TECHNISCHE UNIVERSITEIT EINDHOVEN
Faculteit Wiskunde en Informatica

Tentamen Mathematische Statistiek (2S990) / Mathematische Statistiek 1 (2S210) op donderdag 29 november 2001, 9.00-12.00 uur.

U mag alleen gebruik maken van een onbeschreven Statistisch Compendium (dikt. nr. 2218) en van een zakrekenmachine. De uitwerkingen van de opgaven dienen gemotiveerd, duidelijk geformuleerd en overzichtelijk opgeschreven te worden.

---

1. Zij $X_1, \ldots, X_n$ een steekproef uit een normale verdeling met verwachting μ en variantie 1.

   (a) Toon aan dat $T = \sum_{i=1}^n X_i$ een voldoende steekproefgrootheid (Eng. sufficient statistic) is voor μ.

   (b) Zij c een willekeurig reëel getal. Bereken de verwachting en variantie van de stochast U gedefinieerd door

   $$U = \begin{cases} 0 & \text{indien } X_1 \geq c \\ 1 & \text{indien } X_1 < c \end{cases}.$$

   (c) Laat zien dat de Cramér-Rao ondergrens voor de variantie van een zuivere schatter van $P(X < c)$, waarbij X normaal verdeeld is met verwachting μ en variantie 1, gelijk is aan $\dfrac{1}{2\pi\, n}\, e^{-(c-\mu)^2}$. U mag hierbij gebruiken dat de Cramér-Rao ondergrens voor $g(\mu)$, waarbij g een differentieerbare functie is, gelijk is $(g'(\mu))^2$ maal de Cramér-Rao ondergrens voor μ.

   (d) Bepaal de Maximum Likelihood schatter van $P(X < c)$.

2. Twee types kopieermachines worden in een test met elkaar vergeleken. Men neemt steekproeven van omvang 5 en meet de tijd die nodig is voor elke machine om 10.000 vellen te kopiëren. Dit levert de volgende metingen op:

	meting 1	meting 2	meting 3	meting 4	meting 5
type I	4,2	3,9	3,5	4,0	3,4
type II	3,8	3,6	3,7	4,1	3,3

   (a) Neem aan dat de kopieertijden normaal verdeeld zijn en toets of er verschil is in gemiddelde kopieertijd. Geef duidelijk aan welke hypothese U toetst. U mag aannemen dat de varianties voor beide type kopieermachines gelijk zijn. Neem $\alpha = 0,05$.

   (b) Bereken een 95%-betrouwbaarheidsinterval (gebaseerd op normale verdelingen) voor het quotiënt van de varianties van beide type kopieermachines. Was het terecht om aan te nemen dat de varianties gelijk zijn?

   (c) Voer dezelfde toets uit als bij a) indien men niet aan wil nemen dat de kopieertijden normaal verdeeld zijn.

3. Zij $X_1, \ldots, X_n$een steekproef uit een exponentiële verdeling, met parameter β, d.w.z. $f(x \mid \beta) = \frac{1}{\beta} e^{-x/\beta}$ voor $x \geq 0$.

   (a) Laat zien dat $\overline{X}$ de Maximum Likelihood schatter is voor β en dat $2n\overline{X}/\beta$ verdeeld is volgens een χ^2_{2n}-verdeling. Hint: de χ^2-verdeling is een speciaal geval van de Gammaverdeling.

   (b) Construeer een $100(1-\alpha)\%$-betrouwbaarheidsinterval voor β. U mag de resultaten van (a) gebruiken.

   (c) Leid een benaderend kritiek gebied af voor een gegeneraliseerde likelihood ratio toets voor de hypothese $H_0 : \beta = \beta_0$ tegen het alternatief $H_1 : \beta \neq \beta_0$.

4. De aantallen krassen op een LCD scherm worden gemeten. Uit eerdere onderzoeken is gebleken dat het aantal krassen per cel van een LCD scherm Poisson verdeeld is. De onderstaande aantallen krassen worden gemeten:

3	5	2	4	1	3	4	10

   (a) Toets of het aantal krassen per cel dezelfde intensiteit heeft voor alle cellen. Neem $\alpha = 0,05$.

De helderheid van een cel in een LCD scherm neemt pas zichtbaar af als het aantal krassen groter is dan 5. Als er 3 of 4 krassen zijn, is de helderheid wel minder, maar nog acceptabel. Daarom komt men tot de volgende classificatie van het aantal krassen per cel:

aantal krassen	classificatie
$0, 1, 2$	goed
$3, 4$	acceptabel
$5, 6, \ldots$	slecht

   (b) Neem aan dat de intensiteit λ van het aantal krassen per cel gelijk is aan 3. Laat zien dat de verwachte kansen voor deze 3 categorieën gelijk zijn 0,4232 voor goed, 0,3921 voor acceptabel en 0,1847 voor slecht.

   (c) De data uit het begin van deze opgave reduceert in de classificatie tot 2 goed ,4 acceptabel en 2 slecht. Toets $H_0 : \lambda = 3$ tegen $H_1 : \lambda \neq 3$ door te toetsen of deze 3 getallen passen bij de verwachte aantallen uit onderdeel b). Neem $\alpha = 0,05$.

5. **versie Mathematische Statistiek 2S990** We beschouwen het glijdend histogram $\widehat{N}_n(x)$, d.w.z. de dichtheidschatter gegeven door $\frac{F_n(x+h)-F_n(x-h))}{2h}$ waarbij $F_n(x)$ de empirische verdelingsfunctie is van een steekproef uit een continue verdeling met dichtheid f.

   (a) Bewijs dat het glijdend histogram een kernschatter is, d.w.z. geef expliciet de kernfunctie aan en herschrijf het glijdend histogram in de standaardvorm van een kernschatter.

   (b) Bereken verwachting en variantie van het glijdend histogram.

5. **versie Mathematische Statistiek 1 2S210** Een te hoog cholesterolgehalte in het bloed is slecht voor de gezondheid. Het ministerie van volksgezondheid maakt zich zorgen over het cholesterolgehalte van de Nederlandse bevolking. Om gericht campagne te kunnen voeren, laat het ministerie een cholesterolonderzoek doen in drie verschillende bevolkingsgroepen. Een steekproef uit dit onderzoek leverde het volgende resultaat

Bevolkingsgroep	Cholesterolgehalte
I	403, 311, 269, 336, 259
II	312, 222, 302, 420, 420, 386, 353, 210, 286, 290
III	403, 244, 353, 235, 319, 260

(a) Toets m.b.v. een ANOVA-tabel of er verschil in cholesterolgehalte is tussen de bevolkingsgroepen. Neem $\alpha = 0,05$.

(b) Gebruik een niet-parametrische toets om te onderzoeken of er verschil in cholesterolgehalte is tussen de bevolkingsgroepen. Neem $\alpha = 0,05$.

Puntenwaardering

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

(Totaal 30 punten.)