

Gebruik van een (niet-grafische) rekenmachine is toegestaan. Na de correctie liggen de tentamens ter inzage bij het onderwijsbureau FEW. Dit tentamen bestaat uit 4 opgaven en 1 bijlage. De normering staat bij de vragen aangegeven. Cijfer = $\frac{\text{totaal}+3}{3}$.

SUCCES!

Opgave 1 [5 punten]

Veronderstel dat $X_1, \dots, X_n$ een steekproef is uit de verdeling met dichtheid

$$p_\theta(x) = 2\theta x e^{-\theta x^2} \mathbf{1}_{[0,\infty)}(x),$$

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

- [4 punten] Bepaal de maximum likelihood-schatter voor θ .
- [1 punt] Bepaal de maximum likelihood-schatter voor $E_\theta X_1 = \frac{1}{2} \sqrt{\frac{\pi}{\theta}}$.

Opgave 2 [9 punten]

Veronderstel dat $X_1, \dots, X_n$ een steekproef is uit de verdeling P_θ met dichtheid

$$p_\theta(x) = \frac{2(\theta - x)}{\theta^2} \mathbf{1}_{[0,\theta]}(x),$$

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

- [3 punten] Bepaal de momentenschatter voor θ op basis van deze steekproef.
- [1 punt] Is deze schatter zuiver?
- [1 punt] Zijn zuivere schatters in het algemeen beter dan onzuivere schatters?

Neem in de rest van deze opgave aan dat $\theta = 1$ bekend is en schrijf $F = P_1$. Veronderstel verder dat X is verdeeld volgens F en dat Y is verdeeld als $a + bX$, ofwel dat Y is verdeeld volgens $F_{a,b}$.

- [3 punten] Geef een uitdrukking voor het α -kwantiel $x_\alpha = F^{-1}(\alpha)$ van de verdeling F . als functie van $\alpha \in (0, 1)$.
- [1 punt] Geef een uitdrukking voor het α -kwantiel $y_\alpha = F_{a,b}^{-1}(\alpha)$ van de verdeling $F_{a,b}$.

Opgave 3 [6 punten]

Stel dat we de omvang van fraude met studiefinanciering op gebied van uit- en thuiswonen willen onderzoeken onder Nederlandse WO-studenten die in of voor 2014 zijn begonnen met studeren. We vinden het te veel werk om het aantal gevallen van deze fraude in het hele land te tellen. Daarom maken we een statistisch model voor deze situatie en tellen we het aantal gevallen van deze fraude

onder studenten ingeschreven bij slechts twee universiteiten, de Universiteit Twente (UT) en de VU. We nemen aan dat iedere student in Nederland dezelfde kans heeft om deze fraude te plegen, dat deze kans klein is en dat de studenten onderling onafhankelijk van elkaar wel of geen fraude plegen. Het aantal gevallen van fraude is dan Poisson verdeeld. We nemen als steekproef de studenten aan de UT en de VU en noteren het aantal gevallen van fraude onder UT-studenten als X en het aantal gevallen van fraude onder VU-studenten als Y . Het aantal studenten van de VU is ongeveer 3 keer zo groot als het aantal studenten van de UT. Het totaal aantal studenten in Nederland is ongeveer 25 keer zo groot als het aantal studenten van de UT. We nemen aan dat X Poisson verdeeld is met onbekende parameter $\theta > 0$. De verdeling van Y wordt dan $\text{Poisson}(3\theta)$. Als schatter voor θ kiezen we

$$T_a = aX + (1 - a)\frac{Y}{3}$$

met $a \in [0, 1]$ met a nog te bepalen.

- a. [3 punten] Bereken de mean square error $MSE(\theta; T_a)$ voor willekeurige $a \in [0, 1]$.
- b. [2 punten] Voor welke a is de schatter T_a optimaal? De optimale schatter noteren we als $\hat{\theta}_{opt}$.
- c. [1 punt] Geef een schatter voor de verdeling voor het aantal gevallen van deze fraude in heel Nederland, gebaseerd op $\hat{\theta}_{opt}$.

Opgave 4 [7 punten]

De Gezondheidsraad adviseert 200 gram groenten en twee stuks fruit per dag te eten. Volgens het Centraal Bureau van de Statistiek houdt 26% zich aan de richtlijn voor fruit (zie www.cbs.nl). We vermoeden dat de hoeveelheid fruit per dag afhankelijk is van leeftijd, en dat mensen tussen 10 en 65 jaar zich gemiddeld minder vaak aan de richtlijn houden. Om dit te onderzoeken vragen we 50 willekeurig gekozen personen tussen 10 en 65 jaar naar hun dagelijkse fruitconsumptie. Het aantal personen in deze steekproef dat zich aan de richtlijn houdt noteren we als X . We nemen aan dat X binomiaal verdeeld is met parameters $n = 50$ en $p \in [0, 1]$ onbekend. We toetsen de nulhypothese $H_0 : p \geq 0.26$ versus $H_1 : p < 0.26$. We meten $x = 7$.

- a. [4 punten] Toets de gegeven nulhypothese bij een onbetrouwbaarheidsdrempel $\alpha_0 = 0.05$ (zie Bijlage 1). Vermeld daarbij
 - i.) de toetsingsgrootte
 - ii.) de verdeling (bij benadering) van de toetsingsgrootte onder de (rand van de) nulhypothese
 - iii.) de overschrijdingskans (p -waarde)
 - iv.) de conclusie van de toets.
- b. [1 punt] Geef de interpretatie van een overschrijdingskans (p -waarde).
- c. [2 punten] Bepaal het onderscheidend vermogen in $p = 0.22$.

(NB Volgens Volksgezondheidenzorg.info van het RIVM houdt slechts 5% van de bevolking zich aan de richtlijn omtrent groente en slechts 5 – 10% van de bevolking zich aan de richtlijn omtrent fruit. Deze cijfers zijn lager dan die van het CBS, mogelijk omdat deze cijfers ouder zijn, en/of op een andere manier verzameld zijn.)

Bijlage 1: Tabel normale verdeling

	0	1	2	3	4	5	6	7	8	9
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1	1	1	1	1	1	1	1	1	1

Tabel 1: Verdelingsfunctie van de standaard normale verdeling op het interval $[0, 4]$. De waarde in de tabel is $\Phi(x)$ voor $x = a + b/100$ met a de waarde in de eerste kolom en b het getal in de eerste rij.