Statistiek I

S. Vansteelandt

Academiejaar 2021-2022

U krijgt 4 uur voor het examen. Het examen is relatief lang omdat, ondanks de open boek vorm, een actieve kennis van de leerstof wordt verwacht. Geen van de vragen vergt echter veel rekenwerk. Nagenoeg alle vragen kunnen onafhankelijk van elkaar worden opgelost.

Gelieve uw naam zowel op deze examenkopij te schrijven, als op uw oplossingenblad. **Het is verplicht om alles in te dienen: ook deze examenkopij, uw kladpapier en alle notities die u tijdens het examen maakt.**

Alvast veel succes!

Gesloten boek: u kunt hier maximaal 1u aan werken.

1. (5p) Stel dat er op de diensten intensieve zorgen in Vlaanderen gemiddeld evenveel mensen liggen die wel versus niet gevaccineerd zijn voor COVID-19. Stel verder dat in de ganse Vlaamse bevolking 90% gevaccineerd is voor COVID-19. Hoeveel meer kans loopt een niet-gevaccineerde dan om op de dienst intensieve zorgen te belanden, in vergelijking met een gevaccineerde?

2. (a) (2.5p) Waarom is het belangrijk om van een schatter haar verdelingsfunctie te bepalen? Wees beknopt: 1 informatieve zin kan volstaan.

   (b) (5p) Geef twee redenen waarom men nagenoeg altijd de asymptotische verdeling van een schatter gebruikt, en niet haar exacte verdeling? Duid elk van beide redenen zo mogelijk **beknopt** met een concreet voorbeeld ter illustratie. Het is niet de bedoeling om het voorbeeld uit te werken, maar eerder om een context te geven waar het - volgens de reden die u opgeeft - te verkiezen is om de asymptotische verdeling i.p.v. exacte verdeling te gebruiken.

3. Beschouw absoluut continue toevalsveranderlijken X, Y en Z wiens gezamenlijke dichtheidsfunctie overal strikt positief is. Stel dat $X \perp\!\!\!\perp Y|Z$ en $X \perp\!\!\!\perp Z|Y$. Toon dan de volgende 3 resultaten aan (dewelke heel kort aangetoond kunnen worden):

   (a) (2.5p) Toon aan dat $f_{X|Y}(x|y) = f_{X|Z}(x|z)$ voor alle (x, y, z).

   (b) (2.5p) Gebruikt het vorige resultaat om aan te tonen dat $X \perp\!\!\!\perp Y$.

   (c) (2.5p) Gebruikt het vorige resultaat om aan te tonen dat $X \perp\!\!\!\perp (Y, Z)$.

Open boek:

1. (richttijd: 1u 45 min) Zij $X_1, ..., X_n$ onderling onafhankelijke en identisch verdeelde toevalsveranderlijken met dichtheidsfunctie

$$f_X(x) = \frac{\theta}{2\sqrt{x}} \exp\left(-\theta\sqrt{x}\right)$$

voor $x \geq 0$ en 0 anders, met $\theta > 0$ een ongekende parameter.

   (a) (5p) Toon aan dat de maximum kans schatter $\hat{\theta}$ voor θ gelijk is aan

$$\frac{n}{\sum_{i=1}^{n} \sqrt{X_i}}.$$

   (b) (5p) Toon aan dat $\sqrt{X_i}$ voor $i = 1, ..., n$ Exponentieel verdeeld is met parameter θ (volgens de definitie gegeven in de syllabus).

   (c) (5p) Toon aan dat $\theta/\hat{\theta}$ Gamma-verdeeld is met parameters $\alpha = n$ en $\beta = 1/n$. Ter info: dit is een continue verdeling wiens verdelingsfunctie strikt positief is over $\mathbb{R}^+$ en asymmetrisch.

   (d) (5p) Gebruik voorgaand resultaat om aan te tonen hoe u een exact 95% betrouwbaarheidsinterval voor θ zou kunnen bekomen, indien u iets (bvb. R) ter beschikking had om de dichtheidsfunctie, cumulatieve distributiefunctie of percentielen van de Gamma-verdeling op te vragen.

   (e) (5p)Toon aan dat $\hat{\theta}$ een consistente schatter is voor θ.

   (f) (5p) Toon aan dat

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N(0, \theta^2).$$

   (g) (5p) Gebruik voorgaand resultaat om aan te tonen hoe u een benaderend 95% betrouwbaarheidsinterval voor θ zou kunnen bekomen.

2. (richttijd: 20 min) Beschouw een gerandomiseerde, placebo-gecontroleerde studie naar de cardiovasculaire veiligheid van nieuwe diabetes therapieën in patiënten met type 2 diabetes. De daling in lichaamsgewicht over het verloop van de studie was gemiddeld 4.35 kg (SE 0.3 kg) groter in de semaglutide groep ($n = 822$) dan in de placebo groep ($n = 825$).

   (a) (4p) Bereken een bijhorend 95% betrouwbaarheidsinterval voor het gemiddeld verschil in de daling in lichaamsgewicht tussen beide groepen.

   (b) (6p) Bereken een bijhorend 95% referentie-interval voor de daling in lichaamsgewicht in de semaglutide groep, wetende dat de gemiddelde daling in lichaamsgewicht in de semaglutide groep 4.0 kg bedraagt. Reken daartoe eerst uit wat de variantie[1] is op de daling in lichaamsgewicht in de semaglutide groep. U mag er daarbij van uitgaan dat de daling in lichaamsgewicht even variabel is in alle behandelingsgroepen. U mag tevens gemakshalve onderstellen dat de bekomen schatting van de variantie met de populatievariantie overeenstemt.

[1] Indien u dit niet vindt, noem de variantie dan σ^2.

3. (richttijd: 55 min) Deze vraag (deel b en c) heeft een hogere moeilijkheidsgraad. Stel dat voor i.i.d. toevalsvectoren $(X_i, Y_i, Z_i), i = 1, ..., n$, waarbij X_i enkel de waarden 0 of 1 kan aannemen,

$$E\left\{Y_i(1 - X_i - \theta X_i)|Z_i\right\} = E\left\{Y_i(1 - X_i - \theta X_i)\right\},$$

voor een ongekende parameterwaarde $\theta > 0$.

(a) (4p) Toon aan dat

$$E\left\{Y_i(Z_i - E(Z))(1 - X_i - \theta X_i)\right\} = 0.$$

(b) (5p) Toon aan dat

$$\frac{\sum_{i=1}^n Y_i(Z_i - \overline{Z}_n)(1 - X_i)}{\sum_{i=1}^n Y_i(Z_i - \overline{Z}_n)X_i}$$

met $\overline{Z}_n$ het steekproefgemiddelde van $Z_1, ..., Z_n$, een consistente schatter is voor θ. U mag daarbij $\overline{Z}_n$ door $E(Z)$ vervangen, voor maximaal 3.5 van de te behalen punten.

(c) (6p) Bepaal de asymptotische verdeling van voorgaande schatter voor θ. U mag daarbij $\overline{Z}_n$ door $E(Z)$ vervangen, voor maximaal 4 van de te behalen punten.