

Tentamen **Mathematische Statistiek (2WS05)**,
dinsdag 18 januari 2011, van 14.00–17.00 uur.

Dit is een tentamen met gesloten boek. De uitwerkingen van de opgaven dienen duidelijk en overzichtelijk te worden opgeschreven. Elk onderdeel levert 10 punten op. Het cijfer is het totaal van de behaalde punten gedeeld door 13, afgerond op een geheel getal.

Op elk ingeleverd vel de naam van de student, de code van het college en de datum van het tentamen noteren.

U mag gebruik maken van een onbeschreven Statistisch Compendium en een (grafische) rekenmachine.

1. Een tentamen mathematische statistiek wordt door tien studenten wiskunde en door tien studenten informatica gemaakt. Zij behalen de volgende resultaten:

studenten wiskunde 4,2 5,3 5,7 9,1 6,7 6,3 5,8 5,7 1,8 7,7;

studenten informatica 4,1 5,2 5,0 8,6 6,5 6,2 5,0 5,0 1,4 7,5.

Het faculteitsbestuur wil nagaan of er verschil in cijfers bestaat tussen beide groepen studenten.

- a** Geef een geschikte nulhypothese, toetsingsgrootheid en kritiek gebied bij onbetrouwbaarheidsdrempel $\alpha = 0,05$. Welke veronderstellingen zijn gemaakt?
- b** Schrijf een R-script om de in [a] gemaakte veronderstellingen te verifiëren en de in [a] ontwikkelde toets uit te voeren.
- c** Bereken de overschrijdingskans van de in [a] ontwikkelde toets. Welke conclusie mag het faculteitsbestuur trekken?
2. Zij X_1, \dots, X_n , $n \in \mathbb{N}$, een aselechte steekproef uit een exponentiële verdeling met verwachting $1/\lambda$ voor $\lambda > 0$.
- a** Bewijs dat er een meest aannemelijke schatter $\hat{\lambda}$ van λ bestaat en bepaal deze schatter.
- b** Bepaal de likelihood ratio toets voor $\lambda = 1$ tegen de alternatieve hypothese $\lambda \neq 1$.
- c** Geef een exact betrouwbaarheidsinterval voor λ met onbetrouwbaarheid $\alpha \in (0, 1)$. Motiveer uw antwoord.

Z.O.Z.

3. De stochastische vector $(X_1, X_2)^T$ is absoluut continu verdeeld met kansdichtheid

$$f(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp \left[\frac{-1}{2\sigma^2} ((x_1 - \mu)^2 + (x_2 - \mu)^2) \right],$$

waar $x_1, x_2 \in \mathbb{R}$, $\mu \in \mathbb{R}$ en $\sigma^2 > 0$.

a Is de familie verdelingen van de stochastische variabele Y gedefinieerd door $Y = X_1 - X_2$ volledig? Motiveer uw antwoord.

b Is de familie verdelingen van de stochastische variabele Y gedefinieerd door $Y = X_1 - X_2$ een exponentiële familie? Zo ja, wat is de afdoende statistische grootte? Zo nee, waarom niet?

4. Zij X negatief binomiaal verdeeld voor zekere $r \in \{2, 3, \dots\}$ en onbekende parameter $p \in (0, 1)$.

a Bepaal een momentenschatter voor p .

b Bepaal een UMVZ-schatter voor p . Hint: beschouw de stochast $1/(X - 1)$.

c Is de momentenschatter zuiver? Motiveer uw antwoord.

d Zij X_1, \dots, X_n een aselechte steekproef uit een negatief binomiale verdeling met onbekende parameters $r \in \mathbb{N}$ en $p \in (0, 1)$. Hoe zou u de parameters r en p schatten?

e Zijn de in [d] voorgestelde schatters asymptotisch raak? Motiveer uw antwoord.

Succes!

BEKNOPTE UITWERKING

1. **a** Zij X_i het cijfer van de i -de student wiskunde, Y_i dat van de i -de student informatica. Neem aan dat $(X_i)_{i=1}^{10}$ een aselechte steekproef uit een normale verdeling is met verwachting μ en variantie σ^2 , $(Y_i)_{i=1}^{10}$ een aselechte steekproef uit een normale verdeling met verwachting ν en variantie σ^2 . Onderstel bovendien dat de steekproeven onafhankelijk zijn. Dan is een geschikte nulhypothese $\mu = \nu$. Neem als toetsingsgrootheid

$$\frac{\sqrt{5 \times 18} (\bar{X}_{10} - \bar{Y}_{10})}{\sqrt{\sum_{i=1}^{10} (X_i - \bar{X}_{10})^2 + \sum_{i=1}^{10} (Y_i - \bar{Y}_{10})^2}}.$$

Onder de nulhypothese is deze Student verdeeld met $10 + 10 - 2 = 18$ vrijheidsgraden, zodat het kritieke gebied $(-\infty, -t_{18;0,975}) \cup (t_{18;0,975}, \infty)$ is.

- b** Neem aan dat de vector `wis` (respectievelijk `cs`) de cijfers van de studenten wiskunde (informatica) bevat. Dan voert

```
t.test(x=wis, y=cs, alternative="two.sided", mu=0,
      paired=FALSE, var.equal=TRUE, conf.level=0.95)
```

de toets uit. De normaliteit kan met `qqnorm(wis)` en `qqnorm(cs)` worden geverifieerd, gelijkheid van varianties door `var(wis)` en `var(cs)`.

- c** Voor de gegeven data is $\bar{X}_{10} = 5,83$, $\bar{Y}_{10} = 5,45$, $9S_X^2 = 34,381$, $9S_Y^2 = 34,685$. De toetsingsgrootheid is $3,604997/\sqrt{(69,066)} \approx 0,43$. De tweezijdige overschrijdingskans is

$$2 \min\{P(T_{18} \geq 0,43), P(T_{18} \leq 0,43)\} \approx 0,67,$$

waar T_{18} een stochast is die Student verdeeld is met 18 vrijheidsgraden. De nulhypothese wordt niet verworpen zodat het faculteitsbestuur niet kan concluderen dat er verschil is tussen de twee studentenpopulaties.

2. **a** $L(X_1, \dots, X_n; \lambda) = n \log \lambda - \lambda \sum_{i=1}^n X_i$. De afgeleide naar λ is $n/\lambda - \sum_{i=1}^n X_i$. Nulstellen geeft $\hat{\lambda} = 1/\bar{X}_n$. De tweede afgeleide $-n\lambda^{-2}$ is strikt negatief, dus $\hat{\lambda}$ is het unieke maximum van de log aannemelijkheidsfunctie $L(\cdot; \lambda)$.

- b** Zij $f(x; \lambda) = \lambda e^{-\lambda x}$. De likelihood ratio toetsingsgrootheid is

$$\lambda(X_1, \dots, X_n) = \frac{\prod_{i=1}^n f(X_i; 1)}{\prod_{i=1}^n f(X_i; \hat{\lambda})} = \exp [n + n \log \bar{X}_n - n \bar{X}_n].$$

De nulhypothese wordt verworpen voor kleine waarden, dus voor grote waarden van $n\bar{X}_n - n \log \bar{X}_n$. Onderzoek de functie $g(x) = x - \log x$. Deze is monotoon dalend op het interval $(0, 1)$, monotoon stijgend op $(1, \infty)$. We verwerpen dus voor $2n\bar{X}_n \in (0, a_\alpha) \cup (b_\alpha, \infty)$. Verder is $2n\bar{X}_n$ onder de nulhypothese χ^2 -verdeeld met $2n$ vrijheidsgraden, zodat men a_α en b_α gelijk kan stellen aan de $\alpha/2$ respectievelijk $1 - \alpha/2$ kwantielen van deze verdeling.

- c** Gebruik de verdelingsfunctie: $-2 \sum_{i=1}^n \log(1 - F(X_i)) = 2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2$. Men vindt betrouwbaarheidsinterval

$$\left(\frac{\chi_{2n;\alpha/2}^2}{2 \sum_{i=1}^n X_i}, \frac{\chi_{2n;1-\alpha/2}^2}{2 \sum_{i=1}^n X_i} \right).$$

3. **a** Merk op dat $\mathbb{E}Y = 0$ hoewel Y niet bijna zeker gelijk is aan nul. Derhalve is de familie kansverdelingen niet volledig.
- b** Merk op dat Y normaal verdeeld is met verwachting 0 en variantie $2\sigma^2$. Dit is een exponentiële familie; Y^2 is afdoende.
4. **a** Merk op dat $\mathbb{E}X = r/p$. De momentenschatter is derhalve r/X .
- b** Bereken eerst de verwachting van $1/(X - 1)$:

$$\begin{aligned} \mathbb{E}\left[\frac{1}{X-1}\right] &= \sum_{x=r}^{\infty} \frac{1}{x-1} \binom{x-1}{r-1} p^r (1-p)^{x-r} = \sum_{x=r}^{\infty} \frac{(x-2)!}{(r-1)!(x-r)!} p^r (1-p)^{x-r} \\ &= \frac{p}{r-1} \sum_{z=r-1}^{\infty} \binom{z-1}{r-2} p^{r-1} (1-p)^{z-(r-1)} = \frac{p}{r-1}. \end{aligned}$$

Merk op dat omdat $r \geq 2$ we niet door nul delen. Dus is $(r-1)/(X-1)$ een zuivere schatter voor p . Voorts vormt de negatief binomiale verdeling met gegeven r een exponentiële familie met afdoende grootheid $X-r$. De verzameling $\{Q(p) = \log(1-p) : p \in (0, 1)\} = \mathbb{R}^-$ bevat een rechthoek (bijvoorbeeld $(-7, -6)$), dus er bestaat ten hoogste één zuivere schatter gebaseerd op de afdoende grootheid. Onze $(r-1)/(X-1)$ is gebaseerd op de afdoende grootheid en zuiver, ergo UMVZ.

- c** De momentenschatter is gebaseerd op de afdoende grootheid. Uit [b] volgt dat de UMVZ-schatter de enige zuivere schatter is met deze eigenschap. De momentenschatter is dus onzuiver.
- d** De waardenverzameling van X is $\{r, r+1, \dots\}$. Schat r dus door $\min_{i=1, \dots, n} X_i$. Gegeven deze \hat{r} , moet men de functie $n\hat{r} \log p + (\sum_{i=1}^n X_i - n\hat{r}) \log(1-p)$ optimaliseren naar p . Differentiëren en de afgeleide gelijkstellen aan 0 geeft $\hat{p} = \hat{r}/\bar{X}_n$. De tweede afgeleide naar p is $\frac{-n\hat{r}}{p^2} + \frac{-\sum_i (X_i - \hat{r})}{(1-p)^2} < 0$ dus is \hat{p} het unieke maximum.
- e** Volgens de zwakke wet der grote aantallen convergeert \bar{X}_n in waarschijnlijkheid (verdeling) naar $EX = r/p$. De verdelingsfunctie van $\min_i X_i$ is $G_n(x) = 1 - (1 - F(x))^n$ waar $F(\cdot)$ de verdelingsfunctie van X is. Merk op dat $F(x) = 0$ voor $x < r$, en $F(x) \in (0, 1)$ voor $x \geq r$. Dus $G_n(x) \rightarrow 1$ voor $x \geq r$ en $G_n(x) \rightarrow 0$ voor $x < r$ als $n \rightarrow \infty$. Met andere woorden, $\min_i X_i$ convergeert in verdeling naar r , en is asymptotisch raak. Tenslotte volgt dat \hat{p} (als quotient) ook asymptotisch raak is.