
Tussentoets Inleiding Statistiek (TW2080)
6 oktober 2017, 10:45-12:15u

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 $\mu = \frac{1}{N} \sum_{i=1}^N x_i$. 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 (dus $n = 1$), een realisatie van de stochastische variabele X .

- (a) [3pt] Bereken de meest aannemelijke (=maximum likelihood) schatter voor p op basis van X .
- (b) [2pt] Stel dat p apriori de dichtheid $f(p) = 2p$ heeft, voor $p \in [0, 1]$. Wat is de a posteriori verdeling van p ? Hint: de Beta verdeling met parameters $a > 0$ en $b > 0$ heeft kansdichtheid

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{[0,1]}(x), \quad \text{waar } \Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx.$$

3. De exponentiële verdeling met parameter $\theta > 0$ heeft kansdichtheid

$$f(x|\theta) = \frac{1}{\theta} \exp(-x/\theta) \mathbf{1}_{[0,\infty)}(x).$$

We beschikken over een steekproef $X_1, \dots, X_n$ uit de genoemde verdeling, met θ onbekend en willen θ schatten. Gegeven is dat $E[X_1] = \theta$ en $\text{Var}(X_1) = \theta^2$.

- (a) [2pt] Laat zien dat de Fisher informatie in één waarneming X_1 wordt gegeven door $I(\theta) = \theta^{-2}$.
- (b) [3pt] Beschouw de schatter $\bar{X}_n$, het steekproefgemiddelde, voor θ . Laat zien dat deze schatter de zuivere schatter met de kleinste mogelijke variantie is.
4. 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 .
- (c) [2pt] We willen op basis van $X_1, \dots, X_n$ de nulhypothese $\theta = 1$ toetsen tegen alternatieve hypothese $\theta = 2$. Laat zien dat de meest onderscheidende toets neerkomt op het verwerpen van de nulhypothese voor grote waarden van $\prod_{i=1}^n X_i$.
5. (a) [1pt] Leg uit wat het onderscheidend vermogen (de power) van een toets is.
- (b) [1pt] Stel $X_1, X_2, \dots$ zijn onafhankelijke en gelijkverdeelde (i.i.d.) stochastische variabelen met kansdichtheid die behoort tot een klasse van kansdichtheden $\{f(\cdot|\theta) : \theta \in \mathbb{R}\}$. Noteer de ‘ware parameter’ als θ_0 . Zij $\hat{\theta}_n$ de maximum likelihood schatter gebaseerd op $X_1, \dots, X_n$. Geef de asymptotische verdeling van $\sqrt{n}(\hat{\theta}_n - \theta_0)$, zoals die algemeen geldt onder bepaalde regulariteitsaannames.

Het cijfer C voor de tussentoets wordt als volgt bepaald:

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

Veel succes!

Antwoorden:

1a Er geldt

$$ET_n = \sum_{i=1}^n E[c_i X_i] = \sum_{i=1}^n c_i E[X_i] = \mu \sum_{i=1}^n c_i.$$

Hier wordt de lineariteit van de verwachting gebruikt en ook het feit dat alle X_i 'tjes uniform verdeeld zijn op $\{x_1, \dots, x_N\}$ zodat $E[X_i] = \mu$ voor alle i . De schatter is dus zuiver als en slechts dan als $\sum_{i=1}^n c_i = 1$.

1b Er geldt, vanwege de onafhankelijkheid van de X_i 's en het feit dat $\text{Var}(cY) = c^2 \text{Var}(Y)$, dat

$$\text{Var}(T_n) = \text{Var}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n \text{Var}(c_i X_i) = \sigma^2 \sum_{i=1}^n c_i^2.$$

2a De likelihood is

$$L(p) = (x-1)p^2(1-p)^{x-2}$$

Dus loglik

$$\ell(p) = \log(x-1) + 2 \log p + (x-2) \log(1-p)$$

Afgeleide nemen

$$\ell'(p) = \frac{2}{p} - \frac{x-2}{1-p}$$

Gelijk stellen aan nul geeft $\hat{p} = \frac{2}{x}$. Tekenoverzicht toont aan dat dit een maximum is.

2b De aposteriori dichtheid is evenredig met

$$L(p) \times 2p \mathbf{1}_{[0,1]}(p) \propto p^3(1-p)^{x-2} \mathbf{1}_{[0,1]}(p)$$

Hierin herkennen we de beta-verdeling met parameters 4 en $x-1$.

3a Er geldt

$$I(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right] = \frac{1}{\theta^4} E_\theta [(X - \theta)^2] = \frac{\text{Var}_\theta(X)}{\theta^4} = \frac{1}{\theta^2}.$$

3b De schatter is zuiver omdat $E_\theta[\bar{X}_n] = E_\theta[X_1] = \theta$ voor alle $\theta > 0$, wegens lineariteit van de verwachting. De variantie is $\text{Var}(\bar{X}_n) = n^{-1} \text{Var}(X_1) = \theta^2/n$, wegens onafhankelijkheid van de X_i 's. De Cramér-Rao stelling zegt dat een zuivere schatter nooit een variantie kan hebben die lager is dan $1/(nI(\theta))$, waar $I(\theta)$ de Fisher informatie in één waarneming is. Omdat $1/(nI(\theta)) = \theta^2/n$ ook de variantie van $\bar{X}_n$ is, is er geen zuivere schatter met een lagere variantie.

4a Er geldt

$$f(x|\theta) = \exp((\theta-1)\log x - \log \Gamma(\theta) - x), \quad x \in (0, \infty).$$

Dit betekent dat $f(\cdot|\theta)$ geschreven kan worden als

$$f(x|\theta) = \exp(T(x)c(\theta) + S(x) + d(\theta)), \quad x \in A$$

met $c(\theta) = \theta - 1$, $T(x) = \log x$, $d(\theta) = -\log \Gamma(\theta)$, $S(x) = -x$ en $A = (0, \infty)$. Uiteraard kan ook $c(\theta) = \theta$, $S(x) = -x - \log x$ worden gekozen.

4b Omdat $T(X) = \log X$ voldoende is voor θ op basis van een enkele waarneming, is $\sum_{i=1}^n \log X_i$ voldoende voor θ op basis van de gegeven steekproef. Alternatief geldt voor de simultane dichtheid van $X_1, \dots, X_n$ dat

$$f(x_1, \dots, x_n | \theta) = \frac{(\prod_{i=1}^n x_i)^{\theta-1}}{\Gamma(\theta)^n} e^{-\sum_{i=1}^n x_i} = g\left(\sum_{i=1}^n \log x_i; \theta\right) h(x_1, \dots, x_n)$$

als $x_i > 0$ voor alle i . Hier is

$$g(t; \theta) = \frac{\exp(t(\theta - 1))}{\Gamma(\theta)^n} \text{ en } h(x_1, \dots, x_n) = \exp\left(-\sum_{i=1}^n x_i\right).$$

Voldoendheid volgt dan uit het factoriseringscriterium van Neyman.

4c Schrijf x_i voor een realisatie van X_i . Wegens het Neyman-Pearson lemma geldt dat de meest onderscheidende toets de nulhypothese verwerpt ten gunste van het alternatief voor kleine waarden van de likelihood ratio

$$LR(x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n | 1)}{f(x_1, \dots, x_n | 2)} = \frac{\exp(-\sum_{i=1}^n x_i)}{\prod_{i=1}^n x_i \exp(-\sum_{i=1}^n x_i) / \Gamma(2)^n} = \left(\prod_{i=1}^n x_i\right)^{-1}$$

Verwerpen voor kleine waarden van deze grootheid komt neer op het verwerpen voor grote waarden van $\prod_{i=1}^n x_i$.

5a Het onderscheidend vermogen van een toets (in een bepaald alternatief) is de kans op het verwerpen van de nulhypothese onder het scenario van dat alternatief.

5b Deze verdeling is normaal met verwachting nul en variantie $1/I(\theta_0)$ waar $I(\theta)$ de Fisher informatie m.b.t. θ is in een waarneming X_1 .