

U krijgt de beschikking over een tabellenboekje.

Gebruik van een niet-programmeerbare rekenmachine is toegestaan

1. R.G. te A. besluit deel te nemen aan een quiz. Deze quiz bestaat uit honderd vragen. Om deel te nemen betaalt hij één euro per gestelde vraag. Per goed antwoord krijgt hij zes euro uitbetaald. Laat Y_i de netto uitbetaling bij vraag i zijn (dus de winst na aftrek van één euro). Voor iedere vraag is de kans om het juiste antwoord te geven gelijk aan p . We nemen aan dat de resultaten per vraag onderling onafhankelijk zijn. De stochast $W = \sum_{i=1}^{100} Y_i$ is het aantal gewonnen euro's na het spelen van de quiz (dit aantal is negatief in het geval dat in totaal euro's verloren zijn).

- (a) Formuleer een kansmodel voor het aantal gewonnen/verloren euro's na het spelen van de quiz.
- (b) Bepaal de verwachting van de winst (of het verlies) als functie van p .
- (c) Deze persoon vertelt u het aantal gewonnen/verloren euro's W na afloop van de quiz. Bepaal een zuivere schatter T voor p op basis van W .
- (d) Bepaal de verwachte kwadratische fout van deze schatter T .

2. De stochastische vectoren $(X_1, Y_1), \dots, (X_n, Y_n)$ zijn onderling onafhankelijk met kansdichtheid

$$p_\theta(x, y) = 3\theta^2 y^2 e^{-\theta x - \theta y^3} \quad x > 0 \text{ en } y > 0,$$

waarbij θ een onbekende parameter > 0 is.

- (a) Bepaal de meest aannemelijke schatter voor θ op basis van deze steekproef.
- (b) Laat zien dat de Bayes schatter voor θ relatief ten opzichte van de a-priori dichtheid $\pi(\theta) = e^{-\theta} 1_{(0, \infty)}(\theta)$ gegeven wordt door

$$T = \frac{2n + 1}{\sum_{i=1}^n (X_i + Y_i^3) + 1}.$$

N.B: $\int_0^\infty y^k e^{-cy} dy = \frac{k!}{c^{k+1}}$ voor $c > 0$.

- (c) Bepaal een volledige en voldoende statistiek voor deze familie van kansverdelingen.
3. Als resultaat van een steekproef zijn de getallen 0.04, 0.71, 0.76, 0.94 en 0.96 verkregen.
- (a) Bepaal en teken de empirische verdelingsfunctie op grond van deze waarnemingen.
 - (b) Bepaal de waarde van de Kolmogorov-Smirnov toetsingsgrootte op grond van deze data onder de nulhypothese dat de waarnemingen uit een homogene verdeling op $[0, 1]$ afkomstig zijn. (Als u de vorm van de Kolmogorov-Smirnov toetsingsgrootte niet weet, mag u ook zelf een zinnige toetsingsgrootte bedenken en uitrekenen op basis van bovenstaande data.)

4. Een stalen pijp wordt door een machine van beschermende coating voorzien. Een plek waar de coating te dun is aangebracht wordt een kaaltje genoemd. Indien de machine normaal functioneert is het aantal kaaltjes per 1000 meter pijp een Poisson verdeelde stochastische variabele met verwachting $\lambda = 8$. Op een bepaald moment wordt 1000 meter pijp onderzocht.

U moet in deze opgave verschillende kansen uitrekenen. Voorzover mogelijk dient u daarbij het tabellenboekje te gebruiken. Gebruik de normale benadering voor waarden die niet uit de tabel gehaald kunnen worden.

- (a) De machine wordt opnieuw afgeregeld als het aantal kaaltjes 10 of meer bedraagt. Bepaal de kans dat men bij het onderzoek ten onrechte de machine opnieuw afregelt. Bepaal ook de kans dat men de instelling van de machine ongewijzigd laat, terwijl in werkelijkheid de verwachting van het aantal kaaltjes per 1000 meter gelijk is aan 10.
 - (b) Toets $H_0 : \lambda = 8$ tegen $H_1 : \lambda > 8$ bij onbetrouwbaarheidsdrempel $\alpha = 0.05$. Geef kritieke gebied en formuleer uw conclusie als het aantal kaaltjes in 1000 meter pijp gelijk is aan 12.
 - (c) Bepaal het onderscheidend vermogen voor het toetsingsprobleem uit (b) voor het geval $\lambda = 10$ en $\lambda = 15$.
Schets een grafiek van het onderscheidend vermogen; reken hiertoe eerst het onderscheidend vermogen uit voor $\lambda = 8$, $\lambda = 13.5$ en $\lambda = 17.5$.
5. Een onderzoeker doet onderzoek naar het percentage Nederlanders dat een VU student in het gezin heeft. Hiertoe worden aselect 20 Nederlanders van verschillende gezinnen ondervraagd. Het blijkt dat 1 van deze 20 Nederlanders een gezinslid heeft dat aan de VU studeert. Op grond hiervan wil deze onderzoeker een betrouwbaarheidsinterval opstellen voor p : de proportie gezinnen in Nederland met een gezinslid dat aan de VU studeert.
- (a) De onderzoeker maakt gebruik van de normale benadering van de binomiale verdeling, en komt tot de conclusie dat het tweezijdig 95% betrouwbaarheidsinterval voor p gelijk is aan 0.05 ± 0.0955 , oftewel $[-0.0455, 0.1455]$. Reconstrueer hoe deze onderzoeker tot dit antwoord gekomen is en geef commentaar op deze aanpak.
 - (b) Omdat negatieve waarden voor proporties onzin zijn, besluit de onderzoeker dit interval af te ronden tot $[0, 0.1455]$. Geef aan waarom ook de waarde 0 bij deze uitkomst onmogelijk in een volledig correct betrouwbaarheidsinterval kan zitten.
 - (c) We weten dat er een “exacte” manier is om een betrouwbaarheidsinterval voor p te bepalen. Geef een exact 95% tweezijdig betrouwbaarheidsinterval voor p . (Voorzover waarden niet exact uit het tabellenboekje te halen zijn, of niet eenvoudig exact uit te rekenen zijn, mag u de grenzen van het betrouwbaarheidsinterval zo goed mogelijk afronden.)

Normering:

1(a): 4	2(a): 8	3(a): 5	4(a): 5	5(a): 5
1(b): 3	2(b): 8	3(b): 4	4(b): 8	5(b): 5
1(c): 4	2(c): 6		4(c): 9	5(c): 8
1(d): 8				

Eindcijfer = (totaal + 10)/10