Samenvatting Kansstat

Inhoud

[Hoofdstuk 2 kans ruimte gebeurtenissen en kans 2](#_Toc485636227)

[Hoofdstuk 3 3](#_Toc485636228)

[conditionele kansen 3](#_Toc485636229)

[Onafhankelijkheid 3](#_Toc485636230)

[Hoofdstuk 4 Discrete stochasten 4](#_Toc485636231)

[Hoofdstuk 5 Continue stochasten 5](#_Toc485636232)

[Hoofdstuk 6 Simulatie 5](#_Toc485636233)

[Hoofdstuk 7 verwachting en variantie 6](#_Toc485636234)

[Hoofdstuk 8 Berekeningen met stochasten 7](#_Toc485636235)

[Transformaties 7](#_Toc485636236)

[Jensens ongelijkheid 7](#_Toc485636237)

[Extremen 7](#_Toc485636238)

[Hoofdstuk 9 gezamenlijke / simultane kansverdelingen 8](#_Toc485636239)

[Hoofdstuk 10 covariantie en correlatie 9](#_Toc485636240)

[Hoofdstuk 11.1 & 11.2 sommen van stochasten ????? 10](#_Toc485636241)

[Discrete stochasten 10](#_Toc485636242)

[Continue stochasten 10](#_Toc485636243)

[Hoofdstuk 12 Poisson proces 11](#_Toc485636244)

[Hoofdstuk 13 Wet van de grote getallen 12](#_Toc485636245)

[Hoofdstuk 14 centrale limiet stelling 13](#_Toc485636246)

[Benadering voor de binomiale verdeling 13](#_Toc485636247)

[Hoofdstuk 15 grafische representatie 14](#_Toc485636248)

[Histogrammen 14](#_Toc485636249)

[Kerndichtheidsschatting 14](#_Toc485636250)

[Empirische verdelingsfunctie 14](#_Toc485636251)

[Hoofdstuk 16 numerieke representaties 15](#_Toc485636252)

[Empirische grootheden 15](#_Toc485636253)

[Empirische kwantielen 15](#_Toc485636254)

[Hoofdstuk 17 statistische modellen?? 16](#_Toc485636255)

[Hoofdstuk 19 & 20 onafhankelijke schatters 17](#_Toc485636256)

[Hoofdstuk 21 likelihood functies 17](#_Toc485636257)

[Loglikelihood functie 17](#_Toc485636258)

[Uitzondering: 17](#_Toc485636259)

[Hoofdstuk 23 en 24 betrouwbaarheidsintervallen 18](#_Toc485636260)

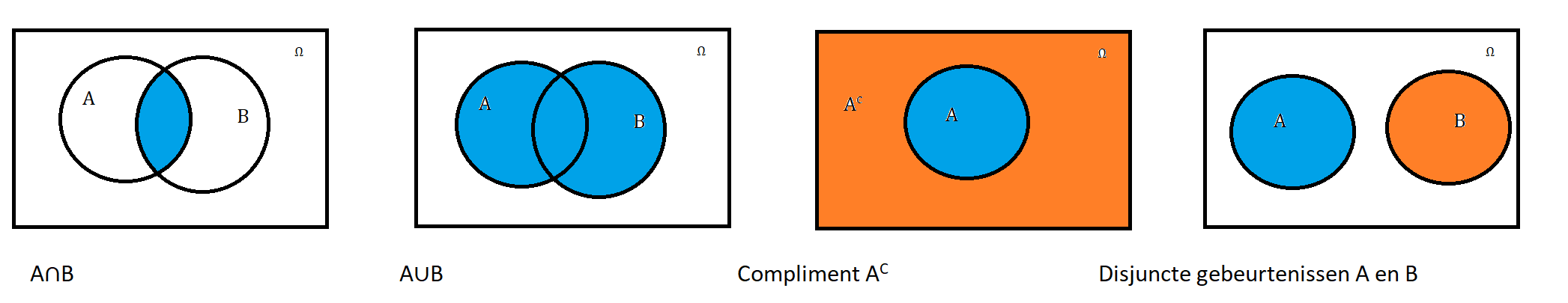
[Hoofdstuk 25 & 26 het toetsen van hypothesen 19](#_Toc485636261)

[Begrippen 19](#_Toc485636262)

[Hoe bereken je de kans op type I en type II fouten? 19](#_Toc485636263)

[Totaal rekenregels voor kansrekening H 2 en H3 20](#_Toc485636264)

# Hoofdstuk 2 kans ruimte gebeurtenissen en kans



* P(∅) = 0
* P(AC) = 1 – P(A)
* P(A) = P(A∩B) + P(A∩Bc)
* P(A1 ∪A2 ∪A3 ∪...) = P(A1) + P(A2) + P(A3) + ... {Geld alleen voor disjuncte gebeurtenissen}
* P(A∪B) = P(A) + P(B)−P(A∩B) {Somregel}
* P(A∪B)c = P(Ac∩Bc) {Morgan’s law}
* P(A∩B)c = P(Ac∪Bc) {Morgan’s law}

# Hoofdstuk 3

## conditionele kansen

***DEF:*** We zeggen: P(A|C), de kans op A is gegeven dat C zich voordoet

Hieruit volgt de *productregel*:

* P(A∩B) = P(B)P(A|B)

En de *wet van de totale kans*:

* P(A) = P(A∩B) + P(A∩Bc)

= P(B)P(A|B) + P(Bc)P(A|Bc)

Hieruit volgt *Bayes rule*:

Als C1 , C2, … , Cn disjuncte gebeurtenissen zijn zodat de vereniging van deze kansen gelijk is aan Ω, kan de kans op willekeurig evenement A worden uitgedrukt in:

* P(A) = P(A|C1) P(C1) + P(A|C2) P(C2) + … + P(A|Cn) P(Cn)
* {te illustreren met een kansboom}

## Onafhankelijkheid

***DEF:*** Twee kansen, A,B, heten onafhankelijk als P(A|B) =P(A). Voor onafhankelijke gebeurtenissen geldt:

* P(A∩B) = P(A) P(B)

***DEF:*** n gebeurtenissen A1, A2, … , An. heten onafhankelijk als voor elk k-tal geldt:

* P(Ai1 ∩Ai2 ∩...∩Aik) = P(Ai1)P(Ai2)···P(Aik) &
* P(Ai1 |Ai2 ∩Ai3 ∩...∩Aik) = P(Ai1)

*LET OP:* disjunct is niet hetzelfde als onafhankelijk

# Hoofdstuk 4 Discrete stochasten

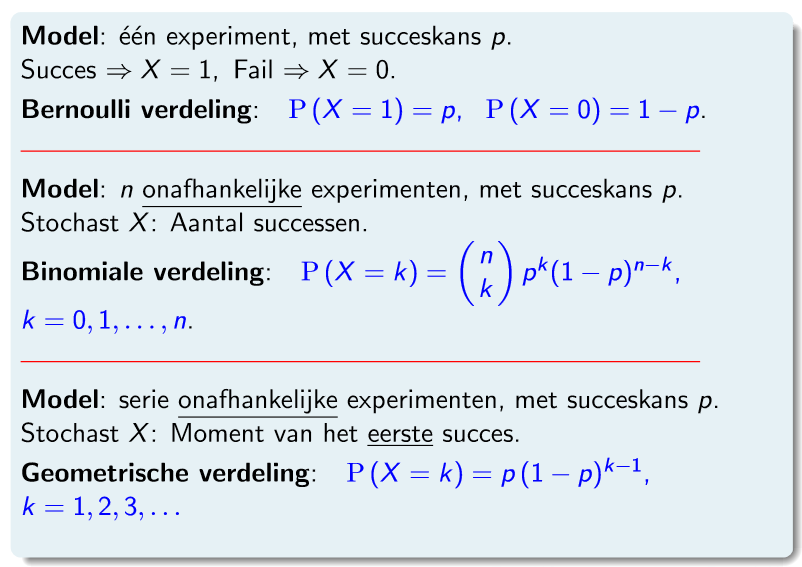
***DEF:*** Een discrete stochast is een functie X : Ω → R op een kansruimte die, hetzij eindig veel waarden a1,...,an, of een (oneindige) rij waarden a1,a2,a3, ... aanneemt.

***DEF:*** de kansmassafunctie p van de stochast X is de functie p: R-> [0,1]. Gedefineerd via p(a) = P(X=a).

***DEF:*** Een verdelingsfunctie van de discrete stochast X is de functie F : R→ [0,1] Gedeﬁnieerd door F(x) = P(X ≤ x).

Opmerkingen:

* F(x) is altijd stijgend
* De uitkomst van p(x) ^van F(x) is altijd op het interval [0,1]
* Het domein van p(x) en F(x) is altijd [-∞,∞]



Waarin

# Hoofdstuk 5 Continue stochasten

DEF: Een continue stochast is een stochast X waarvoor een functie f bestaat waarmee je kansen uit kunt rekenen via:

De functie f(x) heet de kansdichtheid functie van X.

De verdelingsfunctie is op dezelfde manier gedefinieerd als bij discrete stochasten: F(x) = P(X ≤ x). hieruit volgt dat en omgekeerd f (x) = F’(x).

***DEF:*** Als 0 ≤ p ≤ 1 en X is een stochastische variabele, dan is het p-de kwantiel of 100p-de percentiel het kleinste getal qp zodat: F(qp) = P(X ≤ qp) = p

Opmerkingen

* De kansdichtheid is nooit negatief
* -> P(a−δ ≤ X ≤ a + δ) ≈ 2δ·f (a)
* De mediaan m is het 50% percentiel:

# Hoofdstuk 6 Simulatie

Stel F is een stijgende functie van het interval [a, b] op het interval [0,1], en U is uniform verdeeld op [0,1]. Dan heeft X = F−1(U) precies de verdelingsfunctie F.

Voorbeeld exponentiële verdeling:

* U = F(X)
* 1−e−λX = U
* e−λX = 1−U
* −λX = ln(1−U)
* X = −ln(1−U)/λ
* X = −ln(U)/λ

Zodat je de inverse van f(x) = u kan berekenen. Waarin u = P(X ≤ x)

# Hoofdstuk 7 verwachting en variantie

***DEF:*** De verwachting van een discrete stochast X is gedefinieerd als:

***DEF:*** De verwachting van een continue stochast is gedefinieerd als:

***DEF:*** De variantie van een stochast X is gedeﬁnieerd door

* Var(X) = E(X −E[X])2 = σ2

Opmerkingen:

* De verwachting is te interpreteren als ‘het gemiddelde op de lange duur’.
* De variantie is een maat voor de ‘spreiding’ van een stochast X.
* E[aX+b] = a\*E[X] + b
* Var(X) = E[X] – E[X]2
* Var(aX+b) = a2Var(X)
* σ2 is de variantie
* σ is de standaardafwijking of deviatie

# Hoofdstuk 8 Berekeningen met stochasten

## Transformaties

Stel stochast X heeft verdelingsfunctie FX; gevraagd: de v.d.f. van Y = g(X). dan geldt

* FY = P(y< y)
* FY = P(g(X) < y)
* FY =P(X < g-1(y) )
* FY = FX(g-1(y))

**Verandering van eenheden transformatie:**

Als X een continue stochast is met een verdelingsfunctie FX en een kansdichtheidsfunctie fX. als we de eenheden veranderen naar Y = rX+s voor r < 0 dan geldt:

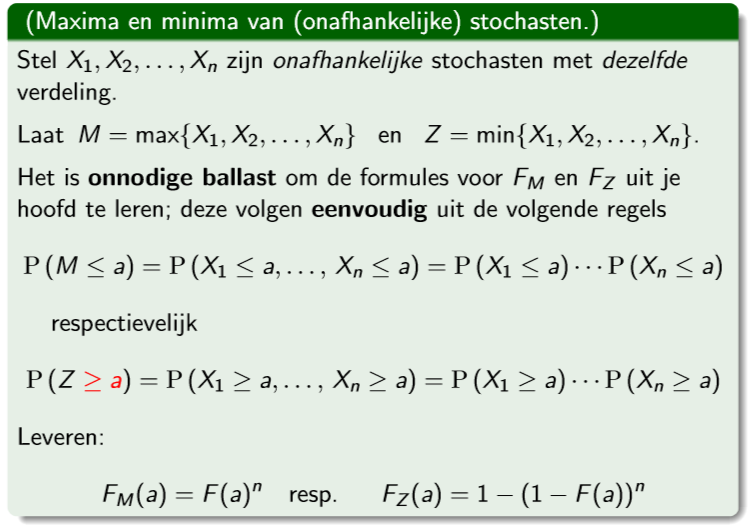
Hiermee zijn alle normale verdelingen om te schrijven naar de standaard normaalverdeling

## Jensens ongelijkheid

Als g(x) een convexe functie (bijvoorbeeld dalparabool) is dan geldt:

Als g(x) een concave functie (bijvoorbeeld Bergparabool) is dan geldt:

## Extremen



# Hoofdstuk 9 gezamenlijke / simultane kansverdelingen

***DEF:*** De gezamenlijke kansmassafunctie van een paar stochasten X, Y is gedeﬁnieerd door:

* F(a,b) = P(X ≤ a, Y ≤ b).

De marginale verdeling van X in een simultane verdeling is gelijk aan de individuele verdeling van X. deze kan berekend worden met:

De verbanden tussen de simultane dichtheidsfuncties en de simultane verdelingsfuncties zijn:

***DEF:*** X1,X2,...,Xn heten onafhankelijk als voor alle a1,...,an geldt:

* P(X1 ≤ a1, ... , Xn ≤ an) = P(X1 ≤ a1) ··· P(Xn ≤ an).

Anders gezegd, als F(a1,a2,...,an) = FX1(a1)FX2(a2)···FXn(an).

Simultane kansverdelingen zijn erg logisch en overzichtelijk in tabelvorm!

# Hoofdstuk 10 covariantie en correlatie

Als X de kansmassafunctie p(a) = P(X = a) heeft en Y =g(X), dan geldt:

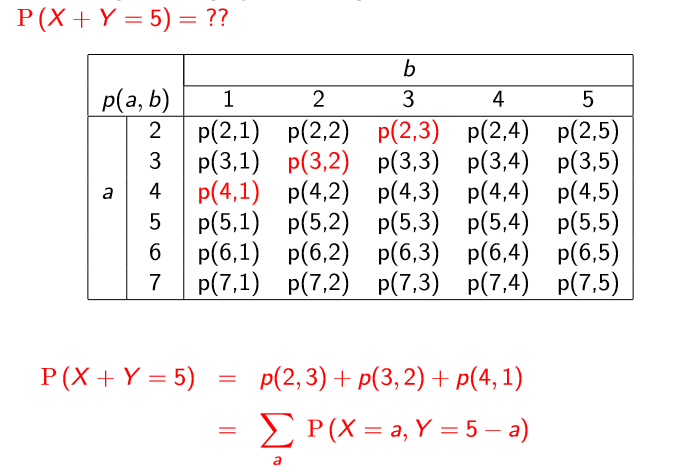
Voor de simultane functies geldt dit op dezelfde manier. De functie g(x) word dan g(x,y,z,…) en de functie p en f worden dan ook van meerdere variabelen. En je moet dan zo vaan integreren/sommeren als er variabelen zijn.

Hieruit volgt dat:

* E[aX +bY + c] = a E[X] + b E[Y] +c
* Var(X + Y) = Var(x) + Var(Y) + 2 E[(x-E[X)(Y-E[Y])]
* Var(X + Y) = Var(x) + Var(Y) + 2 Cov(X,Y)
* Cov(X,Y) = E[(x-E[X)(Y-E[Y])]
* Cov(X,Y) = E[XY]- E[X] E[Y]
* X en Y heten positief / negatief gecorreleerd als naar gelang Cov(X,Y) >0 of < 0
* Als Cov(X,Y) = 0 heten X en Y gecorreleerd
* Voor onafhankelijke stochasten geldt: E[XY ] = E[X]E[Y]
* Cov(rX+s,TY + u) = rt Cov(X,Y)
* Cov(X,Y + Z) = Cov(X,Y) + Cov(X,Z)

# Hoofdstuk 11.1 & 11.2 sommen van stochasten ?????

## Discrete stochasten



Levert:

In het geval X en Y onafhankelijk zijn word dit:

Hieruit volgt dat voor onafhankelijk binomiale verdelingen met dezelfde p geldt:

Bin(n,p) + Bin(m,p) = Bin(n+m, p)

## Continue stochasten

Als X en Y de onafhankelijke stochasten zijn met dichtheden fX en fY, danwordt de dichtheid S = X+Y gegeven door (*LET OP:* dit is de convolutieintegraal):

De som Sn van n onafhankelijke exponenti¨ele variabelen met dezelfde parameter λ heeft een Gamma(n,λ)-verdeling:

Als X ∼N(µ1, σ2 1) en Y ∼N(µ2, σ2 2) onafhankelijke normale variabelen zijn, dan heeft S = X + Y ook een normale verdeling. Uiteraard geldt dan S ∼N(µ1 + µ2, σ2 1 + σ2 2).

U = a1Z1 + a2Z2, V = b1Z1 + b2Z2, dan hebben U en V een zogenoemde bivariate normale verdeling. Eenvoudig te bewijzen: Cov(U,V) = a1b1 + a2b2.

# Hoofdstuk 12 Poisson proces

***DEF:*** Een rij gebeurtenissen op stochastische tijdstippen X1,X2,X3,... is een Poissonproces als

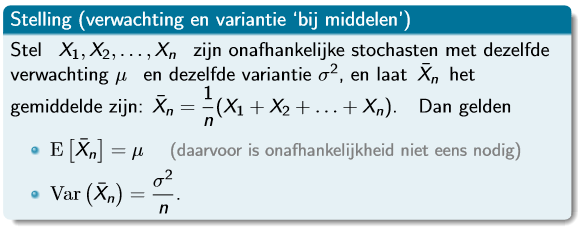
1. Het verwachte aantal gebeurtenissen in een interval van lengte u gelijk is aan λu; λ heet de intensiteit.
2. De aantallen gebeurtenissen N1,N2 in disjuncte tijdsintervallen zijn onafhankelijke stochasten.

Het Poissonproces heeft de onderstaande kansdichtheid functie:

Waarin μ het gemiddeld aantal berichten per tijdseenheid is.

Het poisson proces word veel gebruikt bij aankomsttijden. Hierin is p(2) de kans dat er in een bepaalde tijdsperiode 2 evenementen tegelijke plaatsvinden.

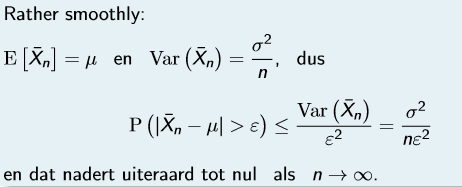
# Hoofdstuk 13 Wet van de grote getallen



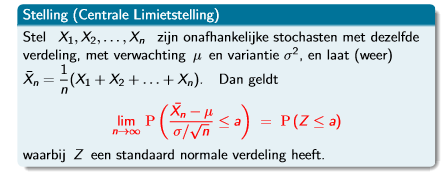
Hieruit volgt een modificatie van Chebychev’s ongelijkheid:

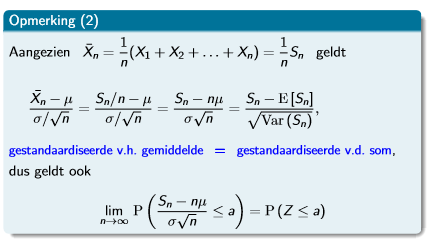
Hieruit volgt de “” regel:

Bewijs dat de fout naar 0 gaat als n naar oneindig gaat:



# Hoofdstuk 14 centrale limiet stelling

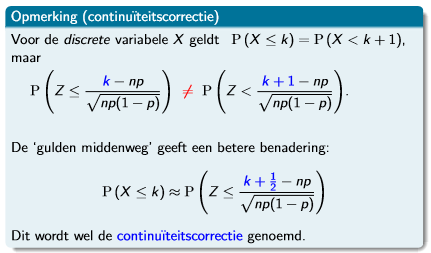




Waarin sn de som van de stochasten X1,X2, … Xn is.

## Benadering voor de binomiale verdeling

De bovenstaande stelling kan gebruikt worden om de binomiale verdeling te benaderen. Hierbij moet er wel rekening gehouden worden met een continuiteitscorrectie:



# Hoofdstuk 15 grafische representatie

## Histogrammen

De hoogte van de histogram kan berekend worden door:

idee: het oppervlakte van de balk van de histogram is gelijk aan de kans dat er een waarde uit dat interval optreed

## Kerndichtheidsschatting

De kerndichtheids schatting is een functie die de kansdichtheid functie benaderd. Het idee is dat er steeds zandhoopjes met het volume gelijk aan de kans van het interval. Deze hoopjes worden opgestapeld en vormen samen de kerndichtheidfunctie.

## Empirische verdelingsfunctie

De empirische verdelingsfunctie is een manier om de verdelingsfunctie van een stochast te schatten. De functie is gelijk aan:

# Hoofdstuk 16 numerieke representaties

## Empirische grootheden









## Empirische kwantielen

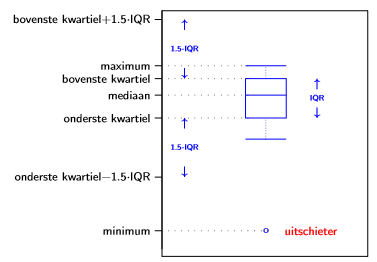
k = floor(p(n+1)) -> kwantiel \* (n+1)

α = p(n+1) – k

qn(p) = xk + α(xk+1 – xk)

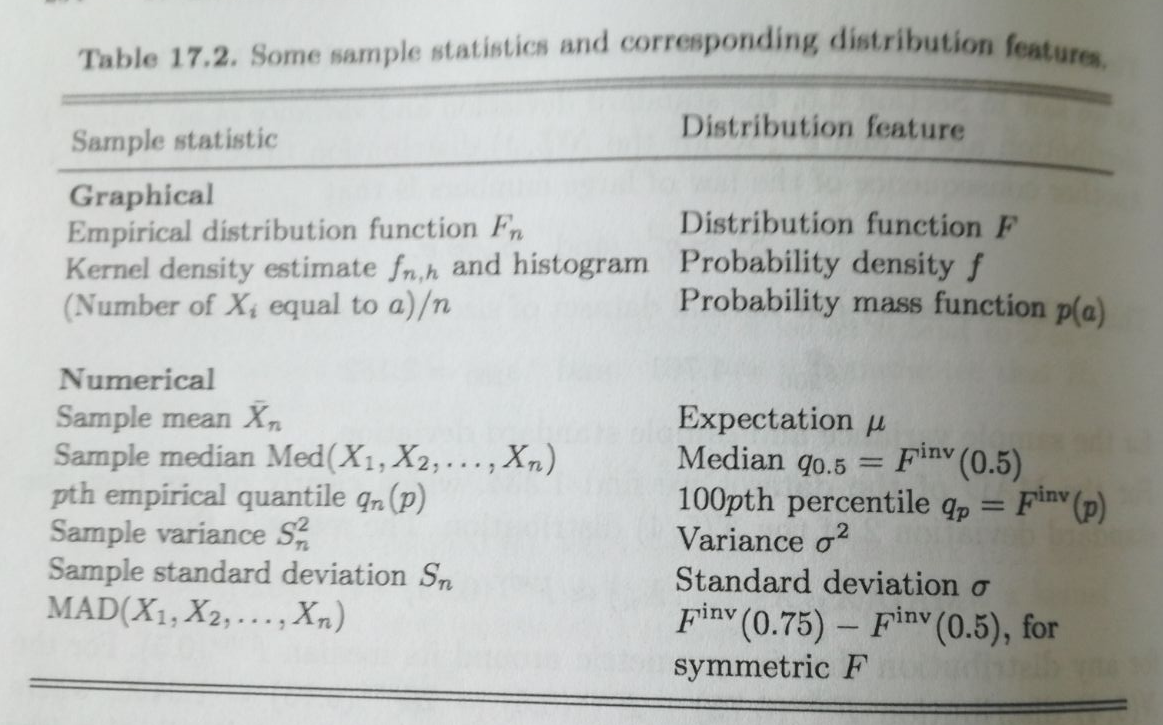
IQR = qn(0.75) – qn(0.25)

met deze waarde is de dataset te visualiseren met een boxplot. Een boxplot word als volg geconstrueerd:



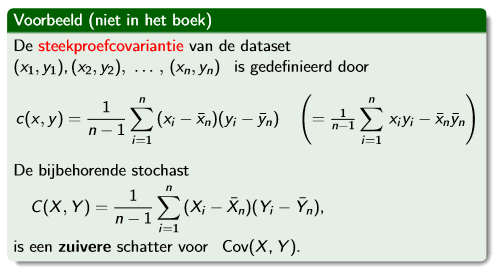
Punten die buiten het onderste kwartiel + 1.5\*IQR worden met rondjes weergegeven.

# Hoofdstuk 17 statistische modellen??



# Hoofdstuk 19 & 20 onafhankelijke schatters

Een schatter heet zuiver (unbiased) voor een parameter x als geldt E[T] = x



Bij het vergelijken van twee zuiveren schatters is de schatter met de kleinste variantie de beste

Bij het vergelijken van niet zuivere schatters is de schatter met de kleinste MSE de beste

MSE(T) =E[(T- θ)2] = Var(T) +E[T − θ]2 = Var(T) + (bias)2.

# Hoofdstuk 21 likelihood functies

DEF: de likelihood functie van een parameter θ L(θ) is gedefineerd als:

L(θ) = P(X1 = x1,X2 = x2,...,Xn = xn) = pθ(x1)pθ(x2)···pθ(xn)

De maximum likelihood schatting van θ is de waarde van θ die de lilelihood L(θ) maximaliseerd. Uit te rekeken door L’(θ) = 0.

Voor continue stochasten geldt dit op dezelfde manier:

L(θ) = P(xi −ε ≤ Xi ≤ xi + ε) = fθ(x1)fθ(x2)···fθ(xn)

## Loglikelihood functie

De loglikelihood functie l(θ) is de log van L(θ) hierdoor is het nulstellen van de afgeleiden veel makkelijker

## Uitzondering:

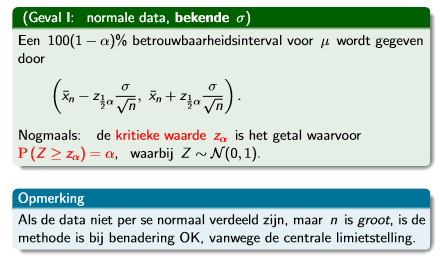


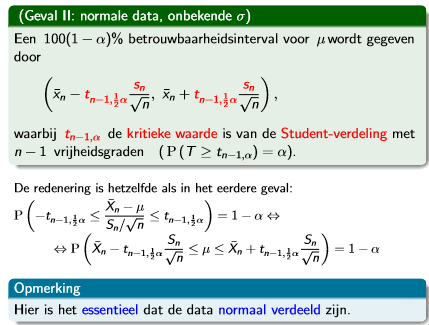
De likelihood functie hiervan is:



De afgeleide hiervan is nooit gelijk aan nul! Maar door de discontinuiteit is L(θ) maximaal voor θ = 1.57

# Hoofdstuk 23 en 24 betrouwbaarheidsintervallen





Er zijn ook eenzijdige betrouwbaarheidsintervallen. Het enige dat hier anders is dat je alpha niet door twee moet delen.

# Hoofdstuk 25 & 26 het toetsen van hypothesen

## Begrippen

*Type I en type II fout:*



*Toetsingsgrootheid (test statistic)(T)*: een steekproefgrootheid op grond waarvan besloten wordt wel of niet te verwerpen.

*p-waarde:* een realisatie t0 van de test statistic T:de kans op een minstens zo afwijkende waarde van T, gegeven dat de nulhypothese waar is.

*signiﬁcantieniveau*: Van te voren afgesproken niveau van de p-waarde waarbij de null hypothese word verworden.

*Kritieke gebied*: de verzameling uitkomsten K waarbij H0 verworpen zal worden: H0 wordt verworpen als de geobserveerde waarde t in K ligt.

*Kritieke waarden:* de waarden op de grenzen van het kritiek gebied.

*Lifcycle van het toetsen van de hypothese:*

* Er is ewen onderzoeksvraag
* Er word data verzameld
* Op grond van data en context word een model geformuleerd.
* De onderzoeksvraag word vertaald naar een hypothese over (de parameters van) het model
  + De p- waarde bij de geobserveerde t van T word bepaald of
  + Het kritieke gebied K word berekend, en ga na of t ∈ K
* Verwerp al of niet de nulhypothese
* Trek een conclusie uit de vorige stappen

## Hoe bereken je de kans op type I en type II fouten?

* De kan op een type I fout is altijd ≤ α
* kans op type II fout kan willekeurig dicht bij 1−α liggen

Als µ de parameter van interesse is kunnen er alleen kansen op type I en II fouten berekend worden voor een specifieke waarde van µ.

De kansen op type I en II fouten kunnen berekend worden met “two sided-tail probabilities en one-sided tail probabilities”

# Totaal rekenregels voor kansrekening H 2 en H3

* P(A) = P(A∩B) + P(A∩Bc)
* P(A1 ∪A2 ∪A3 ∪...) = P(A1) + P(A2) + P(A3) + ... {Geld alleen voor disjuncte gebeurtenissen}
* P(A∩B) = P(A) P(B) {Geldt alleen voor onafhankelijke geb.}
* P(A∪B) = P(A) + P(B)−P(A∩B) {Somregel}
* P(A∪B)c = P(Ac∩Bc) {Morgan’s law}
* P(A∩B)c = P(Ac∪Bc) {Morgan’s law}
* P(A) = P(A|C1) P(C1) + P(A|C2) P(C2) + … + P(A|Cn) P(Cn) {wet van de totale kans}
* {Bayes rule}