Tentamen Inleiding Statistiek

december 2012

Bij het tentamen mogen boek en aantekeningen gebruikt worden.

Opgave 1.

Definieer $f(x) = \theta + 2x(1-\theta)$ voor $0 \leq x \leq 1$, $f(x) = 0$ voor $x < 0$ en voor $x > 1$. Hierbij is $0 \leq \theta \leq 2$ een vaste parameter. U kunt natuurlijk eenvoudig controleren dat f een kansdichtheid is (niet-negatief, integraal 1).

Stel dat ik over n waarnemingen $x_1, \ldots, x_n$ beschik, die ik beschouw als resultaten van onafhankelijk trekkingen uit deze verdeling, waarbij $\theta \in (0,2)$ een onbekende parameter is.

(a) Bereken de methode van momenten-schatter van θ, en geef aan hoe u de meest-aannemelijke schatter van θ zou kunnen berekenen.

(a) Voor grote n en gegeven θ, wat zijn, bij benadering, de kansverdelingen van deze twee schatters? Welke is te preferen?

(c) Hoe zou U een benaderend $100 \times (1-\alpha)\%$ betrouwbaarheids interval voor θ bepalen? (bijv.: $\alpha = 0.05$)

Opgave 2.

Ter herinnering: de gamma verdeling met parameters $\lambda > 0$ en $k > 0$ heeft een kansdichtheid evenredig met $\exp(-\lambda x)x^k$, $x > 0$; de Cauchy verdeling met parameters μ en σ heeft kansdichtheid evenredig met $1/(1 + (x-\mu)^2/\sigma^2)$, $-\infty < x < \infty$.

(a) Hoe ziet een plot uit, van de quantielen van een gamma verdeling uitgezet tegenover de bijbehorende quantielen van een standaard Gauss verdeling? (Een zogenaamde probability plot of QQ plot).

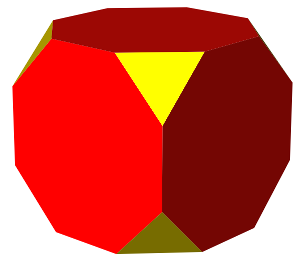
*(b)*En hoe ziet een plot uit, van de quantielen van een Cauchy verdeling tegenover de bijbehorende quantielen van een Gauss verdeling?

In beide gevallen gaat het om kenmerkende aspecten van de verdeling die in de praktijk van nut zou kunnen zijn om uitspraken te doen over de verdeling van een stel experimentele gegevens (ihb, de vraag: wel of niet Gauss verdeeld?), als je een QQ-plot van de data tegenover een Gauss verdeling maakt.

Leg hierbij uit, in beide gevallen, wat het effect van de parameters van de verdeling is op de grafiek.

Opgave 3.

Een vriend van me geeft me een dobbelsteen waarvan de 6 hoeken afgesneden zijn, resulterend in een zestal extra drie-hoekige vlakken. Dus de gemanipuleerde dobbelsteen



heeft 6 achthoekige en 6 driehoekige vlakken (zie plaatje).

Bij het worpen van de dobbelsteen komt het dus te liggen oftewel met een van de 6 achthoekige vlakken boven, of met een van de zes driehoekige vlak boven. De vlakken worden genummerd: de achthoekige kanten van 1 t/m 6, de driehoekige vlakken van 7 t/m 12.

Ik ga de dobbelsteen 100 keer worpen en noteer de 100 uitkomsten: de nummer van de bovenliggende vlak, 100 keer achter elkaar. Ik wil onderzoeken of de dobbelsteen zuiver is, wat in dit geval betekent dat de zes achthoekige kanten allemaal dezelfde kans p hebben, en de zes driehoekige kanten allemaal dezelfde kans q hebben, $6p+6q=1$. Let op: volgens deze definitie is de dobbelsteen zuiver ook in het geval $p \neq q$.

(a) Ontwikkel een statistische toets van de nul-hypothese van zuiverheid (zoals net gedefineerd) versus het alternatief van onzuiver, bij een willekeurige gewenste significantie-niveau α.

(b) Onder de hypothese van zuiverheid, laat zien dat het totaal aantal achthoekige uitkomsten T een voldoende grootheid (*sufficient statistiek*) is voor de parameter p. Wat is de kansverdeling van de data $X_1, \ldots, X_n$ voorwaardelijk op de uitkomst $T = t$ van deze statistiek?

(c) Mijn vriend beweert niet alleen niet alleen zuiver is, maar ook super-zuiver: $q = p$. Onder de hypothese van zuiverheid, ontwikkel een statistische toets van niveau $\alpha = 0.05$ voor de nul-hypothese dat $q = p$ (super-zuiverheid) versus het alternatief dat $q < p$. Wat is het onderscheidingsvermogen van Uw toets tegen het specifieke alternatief $q = 1/24$?

NB: Ik vraag U niet om een volledige numerieke uitwerking, maar om een formule of recept waarmee het verwerpingsgebied van de toets, en zijn bijbehorende onderscheidings-vermogen, bepaald kunnen worden uit α, bijvoorbeeld met eenvoudige functies uit het pakket R.