
Tentamen Inleiding Statistiek (TW2080)
6 november 2015, 13:30-16:30u

Formulebladen, rekenmachines, boeken, aantekeningen etc. zijn bij dit tentamen niet toegestaan.

1. Beschouw een aselechte steekproef van omvang n , met teruglegging, uit een populatie $\{x_1, x_2, \dots, x_N\}$ met (populatie-)gemiddelde $\mu = \frac{1}{N} \sum_{i=1}^N x_i$. Uiteraard geldt $1 \leq n \leq N$. Noteer de bijbehorende stochastische variabelen als $X_1, X_2, \dots, X_n$. Zij $c_1, \dots, c_n \in \mathbb{R}$ nader te kiezen constanten.

- (a) [2pt] Bepaal een conditie op de constanten c_i zodat de schatter

$$T_n = \sum_{i=1}^n c_i X_i$$

zuiver (unbiased) voor μ wordt.

- (b) [2pt] Druk de variantie van T_n uit als functie van de populatievariantie σ^2 , de steekproefomvang n en de c_i 'tjes.
2. De discrete stochastische variabele X heeft de volgende kansmassafunctie

$$P(X = x) = f(x|p) = (x-1)p^2(1-p)^{x-2}, x = 2, 3, \dots \quad \text{waar } p \in [0, 1]. \quad (1)$$

Gegeven is dat $E[X] = 2/p$. We beschikken over één waarneming x , een realisatie van de stochastische variabele X .

- (a) [1pt] Geef de momentenschatter voor p , gebaseerd op X .
- (b) [3pt] Bereken de meest aannemelijke (=maximum likelihood) schatter voor p op basis van X .
- (c) [2pt] Geef de Cramer-Rao ondergrens voor de variantie van een willekeurige zuivere schatter voor p .
- (d) [2pt] Stel dat p apriori de dichtheid $f(p) = 2p$ heeft, voor $p \in [0, 1]$. Wat is de aposteriori verdeling van p ? Hint: de Beta verdeling met parameters a en b heeft kansdichtheid

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{[0,1]}(x).$$

3. Beschouw de volgende familie van kansdichtheden:

$$f(x|\theta) = \frac{x^{\theta-1}}{\Gamma(\theta)} e^{-x} 1_{(0,\infty)}(x), \quad \theta > 0.$$

- (a) [2pt] Laat zien dat deze klasse een exponentiële familie van dichtheden vormt.
 (b) [2pt] We beschikken over onafhankelijke stochastische variabelen $X_1, X_2, \dots, X_n$ uit de verdeling met kansdichtheid $f(\cdot|\theta)$. Laat zien dat de stochastische variabele $T = \sum_{i=1}^n \log X_i$ voldoende is voor θ .
 4. De stochastische variabele X is uniform verdeeld op het interval $[0, \theta]$ ($\theta > 0$). De kansdichtheid van X wordt dan gegeven door

$$f(x|\theta) = \frac{1}{\theta} 1_{[0,\theta]}(x) = \begin{cases} 1/\theta & \text{als } x \in [0, \theta] \\ 0 & \text{als } x \notin [0, \theta]. \end{cases}$$

We toetsen $H_0 : \theta = 3$ tegen $H_1 : \theta \neq 3$ door H_0 te verwerpen als $X \leq 0.5$ of $X \geq 2.5$.

- (a) [2pt] Bereken het significantieniveau van deze toets.
 (b) [2pt] Bereken de kans op het maken van een type II fout als de echte waarde van θ gelijk is aan 3.5.
 (c) [3pt] Bepaal het kritieke gebied van de toets (met als toetsingsgrootheid de waarneming X), als we eisen dat de kans op een type I fout precies 0.1 moet zijn.
 5. We beschikken over de realisatie x van een stochastische variabele X met een normale verdeling met onbekende verwachting $\mu \in \mathbb{R}$ en variantie $\sigma^2 = 1$. Dit betekent dat deze stochastische variabele dichtheid

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}, \quad x \in \mathbb{R}.$$

heeft. We willen de nulhypothese $H_0 : \mu = 3$ toetsen tegen de alternatieve hypothese $H_1 : \mu > 3$.

- (a) [2pt] Laat zien dat de gegeneraliseerde likelihood ratio toets neerkomt op het verwerpen van H_0 ten gunste van H_1 voor grote waarden van X .
 (b) [2pt] Ga na of de resulterende toets uniform meest onderscheidend is.
 6. We beschikken over een dataset $\{x_1, \dots, x_{25}\}$. Deze vatten we op als realisatie van een steekproef $X_1, \dots, X_{25}$ uit een onbekende verdelingsfunctie F . Het vermoeden bestaat dat F behoort tot de klasse van exponentiële verdelingsfuncties, dus dat

$$F(x|\theta) = 1 - e^{-x/\theta} \text{ voor } x \geq 0 \text{ en een zekere } \theta > 0. \quad (2)$$

- (a) [1pt] We willen grafisch nagaan of het vermoeden redelijk is. Geef aan hoe u dit kunt doen; geef eventueel ook aan welke functie in R u hiervoor kunt gebruiken.

- (b) [2pt] We willen nu de nulhypothese toetsen dat de onderliggende verdeling de *standaard* exponentiële verdeling is, dus dat F wordt gegeven door (2) met $\theta = 1$. Beschrijf een goodness of fit toets die kan worden gebruikt om de hypothese te toetsen; dus een toetsingsgrootheid en een manier om de gewenste kritieke waarde voor die grootheid te bepalen.

Het cijfer C voor het tentamen wordt als volgt bepaald:

$$C = 1 + 9 \times \frac{\text{aantal punten}}{\text{totaal te behalen punten}}$$

Veel succes!