UNIVERSITY OF GHENT

Samenvatting

Formules uit de cursus 'Waarschijnlijkheidsrekenen en statistiek'

Auteur:
Nicolas Vanden Bossche

Lesgever:
Prof. Hans De Meyer

Hoofdstuk 1

Het kansbegrip en elementaire kansrekening

1.1 Formules

- $P(A) = \dfrac{\#A}{\#\Omega}$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $p = \dfrac{\text{oppervlakte } A}{\text{oppervlakte } \Omega}$ of nog $P(A) = \dfrac{\text{maat van } A}{\text{maat van } \Omega}$
- $R_n(A) = \dfrac{F_n(A)}{n}$
- $P(A) = \lim\limits_{n \mapsto \infty} R_n(A)$

Hoofdstuk 2

Axioma's van de kanstheorie en het begrip voorwaardelijke kans

2.1 Unie van kansen

- A_k disjunct $\Rightarrow P\left(\bigcup\limits_{k=1}^{\infty} A_k\right) = \sum\limits_{k=1}^{\infty} P(A_k)$

2.2 Voorwaardelijke kansen

- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
- Vermenigvuldigingsregel:
  $P\left(\bigcap\limits_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots$
- Totalekansformule:

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \cdots + P(A_n) \cdot P(B|A_n) \end{aligned}$$

- Regel van Bayes:
  $P(A_i|B) = \dfrac{P(A_i)P(B|A_i)}{P(B)} = \dfrac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)}$

2.3 Onafhankelijke gebeurtenissen

2.3.1 Gewoon onafhankelijk

- $P(A|B) = P(A)$
- $P(A \cap B) = P(A)P(B)$

2.3.2 Voorwaardelijk onafhankelijk

- $P(A \cap B|C) = P(A|C)P(B|C)$

2.4 Inclusie-exclusie

2.4.1 Formule

- Inclusie-exclusieformule:

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &- \cdots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_k\right) \end{aligned}$$

2.4.2 Incidentie

- De kans op ten minste 1 incidentie:

$$\begin{aligned} p_a &= P(A_1 \cup \cdots \cup A_n) = \binom{N}{1}\frac{(N-1)!}{N!} - \binom{N}{2}\frac{(N-2)!}{N!} + \binom{N}{3}\frac{(N-3)!}{N!} - \cdots \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{N+1}\frac{1}{N!} \end{aligned}$$

- De kans op juist k incidenties:
  $p_k = \frac{q_{N-k}}{k!}$ met $q_N = \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{N+1}\frac{1}{N!}$

Hoofdstuk 3

Rijen toevalsexperimenten en Markov-ketens

3.1 Bernoulli

- Kans op m keer succes op n experimenten:
  $p_n(m) = \binom{n}{m} p^m q^{n-m}$

3.2 De Moivre-Laplace

3.2.1 Lokale limietstelling

- De kans dat A zicht juist m keer voordoet op n experimenten:
  Voor $z \mapsto \infty$: $\lim\limits_{n \mapsto \infty} \sqrt{2\pi npq}\; e^{z^2/2}\; p_n(m) = 1$
  met $z = \dfrac{m - np}{\sqrt{npq}}$
  Of nog: $p_n(m) \sim \dfrac{1}{\sqrt{2\pi npq}\; e^{z^2/2}}$

3.2.2 Integrale limietstelling

- Voor S_n het aantal successen in n opeenvolgende experimenten en voor elke $a \leq b$:
  $\lim\limits_{n \mapsto \infty} P\left(a \leq \dfrac{S_n - np}{\sqrt{npq}} \leq b\right) = \dfrac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-z^2/2} \mathrm{d}z$

3.3 Stochastisch convergeren (Stelling van Bernoulli)

- $\lim\limits_{n \mapsto \infty} P\left(\left|\frac{S_n}{n} - p\right| \leq \epsilon\right) = 1, \forall \epsilon > 0$
- Of nog: $\frac{S_n}{n} \xrightarrow{p} p$

3.4 De functie Φ

- $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-z^2/2} \mathrm{d}x$
- $\frac{1}{\sqrt{2\pi}} \int\limits_{a}^{b} e^{-z^2/2} \mathrm{d}z = \Phi(b) - \Phi(a)$

3.5 Limietstelling van Poisson

- $\lim\limits_{n \mapsto \infty} p_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}$ met $\lambda = np$
- Of nog: $p_n(m) \sim \frac{(np)^m}{m!} e^{-np}$

3.6 Markovketens

- $\pi_n = (\pi_1)^n$
- Evenwichtsvergelijkingen:
  $p_j = \sum\limits_{i=1}^{k} p_i p_{ij}$ met $p_i j$ de kans om van i naar j te gaan in 1 stap
- Normaliseringsvergelijking:
  $\sum\limits_{i=1}^{k} p_i = 1$

3.7 Geboorte-dood processen

- Formule:

$$p_i b_i = p_{i+1} d_{i+1} \text{voor} i = 0, ..., m \quad \text{met} \quad b_i = P(A_{i+1}^{(n+1)} | A_i^{(n)})$$
$$\text{en} \quad d_i = P(A_{i-1}^{(n+1)} | A_i^{(n)})$$

- Of nog:
  $p_i = p_0 \frac{b_0 b_1 \cdots b_{i-1}}{d_1 d_2 \cdots d_i}$ met $i = 1,..., m$

3.8 Limietstellingen voor homogene Markov-ketens met $k = 2$

3.8.1 Lokale limietstelling

- Voor $p_n(m)$ de kans op m keer succes in n opeenvolgende experimenten met de volgende parameters:

$$\pi_1 = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix} \text{ met } (\alpha, \beta) \in]0,1[^2,\ p = \frac{\beta}{1-\alpha+\beta},\ q = 1-p$$

- Er geldt:

$$z = (m-np)\frac{1-\alpha+\beta}{\sqrt{m\alpha(1-\alpha)+(n-m)\beta(1-\beta)}}$$

- Of nog (voor z eindig):

$$p_n(m) \sim \sqrt{\frac{1-\alpha+\beta}{2\pi npq(1+\alpha-\beta)}}\ e^{-z^2/2}$$

3.8.2 Integrale limietstelling

- Voor S_n het aantal successen in n opeenvolgende experimenten, voor alle $a \leq b$ en dezelfde parameters als hierboven:

$$\begin{aligned} P\left(a \leq (S_n - np)\sqrt{\frac{1-\alpha+\beta}{npq(1+\alpha-\beta)}} \leq b\right) &\sim \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}\mathrm{d}z \\ &= \Phi(b) - \Phi(a) \end{aligned}$$

Hoofdstuk 4

Toevalsveranderlijken

4.1 Discrete toevalsveranderlijken

4.1.1 Algemene formules

- $P(X \in S) = \sum\limits_{x \in S} p_X(x)$
- $\sum\limits_{x \in W} p_x(x) = 1$
- Voor $Y = g(X)$ geldt: $p_Y(y) = \sum \{x|g(x) = y\} p_X(x)$

4.1.2 Verwachtingswaarde en variantie

Verwachtingswaarde

- $E[X] = \sum\limits_{x \in W} x p_X(x)$
- $\mu_k(X) = E[X^k]$

Variantie

- $\mathrm{var}(X) = E[(X - E[X])^2]$
- $\sigma_X = \sqrt{\mathrm{var}(X)}$

Eigenschappen

- $E[g(X)] = \sum\limits_{x \in W} g(x) p_X(x)$
- Voor $Y = aX + b$ geldt:

$$\begin{aligned} E[Y] &= aE[X] + b \\ \text{var}(Y) &= a^2\text{var}(X) \end{aligned}$$

- $\text{var}(X) = E[X^2] - (E[X])^2$

4.2 Continue toevalsveranderlijken

4.2.1 Algemene formules

- $P(X \in B) = \int\limits_B f_X(x)\mathrm{d}x$
- $P(a \leq X \leq b) = \int\limits_a^b f_X(x)\mathrm{d}x$
- $\int\limits_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$

4.2.2 Verwachtingswaarde en variantie

Verwachtingswaarde

- $E[X] = \int\limits_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$

Variantie

- $\text{var}(X) = E[X^2] - (E[X])^2$

Eigenschappen

- $E[g(X)] = \int\limits_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x$
- Voor $Y = aX + b$ geldt:

$$\begin{aligned} E[Y] &= aE[X] + b \\ \text{var}(Y) &= a^2\text{var}(X) \end{aligned}$$

4.3 Cumulatieve verdelingsfuncties

4.3.1 Discreet

- $F_X(x) = P(X \leq x) = \sum\limits_{k \leq x} p_X(k)$
- $p_X(x) = F_X(x) - F_X(x-1)$

4.3.2 Continu

- $F_X(x) = \int\limits_{-\infty}^{x} f_X(t) \mathrm{d}t$
- $f_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x}(x)$

4.4 Centrale momenten, skewness en kurtosis

4.4.1 (Genormaliseerde) centrale momenten

- Het k-de centrale moment:

  $\mu'_k(X) = E[(X - E[X])^k]$

- Het genormaliseerde k-de centrale moment:

  $\alpha_k(X) = \frac{E[(X - E[X])^k]}{(\mathrm{var}(X))^{k/2}}$

4.4.2 Skewness

- $\alpha_3(X) > 0$: scheef naar rechts
- $\alpha_3(X) < 0$: scheef naar links
- $\alpha_3(X) = 0$: symmetrisch

4.4.3 Kurtosis

- $\alpha_4(X) > 3$: gepiekt
- $\alpha_4(X) < 3$: afgeplat
- $\alpha_4(X) \approx 3$: mesokurtic

Hoofdstuk 5

Gezamenlijke en voorwaardelijke kansverdelingen

5.1 Gezamenlijke discrete toevalsveranderlijken

5.1.1 Algemeen

- $p_{X,Y}(x,y) = P(\{X = x\} \cap \{Y = y\}) = P(X = x, Y = y)$
- Marginaal:

$$\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x,y) \\ p_Y(y) &= \sum_x p_{X,Y}(x,y) \end{aligned}$$

5.1.2 Eigenschappen

- Voor $Z = g(X,Y)$ geldt: $p_Z(z) = \sum\limits_{\{(x,y)|g(x,y)=z\}} p_{X,Y}(x,y)$
- $E[g(X,Y)] = \sum\limits_x \sum\limits_y g(x,y) p_{X,Y}(x,y)$
- $E[aX + bY] = aE[X] + bE[Y]$

5.2 Gezamenlijke continue toevalsveranderlijken

5.2.1 Algemeen

- $P((X,Y) \in B) = \int \int\limits_{(x,y)\in B} f_{X,Y}(x,y) \mathrm{d}x\mathrm{d}y$

- $P(a \leq X \leq b, c \leq Y \leq d) = \int\limits_c^d \int\limits_a^b f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$
- $P(X \in A) = \int\limits_A \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x = \int\limits_A F_X(x)$
- Marginaal:

$$\begin{aligned} f_X(x) &= \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y \\ f_Y(y) &= \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x \end{aligned}$$

5.2.2 Eigenschappen

- Als X en Y gezamenlijk normaal verdeeld zijn met parameters $\sigma_X, \sigma_Y, \mu_X, \mu_Y$ en ρ, dan geldt dat $X \stackrel{d}{=} N(\mu_X, \sigma_X)$ en $Y \stackrel{d}{=} N(\mu_Y, \sigma_Y)$
- $E[g(X,Y)] = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$
- $E[aX + bY] = aE[X] + bE[Y]$

5.3 Gezamenlijke cumulatieve verdelingen

5.3.1 Discreet

- $F_{X,Y}(x,y) = \sum\limits_{k \leq x} \sum l \leq y p_{X,Y}(k,l)$
- Voor $X = \{x_1, x_2, ...\}$ en $Y = \{y_1, y_2, ...\}$:

  $p_{X,Y}(x_k, y_l) = F_{X,Y}(x_k, y_l) - F_{X,Y}(x_{k-1}, y_l) - F_{X,Y}(x_k, y_{l-1}) + F_{X,Y}(x_{k-1}, y_{l-1})$

5.3.2 Continu

- $F_{X,Y}(x,y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f_{X,Y}(s,t)\mathrm{d}s\mathrm{d}t$
- $f_{X,Y}(x,y) = \dfrac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$

5.4 Covariantie en correlatie

5.4.1 Algemeen

- $\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$
- $\rho(X, Y) = \dfrac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$

5.4.2 Eigenschappen

- $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$
- $\text{cov}(aX + bY + c, Z) = a\text{cov}(X, Z) + b\text{cov}(Y, Z)$
- $\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\text{cov}(X, Y)$

5.5 Voorwaardelijke kansverdelingen

5.5.1 Discreet

Algemeen

- $p_{X|A}(x) = P(X = x|A) = \dfrac{P(\{X = x\} \cap A}{P(A)}$
- $p_{X|Y}(x|y) = P(X = x|Y = y) = \dfrac{p_{X,Y}(x, y)}{p_Y(y)}$
- Totalekansformule:

  $p_X(x) = \sum_{i=1}^{n} P(A_i)p_{X|A}(x)$ met $A_1 \cdots A_n$ een partitie voor Ω

Verwachting

- $E[X|A] = \sum_x xp_{X|A}(x)$
- $E[X|Y = y] = \sum_x xp_{X|Y}(x|y)$
- $E[g(X)|A] = \sum_x g(x)p_{X|A}(x)$
- $E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y)$
- Formule van totale verwachting:

  $E[X] = \sum_{i=1}^{n} P(A_i)E[X|A_i]$

5.5.2 Continu

Algemeen

- $f_{X|A}(x) = \dfrac{f_X(x)}{P(X \in A)}$
- $f_{X|Y}(x|y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & , x \in A \\ 0 & \text{, elders} \end{cases}$
- Totalekansformule:

  $f_X(x) = \sum\limits_{i=1}^{n} P(A_i) p_{X|A}(x)$ met $A_1 \cdots A_n$ een partitie voor Ω

Verwachting

- $E[X|A] = \int\limits_{-\infty}^{\infty} x p_{X|A}(x)$
- $E[X|Y = y] = \int\limits_{-\infty}^{\infty} x p_{X|Y}(x|y)$
- $E[g(X)|A] = \int\limits_{-\infty}^{\infty} g(x) p_{X|A}(x)$
- $E[g(X)|Y = y] = \int\limits_{-\infty}^{\infty} g(x) p_{X|Y}(x|y)$
- Formule van totale verwachting:

  $E[X] = \sum\limits_{i=1}^{n} P(A_i) E[X|A_i]$

5.6 Onafhankelijkheden

Voor alle X en Y onafhankelijk geldt:

5.6.1 Discreet

- $p_{X|A}(x) = p_X(x)$
- $p_{X,Y}(x,y) = p_X(x) p_Y(y)$
- $p_{X,Y|A}(x,y) = p_{X|A}(x) p_{Y|A}(y)$
- $p_{X_1 \cdots X_n}(x_1 \cdots x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots$

5.6.2 Continu

- $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- Of nog:

$$
\begin{aligned}
f_{X|Y}(x|y) &= f_X(x), \forall(x,y) : f_Y(y) > 0 \\
f_{Y|X}(y|x) &= f_Y(y), \forall(x,y) : f_X(x) > 0
\end{aligned}
$$

- $f_{X_1 \cdots X_n}(x_1 \cdots x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots$

5.6.3 Algemeen

- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
- $F_{X_1 \cdots X_n}(x_1 \cdots x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots$
- $E[XY] = E[X]E[Y]$
- $\text{cov}(X,Y) = 0$
- $\text{var}(aX + bY) = a^2\text{var} + b^2\text{var}(Y)$

Hoofdstuk 6

Functies van toevalsveranderlijken

6.1 Afgeleide verdelingen

6.1.1 Algemeen

- Haal f_Y uit f_X met $Y = g(X)$:

  1. $F_Y(y) = P(g(X) \leq y) = \int\limits_{\{x|g(x)\leq y\}} f_X(x)\mathrm{d}x$
  2. $f_Y(y) = \dfrac{\mathrm{d}F_Y(Y)}{\mathrm{d}y}$

- $X \stackrel{d}{=} N(0,1) \Rightarrow X^2 \stackrel{d}{=} \text{Gamma}(\frac{1}{2}, \frac{1}{2})$
- $\text{Gamma}(\frac{n}{2}, \frac{1}{2}) = \chi^2_n$ met $E[X] = n$ en $var(X) = 2n$

6.1.2 Lineaire functies

Voor $Y = aX + b$ geldt:

- Continu:

  $$f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$$

- Exponentieel:

  $$f_Y(y) = \frac{\lambda}{|a|} e^{-\lambda(y-b)/a}$$

- Normaal:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-(y-b-a\mu)^2/2a^2\sigma^2}$$

6.1.3 Monotone functies

Voor g strikt monotoon en $Y = g(X)$ waarbij $y = g(x)$ en $x = h(y)$ geldt:

- $f_Y(y) = f_X(h(y)) \left| \frac{\mathrm{d}h(y)}{\mathrm{d}y} \right|$

6.2 Momentgenererende functies

6.2.1 Algemeen

- Functie:

$$\begin{aligned} M_X(s) = E[e^{sx}] &= \sum_x e^{sx} p_X(x) \\ &= \int_{-\infty}^{\infty} e^{sx} f_X(x) \mathrm{d}x \end{aligned}$$

- $Y = aX + b \Rightarrow M_Y(s) = e^{sb} M_X(sa)$
- $\mu_k(X) = E[X^k] = \frac{\mathrm{d}^k}{\mathrm{d}s^k} M_X(s) \Big|_{s=0}$

6.2.2 Onafhankelijkheden

- Voor X en Y onafhankelijk en $Z = X + Y$ geldt:

  $M_Z(s) = M_X(s) M_Y(s)$

- Voor $X_1 \dots X_n$ onafhankelijke Poisson-veranderlijken met parameters $\lambda_1 \dots \lambda_n$:

  $Z = X_1 + \cdots + X_n \stackrel{d}{=} \text{Poisson}(\lambda_1 + \cdots + \lambda_n)$

- Voor $X_1 \stackrel{d}{=} N(\mu_1, \sigma_1^2), \dots, X_n \stackrel{d}{=} N(\mu_n, \sigma_n^2)$ en onafhankelijk:

  $Z = X_1 + \cdots + X_n \stackrel{d}{=} N(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2)$

- Voor $X_1 \stackrel{d}{=} \text{Gamma}(\alpha_1, \beta), \dots, X_n \stackrel{d}{=} \text{Gamma}(\alpha_n, \beta)$ en onafhankelijk:

  $Z = X_1 + \cdots + X_n \stackrel{d}{=} \text{Gamma}(\alpha_1 + \cdots + \alpha_n, \beta)$

6.3 Convolutie

Voor $Z = X + Y$ en X en Y onafhankelijk, geldt:

- $p_Z(z) = \sum\limits_x p_X(x)p_Y(z-x)$
- $f_Z(z) = \int\limits_{-\infty}^{\infty} f_X(x)f_Y(z-x)\mathrm{d}x$

6.4 Randomgeneratoren

6.4.1 Algemeen

- $U_1, U_2, \ldots$, uniform verdeeld op $[0,1]$ noemen we een rij onafhankelijke randomgetallen.
- Multiplicatieve congruentiemethode met multiplicator r, de modulus m en de startwaarde x_0:

  $x_n = rx_{n-1}\text{mod}m, \quad n = 1, 2, \ldots$

6.4.2 Pseudo-inverse van $F_X(x)$

- $F_X^{(-1)}(y) = min\{x \in \bar{\mathbb{R}} | F_X(x) \geq y\}$
- Voor F_X continu en strikt stijgend:

  $F_X^{(-1)} = F_X^{-1}$
- $F_X^{(-1)}(y) \leq u \Leftrightarrow y \leq F_X(u)$
- Voor $F : \bar{\mathbb{R}} \rightarrow [0,1]$ een niet-dalende functie met $F(-\infty) = 0$ en $F(\infty) = 1$:

  $U \overset{d}{=} U[0,1] \Rightarrow X = F_{(-1)}(U)$ heeft F als cumulatieve verdelingsfunctie

6.5 Statistieken

- Een statistiek is een functie $g(X_1, X_2, \ldots, X_n)$
- $\bar{X}_n = \dfrac{\sum\limits_{i=1}^{n} X_i}{n}$ met $E[\bar{X}_n] = E[X]$ en $var(\bar{X}_n) = \dfrac{var(X)}{n}$
- $X \overset{d}{=} N(\mu, \sigma^2) \Rightarrow \bar{X}_n \overset{d}{=} N\left(\mu, \dfrac{\sigma^2}{n}\right)$

- Steekproefvariantie:

$$S_n^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}{n-1}$$

- Steekproefstandaardafwijking:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}{n-1}}$$

- $E[S_n^2] = var(X)$
- Voor de steekproefvariantie S_n^2 van een aselecte steekproef van omvang n, geldt:

$$\begin{aligned} \frac{(n-1)S_n^2}{\sigma^2} &\overset{d}{=} \chi^2_{n-1} \\ S_n^2 \text{ is onafhankelijk van } \bar{X}_n & \\ E[S_n^2] &= \sigma^2 \\ var(S_n^2) &= \frac{2\sigma^4}{n-1} \end{aligned}$$

- Steekproefmoment:

$$\begin{aligned} \text{Het k-de moment } M_k &= \frac{\sum_{i=1}^{n} X_i^k}{n} \\ \text{Het k-de centrale moment } M'_k &= \frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^k}{n} \end{aligned}$$

- Scheefheid:

$$\alpha_3 = \frac{M'_3}{(M'_2)^{3/2}}$$

- Kurtosis:

$$\alpha_4 = \frac{M'_4}{(M'_2)^2}$$

6.6 Ordestatistieken

6.6.1 Algemeen

- Men noemt Y_k de k-de orde statistiek als:

$$Y_1 \leq \cdots \leq Y_n$$

- $F_{Y_k}(y) = \sum\limits_{j=k}^{n} \binom{n}{j} [F_X(y)]^j [1 - F_X(y)]^{n-j}$
- Voor X continu:

  $F_{Y_k}(y) = k\binom{n}{k}[F_X(y)]^{k-1}[1 - F_X(y)]^{n-k} f_X(y)$
- Voor k = 1:

$$\begin{aligned} F_{Y_1}(y) &= 1 - [1 - F_X(y)]^n \\ f_{Y_1}(y) &= n[1 - F_X(y)]^{n-1} f_X(y) \end{aligned}$$

- Voor k = n:

$$\begin{aligned} F_{Y_n}(y) &= [F_X(y)]^n \\ f_{Y_n}(y) &= n[F_X(y)]^{n-1} f_X(y) \end{aligned}$$

6.6.2 Begrippen

- Steekproefbereik $R = Y_n - Y_1$
- Midden van het steekproefbereik $MR = \dfrac{Y_1 + Y_n}{2}$
- Steekproefmediaan $M = \begin{cases} \dfrac{Y_{\frac{n}{2}} + Y_{\frac{n}{2}+1}}{2} & \text{, als n even is} \\ \dfrac{Y_{\frac{n}{2}}}{2} & \text{, als n oneven is} \end{cases}$
- Percentiel $P_p = (1-q)Y_k + qY_{k+1}$ met $p \in]0, 100[$, $k =$ het geheel deel van $\frac{(n+1)p}{100}$ en q het fractioneel deel
- $P_25 = Q_1, P_50 = Q_2 = M, P_75 = Q_3$
- Interkwartielbereik $IRQ = Q_3 - Q_1$

6.6.3 Eigenschappen

- Voor $j < k$ geldt:

$$f_{Y_j,Y_k}(y,z) = \begin{cases} \dfrac{n!}{(j-1)!(k-j-1)!(n-k)!}[F_X(y)]^{j-1} f_X(y) & \\ \qquad \times [F_X(z) - F_X(y)]^{k-j-1} f_X(z)[1 - F_X(z)]^{n-k} & \text{, als } y \leq z \\ 0 & \text{elders} \end{cases}$$

- De kansdichtheidsfunctie van het steekproefbereik R:

$$f_R(r) = \begin{cases} n(n+1)\int_{-\infty}^{\infty} [F_X(x) - F_X(x-r)]^{n-2} f_X(x-r) F_X(x) \mathrm{d}x & \text{, als } r > 0 \\ 0 & \text{, als } r \leq 0 \end{cases}$$

Hoofdstuk 7

Wet van de grote aantallen en de centrale limietstelling

7.1 De ongelijkheden van Markov en Chebyshev

7.1.1 De ongelijkheid van Markov

- $P(X \geq a) \leq \dfrac{E[X]}{a}$
  !! $\bar{X}_n \xrightarrow{p} E[X]$!!

7.1.2 De ongelijkheid van Chebyshev

- $P(|X - E[X]| \geq c) \leq \dfrac{var(X)}{c^2}$

7.2 De centrale limietstelling

- Voor $X_1, X_2, \dots$ met dezelfde verdeling en onafhankelijk waarvoor $E[X_i] = \mu < \infty$ en $var(X_i) = \sigma^2 < \infty$ en > 0:

  $$\forall x \in \mathbb{R} : \lim_{n \mapsto \infty} P\left(\frac{(X_1 - \mu) + \cdots + (X_n - \mu)}{\sqrt{n}\sigma} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} \mathrm{d}z$$

- Of nog:

  $$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \overset{d}{/rightarrow} N(0,1)$$

Hoofdstuk 8

Schatten, schatters en betrouwbaarheidsintervallen

8.1 Schatters

8.1.1 Meest aannemelijke schatter

- $\hat{\theta} = \max\limits_{\theta} f_n(x_1, \ldots, x_n|\theta)$

8.1.2 De momentenmethode

- Voor n parameters te bepalen, gebruik je minstens n van de volgende vgl:

  $\mu_k = m_k$ of nog $E[X^k] = \frac{1}{n} \sum X_i^k$

8.2 Kwaliteit van schatters

- Een schatter δ is zuiver als:

  $\text{bias}(\delta) = E_\theta[\delta] - \theta = 0$ waarbij

  $E_\theta[Z] = E_\theta[g(X_1, \ldots, X_n)] = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1 \ldots x_n) f(x_1 \ldots x_n|\theta) \mathrm{d}x_1 \ldots \mathrm{d}x_n$

- Gemiddelde kwadratische fout:

  $\text{MSE}(\delta) = E_\theta[(\delta(X_1, \ldots, X_n) - \theta)^2] = \text{var}(\delta(X_1, \ldots, X_n)) + bias(\delta)$

8.3 Betrouwbaarheidsintervallen

8.3.1 Voor de verwachtingswaarde

- Voor X normaal verdeeld met ONBEKENDE verwachting en BEKENDE variantie:

$$\begin{aligned} T_1 &= \bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ T_2 &= \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

8.3.2 Voor relatieve frequenties

- Neem $\hat{p} = \dfrac{S_n}{n}$ als schatter:

$$\begin{aligned} T_1 &= \hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ T_2 &= \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \end{aligned}$$

8.3.3 Voor de variantie

- Voor X normaal verdeeld met BEKENDE verwachting en ONBEKENDE variantie:

$$\begin{aligned} T_1 &= \frac{n\hat{\sigma}^2}{\chi^2_{n,\frac{\alpha}{2}}} \\ T_2 &= \frac{n\hat{\sigma}^2}{\chi^2_{n,1-\frac{\alpha}{2}}} \end{aligned}$$

  met $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1} (X_i - \mu)^2$

- Of met steekproefvariantie:

$$\begin{aligned} T_1 &= \frac{(n-1)S_n^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \\ T_2 &= \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \end{aligned}$$

  met $S_n^2 = \frac{1}{n-1} \sum_{i=1} (X_i - \bar{X}_n)^2$

8.3.4 z_α-tabel

γ	α	z_α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.80	0.20	0.8416	0.10	1.2816
0.90	0.10	1.2816	0.05	1.6499
0.95	0.05	1.6449	0.025	1.9600
0.98	0.02	2.0537	0.01	2.3263
0.99	0.01	2.3263	0.005	2.5758
0.999	0.001	3.0902	0.0005	3.2905

Hoofdstuk 9

Het toetsen van hypothesen

9.1 Testen en hun betrouwbaarheid

9.1.1 De testfunctie

- $\phi(x) = \begin{cases} 1, & x \in R(\text{ejection}) \\ 0, & x \in A(\text{cceptance}) \end{cases}$

9.1.2 Fouten van type I en II

- Fout van type I : een ware H_0 wordt verworpen
- Fout van type II: een valse H_0 wordt aangenomen

9.1.3 Tabel

	H_0 waar	H_0 vals
H_0 aanvaarden	$1-\alpha$	β
H_0 verwerpen	α	$1-\beta$

9.2 De verschillende testen

9.2.1 z-test normaal gemiddelde, 1 populatie, σ^2 BEKEND

- Kies $H_0 : \mu = \mu_0$
- Kies α
- Bereken $\bar{x}_n$ en $z = \dfrac{\bar{x}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
- Kies $H_1 =$

- $H_1 : \mu < \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow z < -z_\alpha$
- $H_1 : \mu > \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow z > z_\alpha$
- $H_1 : \mu \neq \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow |z| > z_{\frac{\alpha}{2}}$

9.2.2 t-test normaal gemiddelde, 1 populatie, σ^2 ONBEKEND

- Kies $H_0 : \mu = \mu_0$
- Kies α
- Bereken $\bar{x}_n$, s_n^2 en $t = \dfrac{\bar{x}_n - \mu_0}{\sqrt{s_n^2/n}}$
- Kies $H_1 =$
  - $H_1 : \mu < \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow t < t_{n-1,1-\alpha}$
  - $H_1 : \mu > \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow t > t_{n-1,\alpha}$
  - $H_1 : \mu \neq \mu_0 \Rightarrow H_0$ verwerpen $\Leftrightarrow |t| > t_{n-1,\frac{\alpha}{2}}$

9.2.3 z-test normale gemiddelden, 2 populaties, σ_1^2 en σ_2^2 BEKEND

- Kies $H_0 : \mu_1 - \mu_2 = \delta$
- Kies α
- Bereken $\bar{x}_n$, $\bar{y}_m$ en $z = \dfrac{\bar{x}_n - \bar{y}_m - \delta}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}$
- Kies $H_1 =$
  - $H_1 : \mu_1 - \mu_2 < \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow z < -z_\alpha$
  - $H_1 : \mu_1 - \mu_2 > \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow z > z_\alpha$
  - $H_1 : \mu_1 - \mu_2 \neq \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow |z| > z_{\frac{\alpha}{2}}$

9.2.4 t-test normale gemiddelden, 2 populaties, σ_1^2 en σ_2^2 ONBEKEND

- Kies $H_0 : \mu_1 - \mu_2 = \delta$
- Kies α
- Bereken $\bar{x}_n$, $\bar{y}_m$, s_1^2, s_2^2, $s_p^2 = \dfrac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$ en $t = \dfrac{\bar{x}_n - \bar{y}_m - \delta}{\sqrt{s_p^2 \dfrac{(n+m)}{nm}}}$

- Kies $H_1 =$

  - $H_1 : \mu_1 - \mu_2 < \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow z < -z_\alpha$
  - $H_1 : \mu_1 - \mu_2 > \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow z > z_\alpha$
  - $H_1 : \mu_1 - \mu_2 \neq \delta \Rightarrow H_0$ verwerpen $\Leftrightarrow |z| > z_{\frac{\alpha}{2}}$

9.2.5 Goodness-of-fit

- verwerp $H_0 \Leftrightarrow \chi^2 > \chi^2_{j-1-r,\alpha}$ met r het aantal geschatte parameters
- Hierbij is $\chi^2 = \sum_{j=1}^{n} \frac{(N_j - n\pi_j)^2}{n\pi_j}$ met N_j het aantal dat gemeten is voor j en $\pi_j = N \cdot p_j$

Veel succes op het examen!