

Tentamen Inleiding Statistiek (TW2080)
7 november 2014, 14:00-17:00u

Formulebladen, rekenmachines, boeken, aantekeningen etc. zijn bij deze toets niet toegestaan.

1. Beschouw een eindige populatie van getallen $x_1, x_2, \dots, x_N$ met $x_i \in \{0, 1\}$, $1 \leq i \leq N$. Schrijf $p = \frac{1}{N} \sum_{i=1}^N x_i$ voor het populatiegemiddelde (fractie enen). Uit deze populatie trekken we twee keer aselekt, *zonder* teruglegging een getal. De eerste trekking stellen we voor als stochastische variabele X , de tweede als stochastische variabele Y .

- (a) [2 pt]. Laat zien dat

$$E[Y|X] = \frac{Np - X}{N - 1}.$$

- (b) [3 pt]. Laat zien dat

$$E[XY] = p^2 - \frac{p(1-p)}{N-1}.$$

- (c) [2 pt]. Stel nu dat we meerdere trekkingen (zonder teruglegging) doen, zeg $2 \leq n \leq N$. Noteer nu nog steeds de eerste trekking als X en de *laatste* (dus n -de) trekking als Y . Geef een uitdrukking voor de covariantie tussen X en Y .

2. Een verkoper koopt hard disks (HD) van twee bedrijven A en B. We weten dat de HD geproduceerd door bedrijf A goed werken met kans 0.95 en die geproduceerd door bedrijf B met kans 0.85. Maar we weten niet welk percentage van de stukken geproduceerd is door bedrijf A. Laat dit percentage gelijk zijn aan $0 \leq \theta \leq 1$. We controleren n hard disks. Laat $X_1, \dots, X_n$ de stochastische variabelen zijn die aangeven of het i -de stuk goed functioneert ($X_i = 1$ als de i -de HD werkt, anders $X_i = 0$). We modelleren de X_i 'tjes als onderling onafhankelijke gelijkverdeelde stochastische variabelen.

- (a) [2 pt]. Ga na dat X_i een Bernoulli verdeling met parameter p en laat zien dat p als volgt afhangt van θ : $p = 0.85 + 0.1\theta$.
- (b) [2 pt]. Bepaal een voldoende grootte voor θ .
- (c) [3 pt]. Bepaal de maximum likelihood schatter voor θ .
- (c) [2 pt]. Zij $T = \frac{\bar{X} - 0.85}{0.1}$ en schatter voor θ (merk op dat deze niet noodzakelijkerwijs een element van $[0, 1]$ oplevert, maar T kan desalniettemin als schatter worden gebruikt). Laat zien dat deze T zuiver en consistent is voor θ .
- (d) [2 pt]. Is er een zuivere schatter voor θ met een kleinere variantie dan die van T ?

3. Zij $X_1, \dots, X_n$ een steekproef uit de verdeling met kansdichtheid $f(x|\theta) = e^{-x+\theta} 1_{\{x \geq \theta\}}$, met $\theta \geq 0$. Men wenst de nulhypothese $H_0 : \theta = 0$ te toetsen tegen $H_1 : \theta > 0$.

- (a) [2 pt]. Laat zien dat de (gegeneraliseerde) likelihood-ratio toetsingsgrootheid wordt gegeven door

$$T = e^{-nX_{(1)}} 1_{\{X_{(1)} \geq 0\}},$$

waarbij $X_{(1)}$ het minimum van $X_1, \dots, X_n$ is. De maximum likelihood schatter voor θ wordt gegeven door $X_{(1)}$ (dit hoeft u verder niet te bewijzen).

- (b) [1 pt]. Laat zien dat de toets gebaseerd op de likelihood-ratio grootheid neerkomt op het verwerpen van de nulhypothese als $X_{(1)}$ 'te grote' waarden aanneemt.

- (c) [2 pt]. Laat zien dat de kansdichtheid van $nX_{(1)}$ wordt gegeven door

$$g(x|\theta) = e^{n\theta-x} 1_{\{x \geq n\theta\}}.$$

- (d) [2 pt]. Bereken het kritieke gebied voor de toets gebaseerd op $nX_{(1)}$ bij significantieniveau $\alpha = 0.1$.

4. Beschouw stochastische variabelen

$$X_i = 3t_i + Z_i, \quad i \geq 1,$$

waarbij de t_i 'tjes zijn bekende en gegeven constanten zijn. Verder zijn $Z_1, Z_2, \dots$ alle onafhankelijk met $N(0, \theta)$ verdeling (hierbij is de variantie van Z_1 dus gelijk aan θ), dus

$$f_{X_i|\Theta}(x|\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-(x-3t_i)^2/(2\theta)}.$$

We nemen als apriori dichtheid voor Θ

$$f_{\Theta}(\theta) = \frac{1}{\theta^2} e^{-1/\theta} 1_{\{\theta > 0\}}.$$

- (a) [3 pt]. Laat zien dat de posterior mean schatter gebaseerd op $X_1, \dots, X_n$ gelijk is aan

$$\hat{\theta} = \frac{2 + \sum_{i=1}^n (X_i - 3t_i)^2}{n}.$$

Hint: De stochastische variabele Θ heeft een zogenaamde inverse-gamma verdeling met parameters α en β ($\Theta \sim \text{IG}(\alpha, \beta)$) als zijn kansdichtheid wordt gegeven door

$$f_{\Theta}(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{-\alpha-1} e^{-\beta/\theta} 1_{\{\theta > 0\}}.$$

De verwachting van Θ is gelijk aan $\frac{\beta}{\alpha-1}$ als $\alpha > 1$.

- (b) [1 pt]. Laat zien dat $\hat{\theta}_n$ consistent is voor het schatten van θ .

5. Beschouw een experiment waarin we de effectiviteit van 3 verschillende methoden onderzoeken. De scores zijn als volgt:

Method A	73	77	67	71
Method B	91	81	87	85
Method C	72	77	76	79

We gebruiken de F -toets om te toetsen of er een significant verschil is tussen de 3 methoden. We verkrijgen met behulp van een statistisch software pakket de volgende ANOVA-tabel:

```
> scores.aov <- aov(score ~ method, d)
> summary(scores.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
method	2	416.00	208.00	14.4	0.001568
Residuals	9	130.00	14.44		

- (a) [2 pt]. Schrijf het ANOVA-model als een lineair model met matrix-vector notatie.
- (b) [1 pt]. Is er een significant verschil bij significantieniveau $\alpha = 0.01$?
- (c) [1 pt]. De analyse is gebaseerd op de aanname dat alle verkregen scores realisaties zijn van onafhankelijke stochastische variabelen, elk met variantie σ^2 . Geef een schatting voor σ op grond van de ANOVA-tabel. In het antwoord mag u niet nader uitgewerkte wortels en quotiënten laten staan.

Normering:

Het cijfer C voor het tentamen wordt als volgt bepaald:

$$C = 1 + 9 \times \frac{\text{aantal punten}}{33}$$

Veel succes!

