Opgave 1 Zij $\theta \in \mathbb{R}$, $n \geq 1$ en X_1, X_2, ..., X_n onafhankelijk, identiek verdeelde stochasten met kansdichtheidsfunctie

$$f_\theta(x) = \begin{cases} \frac{1}{4} & \text{als } x \in (\theta-2, \theta+2) \\ 0 & \text{als } x \notin (\theta-2, \theta+2). \end{cases}$$

a 4pt) Bepaal $E(X_1)$ en $Var(X_1)$.
ANTWOORD: X_1 is een uniforme verdeelde stochast op $(\theta-2, \theta+2)$, dus $E(X_1) = \frac{1}{2}(\theta - 2 + \theta + 2) = \theta$, of bereken: $\int\limits_{-\infty}^{\infty} dx f_\theta(x) x = \int\limits_{\theta-2}^{\theta+2} dx \frac{1}{4} x = \frac{1}{8} x^2 |_{\theta-2}^{\theta+2} = \frac{1}{8}((\theta+2)^2 - (\theta-2)^2) = \frac{1}{8}(8\theta) = \theta$.

Omdat X_1 een uniforme stochast is geldt: $Var(X_1) = \frac{(\theta+2-(\theta-2))^2}{12} = \frac{16}{12} = \frac{4}{3}$.
Of bereken $Var(X_1) = E(X_1^2) - E(X_1)^2 = E(X_1^2) - \theta^2$.
$E(X_1^2) = \int\limits_{-\infty}^{\infty} dx f_\theta(x) x^2 = \int\limits_{\theta-2}^{\theta+2} dx \frac{1}{4} x^2 = \frac{1}{12} x^3 |_{\theta-2}^{\theta+2} = \frac{1}{12}((\theta+2)^3 - (\theta-2)^3) = \frac{1}{12}(\theta^3 + 6\theta^2 + 12\theta + 8 - (\theta^3 - 6\theta^2 + 12\theta - 8)) = \frac{1}{12}(12\theta^2 + 16) = \theta^2 + \frac{4}{3}$. Dus $Var(X_1) = E(X_1^2) - \theta^2 = \theta^2 + \frac{4}{3} - \theta^2 = \frac{4}{3}$.

b 4pt) Bepaal de cumulatieve distributiefunctie F_X van X_1.
ANTWOORD: $F_X(x) = P(X \leq x) = \int\limits_{-\infty}^{x} dy f_\theta(y)$

$$F_X(x) = \begin{cases} 0 & \text{als } x \leq -2 + \theta \\ \frac{1}{4}(x - \theta + 2) & \text{als } -2 + \theta < x < 2 + \theta \\ 1 & \text{als } x \geq 2 + \theta \end{cases}$$

Definieer $Y_1 := \min\{X_1, X_2, \ldots, X_n\}$ en $Y_n = \max\{X_1, X_2, \ldots, X_n\}$

c 6pt) Bepaal de kansdichtheidsfunctie van Y_1 en Y_n in termen van de (cumulatieve) distributiefunctie van X_1.
ANTWOORD: $F_{Y_n}(y) := P(Y_n \leq y) = P(\max\{X_1, X_2, \ldots, X_n\} \leq y) = P(X_1 \leq y, X_2 \leq y, \ldots X_n \leq y) = \prod\limits_{i=1}^{n} P(X_i \leq y) = F_X(y)^n$.

Omdat $f_{Y_n}(y) = \frac{d}{dy} F_{Y_n}(y)$ geldt: $f_{Y_n}(y) = n F_X(y)^{n-1} \frac{d}{dy} F_X(y) = n F_X(y)^{n-1} f_X(y)$.

Analoog geldt: $F_{Y_1}(y) := P(Y_1 \leq y) = P(\min\{X_1, X_2, \ldots, X_n\} \leq y) = 1 - P(\min\{X_1, X_2, \ldots, X_n\} > y) = 1 - (P(X_1 > y, X_2 > y, \ldots X_n > y) = 1 - \prod\limits_{i=1}^{n} P(X_i > y) = 1 - (1 - F_X(y))^n$.

Omdat $f_{Y_1}(y) = \frac{d}{dy} F_{Y_1}(y)$ geldt:
$f_{Y_1}(y) = n(1 - F_X(y))^{n-1} \frac{d}{dy} F_X(y) = n(1 - F_X(y))^{n-1} f_X(y)$.

d 2pt) Zijn Y_1 en Y_n onafhankelijk? Licht je antwoord toe.
ANTWOORD: Y_1 en Y_n zijn afhankelijk. Omdat het minimum van n stochasten nooit groter kan zijn dan het maximum, geldt $f_{Y_1, Y_n}(y_1, y_n) = 0$ als $y_1 > y_n$. Maar $f_{Y_1}(y_1) f_{Y_n}(y_n) > 0$, dus de kansdichtheidsfunctie factorizeren niet en dus zijn de stochasten Y_1 en Y_n afhankelijk.

We gaan nu twee schatters voor θ met elkaar vergelijken:

$$T_1 := \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{en } T_2 = \frac{1}{2}(Y_1 + Y_n).$$

e 2pt) Zijn T_1 en T_2 zuivere schatters van θ?
ANTWOORD: T_i is zuiver als $E(T_i) = \theta$.
$E(T_1) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X_1) = \theta$
$E(T_2) = E(\frac{1}{2}(Y_1 + Y_n)) = \theta$.
Dit is in te zien met een symmetrie-argument of door de verwachtingswaarde van Y_1 en Y_n uit te rekenen.

$E(Y_1) = \int_{-\infty}^{\infty} dy f_{Y_1}(y)y = \int_{\theta-2}^{\theta+2} dy\, n(1 - F_X(y))^{n-1} f_X(y) y =$

$\int_{\theta-2}^{\theta+2} dy\, n(1 - \frac{1}{4}(y - \theta + 2))^{(n-1)} \frac{1}{4} y = \int_{\theta-2}^{\theta+2} dy\, n(\frac{1}{2} - \frac{1}{4}y + \frac{1}{4}\theta))^{(n-1)} \frac{1}{4} y$

Substitueer $z = \frac{1}{2} - \frac{1}{4}y + \frac{1}{4}\theta$, dan geldt:

$E(Y_1) = \int_0^1 dz\, n z^{n-1}(2+\theta-4z) = n(2+\theta)\int_0^1 dz\, n z^{n-1} - 4n\int_0^1 dz\, z^n = (2+\theta) - 4\frac{n}{n+1}$.

$E(Y_n) = \int_{-\infty}^{\infty} dy f_{Y_n}(y)y = \int_{\theta-2}^{\theta+2} dy\, n(F_X(y))^{n-1} f_X(y) y = \int_{\theta-2}^{\theta+2} dy\, n(\frac{1}{4}(y-\theta+2))^{(n-1)} \frac{1}{4} y$

Substitueer $z = \frac{1}{4}(y - \theta + 2)$, dan geldt:

$E(Y_n) = \int_0^1 dz\, n z^{n-1}(z + \frac{1}{4}\theta - \frac{1}{2})4 = 4n\int_0^1 dz\, n z^n + 2n(\theta - 2)\int_0^1 dz\, z^{n-1} = 4\frac{n}{n+1} + (\theta - 2)$.

Dus geldt:
$E(T_2) = E(\frac{1}{2}(Y_1+Y_n)) = \frac{1}{2}(E(Y_1)+E(Y_n)) = \frac{1}{2}((2+\theta) - 4\frac{n}{n+1} + 4\frac{n}{n+1} + (\theta-2)) = \theta$.

f 6pt) Bepaal $Var(T_1)$
ANTWOORD: $Var(T_1) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n}Var(X_1)$ want de stochasten $X_1, \ldots, X_n$ zijn onafhankelijk en identiek verdeeld.
$Var(X_1) = \frac{4}{3}$, zie opgave b, en dus geldt: $Var(T_1) = \frac{4}{3n}$.

Gegeven is dat $Var(T_2) = \frac{8}{2+3n+n^2}$.

g 3pt) Welke schatter van θ is het zinvolste.
ANTWOORD: Omdat T_1 en T_2 allebei zuiver zijn kiezen we de schatter met de kleinste variantie, i.e., T_2.

h 5pt) **Bonusopgave:** Laat zien dat $Var(T_2) = \frac{8}{2+3n+n^2}$. U mag gebruiken dat de gezamenlijke kansdichtheidsfunctie van Y_1 en Y_n gegeven wordt door:

$$f_{Y_1,Y_n}(y_1, y_n) = \begin{cases} \frac{n(n-1)}{4^n}(y_n - y_1)^{n-2} & \text{als } \theta - 2 < y_1 < y_n < \theta + 2 \\ 0 & \text{anders} \end{cases}$$

$Var(T_2) = Var(\frac{1}{2}(Y_1 + Y_n)) = \frac{1}{4}Var(Y_1 + Y_2) = \frac{1}{4}(Var(Y_1) + Var(Y_2) + 2Cov(Y_1, Y_2))$.

$Var(Y_1) = E(Y_1^2) - E(Y_1)^2$
$E(Y_1^2) = \int_{-\infty}^{\infty} dy f_{Y_1}(y)y^2 = \int_{\theta-2}^{\theta+2} dy\, n(1 - F_X(y))^{n-1} f_X(y) y^2 =$

$\theta+2$ $\theta+2$

Substitueer $z = \frac{1}{2} - \frac{1}{4}y + \frac{1}{4}\theta$, dan geldt:
$E(Y_1^2) = 4\int_0^1 dz\, nz^{n-1}(2+\theta-4z)^2 = 4\int_0^1 dz\, nz^{n-1}((2+\theta)^2+16z^2-8z(2+\theta)) =$
$4n(2+\theta)^2\int_0^1 dz\, z^{n-1} - 32n(2+\theta)\int_0^1 dz\, z^n + 64n\int_0^1 dz\, z^{n+1} = 4(2+\theta)^2 - 32\frac{n}{n+1}(2+\theta) + 64\frac{n}{n+2}$.
Dus geldt: $VaR(Y_1) = 4(2+\theta)^2 - 32\frac{n}{n+1}(2+\theta) + 64\frac{n}{n+2} - ((2+\theta) - 4\frac{n}{n+1})^2 =$
$4(2+\theta)^2 - 32\frac{n}{n+1}(2+\theta) + 64\frac{n}{n+2} - (2+\theta)^2 + 8(2+\theta)\frac{n}{n+1} - 16(\frac{n}{n+1})^2 =$
$3(2+\theta)^2 - 24\frac{n}{n+1}(2+\theta) + 64\frac{n}{n+2} - 16(\frac{n}{n+1})^2$.

Vanwege symmetrie geldt: $Var(Y_1) = Var(Y_n)$.

Nu moeten we $Cov(Y_1, Y_n)$ nog bepalen.
$Cov(Y_1, Y_n) = E(Y_1Y_n) - E(Y_1)E(Y_n)$
Om $E(Y_1Y_n)$ te bepalen moeten we de gezamenlijke verdeling van Y_1 en Y_n kennen.

$f_{Y_1,Y_n}(y_1, y_n) = f_{Y_1}(y_1)f_{Y_n}(y_n|Y_1 = y_1)$. Stel $\theta - 2 < y_1 < y_n < \theta + 2$. Dan geldt:

Opgave 3 Een bestaand medicijn werkt bij 75% van de patiënten. Een ziekenhuis wil gaan werken met een nieuw medicijn als ze 95% zeker zijn dat het nieuwe medicijn minstens even goed is als het bestaande medicijn. Om dat na te gaan is het nieuwe medicijn toegediend aan 100 patiënten. Het nieuwe medicijn werkte bij 80 patiënten. U mag er vanuit gaan dat $n = 100$ voldoende groot is om limietresultaten te gebruiken.

a 5pt) Formuleer de nulhypothese en de alternatieve hypothese.
ANTWOORD: Zij p de kans dat het nieuwe medicijn werkt bij een patient.
H_0: $p < 0.75$
H_A: $p \geq 0.75$

b 10pt) Is er voldoende bewijs om het nieuwe medicijn te gaan gebruiken?

ANTWOORD: Het aantal patiënten waarbij het medicijn aanslaat is dan $Bin(100, p)$ verdeeld. De schatter $\hat{p}$ van p is $\hat{p} = \frac{80}{100} = \frac{4}{5}$. De variantie van een binomiaalverdeelde stochast is $np(1-p)$ en de variantie van de fractie is $\frac{p(1-p)}{n}$. Een schatter van de variantie is dus $\frac{\hat{p}(1-\hat{p})}{n}$. Omdat dit een continue functie van $\hat{p}$ is en $\hat{p}$ een zuivere schatter van p is zal $\frac{\hat{p}(1-\hat{p})}{n}$ een asymptotische zuivere schatter zijn van $np(1-p)$.

De schatter van de variantie van de fractie is dus $\frac{0.8(1-0.8)}{100} = \frac{0.16}{100}$ en dus is een schatter van de standaarddeviatie $\frac{0.4}{10} = 0.04$.

Een eenzijdig 95% betrouwbaarheidsinterval voor p is dus: $(\hat{p} - z_{0.05}0.04 - \infty) = (0.8 - 1.6450.04, \infty) = (0.7432 - \infty)$.

0.75 is bevat in het 95% betrouwbaarheidsinterval, dus er is niet voldoende bewijs om het nieuwe medicijn te gaan gebruiken.

Opgave 2 Stel dat het IQ van studenten die het vak WISB161 volgen normaal verdeeld is, maar dat het gemiddelde en de variantie van de normale verdeling niet bekend is. Stel dat je wilt testen, met een significantieniveau gelijk aan $\alpha = 0.05$, of het IQ van deze populatie hoger is dan het landelijke gemiddelde van 100. Daartoe meet je het IQ van 5 willekeurig gekozen studenten, en de gevonden IQ-waarden zijn: $\{121, 97, 122, 119, 101\}$.

a 10pt) Bereken een geschikt betrouwbaarheidsinterval voor het gemiddelde IQ van WISB161-studenten en concludeer op basis van dit betrouwbaarheidsinterval of het gemiddelde IQ van deze populatie hoger is dan 100.

Opgave 3 Zij $-2 \leq \theta \leq 2$ een parameter en zij X_θ een continue stochastische variabele op de reële getallen met kansdichtheidsfunctie

$$f_\theta(x) = \begin{cases} \alpha(\theta) + \theta x & \text{voor } x \in (0,1) \\ 0 & \text{als } x \notin (0,1). \end{cases}$$

Stel dat we een willekeurig steekproef $X_1, X_2, \ldots, X_n$ hebben van grootte n hebben waarbij alle X_i onafhankelijk en identiek verdeeld zijn als X_θ.

a 4pt) Gebruik dat f_θ een kansdichtheid is om te bepalen hoe $\alpha(\theta)$ van θ afhangt.

b 2pt) Bepaal $P(X_\theta = \frac{1}{2})$.

c 3pt) Bepaal $E(X_\theta)$.

d 5pt) Bepaal een zuivere schatter van θ op basis van het steekproefgemiddelde.

e 5pt) Bepaal een formule waaraan de meest-waarschijnlijke schatter (maximum likelihood schatter) van θ moet voldoen als we aan mogen nemen dat de absolute waarde van de meest waarschijnlijke schatter kleiner is dan 2.

f 2pt) Stel $n = 2$, $x_1 = \frac{1}{3}$ en $x_2 = \frac{3}{4}$. Bepaal de meest waarschijnlijke schatting van θ.

Opgave 4 Stel dat 100 onderzoeksteams elk een dataset hebben verzameld en daarmee elk een 90%-betrouwbaarheidsinterval voor een parameter θ geconstrueerd hebben. Veronderstel dat de datasets onafhankelijk zijn.

a 5pt) Bepaal de kansfunctie van het aantal keren dat de daadwerkelijke parameterwaarde θ_0 niet in het 90% betrouwbaarheidsinterval valt.

b 10pt) Stel dat je twijfelt of elke onderzoeksgroep in staat is om een correct 90%-betrouwbaarheidsinterval te construeren. Je kiest als nulhypothese dat elke onderzoeksgroep correct een 90%-betrouwbaarheidsinterval kan construeren en de alternatieve hypothese is dat niet alle groepen dit kunnen. Om deze test uit te voeren heb je zelf θ_0 met grote nauwkeurigheid bepaalt en op basis hiervan geconcludeerd dat θ_0 bevat was in 80 van de 100 90% betrouwbaarheidsintervallen.

Voer een geschikte test uit met een significantieniveau van $\alpha = 0.0456$. Je mag ervan uitgaan dat de steekproefgrootte groot genoeg is om asymptotsiche benaderingen te mogen gebruiken.

Opgave 5 Zij $X_1, X_2, \ldots, X_n$ onafhankelijke, identiek verdeelde Bernoulli stochasten met parameter p. Definieer een run als een maximale hoeveelheid opeenvolgende enen. In het onderstaande realisatie met $n = 12$ zijn er 3 runs.

$$0\ \ 0\ \ 0\ \ \underbrace{1\ \ 1}_{\text{run 1}}\ \ 0\ \ \underbrace{1}_{\text{run 2}}\ \ 0\ \ 0\ \ \underbrace{1\ \ 1\ \ 1}_{\text{run 3}}$$

Zij $I_i = \begin{cases} 1 & \text{als een run begint op positie } i \\ 0 & \text{anders} \end{cases}$ voor $1 \leq i \leq n$ en zij $R = \sum_{i=1}^{n} I_i$ het aantal runs.

a 5pt) Bepaal de marginale kansverdeling van $I_1, I_2, \ldots, I_n$.

b 5pt) Bepaal de verwachtingswaarde van R.

c 5pt) **Bonusopgave** Bepaal de variantie van R.

Opgave 6 Zij $X_1,\ X_2,\ \ldots,\ X_n$ onafhankelijke identiek verdeelde stochasten met kansdichtheidsfunctie

$$f_\theta(x) = \begin{cases} \frac{1}{1+\theta} e^{-\frac{x}{1+\theta}} & \text{als } x > 0 \\ 0 & \text{als } x \leq 0 \end{cases}$$

met $1 \leq \theta \leq 2$. Zij T_1 een schatter van θ gedefinieerd door $T_1(X_1, X_2, \ldots, X_n) = \bar{X}_n - 1$ met $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$.

a 6pt) Bereken de gemiddelde kwadratische fout van T_1.

b 6pt) De schatter T_1 kan schattingen produceren die buiten het interval $[1, 2]$ liggen. We definiëren daarom een tweede schatter T_2 door

$$T_2(X_1, X_2, \ldots, X_n) = \min\{2, \max\{1, T_1(X_1, X_2, \ldots, X_n)\}\}$$

met andere woorden, als de schatting $t_1(x_1, x_2, \ldots, x_n)$ groter is dan 2, dan nemen we 2 als schatting, als de schatting kleiner is dan 1, dan nemen we 1 als schatting. Laat zien of de gemiddelde kwadratische fout van T_2 groter, kleiner, of gelijk is aan de gemiddelde kwadratische fout van T_1. Merk op, ook als je vraag **a** niet hebt kunnen beantwoorden kun je vraag **b** beantwoorden.

c 2pt) Beargumenteer welke schatter de voorkeur heeft?

Opgave 7 Zij X uniform verdeeld op $(0, 1)$ en Y een Pareto verdeling met parameter $\alpha = 1$, i.e., de kansdichtheidsfunctie van Y wordt gegeven door:

$$f_Y(y) = \begin{cases} \frac{1}{y^2} & \text{als } y \geq 1 \\ 0 & \text{anders.} \end{cases}$$

X en Y zijn onafhankelijke stochasten. Definieer de stochast Z door $Z = X^Y$.

a 10pt) Bepaal de cumulatieve verdelingsfunctie $F_Z(z)$.

Einde.