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

Veel succes !

Vraag 1 : Varia problemen

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

(b) We hebben een muntstuk dat kop geeft met kans p. We werpen deze munt N keer en noteren X het aantal keer kop en Y het aantal keer munt. Zijn X en Y onafhankelijk ?

Zijn X en Y onafhankelijk als het aantal worpen, N, zelf een stochastische veranderlijke is en een Poisson-verdeling met parameter $\lambda > 0$ volgt ?

(c) Kan een test met een power van 50% een p-waarde van 0.005 geven ? Motiveer uw antwoord.

(d) De dichtheidsfunctie van X en Y wordt gegeven door :

$$f(x, y) = \exp[-(x + y)], \quad 0 \le x < \infty,\ 0 \le y < \infty.$$

Bereken $\mathrm{P}(X < Y)$ en $\mathrm{P}(X < a)$, $a \in \mathbb{R}$.

(e) Toon dat

$$\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} \{X_i - E(X_i|S_i)\}\{Y_i - \theta_z Z_i\}$$

in kans naar 0 convergeert als je weet dat de toevalsvectoren (S_i, X_i, Z_i, Y_i) onafhankelijk en identisch verdeeld zijn, $E[Y_i|X_i, Z_i, S_i, U_i] = U_i + \theta_z Z_i$, en U_i onafhankelijk is van X_i conditioneel op S_i.

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 $X_1, \ldots, X_n$ een reeks i.i.d. stochastische veranderlijken zodat $\nu_n(X_n - a) \xrightarrow{\mathcal{D}} Z$ 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 $\nu_n(g(X_n) - g(a)) \xrightarrow{\mathcal{D}} g'(a)\, Z$ 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

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

als $n \to \infty$.

(f) Zij $X_1, \ldots, X_n$ 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

De Weibull verdeling wordt veel gebruikt in overlevingsanalyse en meteorologie. Haar dichtheidsfunctie wordt gegeven door

$$x \mapsto \frac{k}{\lambda^k} x^{k-1} \exp\left[- \left(\frac{x}{\lambda} \right)^k \right], x \geq 0,$$

met schaal parameter $\lambda > 0$ en vaste vorm parameter $k > 0$.

(a) Stel dat we een reeks i.i.d. observaties $X_1, \ldots, X_n$ hebben die afkomstig zijn van een Weibull populatie met vorm parameter k. Geef de likelihood functie, de score functie voor λ en de maximum likelihood schatter voor λ.

(b) Toon dat de MLE een consistente schatter voor de parameter λ is. Hiervoor is het nodig om $\mathrm{E}[X_1^k]$ te berekenen.

(c) Stel dat X een stochastische veranderlijke is die een Weibull verdeling volgt met $\lambda = 1$. Welke verdeling volgt dan de getransformeerde veranderlijke X^k ?

(d) Zij U uniform verdeeld over $[0, 1]$. Bewijs dat dan $(-\log(U))^{1/k}$ een Weibull verdeling volgt.

Vraag 4 : Toetsen van hypothesen

Stel dat we beschikken over 2 groepen (A en B) normaal verdeelde gegevens die allen onderling onafhankelijk van elkaar zijn. Gegeven zijn de steekproefgrootten $n_A = 9$ en $n_B = 5$, de steekproefvarianties $S_A^2 = 2$ en $S_B^2 = 14$ en de steekproefgemiddelden $\bar{x}_A = 2$ en $\bar{x}_B = 6$.

(a) Geef een 90% betrouwbaarheidsinterval voor de verhouding σ_B / σ_A.

(b) Kunnen we aantonen dat σ_B^2 meer dan 2 keer zo groot is als σ_A^2 ?

(c) Hoe groot moet het verschil tussen de populatiegemiddelden μ_A en μ_B minstens zijn om met een kans van minstens 40% (ongeveer) te kunnen besluiten dat deze verschillen ?
Voor deze vraag mag je ervan uitgaan dat $\sigma_A^2 = \sigma_B^2$.

Vraag 5 : De random walker

Na een avond met enkele tequila's kan een van de gasten niet meer heel goed lopen. Hij wandelt in een rechte lijn, maar op volgende manier : met kans 1/2 doet hij een stap naar voor, met kans 1/2 een stap naar achter. Elke van zijn stappen heeft een lengte van 0.75m. Bereken via een voldoende approximatie het aantal stappen dat de gast nodig heeft om met een kans van 95% meer dan 3m van zijn startpunt weg te zijn. Verklaar de aannamen die u doet voor deze berekening.

Hint : modeleer het probleem aan de hand van kansvariabelen S_i die stap i voorstellen.