.$$

De marginale verdelingsdichtheid is dus

$$(2;13) \quad f_1(x_1) = \frac{dF_1(x_1)}{dx_1} = \int_{-\infty}^{+\infty} f(x_1, v_2) dv_2.$$

Voor de marginale verdelingsfunctie en de marginale verdelingsdichtheid van $\underline{x}_2$ gelden analoge formules.

Voorbeeld 2;4

De marginale verdeling van $\underline{x}$ behorende bij (2;9) laat zich eenvoudig als volgt berekenen. Stellen wij

$$(2;14) \quad u = \frac{x - \mu_1}{\sigma_1} \quad \text{en} \quad v = \frac{y - \mu_2}{\sigma_2}$$

dan is $dx = \sigma_1 du$ en $dy = \sigma_2 dv$.

Substitueren wij dit in (2;9) en stelt men voor: "op een constante factor na gelijk aan", dan is dus

$$f(x,y) = e^{-\frac{1}{2\sigma_1^2\sigma_2^2} (u^2 - 2\rho uv + v^2)}$$

Volgens (2;13) geldt dus voor de marginale verdelingsdichtheid $f_1(x)$ van $\underline{x}$:

$$\begin{aligned} f_1(x) &:: e^{-\frac{u^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_1^2\sigma_2^2} (u^2 - 2\rho uv + v^2)} \sigma_2 dv = \\ &= e^{-\frac{1}{2} \frac{u^2}{\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_1^2\sigma_2^2} (v - \rho u)^2} \sigma_2 dv \end{aligned}$$

Door substitutie van $w = v - \rho u$ in de integraal blijkt, dat deze - daar de grenzen door de substitutie niet veranderen - niet meer van u afhangt, zodat dus geldt

$$f_1(x) :: e^{-\frac{1}{2} \frac{u^2}{\sigma_1^2}} = e^{-\frac{1}{2} \frac{(x - \mu_1)^2}{\sigma_1^2}}$$

Dit betekent echter, dat de marginale verdeling van $\underline{x}$ een $N(\mu_1, \sigma_1)$ -verdeling is. De nog ontbrekende factor kan aangevuld worden met behulp van de voorwaarde (1;14), die uiteraard ook voor marginale verdelingen geldt. Uit formule (1;18) blijkt, dat het resultaat moet zijn

$$(2;15) \quad f_1(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x - \mu_1)^2}{\sigma_1^2}}$$

Een analoge formule geldt voor $f_2(y)$, de marginale verdelingsdichtheid van $\underline{y}$.

Voorwaardelijke verdelingen.

Bij discrete verdelingen kunnen de voorwaardelijke verdelingen direct volgens de definitie van voorwaardelijke kansen berekend worden. Een voorbeeld hiervan was (2;5). In het continue geval doet zich de complicatie voor, dat de voorwaarde, waaronder men werkt, zelf een kans 0 bezit, zodat de definitie niet zonder meer bruikbaar is. Bezitten $\underline{x}_1$ en $\underline{x}_2$ een continue twee-dimensionale verdeling,

dan is dus $P[\underline{x}_2 = x_2] = 0$ voor iedere x_2 , zodat voorwaardelijke kansen van de vorm

$$P[\underline{x}_1 \leq x_1 | \underline{x}_2 = x_2]$$

vooralsnog niet gedefinieerd zijn. Deze kans wordt nu gedefinieerd als

$$(2;16) \quad F(x_1 | x_2) = P[\underline{x}_1 \leq x_1 | \underline{x}_2 = x_2] = \lim_{\Delta x_2 \downarrow 0} P[\underline{x}_1 \leq x_1 | x_2 \leq x_2 \leq x_2 + \Delta x_2]$$

Volgens de definitie van een voorwaardelijke kans en (2;8) is dit gelijk aan

$$\lim_{\Delta x_2 \downarrow 0} \left\{ \int_{-\infty}^{x_1} dv_1 \int_{x_2}^{x_2 + \Delta x_2} dv_2 f(v_1, v_2) \right\} / \left\{ \int_{-\infty}^{+\infty} dv_1 \int_{x_2}^{x_2 + \Delta x_2} dv_2 f(v_1, v_2) \right\}$$

en hierin is, na deling van teller en noemer door Δx_2 :

$$\lim_{\Delta x_2 \downarrow 0} \int_{-\infty}^{x_1} dv_1 \int_{x_2}^{x_2 + \Delta x_2} dv_2 f(v_1, v_2) / \Delta x_2 = \int_{-\infty}^{x_1} f(v_1, x_2) dv_1 = \frac{\partial}{\partial x_2} \int_{-\infty}^{x_1} \int_{-\infty}^{+\infty} f(v_1, v_2) dv_1 dv_2 = \frac{\partial}{\partial x_2} F(x_1, x_2),$$

terwijl voor de noemer geldt volgens (2;12)

$$\begin{aligned} \lim_{\Delta x_2 \downarrow 0} \int_{-\infty}^{+\infty} dv_1 \int_{x_2}^{x_2 + \Delta x_2} dv_2 f(v_1, v_2) / \Delta x_2 &= \int_{-\infty}^{+\infty} f(v_1, x_2) dv_1 = \frac{\partial}{\partial x_2} \int_{-\infty}^{+\infty} dv_1 \int_{-\infty}^{+\infty} dv_2 f(v_1, v_2) = \\ &= \frac{\partial}{\partial x_2} F_2(x_2) = f_2(x_2), \end{aligned}$$

als F_2 en f_2 de marginale verdelingsfunctie resp. -dichtheid van $\underline{x}_2$ voorstellen. Derhalve is

$$(2;17) \quad F(x_1 | x_2) = \frac{\frac{\partial}{\partial x_2} F(x_1, x_2)}{f_2(x_2)}$$

en

$$(2;18) \quad f(x_1 | x_2) = \frac{\partial}{\partial x_1} F(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}.$$

(De lezer controle of $f(x_1 | x_2)$ inderdaad een kansverdeling is, d.w.z. of $\int_{-\infty}^{+\infty} f(x_1 | x_2) dx_1 = 1$ is).

Voorbeeld 2;5

Wij berekenen nu de voorwaardelijke verdelingsdichtheid $f(x|y)$ van de twee-dimensionale normale verdeling (2;9). Deze is, volgens (2;18) gelijk aan

$$\frac{f(x, y)}{f_2(y)}$$

waarin, volgens (2;15)

$$f_2(y) = e^{-\frac{1}{2} \frac{(y - \mu_2)^2}{\sigma_2^2}}$$

is ($::$ betekent: op een constante factor na gelijk aan).

Voeren wij weer de substitutie (2;14) uit, dan komt er

$$\begin{aligned} f(x|y) &:: e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) + \frac{1}{2}v^2} \\ &= e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + \rho^2 v^2)} = e^{-\frac{1}{2(1-\rho^2)}(u - \rho v)^2} \\ &= e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x - \mu_1}{\sigma_1} - \rho \frac{y - \mu_2}{\sigma_2}\right)^2} \\ &= e^{-\frac{\left|x - \left\{\rho \frac{\sigma_1}{\sigma_2}(y - \mu_2) + \mu_1\right\}\right|^2}{2(1-\rho^2)\sigma_1^2}} \end{aligned}$$

waarin nu y , wegens de voorwaarde $\underline{y}=y$, als een gegeven constante optreedt. Derhalve volgt uit (1;18):

$$(2;19) \quad f(x|y) = \frac{1}{\sigma_1 \sqrt{2\pi(1-\rho^2)}} e^{-\frac{\left|x - \left\{\rho \frac{\sigma_1}{\sigma_2}(y - \mu_2) + \mu_1\right\}\right|^2}{2(1-\rho^2)\sigma_1^2}}.$$

In sommige praktijksituaties zijn van een twee-dimensionale verdeling $F(x_1, x_2)$ de marginale verdeling $F_1(x_1)$ en de voorwaardelijke verdeling $F(x_2|x_1)$ gegeven, terwijl de marginale verdeling $F_2(x_2)$ berekend moet worden.

Voor een discrete verdeling, waarin $\underline{x}_1$ de waarden $1, \dots, m$ en $\underline{x}_2$ de waarden $1, \dots, n$ aan kan nemen, geschiedt een dergelijke berekening met de formule (vgl. I;3;17).

$$(2;20) \quad P[\underline{x}_2 = j] = \sum_{i=1}^m P[\underline{x}_1 = i \wedge \underline{x}_2 = j] = \sum_{i=1}^m P[\underline{x}_1 = i] P[\underline{x}_2 = j | \underline{x}_1 = i]$$

Bij continue verdelingen kan de gezochte marginale verdelingsdichtheid gevonden worden via (vgl. (2;13) en (2;18)):

$$(2;21) \quad f_2(x_2) = \int_{-\infty}^{+\infty} f(v_1, x_2) dv_1 = \int_{-\infty}^{+\infty} f_1(v_1) f(x_2 | v_1) dv_1$$

Een voorbeeld van een dergelijke berekening volgt hier voor het geval de ene marginale verdeling discreet is en de andere continu.

Voorbeeld 2;6

Een leverancier verkoopt een artikel uit voorraad aan zijn klanten, terwijl hij het artikel zelf op order inkoopt bij een fabriek. De frequenties van de levertijden van de fabriek naar de leverancier sluiten goed aan bij een Erlang-verdeling. De verdelingsfunctie van het aantal klanten in een periode van T maanden is een Poissonverdeling met parameter μT . Hoe groot is de kans dat er tijdens de levertijd van de volgende bestelling hoogstens n klanten komen?

Oplossing

De verdelingsfunctie van de levertijd is (zie (1;26))

$$(2;22) \quad F(T) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^T e^{-\lambda t} t^{\alpha-1} dt \quad (\alpha \text{ geheel positief})$$

en de verdelingsfunctie van het aantal klanten dat tijdens een vaste levertijd van T maanden komt is (zie (1,9)):

$$(2;23) \quad P[k \leq n | t = T] = \sum_{k=0}^n \frac{(\mu T)^k e^{-\mu T}}{k!}$$

Om de verdelingsfunctie van het aantal klanten in de komende levertijd te vinden, moeten wij volgens (2;21) de in (2;23) opgegeven kans integreren over T; deze verdelingsfunctie is dus (vgl. (1;26)):

$$\begin{aligned} (2;24) \quad P[k \leq n] &= \int_0^\infty \left(\sum_{k=0}^n \frac{(\mu t)^k e^{-\mu t}}{k!} \right) f(t) dt = \\ &= \sum_{k=0}^n \int_0^\infty \frac{(\mu t)^k e^{-\mu t}}{k!} \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda t} t^{\alpha-1} dt = \\ &= \sum_{k=0}^n \frac{\mu^k \lambda^\alpha}{k! \Gamma(\alpha)} \int_0^\infty e^{-(\lambda+\mu)t} t^{\alpha+k-1} dt = \sum_{k=0}^n \frac{\mu^k \lambda^\alpha}{k! \Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+k)}{(\lambda+\mu)^{\alpha+k}} = \\ &= \sum_{k=0}^n \binom{\alpha+k-1}{\alpha-1} \cdot \left(\frac{\lambda}{\lambda+\mu} \right)^\alpha \left(\frac{\mu}{\lambda+\mu} \right)^k \end{aligned}$$

Uit een vergelijking van het hier gevonden resultaat met (1;8) blijkt dat (2;24) een negatief binomiale verdeling voorstelt met parameters α en $\frac{\lambda}{\lambda+\mu}$. In een geval, waarin de levertijden een Erlang-verdeling volgen en het aantal klanten in een periode van T maanden een Poissonverdeling volgt met parameter μT , is de onvoorwaardelijke verdeling van het aantal klanten tijdens de levertijd dus een negatief binomiale verdeling.

Alle in deze paragraaf ingevoerde begrippen kunnen uitgebreid worden tot ruimten van n dimensies. Wij zullen dit hier niet behandelen en volstaan met het volgende voorbeeld.

Voorbeeld 2;7

Laat $x_i (i=1,2,\dots,n)$ het aantal malen kruis voorstellen bij de i^e worp uit een rij onafhankelijke worpen met een munt. Iedere x_i kan dus de waarde 0 of 1 aannemen en de kansen hierop stellen wij q resp. p . Dan geldt dus

$$P[x_i=1]=p, \quad P[x_i=0]=q \quad (i=1,2,\dots,n)$$

en wegens de onafhankelijkheid

$$(2;25) \quad P[x_1=x_1 \wedge x_2=x_2 \wedge \dots \wedge x_m=x_m] = p^{\sum x_i} q^{n-\sum x_i},$$

waarin $x_i (i=1,2,\dots,n)$ de waarde 0 of 1 bezit. Stellen wij

$$(2;26) \quad x = \sum_{i=1}^n x_i,$$

dan is x het aantal malen kruis bij n worpen, waarvoor dus (1;6), de binomiale verdeling, geldt.

Deze opbouw van een binomiaal verdeelde grootheid x met parameters n en p als som van een aantal onafhankelijke grootheden x_i met dezelfde p en $n=1$ zal ons later nog te pas komen.

3. Onafhankelijkheid van stochastische grootheden.

In paragraaf I,4 is het begrip stochastische onafhankelijkheid gedefiniëerd voor gebeurtenissen, door middel van de productformules (4;8) en (4;13). Aequivalent hiermee is, dat voorwaardelijke en onvoorwaardelijke kansen gelijk zijn (formule (4;7)). Wij sluiten hier bij die definities aan, waarbij dan de beschouwde gebeurtenissen zijn: $\underline{x}_1 \leq x_1, \underline{x}_2 \leq x_2, \text{ enz.}$

Definitie 3;1

Een aantal stochastische grootheden $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ zijn onafhankelijk verdeeld (stochastische onafhankelijkheid) indien zij een simultane kansverdeling bezitten, waarvan de verdelingsfunctie voldoet aan

$$(3;1) \quad F(x_1, x_2, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n),$$

waarin F de simultane verdelingsfunctie is en F_i ($i=1, 2, \dots, n$) de marginale verdelingsfunctie van $\underline{x}_i$.

Opmerking

Vergelijking (3;1) is identiek met:

$$(3;2) \quad P[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 \leq x_2 \wedge \dots \wedge \underline{x}_n \leq x_n] = \prod_{i=1}^n P[\underline{x}_i \leq x_i].$$

Voor de eenvoud beschouwen wij verder in het bijzonder het geval van een twee-dimensionale verdeling van stochastische grootheden $\underline{x}_1$ en $\underline{x}_2$. De uitbreiding tot meer dan 2 dimensies is gemakkelijk aan te geven en de bewijzen verlopen analoog. De formules (3;1) en (3;2) gaan voor $n=2$ over in

$$(3;1') \quad F(x_1, x_2) = F_1(x_1) F_2(x_2)$$

en

$$(3;2') \quad P[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 \leq x_2] = P[\underline{x}_1 \leq x_1] P[\underline{x}_2 \leq x_2].$$

Stelling 3;1

Zijn $\underline{x}_1$ en $\underline{x}_2$ stochastisch onafhankelijk, dan geldt

$$(3;3) \quad P[\underline{x}_1 \leq x_1 | \underline{x}_2 \leq x_2] = P[\underline{x}_1 \leq x_1].$$

Bewijs

Volgens de definitie van een voorwaardelijke kans, tezamen met (3;1'), geldt

$$P[\underline{x}_1 \leq x_1 | \underline{x}_2 \leq x_2] = \frac{F(x_1, x_2)}{F_2(x_2)} = \frac{F_1(x_1) F_2(x_2)}{F_2(x_2)} = F_1(x_1) = P[\underline{x}_1 \leq x_1].$$

Stelling 3;2

Zijn $\underline{x}_1$ en $\underline{x}_2$ discreet en onafhankelijk verdeeld, dan geldt

$$(3;4) \quad \begin{cases} P[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 = x_2] = P[\underline{x}_1 \leq x_1] P[\underline{x}_2 = x_2], \\ P[\underline{x}_1 = x_1 \wedge \underline{x}_2 \leq x_2] = P[\underline{x}_1 = x_1] P[\underline{x}_2 \leq x_2] \end{cases}$$

en

$$(3;5) \quad P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2] = P[\underline{x}_1 = x_1] P[\underline{x}_2 = x_2].$$

Bewijs

Is $x_2' < x_2$ de grootste waarde $< x_2$, die door $\underline{x}_2$ aangenomen kan worden, dan is $P[\underline{x}_2 < x_2] = P[\underline{x}_2 \leq x_2']$ en daar (3;2') voor alle mogelijke waarden van x_1 en x_2 geldt is dus tevens

$$(3;6) \quad P[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 < x_2] = P[\underline{x}_1 \leq x_1] P[\underline{x}_2 < x_2].$$

Is er geen dergelijke waarde x_2' , dan zijn linker- en rechterlid van (3;6) beide = 0, zodat de geldigheid behouden blijft.

Laat nu

A_1 resp. A_2 voorstellen $\underline{x}_1 \leq x_1$ resp. $\underline{x}_2 \leq x_2$,

B_1 resp. B_2 $\underline{x}_1 < x_1$ resp. $\underline{x}_2 < x_2$,

C_1 resp. C_2 $\underline{x}_1 = x_1$ resp. $\underline{x}_2 = x_2$.

Dan is $A_2 \equiv B_2 \vee C_2$, dus $A_1 \wedge A_2 \equiv (A_1 \wedge B_2) \vee (A_1 \wedge C_2)$, zodat

$$(3;7) \quad \begin{aligned} P[A_1 \wedge A_2] &= P[A_1] P[A_2] = P[A_1] P[B_2 \vee C_2] = \\ P[A_1] \{P[B_2] + P[C_2]\} &= P[A_1] P[B_2] + P[A_1] P[C_2]. \end{aligned}$$

Anderzijds is volgens (3;6):

$$P[A_1 \wedge B_2] = P[A_1] P[B_2],$$

zodat

$$\begin{aligned} P[A_1 \wedge A_2] P[(A_1 \wedge B_2) \vee (A_1 \wedge C_2)] &= P[A_1 \wedge B_2] + P[A_1 \wedge C_2] = \\ P[A_1] P[B_2] + P[A_1 \wedge C_2]. \end{aligned}$$

Uit deze relatie en (3;7) volgt direct, dat

$$(3;8) \quad P[A_1 \wedge C_2] = P[A_1] P[C_2],$$

hetgeen overeenkomt met de eerste der twee vergelijkingen (3;4). De tweede wordt uiteraard op dezelfde wijze bewezen.

Om (3;5) te bewijzen ontwikkelen wij nu $P[A_1 \wedge C_2]$ op analoge wijze met behulp van $A_1 = B_1 \vee C_1$. Dit geeft:

$$\begin{aligned} P[A_1 \wedge C_2] &= P[(B_1 \wedge C_2) \vee (C_1 \wedge C_2)] = P[B_1 \wedge C_2] + P[C_1 \wedge C_2] = \\ &= P[B_1] P[C_2] + P[C_1 \wedge C_2], \end{aligned}$$

daar (3;4) niet alleen geldt met $\underline{x}_1 \equiv x_1$, maar ook met $\underline{x}_1 \equiv x'_1$ als x'_1 de grootste waarde $< x_1$ is, die $\underline{x}_1$ aan kan nemen (vergelijk (3;6)). Anderzijds is volgens (3;8)

$$\begin{aligned} P[A_1 \wedge C_2] &= P[C_2] P[A_1] = P[C_2] \{P[B_1] + P[C_1]\} = \\ &= P[B_1] P[C_2] + P[C_1] P[C_2], \end{aligned}$$

wat, tezamen met de vorige relatie, leidt tot

$$P[C_1 \wedge C_2] = P[C_1] P[C_2]$$

dus tot (3;5).

Stelling 3;3

Zijn $\underline{x}_1$ en $\underline{x}_2$ continu en onafhankelijk verdeeld, dan geldt

$$(3;9) \quad f(x_1, x_2) = f_1(x_1) f_2(x_2),$$

waarin f de simultane en f_1 en f_2 de marginale verdelingsdichtheden voorstellen.

Bewijs

Deze stelling volgt onmiddellijk uit definitie 2;3

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 \{F_1(x_1) F_2(x_2)\}}{\partial x_1 \partial x_2} = f_1(x_1) f_2(x_2).$$

Opmerkingen

In paragraaf I;4 is opgemerkt, dat de bijzondere productregel (4;14) niet voldoende is om stochastische onafhankelijkheid te garanderen. Daartoe was (4;13) nodig. In definitie 3;1 ligt echter een analoge relatie als (4;13) opgesloten. Immers vult men in (3;1) bijv. in $x_n = \infty$, dan gaat deze relatie over in dezelfde relatie, maar nu voor de eerste $n-1$ stochastische grootheden. Wij hebben hier dus aan (3;1) genoeg voor de definitie van stochastische onafhankelijkheid. Anderzijds kan men door een tegenvoorbeeld aantonen dat voor meer dan twee stochastische grootheden paarsgewijze onafhankelijkheid niet voldoende is voor volledige onafhankelijkheid.

De relaties (3;5) en (3;9) (voor n dimensies opgeschreven) kunnen ook dienen om stochastische onafhankelijkheid te definieren. Dan kan (3;1) daaruit gemakkelijk afgeleid worden. De hierboven gevolgde methode bezit het voordeel, dat het continue en het discrete geval (en ook het hier niet behandelde gemengde geval) in één universele formule samengevat worden.

Uit de stellingen 3;2 en 3;3 volgt, dat voor onafhankelijke grootheden geldt:

$$(3;10) \quad P[x_1 = x_1 | x_1 = x_1] = P[x_1 = x_1] \quad (\text{discontinue geval})$$

en (volgens (2;18))

$$(3;11) \quad f(x_1 | x_1) \cdot f(x_1) \quad (\text{continue geval}).$$

Of, algemeen gezegd, bij stochastisch onafhankelijke grootheden is de voorwaardelijke verdeling van een aantal daarvan, onder voorwaarden, die alleen op de overige grootheden betrekking hebben, gelijk aan de marginale verdeling van de eerstgenoemde grootheden.

Voorbeelden

Bij vele experimentele onderzoeken richt men de experimenten zo in, dat in het wiskundige model met onafhankelijke stochastische grootheden gewerkt kan worden. De wiskundige moeilijkheden worden daardoor sterk verminderd, omdat dan de eenvoudige relaties, die in dit hoofdstuk zijn afgeleid, gelden. Als voorbeeld van de vereenvoudigingen, die optreden, beschouwen wij de twee-dimensionale normale verdeling (2;9). De marginale verdelingen van $\underline{x}$ en $\underline{y}$ zijn, volgens (2;15)

$$f_1(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x - \mu_1)^2}{\sigma_1^2}} \quad \text{en} \quad f_2(y) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y - \mu_2)^2}{\sigma_2^2}}.$$

Zijn $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk, dan geldt dus volgens (3;9)

$$(3;12) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2} \left\{ \frac{(x - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right\}},$$

hetgeen overeenkomt met (2;9) wanneer daarin $\rho=0$ gesteld wordt. De simultane verdelingsdichtheid is nu dus veel eenvoudiger dan in het algemene geval: ten eerste is er één parameter minder en

ten tweede kunnen allerlei berekeningen gemakkelijker uitgevoerd worden, omdat $f(x,y)$ uit twee factoren bestaat, die ieder slechts één der beide grootheden bevatten. Voor meer dan 2 grootheden is de vereenvoudiging van dezelfde aard, terwijl het effect bij het oplossen van vele problemen nog sterker is.

Dat $\underline{x}$ en $\underline{y}$, als zij een simultane normale verdeling bezitten, dan en slechts dan onafhankelijk verdeeld zijn als $\rho=0$ is, volgt ook uit de formules (2;15) en (2;19), die de marginale en de voorwaardelijke verdelingen van $\underline{x}$ voorstellen. Deze zijn dan en slechts dan identiek, als $\rho=0$ is.

Een tweede voorbeeld van stochastisch onafhankelijke grootheden verkrijgt men door de resultaten te beschouwen van n op dezelfde wijze uitgevoerde worpen met een dobbelsteen. In dat geval kan men de worp-resultaten aangeven met $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ en het ligt voor de hand, daar een dobbelsteen geen geheugen heeft, als modelveronderstelling in te voeren, dat deze grootheden stochastisch onafhankelijk zijn en alle dezelfde (marginale) verdeling bezitten.

Indien wij voor de eenvoud veronderstellen, dat de dobbelsteen zuiver is, en indien $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ een willekeurig stel mogelijke uitkomsten is, geldt volgens (3;5)

$$(3;13) \quad P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \dots \wedge \underline{x}_n = x_n] = \prod_{i=1}^n P[\underline{x}_i = x_i] = 6^{-n}$$

en deze kans is dus voor alle mogelijke n -tallen uitkomsten dezelfde.

Niet alleen bij worpen met dobbelstenen of munten is de onderstelling dat de experimenten stochastisch onafhankelijk zijn vaak gerechtvaardigd. Wanneer men steekproeven zorgvuldig kiest (bijvoorbeeld met behulp van aselechte getallen (vgl. voorbeeld 5;3 in hoofdstuk I)), dan kan men veelal uitgaan van de onderstelling dat de meetresultaten door stochastisch onafhankelijke grootheden voorgesteld kunnen worden. Terloops merken wij hierbij op, dat men een dergelijke onderstelling niet klakkeloos behoeft te aanvaarden, doch experimenteel kan onderzoeken. De statistiek beschikt dan over de middelen om op grond van waarnemingsresultaten de gemaakte onderstelling op zijn juistheid te toetsen.

In sommige gevallen is het mogelijk door een geschikte trans-

formatie stochastisch afhankelijke grootheden te vervangen door stochastisch onafhankelijke grootheden. Wij besluiten deze paragraaf met een illustratie van een dergelijke transformatie.

Beschouw een reeks worpen met een munt. Worden deze telkens op dezelfde wijze uitgevoerd, onafhankelijk van de bij vorige worpen verkregen resultaten, dan kan men de worpresultaten als stochastisch onafhankelijk beschouwen, met eenzelfde kans p op kruis voor iedere worp. Stellen nu $\underline{x}_1, \underline{x}_2$, enz. de nummers voor van de worpen, waarbij voor de eerste maal, voor de tweede maal, enz. kruis verkregen wordt, dan geldt uiteraard

$$(3;14) \quad P[\underline{x}_1 < \underline{x}_2 < \dots] = 1,$$

daar de tussen de vierkante haken genoteerde eventualiteit zeker is. Deze stochastische grootheden zijn dus stochastisch afhankelijk, daar de waarden, die bijv. $\underline{x}_2$ aan kan nemen, afhangen van de waarde, die $\underline{x}_1$ aanneemt (en andersom). Beschouwt men echter de grootheden:

$$(3;15) \quad \underline{y}_1 = \underline{x}_1, \quad \underline{y}_2 = \underline{x}_2 - \underline{x}_1, \quad \underline{y}_3 = \underline{x}_3 - \underline{x}_2, \quad \text{enz.},$$

dan zijn deze grootheden weer onafhankelijk en zij bezitten alle dezelfde negatief binomiale verdeling (formule (1;8), hier met $k=1$). Dit volgt ogenblikkelijk uit de overweging, dat na iedere worp, waarbij kruis verkregen is, een "nieuwe" rij van onafhankelijke worpen "begint", waarop formule (1;8) van toepassing is. Bij verschillende berekeningen omtrent de kansverdeling van $\underline{x}_k$ (met $k > 1$) is het dan ook eenvoudiger om van de $\underline{y}$'s uit te gaan.

4. Mathematische verwachting en momenten

Het begrip mathematische verwachting, dat in deze paragraaf wordt besproken, is zowel voor de theorie als voor de toepassingen van groot belang. De momenten van een kansverdeling, die o.a. gebruikt kunnen worden om een globale beschrijving van ligging en vorm daarvan te geven, zijn speciale gevallen.

Wij behandelen eerst het ééndimensionale geval.

Definitie 4;1

Is $\underline{x}$ een discreet verdeelde stochastische grootheid, die de waarden $x_1, x_2, \dots$ aan kan nemen, met

$$(4;1) \quad p_i = P[\underline{x} = x_i] \quad (\sum p_i = 1)$$

en is $\varphi(x)$ een functie van x , dan wordt de mathematische verwachting van $\varphi(\underline{x})$ - notatie: $\mathbb{E}\varphi(\underline{x})$ - gedefinieerd door

$$(4;2) \quad \mathbb{E}\varphi(\underline{x}) \stackrel{\text{def}}{=} \sum p_i \varphi(x_i).$$

Bezit $\underline{x}$ een continue verdeling met verdelingsdichtheid $f(x)$, dan luidt de definitie

$$(4;3) \quad \mathbb{E}\varphi(\underline{x}) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \varphi(x) f(x) dx$$

Opmerking

De functie $\varphi(\underline{x})$ van $\underline{x}$ is, daar $\underline{x}$ stochastisch is, zelf in het algemeen ook een stochastische grootheid. De verwachting $\mathbb{E}^1$ is een "operator", die deze stochastische grootheid in een getal omzet. Immers de rechterleden van (4;2) en (4;3) zijn getallen.

Voor gemengde kansverdelingen (vergelijk blz.10) is $\mathbb{E}\varphi(\underline{x})$ een lineaire combinatie, met coëfficiënten λ en $(1-\lambda)$ van de rechterleden van (4;2) en (4;3).

Definitie 4;2

Het k^{e} moment ($k=0,1,2,\dots$) van een stochastische grootheid $\underline{x}$ is de verwachting van $\underline{x}^k$.

Notatie: μ_k of, indien nodig, $\mu_k(\underline{x})$. Dus

1) Ter verkorting wordt de toevoeging "mathematische" vaak weggelaten.

$$(4;4) \quad \mu_k \stackrel{\text{def}}{=} \sum x^k$$

Opmerkingen

Voor iedere kansverdeling geldt: $\mu_0=1$. Stellen wij de kansverdeling van $\underline{x}$ aanschouwelijk voor als een massaverdeling over de x -as, in het discrete geval met massapunten van massa p_i in de punten x_i en in het continue geval als een continu aangebrachte massa met dichtheid $f(x)$, dan stelt μ_0 de totale massa voor. Het eerste moment μ_1 is de abscis van het zwaartepunt van de massaverdeling en μ_2 is het traagheidsmoment om de oorsprong. De beschrijvende waarde van deze momenten voor de kansrekening is dezelfde als die in de mechanica. Aan deze analogie is het gebruik van de term "momenten" toe te schrijven.

Bij μ_1 wordt de index 1 vaak weggelaten.

Voorbeeld 4;1

Indien $\underline{x}$ kansen p resp. q ($p+q=1$) bezit op de waarden 0 resp. 1 (vgl. (1;5) en fig. 1;1), dan geldt

$$(4;5) \quad \mu_k = q \quad \text{voor } k=1,2,\dots$$

Immers dan is

$$\mu_k = q \cdot 1^k + p \cdot 0^k = q \quad (k > 0)$$

Voorbeeld 4;2

Indien $\underline{x}$ een negatief binomiale verdeling bezit met $k=1$ (vgl. (1;8)), dan is

$$P[\underline{x}=x] = pq^{x-1} \quad (x=1,2,\dots)$$

en dus

$$\sum_{x=1}^{\infty} x P[\underline{x}=x] = \sum_{x=1}^{\infty} x pq^{x-1} = p \sum_{x=1}^{\infty} x q^{x-1}$$

Nu is

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}$$

en nemen wij links en rechts de afgeleide, dan vinden wij

$$\sum_{x=1}^{\infty} x q^{x-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

Derhalve is

$$(4;6) \quad \mu = \mathcal{E} \underline{x} = \frac{1}{p}$$

Voorbeeld 4;3

Is $\underline{x}$ normaal verdeeld met parameters μ en σ (vgl. (1;18)), dan is

$$\mathcal{E} \underline{x} = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx$$

Substitueer $u = \frac{x-\mu}{\sigma}$: $x = u\sigma + \mu$, $dx = \sigma du$.

$$\begin{aligned} \mathcal{E} \underline{x} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (u\sigma + \mu) e^{-\frac{1}{2} u^2} du = \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-\frac{1}{2} u^2} du + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} u^2} du \end{aligned}$$

De eerste term valt weg wegens symmetrie van het argument onder de integraal en de tweede term is volgens de appendix bij paragraaf 1 gelijk aan μ . Dus is

$$(4;7) \quad \mathcal{E} \underline{x} = \mu$$

Wij merken op dat de parameter μ van de normale verdeling dus gelijk is aan het eerste moment van die verdeling, of zoals men gemakshalve vaak zegt, gelijk is aan "de verwachting" van de normale verdeling. Bij wijze van illustratie berekenen wij ook nog het tweede moment $\mathcal{E} \underline{x}^2$; dit kan als volgt gedaan worden:

$$\begin{aligned} \mu_2 &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u + \mu)^2 e^{-\frac{1}{2} u^2} du = \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-\frac{1}{2} u^2} du + \mu^2, \end{aligned}$$

daar nu de middelste term wegvalt. Nu is, met partieel integreren,

$$\int_{-\infty}^{+\infty} u^2 e^{-\frac{1}{2} u^2} du = \int_{-\infty}^{+\infty} u d(e^{-\frac{1}{2} u^2}) = \left[-u e^{-\frac{1}{2} u^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{1}{2} u^2} du = \sqrt{2\pi},$$

hetgeen na invullen geeft

$$(4;8) \quad \mu_2 = \mathcal{E} \underline{x}^2 = \sigma^2 + \mu^2$$

Voorbeeld 4;4

Bezit $\underline{x}$ een γ -verdeling met parameters α en λ (vgl. (1;25)), dan is

$$(4;9) \quad \mathcal{E} \underline{x} = \int_0^{\infty} x \cdot \frac{\lambda^{\alpha}}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} e^{-\lambda x} x^{\alpha} dx =$$

$$\frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}.$$

Verder is

$$(4;10) \quad \mathcal{E} \underline{x}^2 = \int_0^{\infty} x^2 \cdot \frac{\lambda^{\alpha}}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} e^{-\lambda x} x^{\alpha+1} dx =$$

$$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Een overzicht van de eerste momenten van de in paragraaf 1 genoemde verdelingen is opgenomen in tabel 4;I op blz. 41.

Definitie 4;3

Het k^e gereduceerde moment ($k=1,2,\dots$) van een stochastische grootheid $\underline{x}$ is de verwachting van $(\underline{x}-\mu)^k$.

Notatie: $\tilde{\mu}_k$ of, indien nodig $\tilde{\mu}_k(\underline{x})$. Verder noemt men $\tilde{\underline{x}} = \underline{x} - \mu$, de gereduceerde $\underline{x}$. Dus:

$$(4;11) \quad \tilde{\mu}_k = \mu_k(\tilde{\underline{x}}) = \mathcal{E} \tilde{\underline{x}}^k = \mathcal{E} (\underline{x} - \mu)^k.$$

Opmerkingen

Het eerste gereduceerde moment is altijd 0. Het tweede gereduceerde moment wordt de variantie van $\underline{x}$ genoemd en met σ^2 , zo nodig $\sigma^2(\underline{x})$, aangegeven:

$$(4;12) \quad \sigma^2(\underline{x}) = \mathcal{E} (\underline{x} - \mu)^2 = \mathcal{E} \tilde{\underline{x}}^2 = \tilde{\mu}_2.$$

De variantie komt overeen met het traagheidsmoment om het zwaartepunt. Voor iedere kansverdeling is $\sigma^2 \geq 0$ en alleen dan 0 als $\underline{x}$ één waarde met kans 1 aanneemt. De wortel uit de variantie wordt de spreiding of standaardafwijking van $\underline{x}$ genoemd.

De berekening van de gereduceerde momenten, waarvan vooral de variantie belangrijk is, geschiedt het snelst met behulp van

verderop af te leiden stellingen over de mathematische verwachting.

Het geval van meerdere dimensionale verdelingen behandelen wij voor het gemak in hoofdzaak voor 2 dimensies. Voor meer dan 2 verloopt het geheel analoog.

Definitie 4;4

Bezitten $\underline{x}$ en $\underline{y}$ een simultane discrete verdeling, waarbij het punt $\underline{P} = (\underline{x}, \underline{y})$ met kans p_j ($j=1,2,\dots$) in het punt $P_j = (x_j, y_j)$ terecht komt en is $\varphi(x,y)$ een functie van x en y , dan is de mathematische verwachting van $\varphi(\underline{x}, \underline{y})$:

$$(4;13) \quad \mathbb{E}\varphi(\underline{x}, \underline{y}) \stackrel{\text{def}}{=} \sum_j p_j \varphi(x_j, y_j)$$

Is de verdeling continu met verdelingsdichtheid $f(x,y)$, dan luidt de definitie

$$(4;14) \quad \mathbb{E}\varphi(\underline{x}, \underline{y}) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x,y) f(x,y) dx dy$$

Definitie 4;5

De voorwaardelijke verwachting van $\varphi(\underline{x}, \underline{y})$, onder de voorwaarde $\underline{y}=\underline{y}$, is de verwachting van $\varphi(\underline{x}, \underline{y})$ met betrekking tot de voorwaardelijke verdeling van $\underline{x}$ onder dezelfde voorwaarde.

Notatie: $\mathbb{E}(\varphi(\underline{x}, \underline{y}) | \underline{y}=\underline{y})$ of $\mathbb{E}(\varphi(\underline{x}, \underline{y}) | \underline{y})$, of kortweg: $\mathbb{E}\varphi(\underline{x}, \underline{y})$.

Dus, volgens (2;18), voor het continue geval:

$$(4;15) \quad \mathbb{E}\varphi(\underline{x}, \underline{y}) = \int_{-\infty}^{+\infty} \varphi(\underline{x}, \underline{y}) \frac{f(\underline{x}, \underline{y})}{f_2(\underline{y})} d\underline{x}$$

Voorbeeld 4;5

Volgens (2;19) is de voorwaardelijke verdeling van $\underline{x}$, onder de voorwaarde $\underline{y}=\underline{y}$, als $\underline{x}$ en $\underline{y}$ een simultane normale verdeling (2;9) bezitten, ook normaal. Uit (2;19) en (4;7) volgt nu direct:

$$(4;16) \quad \mathbb{E}(\underline{x} | \underline{y}) = \underline{\mu}_1 + \rho \frac{\sigma_1}{\sigma_2} (\underline{y} - \underline{\mu}_2)$$

Voorbeeld 4;6

Hoe groot is de verwachting van het aantal klanten in voorbeeld 2;6 voor het geval de levertijd T maanden bedraagt?

Oplossing

Gevraagd wordt de verwachting van het aantal klanten $\underline{k}$, dat binnenkomt in een periode van T maanden. De voorwaardelijke kansverdeling van $\underline{k}$ is volgens (2;23)

$$P[k \leq n | t = T] = \sum_{k=0}^n \frac{(\mu T)^k e^{-\mu T}}{k!}$$

De voorwaardelijke verwachting is dus

$$\begin{aligned} (4;17) \quad \mathcal{E}(k | T) &= \sum_{k=0}^{\infty} k \cdot \frac{(\mu T)^k e^{-\mu T}}{k!} = \mu T \sum_{k=1}^{\infty} \frac{(\mu T)^{k-1} e^{-\mu T}}{(k-1)!} \\ &= \mu T \sum_{k=0}^{\infty} \frac{(\mu T)^k e^{-\mu T}}{k!} = \mu T \end{aligned}$$

De verwachting van het aantal klanten is dus evenredig met de lengte van het tijdsinterval.

Uit (4;17) volgt dat het eerste moment van een Poisson-verdeling met parameter μT gelijk is aan μT . Substitueert men voor μT de gebruikelijke parameter λ (vgl. (1;9)), dan is dus tevens bewezen dat voor de Poisson-verdeling (1;9) geldt

$$(4;18) \quad \mu_1 = \mathcal{E}k = \lambda$$

Stelling 4;1

Is φ een functie van $\underline{x}$ alleen: $\varphi(\underline{x})$, dan is de verwachting van $\varphi(\underline{x})$ gelijk aan de verwachting van φ over de marginale verdeling van $\underline{x}$.

Bewijs

Wij geven het bewijs alleen voor het continue geval; voor het discrete verloopt het analoog.

Volgens (4;14) is

$$\begin{aligned} \mathcal{E}\varphi(\underline{x}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(\underline{x}) f(\underline{x}, \underline{y}) d\underline{x} d\underline{y} = \\ &= \int_{-\infty}^{+\infty} \varphi(\underline{x}) d\underline{x} \int_{-\infty}^{+\infty} f(\underline{x}, \underline{y}) d\underline{y} = \int_{-\infty}^{+\infty} \varphi(\underline{x}) f_1(\underline{x}) d\underline{x}, \end{aligned}$$

waarin $f_1(\underline{x})$ (vergelijk (2;13)) de marginale verdelingsdichtheid van $\underline{x}$ is.

Voorbeeld 4;7

Voor de tweedimensionale normale verdeling, die gegeven wordt

door (2;9), is de marginale verdeling van $\underline{x}$ volgens (2;15) een normale, met parameters μ_1 en σ_1 . Volgens (4;7) en stelling 4;1 is dus voor deze verdeling

$$(4;19) \quad \mathcal{E}\underline{x} = \mu_1, \quad \mathcal{E}\underline{x}^2 = \sigma_1^2 + \mu_1^2$$

en analoog voor $\underline{y}$.

Stelling 4;2

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan geldt voor de verwachting van een produkt $\varphi(\underline{x})\psi(\underline{y})$ van functies van x en y afzonderlijk:

$$(4;20) \quad \mathcal{E}\varphi(\underline{x})\psi(\underline{y}) = \mathcal{E}\varphi(\underline{x}) \mathcal{E}\psi(\underline{y}).$$

Bewijs (voor het continue geval):

Met behulp van (3;9) en stelling 4;1 verkrijgen wij:

$$\begin{aligned} \mathcal{E}\varphi(\underline{x})\psi(\underline{y}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x)\psi(y) f(x,y) dx dy = \\ &= \int_{-\infty}^{+\infty} \varphi(x) f_1(x) dx \int_{-\infty}^{+\infty} \psi(y) f_2(y) dy = \mathcal{E}\varphi(\underline{x}) \mathcal{E}\psi(\underline{y}). \end{aligned}$$

Voorbeeld 4;8

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan is

$$(4;21) \quad \mathcal{E}\underline{x}\underline{y} = \mathcal{E}\underline{x} \mathcal{E}\underline{y}.$$

Door middel van een tegenvoorbeeld kan men aantonen, dat het omgekeerde niet het geval is.

Stelling 4;3

De operator $\mathcal{E}$ is een lineaire operator, d.w.z.

$$(4;22) \quad \mathcal{E}\{a\varphi(\underline{x},\underline{y}) + b\psi(\underline{x},\underline{y}) + c\} = a\mathcal{E}\varphi(\underline{x},\underline{y}) + b\mathcal{E}\psi(\underline{x},\underline{y}) + c,$$

als $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten, φ en ψ functies en a , b en c constanten zijn. De stelling geldt analoog voor meer dan 2 stochastische grootheden.

Bewijs

De geldigheid van (4;22) volgt onmiddellijk uit definitie 4;1.

Toepassing

Deze stelling maakt het mogelijk de gereduceerde momenten in de gewone uit te drukken. Bijvoorbeeld (vergelijk (4;12)):

$$\begin{aligned} \sigma^2(\underline{x}) &= \mathcal{E}(\underline{x} - \mu)^2 = \mathcal{E}(\underline{x}^2 - 2\mu\underline{x} + \mu^2) = \\ &= \mathcal{E}\underline{x}^2 - 2\mu\mathcal{E}\underline{x} + \mu^2 = \mu_2 - 2\mu^2 + \mu^2 = \mu_2 - \mu^2, \end{aligned}$$

hetgeen dus leidt tot de volgende zeer nuttige formule ter berekening van de variantie

$$(4;23) \quad \sigma^2 = \mu_2 - \mu^2 \quad \text{of} \quad \sigma^2(\underline{x}) = \mathcal{E} \underline{x}^2 - (\mathcal{E} \underline{x})^2.$$

Passen wij deze formule op de variantie van de normale verdeling toe, dan vinden wij met behulp van (4;7) en (4;8)

$$(4;24) \quad \sigma^2(\underline{x}) = \mathcal{E} \underline{x}^2 - (\mathcal{E} \underline{x})^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

De parameter σ is dus gelijk aan de wortel van de variantie, dat is de standaardafwijking. Daar bovendien de parameter μ , zoals wij zagen gelijk is aan de verwachting $\mathcal{E} \underline{x}$, wordt een normale verdeling door verwachting en variantie dus volledig bepaald.

Als tweede illustratie passen wij formule (4;23) toe om de variantie van een γ -verdeling te berekenen; wij krijgen met behulp van (4;9) en (4;10):

$$(4;25) \quad \sigma^2(\underline{x}) = \mathcal{E} \underline{x}^2 - (\mathcal{E} \underline{x})^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}.$$

Opmerking

Uit (4;23) volgt direct de ongelijkheid

$$(4;26) \quad \mu_2 \geq \mu^2 \quad \text{of} \quad \mathcal{E} \underline{x}^2 \geq (\mathcal{E} \underline{x})^2,$$

waarbij de gelijkheidtekens alleen gelden als $\underline{x}$ met kans 1 een bepaalde waarde aanneemt.

Stelling 4;4. Optelregel voor varianties

Zijn $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk verdeeld¹⁾, dan geldt

$$(4;27) \quad \sigma^2(\underline{x} + \underline{y}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}).$$

De stelling geldt ook voor meer dan twee grootheden.

Bewijs

Volgens (4;22) is

1) Wanneer wij deze onderstelling maken, dan nemen wij steeds bovendien stilzwijgend aan, dat $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten.

$$\mathcal{E}(\underline{x} + \underline{y}) = \mathcal{E}\underline{x} + \mathcal{E}\underline{y},$$

dus

$$\{\mathcal{E}(\underline{x} + \underline{y})\}^2 = (\mathcal{E}\underline{x})^2 + (\mathcal{E}\underline{y})^2 + 2 \mathcal{E}\underline{x} \mathcal{E}\underline{y}.$$

Verder is, eveneens volgens (4;22)

$$\mathcal{E}(\underline{x} + \underline{y})^2 = \mathcal{E}(\underline{x}^2 + \underline{y}^2 + 2\underline{x}\underline{y}) = \mathcal{E}\underline{x}^2 + \mathcal{E}\underline{y}^2 + 2 \mathcal{E}\underline{x} \mathcal{E}\underline{y},$$

daar, wegens de onafhankelijkheid (4;21) toegepast kan worden.

Volgens (4;23) is

$$\sigma^2(\underline{x} + \underline{y}) = \mathcal{E}(\underline{x} + \underline{y})^2 - \{\mathcal{E}(\underline{x} + \underline{y})\}^2;$$

substitueren wij de bovenstaande resultaten hierin, dan komt er

$$\sigma^2(\underline{x} + \underline{y}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 + \mathcal{E}\underline{y}^2 - (\mathcal{E}\underline{y})^2 = \sigma^2(\underline{x}) + \sigma^2(\underline{y}),$$

weer volgens (4;23).

Opmerking

De voorwaarde van onafhankelijkheid is voldoende, doch niet noodzakelijk. Nodig en voldoende is blijkens het bewijs, dat (4;21) geldt. Deze stelling geeft tevens een belangrijk verschil aan tussen verwachtingen en varianties. Voor optelbaarheid van verwachtingen behoeft (4;21) niet te gelden.

De stellingen 4;3 en 4;4 zijn zowel voor de theoretische statistiek als voor de toepassingen van groot belang. Wij geven twee toepassingen.

Voorbeeld 4;9

Bereken de verwachting en de variantie van de binomiale verdeling.

Oplossing

Volgens voorbeeld 2;7 kan een binomiale grootheid $\underline{x}$ (vergelijk (1;6)) beschouwd worden als de som van n onafhankelijk verdeelde grootheden $\underline{x}_i$, die ieder een kans p resp. q bezitten op de waarden 1 resp. 0.

Voor ieder daarvan geldt

$$\mathcal{E}\underline{x}_i = p \quad \text{en} \quad \mathcal{E}\underline{x}_i^2 = p,$$

dus (vergelijk (4;23))

$$\sigma^2(\underline{x}_i) = p - p^2 = pq.$$

Voor

$$\underline{x} = \sum_{i=1}^n \underline{x}_i$$

geldt dus volgens stelling 4;3

$$(4;28) \quad \mu = \mathbb{E}\underline{x} = np$$

en volgens (4;27)

$$(4;29) \quad \sigma^2 = \sigma^2(\underline{x}) = npq$$

Voorbeeld 4,10

Het bureau voor marktanalyse genoemd in voorbeeld 5;9 van hoofdstuk I vraagt zich af hoeveel bezoeken men gemiddeld zal moeten afleggen om k deelnemers aan het consumentenpanel te vinden. Bovendien wil men graag weten of het aantal bezoeken, nodig om k deelnemers te krijgen, sterk zal variëren en daarom vraagt men ook hoe groot de spreiding van dit aantal bezoeken bedraagt.

Oplossing

Het aantal bezoeken $\underline{n}$ nodig om precies k deelnemers te vinden volgt, zoals wij in voorbeeld 5;9 zagen, een negatief binomiale verdeling. De stochastische grootheid $\underline{n}$ kan beschouwd worden als de som van k onafhankelijke grootheden $\underline{n}_1, \underline{n}_2, \dots, \underline{n}_k$ waarvan $\underline{n}_1$ het aantal experimenten (= bezoeken) voorstelt tot en met het eerste succes, $\underline{n}_2$ het daarop volgende aantal experimenten tot en met het tweede succes, enz. Daar alle experimenten onafhankelijk zijn, geldt dit ook voor de grootheden $\underline{n}_i$ ($i=1,2,\dots,k$) en deze bezitten bovendien alle dezelfde verdeling als $\underline{n}_1$. Het eerste moment daarvan is reeds berekend (vgl. (4;6)); het tweede kan op analoge wijze afgeleid worden. Differentieert men namelijk de betrekking

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$$

tweemaal naar q, dan vindt men

$$\sum_{n=0}^{\infty} n(n-1)q^{n-2} = \frac{2}{(1-q)^3},$$

waaruit, samen met (4;6) volgt

$$(4;30) \quad \mathbb{E} n_i^2 = \sum_{n_i=0}^{\infty} n_i^2 p q^{n_i-1} = \frac{1+q}{p^2}.$$

Passen wij nu (4;23) toe, dan blijkt

$$(4;31) \quad \sigma^2(n_i) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

en dus leiden stelling 4;3 resp. stelling 4;4 tot

$$(4;32) \quad \mu = \mathbb{E} n = \frac{k}{p} \text{ en } \sigma^2 = \sigma^2(n) = \frac{kq}{p^2}.$$

In het geciteerde voorbeeld is voor $p = \frac{2}{5}$ en $k = 10$ het verwachte aantal bezoeken dus 25, terwijl de spreiding

$$\sqrt{\frac{kq}{p^2}} = \sqrt{\frac{10 \cdot \frac{3}{5}}{(\frac{2}{5})^2}} = 6,12 \text{ bedraagt.}$$

Stelling 4;5

Zijn a en b constanten en is $\underline{x}$ een stochastische grootheid, dan is

$$(4;33) \quad \sigma^2(a\underline{x} + b) = a^2 \sigma^2(\underline{x})$$

Bewijs

$$\mathbb{E}(a\underline{x} + b) = a \mathbb{E}\underline{x} + b.$$

De bij $a\underline{x} + b$ behorende gereduceerde grootheid is dus

$$a\underline{x} + b - a \mathbb{E}\underline{x} - b = a(\underline{x} - \mathbb{E}\underline{x}) = a\tilde{\underline{x}}.$$

Dus

$$\sigma^2(a\underline{x} + b) = \mathbb{E}(a\tilde{\underline{x}})^2 = \mathbb{E}a^2\tilde{\underline{x}}^2 = a^2 \mathbb{E}\tilde{\underline{x}}^2 = a^2 \sigma^2(\underline{x}).$$

De belangrijkste toepassing hiervan is vervat in:

Stelling 4;6

Is $\underline{x}$ een stochastische grootheid met verwachting μ en variantie σ^2 , dan bezit de stochastische grootheid

$$(4;34) \quad \underline{x}^* \stackrel{\text{def}}{=} \frac{\underline{x} - \mu}{\sigma} = \frac{\tilde{\underline{x}}}{\sigma}$$

verwachting 0 en variantie 1.

Het bewijs wordt aan de lezer overgelaten.

Opmerkingen

De grootheid $\underline{x}^*$ wordt de bij $\underline{x}$ behorende gestandaardiseerde grootheid of kortweg de gestandaardiseerde $\underline{x}$ genoemd (vergelijk ook blz. 5).

Dit standaardiseringsproces stelt ons o.a. in staat om iedere normale verdeling te beheersen met behulp van een tabel van de gestandaardiseerde normale verdeling. Zoals intuïtief duidelijk is, maar ook via een eenvoudige transformatie bewezen kan worden (vgl. stelling 2;2 uit hoofdstuk III) geldt namelijk dat de gestandaardiseerde grootheid $\underline{x}^*$ normaal verdeeld is, wanneer $\underline{x}$ een normale verdeling bezit.

Voorbeeld 4;11

$\underline{x}$ is $N(4;0,5)$ -verdeeld; dus normaal met verwachting 4 en spreiding 0,5. Gevraagd wordt te berekenen

$$P[\underline{x} \geq 4,75].$$

De berekening verloopt als volgt:

$$P[\underline{x} \geq 4,75] = P\left[\frac{\underline{x} - \mu}{\sigma} \geq \frac{4,75 - \mu}{\sigma}\right] = P[u \geq 1,5],$$

waarin u de gestandaardiseerde normaal verdeelde grootheid voorstelt. Uit de tabel van de normale verdeling blijkt, dat de gezochte kans 0,067 bedraagt.

Voorbeeld 4;12¹⁾

Een verpakkingsmachine, die pakjes thee van nominaal 100 gram maakt, verpakt gemiddeld 101,0 gram in ieder pakje. De verdeling van het gewicht aan thee in een pakje is normaal en bezit een spreiding van 0,60 gram.

Beantwoord de volgende vragen:

- a) Hoeveel procent van de pakjes bevat minder dan het nominale gewicht?
- b) Tot welk bedrag zou men het gemiddelde gewicht, bij gelijkblijvende spreiding, op moeten voeren om dit percentage tot 2% -----

1) Dit voorbeeld is ontleend aan J. HEMELRIJK en DORALINE WABEKE, Elementaire statistische opgaven met uitgewerkte oplossingen, Noorduijn en Zoon N.V., Gorinchem 1957. Dit boekje bevat 100 opgaven, waarvan een groot deel op praktijkgevallen is gebaseerd. Voorbeeld 4,12 is opgave 42.

terug te brengen?

c) Hoeveel procent van de pakjes bevat dan meer dan 103,0 gram?

Oplossing

Het gewicht van de inhoud is $N(101,0; 0,60)$ verdeeld.

a) De kans op pakjes met minder dan het nominale gewicht bedraagt

$$P[\underline{x} \leq 100] = P\left[\frac{\underline{x} - 101,0}{0,60} \leq \frac{100 - 101,0}{0,60}\right] = P[\underline{u} \leq -1,67] = 0,0475$$

en het gezochte percentage is dus 4,75%.

b) Het percentage van 4,75% kan bij gelijkblijvende spreiding alleen verminderd worden door het gemiddelde te verhogen, bijvoorbeeld tot μ' . Gegeven is dan, dat moet gelden

$$P[\underline{x} \leq 100] = P\left[\frac{\underline{x} - \mu'}{0,60} \leq \frac{100 - \mu'}{0,60}\right] = P\left[\underline{u} \leq \frac{100 - \mu'}{0,60}\right] = 0,02$$

Uit de tabel van de normale verdeling lezen wij af, dat dit percentage ongeveer bereikt wordt bij $u = -2,06$ en dus is

$$\frac{100 - \mu'}{0,60} = -2,06$$

of $\mu' = 101,24$. Het gemiddelde gewicht moet dus op de waarde van 101,24 ingesteld worden.

c) Voor de laatste vraag berekenen wij

$$P[\underline{x} \geq 103] = P\left[\underline{u} \geq \frac{103 - 101,24}{0,60}\right] = P[\underline{u} \geq 2,93] = 0,0017,$$

m.a.w. stellen wij de verpakkingsmachine in op 101,24 gram, dan bevat 0,17% van de pakjes meer dan 103 gram thee.

Stelling 4;7

Indien $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten en $\varphi(\underline{x}, \underline{y})$ is een functie van $\underline{x}$ en $\underline{y}$, terwijl

$$(4;35) \quad g(\underline{y}) \stackrel{\text{def}}{=} \mathcal{L}(\varphi(\underline{x}, \underline{y}) | \underline{y} = \underline{y})$$

de voorwaardelijke verwachting van $\varphi(\underline{x}, \underline{y})$ onder de voorwaarde $\underline{y} = \underline{y}$ is, dan geldt

$$(4;36) \quad \mathcal{L}\varphi(\underline{x}, \underline{y}) = \mathcal{L}g(\underline{y})$$

Bewijs (voor het continue geval)

Volgens definities 4;4 en 4;5 geldt

$$\begin{aligned} \xi \varphi(x,y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x,y) f(x,y) dx dy = \\ &= \int_{-\infty}^{+\infty} f_2(y) dy \int_{-\infty}^{+\infty} \varphi(x,y) \frac{f(x,y)}{f_2(y)} dx = \\ &= \int_{-\infty}^{+\infty} g(y) f_2(y) dy = \xi g(y). \end{aligned}$$

Deze stelling houdt dus in dat de verwachting van een functie ten opzichte van een tweedimensionale kansverdeling berekend kan worden door eerst de voorwaardelijke verwachting t.o.v. één van de variabelen te bepalen en vervolgens van de gevonden functie de verwachting t.o.v. de andere variabele te berekenen. Wij kunnen deze stelling bijvoorbeeld gebruiken in een geval, zoals in voorbeeld 2;6 geschetst is. Men berekent dan eerst de voorwaardelijke verwachting bij gegeven waarde T van $\underline{t}$; deze is volgens (4;17) gelijk aan μT . Neemt men van $\mu \underline{T}$ de verwachting t.o.v. de marginale verdeling van $\underline{T}$, dan vindt men (vgl. (4;9)) voor de verwachting van het aantal klanten gedurende de levertijd $\frac{\alpha \mu}{\lambda}$. In dit geval hebben wij de verdeling van het aantal klanten reeds berekend (vgl. voorbeeld (2;6)) en dan vormt deze weg geen vereenvoudiging¹⁾. Anders is het echter gesteld in voorbeeld 4;13.

1) Past men formule (4;32) toe op de verdeling (2;24), dan vindt men een andere waarde dan $\frac{\alpha \mu}{\lambda}$. De in (4;32) gegeven verwachting behoort bij de negatief binomiale verdeling, zoals deze in (1;8) gegeven is. Hierin is $\underline{x}$ het aantal worpen tot en met het k^{de} succes. Men kan echter ook de kans opgeven dat er $\underline{x}$ mislukkingen hebben plaatsgevonden voor het k^{de} succes, waarbij $\underline{x}$ dus waarden vanaf 0 en niet vanaf k aan kan nemen. Dit is de vorm waarin de verdelingsfunctie (2;24) opgeschreven staat. Bij deze vorm is de verwachting van de negatief binomiale verdeling met parameters k en p gelijk aan $\frac{kq}{p}$, terwijl de variantie $\frac{kq}{p^2}$ blijft (vgl. (4;32)). Is $k=\alpha$ en $p=\frac{\lambda}{\lambda+\mu}$, dan vindt men voor de gezochte verwachting wederom $\frac{\alpha \mu}{\lambda}$.

Een grootheid, die evenals verwachting en variantie van groot belang is, is de covariantie van twee stochastische groot-heden $\underline{x}$ en $\underline{y}$. Deze wordt geschreven als $\text{cov}(\underline{x}, \underline{y})$ en gedefinieerd door

$$(4;37) \quad \text{cov}(\underline{x}, \underline{y}) \stackrel{\text{def}}{=} \mathcal{E} \tilde{\underline{x}} \tilde{\underline{y}} = \mathcal{E}(\underline{x} - \mathcal{E}\underline{x})(\underline{y} - \mathcal{E}\underline{y})$$

Voorbeeld 4;13

Wij berekenen nu, met behulp van stelling 4;7 de covariantie van $\underline{x}$ en $\underline{y}$, voor het geval van de tweedimensionale normale verdeling (formule (2;9)).

Volgens (4;16) is

$$\mathcal{E}(\underline{x} - \mu_1 | \underline{y}) = \rho \frac{\sigma_1}{\sigma_2} (\underline{y} - \mu_2),$$

dus

$$g(\underline{y}) = \mathcal{E}\{(\underline{x} - \mu_1)(\underline{y} - \mu_2) | \underline{y}\} = \rho \frac{\sigma_1}{\sigma_2} (\underline{y} - \mu_2)^2$$

Volgens (4;36) is dus

$$\begin{aligned} \mathcal{E}(\underline{x} - \mu_1)(\underline{y} - \mu_2) &= \mathcal{E} \rho \frac{\sigma_1}{\sigma_2} (\underline{y} - \mu_2)^2 = \rho \frac{\sigma_1}{\sigma_2} \mathcal{E}(\underline{y} - \mu_2)^2 = \\ &= \rho \frac{\sigma_1}{\sigma_2} \sigma_2^2 = \rho \sigma_1 \sigma_2 \end{aligned}$$

Dus

$$(4;38) \quad \text{cov}(\underline{x}, \underline{y}) = \rho \sigma_1 \sigma_2$$

In dit voorbeeld is stelling 4;7 van groot nut geweest, aangezien men zonder kennis van deze stelling $\mathcal{E}(\underline{x} - \mu_1)(\underline{y} - \mu_2)$ geheel had moeten uitschrijven, hetgeen een ingewikkelde vorm tot resultaat heeft.

Stelling 4;8

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan is

$$(4;39) \quad \text{cov}(\underline{x}, \underline{y}) = 0.$$

Bewijs

Men houde in het oog, dat $\mathcal{E}\underline{x}$ en $\mathcal{E}\underline{y}$ constanten zijn. Volgens stelling 4;3 is

$$\begin{aligned} \mathcal{E}(\underline{x} - \mathcal{E}\underline{x})(\underline{y} - \mathcal{E}\underline{y}) &= \mathcal{E}(\underline{x}\underline{y} - \underline{x}\mathcal{E}\underline{y} - \underline{y}\mathcal{E}\underline{x} + \mathcal{E}\underline{x}\mathcal{E}\underline{y}) = \\ &= \mathcal{E}\underline{x}\underline{y} - \mathcal{E}\underline{x}\mathcal{E}\underline{y} - \mathcal{E}\underline{y}\mathcal{E}\underline{x} + \mathcal{E}\underline{x}\mathcal{E}\underline{y} = \mathcal{E}\underline{x}\underline{y} - \mathcal{E}\underline{x}\mathcal{E}\underline{y} = 0 \end{aligned}$$

volgens (4;21).

Stelling 4;9

Bezitten $\underline{x}$ en $\underline{y}$ een simultane verdeling, dan geldt

$$(4;40) \quad \sigma^2(a\underline{x} + b\underline{y}) = a^2\sigma^2(\underline{x}) + b^2\sigma^2(\underline{y}) + 2ab\text{cov}(\underline{x}, \underline{y}).$$

Bewijs

De bij $a\underline{x} + b\underline{y}$ behorende gereduceerde grootheid is $a\tilde{\underline{x}} + b\tilde{\underline{y}}$,
daar $\mathcal{L}(a\underline{x} + b\underline{y}) = a\mathcal{L}\underline{x} + b\mathcal{L}\underline{y}$ is.

Derhalve is

$$\sigma^2(a\underline{x} + b\underline{y}) = \mathcal{L}(a\tilde{\underline{x}} + b\tilde{\underline{y}})^2 = a^2\mathcal{L}\tilde{\underline{x}}^2 + b^2\mathcal{L}\tilde{\underline{y}}^2 + 2ab\mathcal{L}\tilde{\underline{x}}\tilde{\underline{y}}.$$

Stelling 4;10

Bezitten $\underline{x}$ en $\underline{y}$ een simultane verdeling, dan geldt

$$(4;41) \quad |\text{cov}(\underline{x}, \underline{y})| \leq \sigma(\underline{x})\sigma(\underline{y}).$$

Bewijs

Stel in (4;40) $b=1$. Daar een variantie steeds ≥ 0 is, geldt

$$0 \leq \sigma^2(a\underline{x} + \underline{y}) = \sigma^2(\underline{x})a^2 + 2\text{cov}(\underline{x}, \underline{y})a + \sigma^2(\underline{y}).$$

Dit geldt voor iedere a , dus moet de discriminant van deze kwadratische vorm in a niet positief zijn:

$$\frac{1}{4}D = \{\text{cov}(\underline{x}, \underline{y})\}^2 - \sigma^2(\underline{x})\sigma^2(\underline{y}) \leq 0.$$

Hieruit volgt (4;41).

De correlatiecoëfficiënt van $\underline{x}$ en $\underline{y}$ wordt gedefinieerd door

$$(4;42) \quad \rho(\underline{x}, \underline{y}) \stackrel{\text{def}}{=} \frac{\text{cov}(\underline{x}, \underline{y})}{\sigma(\underline{x})\sigma(\underline{y})}$$

Volgens (4;41) geldt

$$(4;43) \quad -1 \leq \rho(\underline{x}, \underline{y}) \leq +1$$

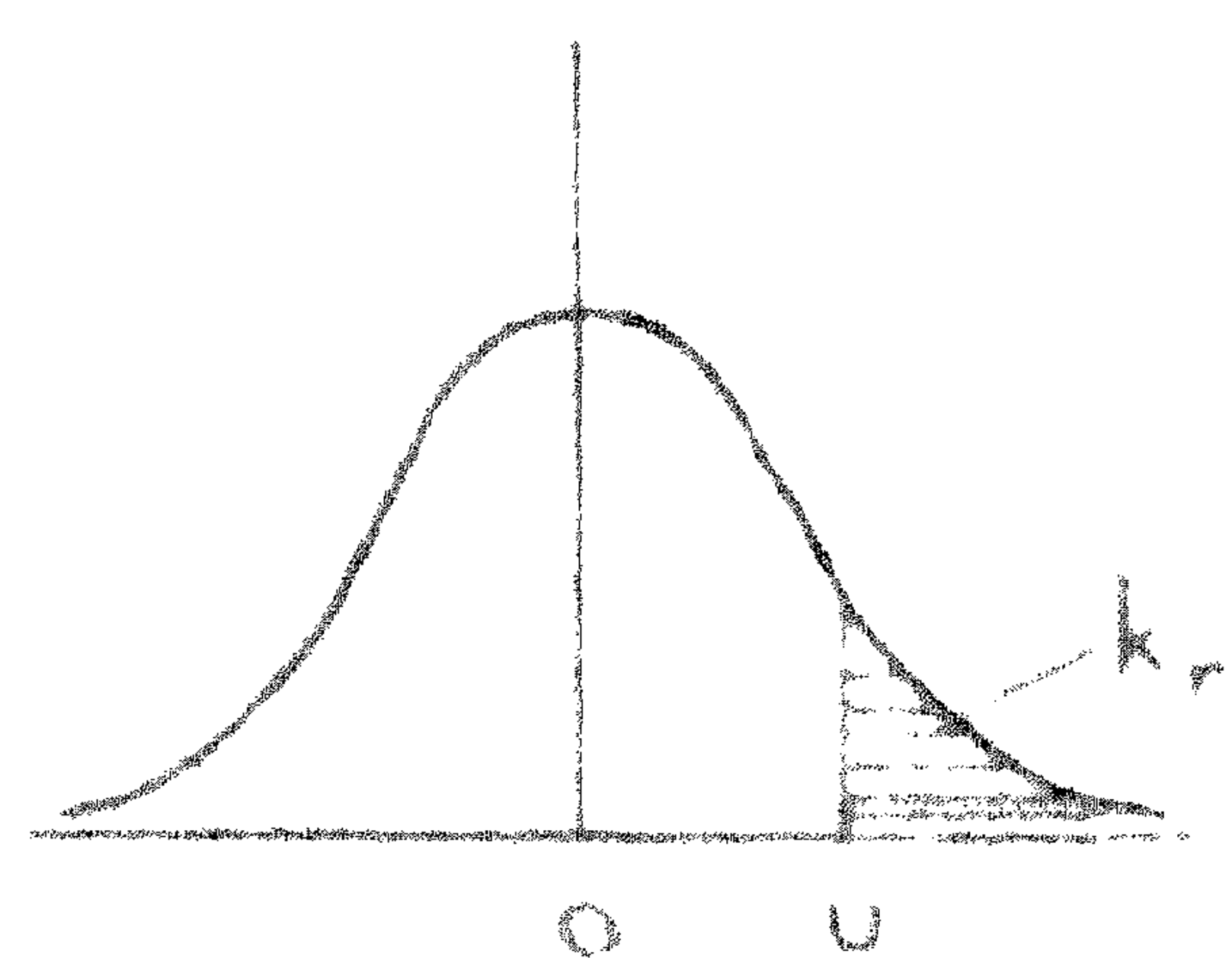
en volgens (4;38) is $\rho(\underline{x}, \underline{y})$ voor een tweedimensionale normale verdeling gelijk aan de parameter ρ uit formule (2;9).

Tabel 4;I

Overzicht van de kansen resp. verdelingsdichtheden, verwachtingen en varianties van enige belangrijke verdelingsfuncties.

naam van de verdeling	kansen resp. verdelingsdichtheden	verwachting	variantie
alternatief	$P[x=0]=p \quad P[x=1]=1-p=q$	q	pq
binominale verdeling	$P[x=k] = \binom{n}{k} p^k q^{n-k} \quad (k=0, \dots, n)$	np	npq
hypergeometrische verdeling	$P[x=k] = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{M+N}{n}} \quad (\max(n-N, 0) \leq k \leq \min(M, n))$	$\frac{Mn}{M+N}$	$\frac{nMN(M+N-n)}{(M+N)^2(M+N-1)}$
negatief binominale verdeling	$P[x=x] = \binom{x-1}{k-1} p^k q^{x-k} \quad (x=k, k+1, \dots)$	$\frac{k}{p}$	$\frac{kq}{p^2}$
Poissonverdeling	$P[x=k] = \frac{\lambda^k e^{-\lambda}}{k!} \quad (k=0, 1, \dots)$	λ	λ
(discrete) homogene verdeling	$P[x=k] = \frac{1}{n+1} \quad (k=0, \dots, n)$	$\frac{n}{2}$	$\frac{1}{12} n(n+2)$
(continue) homogene verdeling	$\frac{1}{\theta_2 - \theta_1}$ voor $\theta_1 \leq x \leq \theta_2$	$\frac{1}{2}(\theta_1 + \theta_2)$	$\frac{1}{12}(\theta_2 - \theta_1)^2$
normale verdeling	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty \leq x \leq +\infty)$	μ	σ^2
exponentiële verdeling	$\lambda e^{-\lambda x} \quad (0 \leq x < \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
χ^2 -verdeling	$\frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (0 \leq x < \infty)$	n	$2n$
gamma-verdeling	$\frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \quad (0 \leq x < \infty)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
beta-verdeling	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (0 \leq x \leq 1)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
logarithmisch normale verdeling	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \cdot \frac{1}{x} \quad (0 \leq x < \infty)$	$e^{\mu + \frac{1}{2}\sigma^2}$	$e^{2\mu + 2\sigma^2} - 2\mu\sigma^2 - e^{\mu + \frac{1}{2}\sigma^2}$

1) Vergelijk voetnoot 1) op blz. 38.



Tabel 4; II
N(0,1)-verdeling

Waarden van $k_r \cdot 10^4$ voor $u = 0,00(0,01)3,49$.

$$k_r = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-\frac{1}{2}v^2} dv$$

u	0	1	2	3	4	5	6	7	8	9
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,2	4207	4168	4129	4090	4052	4013	3974	3935	3897	3859
0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2,8	0025	0025	0024	0023	0023	0022	0021	0021	0020	0019
2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3,0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3,3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3,4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Een 5 betekent, dat bij afronding naar het dichtsbijgelegen oneven cijfer afgerond moet worden. Een 5 wordt naar het dichtsbijzijnde even cijfer afgerond.