

Tentamen Inleiding Statistiek (TW2080)

16 april 2014, 14:00-17:00u

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

1. Beschouw de volgende populatie van getallen: $\{x_1, \dots, x_N\}$. Het populatiegemiddelde en de populatieverwachting worden gedefinieerd door

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \text{ en } \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2.$$

We doen n trekkingen *met* teruglegging uit de populatie en noteren de bijbehorende stochastische variabelen als $X_1, \dots, X_n$. Definieer

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \text{ met } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

- ⌘ (a) Laat zien dat S_n^2 een *zuivere* schatter (unbiased estimator) voor σ^2 is.
⌘ (b) Laat zien dat $S_n = \sqrt{S_n^2}$ geen zuivere schatter is voor σ , tenzij $\text{Var}(S_n) = 0$.
⌘ (c) Stel dat $x_i \in \{0, 1\}$ voor $i = 1, 2, \dots, N$. Laat zien dat in dat geval

$$\sigma^2 = \mu(1 - \mu).$$

2. We beschikken over een steekproef $X_1, \dots, X_n$ uit de verdeling met kansdichtheid

$$f(x|\theta) = \theta x^{\theta-1} 1_{(0,1)}(x).$$

Hier is $\theta > 0$ een onbekende parameter.

- ⌘ (a) Laat zien dat de familie $\{f(\cdot|\theta) : \theta > 0\}$ een exponentiële familie van dichtheden is.
⌘ (b) Laat zien dat $\sum_{i=1}^n \log X_i$ (of, equivalent, $\prod_{i=1}^n X_i$) een voldoende grootte is voor θ .
⌘ (c) Leid de maximum likelihood schatter voor θ af, gebaseerd op $X_1, \dots, X_n$.
⌘ (d) Leid de momentenschatter af voor θ , weer gebaseerd op $X_1, \dots, X_n$.
⌘ (e) In dit onderdeel is $n = 1$. Iemand komt met een schatter voor θ aan:

$$\hat{\theta} = \frac{X_1}{1 - X_1}.$$

Laat zien dat de verwachting van deze schatter oneindig is.

3. De exponentiële verdeling met parameter θ heeft kansdichtheid

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

Gegeven is dat voor een stochastische variabele X met dichtheid $f(\cdot|\theta)$,

$$EX = \theta \text{ en } \text{Var}(X) = \theta^2.$$

We beschikken over een steekproef $X_1, \dots, X_n$ uit de genoemde verdeling, met θ onbekend en willen θ schatten.

- (a) Laat zien dat de schatter $\bar{X}_n$, het steekproefgemiddelde, zuiver is voor θ en bepaal de variantie van deze schatter.
 - (b) Laat zien dat de Fisher informatie in één waarneming X_1 wordt gegeven door $I(\theta) = \theta^{-2}$.
 - (c) Is er een zuivere schatter voor θ met een kleinere variantie dan $\bar{X}_n$? Motiveer uw antwoord.
4. Zij X een binomiaal verdeelde stochastische variabele met parameters $n = 5$, bekend, en $p \in [0, 1]$, onbekend. De kansfunctie wordt dus gegeven door

$$f(x|p) = P_p(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

- (a) We toetsen de nulhypothese $H_0 : p = 1/2$ tegen het alternatief $H_1 : p = 3/4$. Laat zien dat de likelihood ratio toets voor dit probleem leidt tot de beslisregel om H_0 te verwerpen ten gunste van H_1 bij te grote waarden van X .
 - §(b) We gebruiken nu de volgende beslisregel om nulhypothese $H_0 : p = 1/2$ te toetsen tegen (*samengesteld*) alternatief $H_1 : p > 1/2$: verwerp H_0 ten gunste van H_1 als $X = 5$. Geef het significantieniveau α van deze toets. Herinner u dat $n = 5$.
 - (c) Bereken het onderscheidend vermogen (power functie) van de in het vorige onderdeel beschreven toets als functie van $p \in (1/2, 1]$. Is deze toets uniform meest onderscheidend over alle toetsen met hetzelfde significantieniveau? Motiveer uw antwoord.
 - (d) Gegeven is dat de ML schatter voor p gebaseerd op X gelijk is aan X/n . Laat zien dat de GLR toets voor de nulhypothese $H_0 : p = 1/2$ tegen het (*tweezijdige, samengestelde*) alternatief $H_1 : p \neq 1/2$ leidt tot de beslisregel om H_0 ten gunste van H_1 te verwerpen voor grote waarden van $(X/n) \log(X/n) + (1 - X/n) \log(1 - X/n)$ en dus voor grote waarden van $|X/n - 1/2|$.
5. Er is een experiment uitgevoerd met een bloeddrukverlagend medicijn. Vijftig personen met een te hoge bovendruk kregen het medicijn gedurende vier weken toegediend. In de week voorafgaand aan het onderzoek is de bloeddruk (bovendruk) bij de personen gemeten. De metingen worden aangeduid met $x_1, \dots, x_n$ (met $n = 50$). In de week na afloop van het experiment worden de bloeddrukken ook weer gemeten, bij dezelfde personen. Deze ‘na-metingen’ worden aangeduid met $y_1, \dots, y_n$.
- (a) Zijn de gegevens gepaard of ongepaard? Motiveer uw antwoord.
 - §(b) We willen nagaan of ‘het medicijn werkt’. Om deze vraag te beantwoorden zien we de ‘bloeddruktoename’ voor de n personen, $z_1, \dots, z_n$ met $z_i = y_i - x_i$ als realisatie van een steekproef $Z_1, \dots, Z_n$ uit een normale verdeling met verwachting μ en variantie σ^2 . Als nulhypothese wordt gekozen $H_0 : \mu = 0$. Formuleer de alternatieve hypothese, die correspondeert met de wenselijke werking (bloeddrukverlaging) van het medicijn.
 - §(c) De correcte toets wordt uitgevoerd met behulp van het statistisch programma R. De p -waarde behorend bij de gegeven dataset blijkt gelijk te zijn aan 0.07. Leg uit of op basis van deze analyse de nulhypothese verworpen wordt bij significantieniveau $\alpha = 0.05$ of niet.
 - §(d) Beschrijf een grafische methode en een formele toets die kunnen worden gebruikt om na te gaan of de normaliteitsaannname op $Z_1, \dots, Z_n$ redelijk is. Indicatie: noem de methode en gebruik twee a drie zinnen ter omschrijving.

- (e) In plaats van het toetsen op verlaging van de bloeddruk, kan het ook interessant zijn de bloeddruk na behandeling te modelleren als functie van de bloeddruk voor de behandeling. Gegeven de bloeddruk voor de behandeling (dus de realisaties x_i van X_i), kan het volgende lineaire regressiemodel voor de bloeddruk na behandeling, Y_i , worden gebruikt:

$$Y_i = \beta + x_i + \epsilon_i, \quad \text{waar } \epsilon_1, \dots, \epsilon_n \sim^{iid} N(0, \sigma^2).$$

Druk de kleinste kwadratenschatting voor β uit in de data $x_1, \dots, x_n$ en $y_1, \dots, y_n$.

Normering:

1a: 3	1c: 2	2b: 2	2d: 2	3a: 1	3c: 1	4b: 2	4d: 3	5b: 1	5d: 2
1b: 2	2a: 2	2c: 3	2e: 1	3b: 2	4a: 2	4c: 2	5a: 1	5c: 1	5e: 2

Het cijfer C voor het tentamen wordt als volgt bepaald:

vroegetoetsen: 14P (18P?)

$$C = 1 + 9 \times \frac{\text{aantal punten}}{37}$$

Hw4: 4P

Veel succes!

