Statistiek I, Bachelor Wiskunde (Prof. C. Ley)
Examen September 2018

Veel succes !

Vraag 1 : Varia problemen

(a) We werpen (eerlijke) munten op tot het moment dat we de vierde keer munt gooien. De stochastische veranderlijke X is het aantal keer dat we kop hebben geworpen. Wat is de verdeling van X ?

(b) Zij X en Y twee onafhankelijke toevalsveranderlijken waar X een Poisson verdeling met parameter $\lambda > 0$ volgt en Y een Bernoulli verdeling met parameter $p > 0$. Bereken de verdeling van $P = XY$ en $S = X + Y$. Zijn P en S onafhankelijk van elkaar ?

(c) De gezamenlijke dichtheidsfunctie van X en Y wordt gegeven door

$$f(x,y) = \begin{cases} \frac{6}{7}(x^2 + \frac{xy}{2}) & \text{als } 0 < x < 1, 0 < y < 2 \\ 0 & \text{anders.} \end{cases}$$

Bepaal de dichtheidsfunctie van X. Zijn de toevalsveranderlijken X en Y onafhankelijk van elkaar ?

(d) Zij X de kansvariabele die voorstelt ofdat een student Wiskunde slaagt of niet voor het examen Statistiek I, X neemt de waarde 1 aan indien de student slaagt en 0 anders. We weten dat 60% van de studenten Wiskunde mannen zijn en dat zij een slaagpercentage p hebben. De vrouwelijke studenten Wiskunde hebben een slaagpercentage q. Bepaal de variantie van X in functie van p en q.

(e) Zij de kansvariabele X uniform verdeeld tussen a en b ($a, b \in \mathbb{R}$). Bepaal de momentgenererende functie $\psi_X(t)$ en **gebruik deze** om de verwachtingswaarde van X te bepalen.

Vraag 2 : De Delta Methode

Zij $X_1, \ldots, X_n$ een reeks i.i.d. stochastische veranderlijken in $\mathbb{R}$ met gelijke verdeling P_θ, $\theta \in \Theta$. Stel dat we over een statistiek $T_n = T_n(X_1, \ldots, X_n)$ bevatten zodat

$$\nu_n(T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(\theta)),$$

als $n \to \infty$ (waar $\nu_n \to \infty$ als $n \to \infty$). De afhankelijkheid van σ^2 op θ is niet wenselijk.

(a) Verklaar waarom deze afhankelijkheid een nadeel is als we betrouwbaarheidsintervallen voor θ wensen op te stellen.

(b) De Delta-methode kan een oplossing leveren ! Stel dat we een diffeomorfisme $g : \Theta \to \Theta$ vinden zodat

$$\nu_n(g(T_n) - g(\theta)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2)$$

als $n \to \infty$, waar c niet van θ afhangt. We zeggen dat zo'n transformatie de "variantie stabiliseert". Stel het $(1-\alpha)$-betrouwbaarheidsinterval voor θ op dat men hierdoor kan krijgen.

(c) Welke eigenschap moet de functie g bezitten om de variantie kunnen te stabiliseren ?

(d) Zij $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(0, \theta = \sigma^2)$. Toon dat

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\sigma^4),$$

als $n \to \infty$, en geef een variantie-stabiliserende transformatie g voor dit probleem. Geef ook het betrouwbaarheidsinterval voor σ^2 op het niveau $1 - \alpha$.

(e) Nu kijken we naar een limitatie van de Delta methode. Zij $X_1, \dots, X_n$ een reeks i.i.d. stochastische veranderlijken zodat $\nu_n(X_n - a) \overset{\mathcal{D}}{\to} Z$ als $\nu_n \to \infty$. Zij $g : \mathbb{R} \to \mathbb{R}$ en afleidbare functie in a. Via de Delta methode krijgen we dan dat $\nu_n(g(X_n) - g(a)) \overset{\mathcal{D}}{\to} g'(a)\, Z$ als $n \to \infty$. In de situatie waar $g'(a) = 0$ blijft dit nog steeds correct, maar is weinig zinvol. Stel nu dat g twee keer afleidbaar is in a, met $g'(a) = 0$ en $g''(a) \neq 0$. Bewijs dat dan

$$\nu_n^2(g(X_n) - g(a)) \overset{\mathcal{D}}{\to} \frac{1}{2} g''(a) Z^2$$

als $n \to \infty$.

(f) Zij $X_1, \dots, X_n$ i.i.d. met $\mathrm{E}[X_1] = 0$ en $\mathrm{Var}[X_1] = \sigma^2$. Geef de asymptotische verdeling van $\sin(\bar{X}^{(n)})$ en $\cos(\bar{X}^{(n)})$, met $\bar{X}^{(n)} := \frac{1}{n} \sum_{i=1}^n X_i$.

Vraag 3 : Maximum likelihood schatter

Een heel bekende verdeling is de Gumbel verdeling met dichtheidsfunctie

$$x \mapsto \exp(-(x - \mu) - \exp(\mu - x)), x \in \mathbb{R},$$

met locatie parameter $\mu \in \mathbb{R}$.

(a) Stel dat we een reeks i.i.d. observaties $X_1, \dots, X_n$ hebben die afkomstig zijn van een Gumbel populatie. Geef de likelihood functie voor μ en de maximum likelihood schatter voor μ.

(b) Toon dat deze MLE een consistente schatter voor de parameter μ is.
(Hint : $\displaystyle\int_{-\infty}^{\infty} exp(-2x - exp(-x))dx = 1$)

(c) Stel dat Z een stochastische veranderlijke is die een exponentiele verdeling $Exp(\lambda)$ volgt met $\lambda = 1$. Toon dat dan $-\log(Z)$ een Gumbel verdeling volgt.

Vraag 4 : Toetsen van hypothesen

Om het effect van de bloeddrukverlagende medicatie Lisinopril en Amlodipine te vergelijken, neemt men 2 groepen proefpersonen (groep A resp. groep B) en meet men van elk individu de bloeddruk voor en na de behandeling met Lisinopril (resp. Amlodipine). In de onderstaande tabellen vind je de (fictieve) resultaten van de studie.

Groep A (Lisinopril) :

voor	na
99	95
110	100
105	107
108	102
104	102

Groep B (Amlodipine) :

voor	na
99	98
92	94
98	96
98	99
94	95
95	97
105	100
103	101

Probeer de volgende vragen te beantwoorden. Maak indien nodig extra veronderstellingen en rapporteer deze.

(a) Bepaal een 95% betrouwbaarheidsinterval voor de gemiddelde afname van de bloeddruk voor de groep die Lisinopril toegediend kreeg.

(b) Kan je op het 5% significantieniveau aantonen dat de gemiddelde afname in bloeddruk voor deze behandelingsgroep verschilt van 0 ?

(c) Test op het 5% significantieniveau of de gemiddelde verlaging van de bloeddruk groter is in behandelingsgroep A dan in behandelingsgroep B.

Vraag 5 : De basketspeler

Een basketspeler heeft een success-rate van 80% voor elk van zijn worpen, en dit onafhankelijk van stress of moeheid. Zijn success-rate blijft dus steeds gelijk. Een journalist wil een artikel over de speler schrijven en daarvoor de kans berekenen, dat van 100 worpen niet meer dan 88 ingaan. Als wiskundige wordt u gevraagd om de journalist bij de voorbereiding van zijn artikel te helpen. Geef dus de precieze formule voor deze kans.

Deze formule is niet gemakkelijk te berekenen zonder Computer-programma. Daarom moet u nu ook een approximatieve waarde leveren met behulp van de Centrale Limietstelling. Geef de nodige voorwaarden voor de Centrale Limietstelling en beargumenteer of deze al dan niet voldaan zijn. Benader vervolgens deze kans via de CLT en rond deze af op 4 beduidende cijfers.