Tentamen Kansrekening en Statistiek voor Technische Aardwetenschappen, 7 april 2006.

Het tentamen bestaat uit twee delen. Mensen die tussentoets goed hebben gemaakt (hogere cijfer dan 5), moeten alleen Deel II maken, en ze hebben daar anderhalf uur voor. Mensen die tussentoets niet of niet goed hebben gemaakt, of die niet gelukkig zijn met hun toetscijfer, moeten beide delen doen, en ze hebben daar 3 uur voor.
Alle antwoorden moeten zijn voorzien van een (korte) argumentatie, dus laat bij elke vraag zien hoe je aan het antwoord bent gekomen. Een vraag waarbij alleen het antwoord is vermeld, wordt niet goed gerekend. Voor meer duidelijkheid omcirkel de eindelijke antwoorden. Lever Delen I en II apart in.

Gedurende het tentamen mag of een formuleblad of een zelfgemaakt uitreksel van hooguit twee A4 gebruikt worden en een rekenmachine.

Deel II

1. Zij θ een onbekende parameter van een kansverdeling en zij $T = T(X_1, X_2, ..., X_n)$ een schatter voor θ op basis van de steekproef $X_1, ..., X_n$. Geef de definitie van een zuivere schatter en van de *bias*, dus onzuiverheid van een schatter.

2. Beschouw de volgende situatie: onderzoekers verzamelen gegevens van patienten die lijden aan een bepaalde soort kanker. Men is geinteresseerd in de gemiddelde tijd tot het overlijden aan de gevolgen van deze soort kanker. Sommige patienten gaan inderdaad dood als gevolg van deze kanker, en voor deze patienten kennen wij de zogenaamde "time to event" T, dat is de tijd vanaf het moment dat deze kanker bij een patient wordt geconstateerd tot de tijd van de dood. Echter, voor sommige patienten kennen wij deze "time to event" niet, bijv. omdat een patient naar een andere stad verhuist, of niet meer gevolgd wordt. Voor deze patienten kennen wij de zogenaamde "censored time", of de tijd onder observatie T^*. Dwz, de enige wat wij over deze patienten weten is dat de werkelijke "time to event" T (tijd tot de dood) groter is dan T^*.
   a) Onderzoekers besluiten om alleen de werkelijke geregestreerde "times to event" te middelen om aan de schatter te komen voor de gemiddelde overlevingsduur (dus gecensureerde tijden T^* worden helemaal niet meegenomen in de statistische analyse). Is dan de schatter voor de gemiddelde overlevingstijd zuiver of onzuiver? Als die onzuiver is, is dan de bias negatief of positief? Met andere worden, is dan de prognose gebaseerd op deze berekeningen te optimistisch of te pessimistisch en waarom?
   b) Een andere groep onderzoekers gaat anders te werk. Ze gaan *alle* tijden middelen, dus de werkelijke "times to event" T en ook de gecensureerde T^* (dus ze doen alsof T^* tijden tot de dood weergeven). Is dan de schatter zuiver of niet? Zo niet, is de bias positief of negatief, en is die groter of kleiner dan bij de onderzoeksmethode in de onderdeel a)? Is de prognose te pessimistisch of te optimistisch en meer of minder zo dan in a)? Waarom?

3. Een creditcardmaatschappij controleert voor iedere credit card dagelijks transacties om fraude te voorkomen. Aan het eind van de dag wordt op grond van verschillende indicatoren getoetst of er met de credit card gefraudeerd is. In feite, toetst men voor elke credit card de hypothese H_0: "Normaal card gebruik door de eigenaar", tegen H_1: "Credit card fraude door niet-eigenaar", en als H_0 wordt verworpen, dan wordt credit card geblokkeerd. De toets wordt uitgevoerd op significantieniveau $\alpha = 0.99$.
   Welke van de onderstande beweringen zijn juist:
   a) 99% van de gestolen (en gebruikt) credit cards worden geblokkeerd.
   b) 1% van de gestolen (en gebruikt) credit cards worden niet geblokkeerd.
   c) met 99% van de geblokkeerd credit cards is werkelijk gefraudeerd.
   d) 1% van de door eigenaar gebruikt credit cards worden toch geblokkeerd.

4. Het waterniveau bij hoog water wordt dagelijks gemeten (in meters) bij een meetstation op Ijsselmeer, en het is verondersteld dat deze waterhoogtes exponentiel verdeeld zijn met parameter $\lambda = 0.83$.
   a) Wat is de verwachtingswaarde en de variantie van het waterhoogte?

b) Men verzamelt de metingen gedurende het hele jaar: $H_1, H_2, ..., H_{365}$, dus 365 dagen, en berekent het gemiddelde waterhoogte $\bar{H}_{365}$. Wat is de verwachtingswaarde en de variantie van het gemiddelde waterhoogte $\bar{H}_{365}$?

c) Met behulp van de Chebyshev's ongelijkheid, geef een afschatting van de kans dat het jaarlijkse gemiddelde waterniveau meer dan 0.10 m afwijkt van het verwachtingswaarde.

d) Geef een afschatting van dezelfde kans met behulp van de Centrale Limietstelling.

5. In medische literatuur wordt beweert dat het gemiddelde geborte gewicht van babies 3.2 kg is. Wij willen dat checken en daarvoor hebben wij geborte gewichten verzameld van 16 babies geboren tijdens Maart 2006 in The Portland Hospital in Londen (zie www.theportlandhospital.nl/webbabies). Dat zijn de gevonden gewichten:
3.58 3.44 3.3 3.4 3.93 3.3 2.54 2.6 3.76 2.68 3.5 3 3.02 3.06 3.9 2.9
Op grond van deze metingen (en de aanname van normaal verdeelde gewichten) willen wij toetsen

$$H_0 : \mu = 3.2 \quad \text{tegen} \quad H_1 : \mu \neq 3.2$$

bij significantieniveau $\alpha = 0.05$.
a) Wat is de toetsingsgrootheid T en wat is zijn (theoretische) verdeling? Bepaal zijn waarde op grond van de data.
b) Wordt de nulhypothese verworpen of niet? Waarom?
c) Geef het 95% betrouwbaarheidsinterval voor μ op grond van de data.
d) Eigenlijk zijn wij meer geinteresseerd of het werkelijke gebortegewichten *groter* zijn geworden in de laaste tijd ivm overgewicht, dus wij willen toetsen

$$H_0 : \mu = 3.2 \quad \text{tegen} \quad H_1 : \mu > 3.2$$

bij significantieniveau $\alpha = 0.10$. Wat is nu het kritiek gebied van deze toets? Wordt nu H_0 verworpen of niet?

Deel I

1. Van twee dozen met elk 10 aanstekers bevat de één 1 defecte aansteker, de andere 2. Je kiest op aselecte wijze een doos en trekt daaruit een aansteker.
   a) Wat is de kans dat die werkt?
   b) Nu test je de getrokken aansteker en die blijkt te werken. Wat is de kans dat het uit de tweede doos komt?
   c) Vervolgens trek je de volgende aansteker uit dezelfde doos als de eerste aansteker. Wat is de kans dat die ook werkt?

2. De gezamenlijke verdeling van de discrete stochasten X en Y is gegeven in de volgende (onvolledige) tabel:

	X			
Y	1	2	3	$P(Y)$
0		0.1		0.4
1	0.3			
$P(X)$				1

   Verder is gegeven dat $P(X = 1|Y = 0) = 0.4$ en $P(X = 2|Y = 1) = 0.4$.
   a) Maak de tabel compleet.
   b) Bereken de verwachtingswaarden en de varianties van X en van Y.
   c) Bereken de covariantie tussen X en Y. Zijn deze stochasten onafhankelijk? Ongecorreleerd?

3. Een examen bestaat uit 5 meerkeuzevragen. Bij elke vraag zijn 4 alternatieven gegeven, waarvan slechts één juist is. Stel dat een student bij elke vraag geheel willekeurig één van de gegeven 4 alternatieven kiest en laat X het totale aantal correct beantwoorde vragen zijn.
   a) Bepaal de kansverdeling van X, dwz geef het waardebereik van X aan en de bijbehorende kansen. Is dat een bekende kansverdeling?
   b) Om te slagen voor het examen moeten *minstens* 4 vragen goed beantwoord wordent. Wat is de kans om voor het examen te slagen?

4. Aan een bar staan 4 barkrukken op een rij. De eerste persoon die binnenkomt kiest willekeurig één van de vier krukken. De tweede die binnenkomt kiest willekeurig één van de overgebleven drie krukken.
   a) Wat is de kans dat deze personen naast elkaar zitten?
   b) Kan je een algemene formule voor deze kans afleiden als het niet om vier maar n krukken gaat?

5. De continue stochast X heeft de volgende dichtheid:

$$f(x) = \begin{cases} c(x+1), & -1 \le x \le 1, \\ 0, & \text{elders,} \end{cases}$$

   waar c een bepaalde constante is. Bereken:
   a) de waarde van c,
   b) de verdelingsfunctie $F(x)$,
   c) de kans $P(-0.5 < X \le 0.5)$,
   d) de verwachtingswaarde EX en de variantie $\mathrm{Var}(X)$.

6. Zij U_1 en U_2 onafhankelijke continue uniforme stochasten, U_1 op interval $[0, 1]$ en U_2 op interval $[-2, 2]$. Defineer een nieuwe stochast $S = 2 + 3U_1 - 2U_2$. Wat is de verwachtingswaarde, variantie en standaard deviatie van S?

7. Een stochast X_1 heeft normale verdeling met verwachtingswaarde $\mu_1 = 10$ en standaard deviatie $\sigma_1 = 5$, en stochast X_2 heeft normale verdeling met verwachtingswaarde $\mu_2 = -10$ en standaard deviatie $\sigma_2 = 5$. Verder het is bekend dat X_1 en X_2 gecorreleerd zijn met $\mathrm{Cov}(X_1, X_2) = 10$.
   a) Wat is correlatie coefficient $\rho(X_1, X_2)$?
   b) Wat is de verdeling van de stochast $Y = X_1 + X_2$? Bereken $E(Y)$, $\mathrm{Var}(Y)$ en de standaarddeviatie van Y?
   b) Wat is de kans $P(-10 < Y < 10)$? (Op de volgende bladzijde staat de tabel van Normale $N(0, 1)$ verdeling afgedrukt.)

z	0	1	2	3	4	5	6	7	8	9
0.0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0.1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0.2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0.3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0.4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0.5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0.6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0.7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0.8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0.9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1.0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1.1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1.2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1.3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1.4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1.5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1.6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1.7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1.8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1.9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2.0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2.1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2.2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2.3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2.4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2.5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2.6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2.7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2.8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2.9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3.0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3.1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3.2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3.3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3.4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Tabel 1: *Rechteroverschrijdingskansen* $\mathbf{P}(Z \geq z)$ *van de standaard normale verdeling.*

$k \backslash \alpha$	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
∞	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Tabel 2: *Kritieke waarden* $t_{k,\alpha}$ *van de Student verdeling*