Inleiding Kansrekening en Statistiek (WISB 161)
Tentamen

Sjoerd Dirksen

28 januari 2021, 13:30-16:30

Dit tentamen bestaat uit 4 vragen, die in totaal 45 punten waard zijn. Als het je niet lukt om een onderdeel van een vraag te beantwoorden, dan is het nog steeds mogelijk om de daarop volgende onderdelen te beantwoorden. Schrijf op elk ingeleverd blad jouw naam en studentnummer en nummer de pagina's. Voorzie elk antwoord van een zorgvuldige motivatie. Vereenvoudig jouw antwoorden voor zover mogelijk. Je mag gebruik maken van de tabel aan het einde van het tentamen.

Vraag 1 [9 punten]

Zij (X, Y) een discrete kansvector die waarden aanneemt in $\{1, 2, 3\} \times \{1, 2, 3\}$. De waarden van de kansfunctie van (X, Y) worden gegeven in de onderstaande tabel:

			X	
		1	2	3
	1	$*$	$*$	$*$
Y	2	$*$	0	$*$
	3	0	$*$	0

In het bijzonder is $\mathbb{P}(X = 1, Y = 3) = 0$. Stel nu dat

$$\mathbb{P}(Y = 1 \mid X = i) = \frac{2}{3} \text{ en } \mathbb{P}(X = i \mid Y = 1) = \frac{i}{6} \text{ voor } i = 1, 2, 3.$$

Bepaal de missende waarden, gemarkeerd met een $*$, in de tabel.

Vraag 2 [8 punten]

Zij Y_n binomiaal verdeeld met parameters n en $\frac{3}{4}$. Laat met behulp van de centrale limietstelling zien dat

$$\lim_{n\to\infty} \mathbb{P}(4Y_n \leq 3n) = \frac{1}{2}.$$

Vraag 3 [16 punten]

Docent Sieuwerd geeft online het vak *Inleiding Kansrekening en Statistiek* voor Bedrijfskunde. Hij is geïnteresseerd in de kans p^* dat een doorsnee student een meerkeuze quizvraag correct beantwoordt. In het bijzonder wil hij toetsen of de studenten iets opgestoken hebben en beter presteren dan door simpelweg te raden. Hij wil daarom tijdens de hoorcolleges via *Mentimeter* n quizzes afnemen met steeds vier mogelijke antwoorden. Hij noteert bij elke quiz hoeveel foute antwoorden gegeven worden voordat voor het eerst het goede antwoord gekozen wordt. Laat X_i het aantal foute antwoorden in quiz i zijn voordat het eerste goede antwoord gegeven wordt. Sieuwerd gaat er van uit dat de antwoorden die de studenten geven onafhankelijk van elkaar zijn en dat de uitkomsten van verschillende quizzes ook onafhankelijk van elkaar zijn. Bovendien is hij een beetje ijdel en neemt daarom aan dat er oneindig veel studenten in zijn virtuele collegezaal zitten.

(a) Laat zien dat, voor een gegeven $1 \le i \le n$, $X_i + 1$ geometrisch verdeeld is met parameter p^*.

Opmerking: in de opgaven hebben we gezien dat

$$\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i+1)}$$

een consistente schatter van p^* is. Deze uitspraak mag in het vervolg zonder bewijs gebruikt worden.

(b) Sieuwerd bekijkt het toetscriterium

$$\text{verwerp } H_0 \text{ als } \sum_{i=1}^{n}(X_i+1) \le tn, \tag{1}$$

voor een $t < 4$. Leg uit waarom (1) als toetscriterium gebruikt kan worden voor de hypothesetoets

$$H_0 : p^* = \frac{1}{4} \qquad \text{v.s.} \qquad H_1 : p^* = \frac{3}{4}.$$

(c) Laat zien dat de kans op een type I fout naar 0 convergeert als $n \to \infty$.

(d) Zij $t = 2$ en neem aan dat Sieuwerd 100 quizzes afneemt. Gebruik Chebyshev's ongelijkheid om te laten zien dat de kans op een type II fout niet groter is dan $\frac{1}{100}$.

(e) De onderwijscoördinator verlost Sieuwerd van zijn ijdelheid door hem er op te wijzen dat er maar 200 studenten voor zijn vak aangemeld zijn. Welke kansfunctie zou Sieuwerd voor X_1 kunnen gebruiken? Druk jouw antwoord uit in termen van p^*.

Vraag 4 [12 punten]

Sifan is een fanatieke hardloopster. Zij gaat elke week precies een uur hardlopen en probeert in dit uur zo ver mogelijk te lopen. Haar vriendin Josefine volgt het vak *Inleiding Kansrekening en Statistiek* en probeert te modelleren hoe vaak Sifan haar persoonlijk record in de toekomst zal kunnen verbeteren. Zij modelleert de afstanden die Sifan in n (toekomstige) weken loopt als een continue kansvector $(W_1, \dots, W_n)$ en zij neemt aan dat de afstanden $W_1, \dots, W_n$ onafhankelijk en identiek verdeeld zijn. Josefine negeert de afstanden die Sifan in het verleden heeft gelopen. Merk op dat Sifan een nieuw persoonlijk record vestigt in week k als

$$W_k \ge \max\{W_1, \dots, W_{k-1}\}.$$

(a) Laat eerst zien dat

$$\mathbb{P}(Y \le Z) = \int_{-\infty}^{\infty} F_Y(z) f_Z(z) dz$$

als (Y, Z) een continue kansvector is en Y en Z onafhankelijk zijn.

(b) Laat met behulp van deel (a) zien dat de kans dat Sifan in week n een nieuw persoonlijk record vestigt gelijk is aan $\frac{1}{n}$.

(c) Laat zien dat het verwachte aantal nieuwe persoonlijk records dat tot en met week n gevestigd wordt gelijk is aan $\sum_{k=1}^{n} \frac{1}{k}$.

Opmerking: het is bekend dat $\lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} = \infty$.

Een aantal belangrijke kansverdelingen

Discrete kansverdelingen

Naam	Kansfunctie	Verwachtingswaarde	Variantie
$\mathrm{Ber}(p)$	$\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$	p	$p(1-p)$
$\mathrm{Binom}(n,p)$	$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ \ k=0,\ldots,n$	np	$np(1-p)$
$\mathrm{Geom}(p)$	$\mathbb{P}(X=k)=p(1-p)^{k-1},\ \ k\in\mathbb{N}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{Pois}(\lambda)$	$\mathbb{P}(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ \ k\in\mathbb{N}\cup\{0\}$	λ	λ

Continue verdelingen

Naam	Kansdichtheid	Verwachtingswaarde	Variantie
$\mathrm{Unif}(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a} & \text{als } a<x<b\\ 0 & \text{elders}\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\lambda)$	$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x} & \text{als } x>0\\ 0 & \text{elders}\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathrm{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-(x-\mu)^2/(2\sigma^2))$	μ	σ^2