Korteweg-de Vries Instituut	Tentamen Stochastiek 2
UvA	22 december 2017

- **U mag geen rekenmachines, telefoons, laptops, of andere hulpmiddelen gebruiken.**
- **Zet uw naam en studentnummer duidelijk bovenaan op *alle* vellen die u inlevert.**

1. Laat $X_1, \ldots, X_n$ een rij onafhankelijke stochastische variabelen zijn met kansdichtheid p_θ gegeven door

$$p_\theta(x) = 2\theta x^{-1-2\theta}, \quad x \geq 1,$$

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

   (a) Bepaal een momentenschatter voor θ onder de extra aanname dat $\theta > 1/2$.

   (b) Bepaal de maximum likelihood schatter voor θ.

   (c) Bereken de waargenomen Fisher informatie.

   (d) Geef een benaderend betrouwbaarheidsinterval voor θ, met onbetrouwbaarheid α.

   (e) Beschouw het toetsingsprobleem $H_0 : \theta = 1/2$, tegen $H_1 : \theta \neq 1/2$. Bepaal de likelihood ratio statistiek voor dit probleem en geef het kritieke gebied van de likelihood ratio toets, met onbetrouwbaarheid bij benadering α. Licht uw antwoord toe.

2. Een opiniepeiler wil onderzoeken welke fractie p van de Nederlanders voor een verbod op knalvuurwerk is. Zij vraagt dat aan 900 mensen. Zij X het aantal mensen dat de vraag bevestigend beantwoordt.

   (a) Als we aannemen dat de 900 mensen willekeurig en onafhankelijk gekozen zijn, wat is dan een voor de hand liggend model voor de verdeling van X?

   (b) De opiniepeiler wil aantonen dat meer dan 50% van de mensen voor een verbod is. Formuleer dit als een toetsingsprobleem.

   (c) Stel, van de 900 mensen zeggen er x voor een verbod te zijn. Geef voor deze situatie een formule voor de (benaderde) p-waarde voor het toetsingsprobleem van onderdeel (b). (U mag hierbij gebruik maken van de gebruikelijke benadering.)

3. Een natuurkundige doet n onafhankelijke metingen van een onbekende constante μ. Hij veronderstelt dat de metingen $X_1, \ldots, X_n$ een rij onafhankelijke, $N(\mu, \sigma^2)$-verdeelde stochastische variabelen vormen, voor één of andere $\sigma^2 > 0$.

   (a) Stel dat de natuurkundige aanneemt dat $\sigma^2 = 1$. Leidt een 95% betrouwbaarheidsinterval voor μ af.

   (b) Stel nu dat de natuurkundige geen aanname wil doen over de waarde van σ^2. Leidt voor deze situatie ook een 95% betrouwbaarheidsinterval voor μ af.

   (c) Wat is het belangrijkste verschil tussen de intervallen uit (a) en (b)? Waardoor wordt dat verklaard?

   (d) De natuurkundige doet ook een Bayesiaanse analyse onder de aanname dat $\sigma^2 = 1$. Als a-priori verdeling op μ gebruikt hij een standaard normale verdeling. Laat zien dat de a-posteriori verdeling voor μ de

   $$N\Big(\frac{\sum X_i}{1+n}, \frac{1}{1+n}\Big)\text{-verdeling}$$

   is.

   (e) De natuurkundige gebruikt de a-posteriori verdeling om een 95% credible interval voor μ te bepalen. Geef een uitdrukking voor zo'n interval.

   (f) Laat zien dat het verschil tussen de intervallen uit onderdelen (a) en (e) verdwijnt als $n \to \infty$. Welke algemene stelling doet hier een uitspraak over?

4. Beschouw observaties $(x_1, Y_1), \ldots, (x_n, Y_n)$, waarin de x-variablen $x_1, \ldots, x_n$ deterministische reële getallen zijn die niet allemaal gelijk zijn en $Y_1, \ldots, Y_n$ stochastische variabelen die voldoen aan de "regressierelatie"

   $$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

   met $\beta_0, \beta_1 \in \mathbb{R}$ en $\varepsilon_1, \ldots, \varepsilon_n$, onafhankelijk en $N(0, \sigma^2)$-verdeeld voor $\sigma^2 > 0$. (De ε_i worden niet geobserveerd.)

   (a) Stel dat alle parameters β_0, β_1 en σ^2 onbekend zijn. Geef een uitdrukking voor de likelihood voor $Y_1, \ldots, Y_n$ en laat zien dat de maximum likelihood schatter (MLS) voor het paar (β_0, β_1) gelijk is aan het punt waar de functie

   $$(\beta_0, \beta_1) \mapsto \sum (Y_i - \beta_0 - \beta_1 x_i)^2$$

   minimaal is.

(b) Stel dat de x-variabelen zo genormaliseerd zijn dat voor het gemiddelde geldt dat $\bar{x} = n^{-1} \sum_{i=1}^n x_i = 0$. Laat zien dat de MLS $\hat{\beta}_1$ voor β_1 in dat geval gegeven wordt door

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

(c) We willen aantonen dat de Y-variabelen op een significante manier van de x-variabelen afhangen. Formuleer dit, in de context van het gepostuleerde model, als een toetsingsprobleem.

(d) Stel nu dat de variantie σ^2 van de "fouten" ε_i bekend is en neem weer aan dat $\bar{x} = 0$. Leidt een toets van niveau $\alpha \in (0, 1)$ af voor de toets van onderdeel (c). (Hint: kijk naar onderdeel (b) om een geschikte toetsingsgrootheid te vinden.)

Maximale punten:

1: 2,3,2,2,3. **2**: 1,1,3. **3**: 3,3,1,3,1,2. **4**: 2,3,1,3. (totaal: 39 punten)

Berekening Eindcijfer:

T = 1 + 9×"fractie behaalde punten".
Als T < 5, dan **Eindcijfer** = T.
Als T ≥ 5, dan **Eindcijfer** = 0.5×T + 0.3×Tussentoets + 0.2×Inleverhuiswerk.