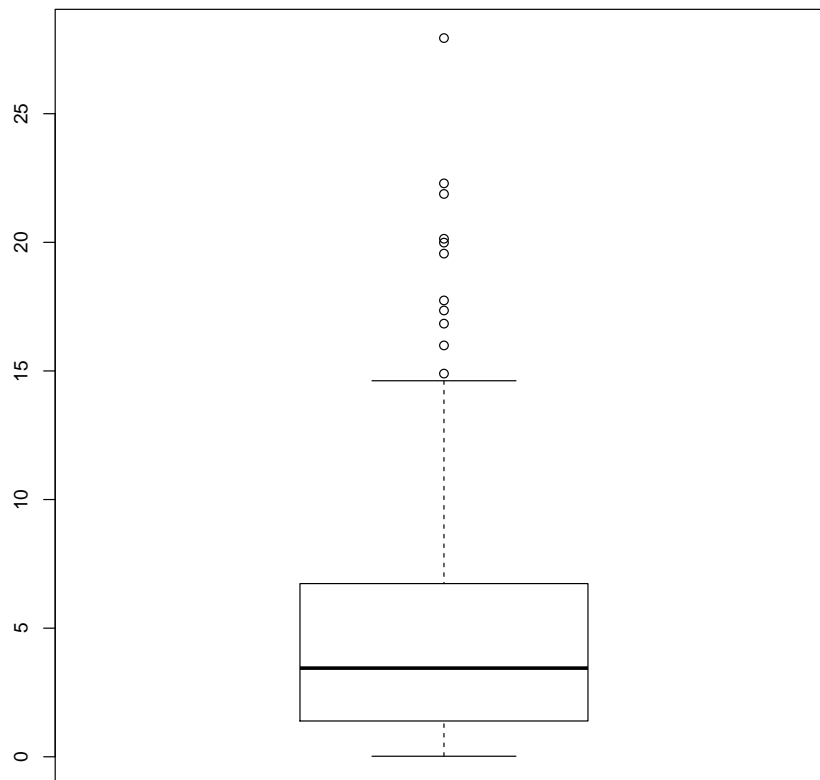


Het is toegestaan een (grafische) rekenmachine te gebruiken en het bijgeleverde formuleblad. Het is *niet* toegestaan een boek, aantekeningen, telefoons of apparaten met internetverbinding te gebruiken.

1. Bijgaand ziet u een boxplot van een steekproef uit een $\text{Exp}(\lambda)$ verdeling. De mediaan van deze steekproef is 3.44.
 - (a) Geef een (redelijke) schatting voor λ op basis van de mediaan.
 - (b) Geef een (redelijke) schatting voor de omvang van de steekproef. Hint: er zijn 11 uitbijters.



2. We beschouwen een steekproef X_1, \dots, X_n uit een verdeling met dichtheid

$$f_\lambda(x) = \begin{cases} \frac{2}{\lambda^2}(\lambda - x) & 0 \leq x \leq \lambda, \\ 0 & \text{elders,} \end{cases}$$

met parameter $\lambda > 0$. Om λ te schatten gebruiken we de schatter

$$T_a = a \cdot \bar{X}_n,$$

waarbij a een nader te bepalen constante is.

- (a) Voor welke keuze van a is T_a een zuivere schatter?
 (b) Voor welke keuze van a is $\text{MSE}(T_a)$ minimaal?
3. We beschouwen een rij onafhankelijke dobbelsteenworpen $X = (X_1, \dots, X_n)$. Noteer $p_k := P(X_1 = k)$. De parameterruimte is dan

$$\Theta = \left\{ (p_1, \dots, p_6) \in [0, 1]^6 : \sum_i p_i = 1 \right\}.$$

We willen toetsen of de dobbelsteen zuiver is en nemen

$$H_0 : \theta \in \Theta_0 \quad \text{met } \Theta_0 = \left\{ \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right) \right\},$$

$$H_1 : \theta \in \Theta_1 \quad \text{met } \Theta_1 = \Theta \setminus \Theta_0.$$

We beschouwen twee toetsen, beide met onbetrouwbaarheidsdrempel $\alpha = 0.05$, gebaseerd op de toetsingsgrootheden

$$T_1 = |\{i : X_i = 1\}|, \quad T_2 = |\{i : X_i \leq 3\}|.$$

Een waarneming X kan samengevat worden door alleen de frequenties $X^{(1)}, \dots, X^{(6)}$ van de uitkomsten $1, \dots, 6$ te geven. Stel dat voor $n = 100$ de samengevatte uitkomst gelijk is aan

$$(X^{(1)}, \dots, X^{(6)}) = (13, 13, 13, 21, 20, 20).$$

Gebruik de normale benadering voor binomiale kansen.

- (a) Bepaal het kritieke gebied voor de eerste toets en trek uw conclusie.
 (b) Bepaal de p -waarde van de uitkomst als de tweede toets gebruikt wordt. Trek ook hier uw conclusie.
 (c) Zij $\pi(\theta, T_i)$ het onderscheidend vermogen van de toets op basis van T_i . Laat zien dat er $\theta \in \Theta_1$ bestaat waarvoor

$$\pi(\theta, T_1) > \pi(\theta, T_2).$$

4. Beschouw een steekproef X_1, \dots, X_n uit een verdeling met dichtheid

$$f_\theta(x) = \sqrt{\frac{\theta}{\pi}} e^{-\theta x^2}, \quad x \in \mathbb{R},$$

waarbij $\theta > 0$ een onbekende parameter is. Deze verdeling voldoet aan $\text{Var}(X_1) = \frac{1}{2\theta}$.

- (a) Bepaal de maximum likelihoodschatter voor $\text{Var}(X_1)$.
- (b) Neem nu als a priori verdeling voor θ een $\text{Exp}(\beta)$, dus $\pi(\theta) = \beta e^{-\beta\theta}$. Bepaal de Bayes schatter voor $\text{Var}(X_1)$.

Hint: U mag gebruiken dat $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

EINDE

Puntenverdeling volgens onderstaande tabel.

Opgave	1	2	3	4	Gratis	Totaal
Punten	20	20	30	20	10	100

Indien u p punten scoort en de cijfers b_1 en b_2 voor de bonusopdrachten behaald hebt, dan is uw eindcijfer (voor afronding)

$$\min\left(\frac{2p + b_1 + b_2}{20}, 10\right).$$

Formuleblad bij Statistiek

DIT BLAD IS NIET EEN SAMENVATTING OF OVERZICHT, EN DIENT SLECHTS ALS HULPMIDDEL.

Kansverdelingen

1. **Alternatieve of Bernoulli verdeling:** $\text{Alt}(p)$ of $\text{Ber}(p)$.

$$P(X = 1) = p \text{ en } P(X = 0) = 1 - p. \quad \text{EX} = p; \quad \text{Var}(X) = p(1 - p).$$

2. **Binomiale verdeling:** $\text{Bin}(n, p)$.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ voor } k = 0, 1, \dots, n. \quad \text{EX} = np; \quad \text{Var}(X) = np(1 - p).$$

3. **Geometrische verdeling:** $\text{Geo}(p)$.

$$P(X = k) = p(1 - p)^{k-1} \text{ voor } k = 1, 2, \dots. \quad \text{EX} = 1/p; \quad \text{Var}(X) = (1 - p)/p^2.$$

4. **Poisson-verdeling:** $\text{Pois}(\mu)$.

$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu} \text{ voor } k = 0, 1, \dots. \quad \text{EX} = \mu; \quad \text{Var}(X) = \mu.$$

5. **Exponentiële verdeling:** $\text{Exp}(\lambda)$.

$$f(x) = \lambda e^{-\lambda x} \text{ en } F(x) = 1 - e^{-\lambda x} \text{ voor } x \geq 0. \quad \text{EX} = 1/\lambda; \quad \text{Var}(X) = 1/\lambda^2.$$

6. **Homogene of Uniforme verdeling** op $[a, b]$: $\text{hom}[a, b]$ of $U(a, b)$.

$$f(x) = \frac{1}{b - a} \text{ en } F(x) = \frac{x - a}{b - a} \text{ voor } a \leq x \leq b. \quad \text{EX} = \frac{1}{2}(a + b); \quad \text{Var}(X) = \frac{1}{12}(b - a)^2.$$

7. **Normale verdeling:** $N(\mu, \sigma^2)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ en } F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt. \quad \text{EX} = \mu; \quad \text{Var}(X) = \sigma^2.$$

8. **Gamma verdeling:** $\Gamma_{\alpha, \lambda}$ ($\alpha, \lambda > 0$).

$$f(x) = \frac{x^{\alpha-1} \lambda^\alpha e^{-\lambda x}}{\Gamma(\alpha)} 1_{[0, \infty)}(x) \text{ met } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \quad \text{EX} = \alpha/\lambda; \quad \text{Var}(X) = \alpha/\lambda^2.$$

9. **Beta verdeling:** $B_{\alpha, \beta}$ ($\alpha, \beta > 0$).

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} 1_{[0, 1]}(x) \text{ met } B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

$$\text{EX} = \frac{\alpha}{\alpha + \beta}; \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Covariantie en correlatie

1. Definitie covariantie: $\text{Cov}(X, Y) = E(X - EX)(Y - EY)$.

$$\text{Eigenschappen: } \text{Cov}(X, Y) = EXY - EX \cdot EY, \quad \text{Cov}(rX + s, tY + u) = rt \text{Cov}(X, Y).$$

2. Definitie correlatiecoëfficiënt: $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$.

3. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.

Empirische verdelingsfunctie

$$\text{Voor een dataset van } n \text{ elementen: } F_n(x) = \frac{\text{aantal elementen in de dataset } \leq x}{n}.$$

Centrale limietstelling

Als X_1, X_2, \dots een rij onafhankelijke stochasten is met alle dezelfde verdeling en met verwachting μ and variantie σ^2 , dan geldt:

Centrale limietstelling:

$$\lim_{n \rightarrow \infty} P \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq a \right) = P(Z \leq a),$$

waarbij Z een $N(0, 1)$ verdeling heeft.

Locatie-schaal families

De locatie-schaal familie van dichtheden behorende bij een dichtheid f , wordt gegeven door

$$\left\{ f_{a,\lambda}(\cdot) = \frac{1}{\lambda} f \left(\frac{\cdot - a}{\lambda} \right) \mid a \in \mathbb{R}, \lambda > 0 \right\}.$$

Als $Z \sim f$ en $X = \lambda Z + a$, dan $X \sim f_{a,\lambda}$.

Kwantielen

Het α -onderkwantiel van een stochast X met verdelingsfunctie F wordt gedefinieerd als

$$F^{-1}(\alpha) = \inf\{x \in \mathbb{R} \mid P(X \leq x) \geq \alpha\}.$$

Als $X = \lambda Z + a$, dan $F_X^{-1}(\alpha) = \lambda F_Z^{-1}(\alpha) + a$. Voor een dataset x_1, \dots, x_n schatten we het α -onderkwantiel \hat{x}_α door $k = \lfloor \alpha(n+1) \rfloor$, $\eta = \alpha(n+1) - k$ en $\hat{x}_\alpha = (1-\eta)x_{(k)} + \eta x_{(k+1)}$.

Schatters

De onzuiverheid (bias) van een schatter T voor θ : $ET - \theta$.

Als T_1 en T_2 zuivere schatters voor θ zijn, dan heet T_2 efficiënter dan T_1 als $\text{Var}(T_2) < \text{Var}(T_1)$.

De 'mean squared error' van een schatter T voor θ : $\text{MSE}(T) = E(T - \theta)^2$.

Eigenschap: $\text{MSE}(T) = \text{Var}(T) + (ET - \theta)^2$.

Gestandaardiseerd en gestudentiseerd steekproefgemiddelde

Als X_i een $N(\mu, \sigma^2)$ verdeling hebben voor $i = 1, \dots, n$, en onafhankelijk zijn, dan hebben

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \quad \text{en} \quad \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

een $N(0, 1)$, respectievelijk $t(n-1)$ -verdeling. Hierbij is $t(n-1)$ de t -verdeling met $n-1$ vrijheidsgraden. Verder geldt

$$(n-1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n-1).$$

Twee steekproeven

Bij de twee steekproeven t -toets:

Gelijke varianties: $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m} \right)$.

Ongelijke varianties: $S_d^2 = \frac{S_X^2}{n} + \frac{S_Y^2}{m}$.

a	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: *Rechteroverschrijdingskans* $1 - \Phi(a) = P(Z \geq a)$ van de $N(0, 1)$ -variabele Z .

$m \setminus p$	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: Right critical values $t_{m,p}$ of the t -distribution with m degrees of freedom corresponding to right tail probability p : $P(T_m \geq t_{m,p}) = p$. The last row in the table are right critical values of the $N(0, 1)$ distribution: $t_{\infty,p} = z_p$

m	$\alpha = 0.995$	0.99	0.975	0.95	0.90	0.1	0.05	0.025	0.01	0.005
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.6
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.3	12.8
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.1	13.3	14.9
5	0.41	0.55	0.83	1.15	1.61	9.24	11.1	12.8	15.1	16.8
6	0.68	0.87	1.24	1.64	2.20	10.6	12.6	14.4	16.8	18.5
7	0.99	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
70	43.2	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	104.2
100	67.3	70.0	74.2	77.9	82.4	118.5	124.3	129.6	135.8	140.2

Tabel 3: Right critical values $\chi_p^2(m)$ of the χ^2 -distribution with m degrees of freedom corresponding to right tail probability p : $P(\chi^2(m) \geq \chi_p^2(m)) = p$.

Volledige oplossingen:

1a De verdelingsfunctie van de exponentiële verdeling wordt gegeven door

$$F(x) = 1 - e^{-\lambda x},$$

en de mediaan m van de verdeling voldoet aan $F(m) = \frac{1}{2}$. Als we dit oplossen dan vinden we

$$m = \frac{\log(2)}{\lambda}.$$

Een schatting voor m is 3.44. Een schatting voor λ is dus

$$\hat{\lambda} = \frac{\log(2)}{3.44} \approx 0.201.$$

1b Methode 1 Een manier om dit te doen is om de locatie w van de bovenste whisker af te lezen uit het plaatje: $w \approx 14.5$. Als $X \sim \text{Exp}(\lambda)$, dan

$$P(X > w) = 1 - F(w) = e^{-\lambda w} \approx e^{-\hat{\lambda} w} \approx e^{-0.201 \cdot 14.5} \approx 0.054.$$

Aangezien er 11 datapunten boven de bovenste whisker liggen, kunnen we de omvang van de steekproef schatten met $\hat{n} = 11/0.054 \approx 203$.

Methode 2 Een andere methode is de theoretische locatie w_t van w te gebruiken, dan is het niet nodig om iets af te lezen uit het plaatje. Daartoe bepalen we eerst het eerste en derde kwartiel van de verdeling:

$$x_{1/4} = F^{-1}(1/4) = \frac{\log(4/3)}{\lambda}, \quad x_{3/4} = F^{-1}(3/4) = \frac{\log(4)}{\lambda}.$$

Nu geldt

$$w_t = x_{3/4} + \frac{3}{2}(x_{3/4} - x_{1/4}) = \frac{4 \log(2) + 3 \log(3)}{2\lambda}.$$

De kans op een outlier is dan (bij benadering)

$$P(X > w_t) = e^{-\lambda w_t} = \frac{\sqrt{3}}{36} \approx 0.048,$$

waarmee de omvang van de steekproef geschat wordt op $\hat{n} = 36 * 11/\sqrt{3} \approx 229$. Het aardige van deze methode is dat je ziet dat de schatting voor n niet afhangt van λ . Dus je kunt de omvang van de steekproef schatten door alleen gebruik te maken van het aantal outliers.

(In werkelijkheid is het plaatje gemaakt met $\lambda = 1/5$ en $n = 200$.)

2a Bepaal eerst de verwachting van X_1 en dan van T_a .

$$\begin{aligned} \mathbb{E}[X_1] &= \int_{-\infty}^{\infty} x f_{\lambda}(x) dx = \frac{2}{\lambda^2} \int_0^{\lambda} \lambda x - x^2 dx = \frac{\lambda}{3}. \\ \mathbb{E}[T_a] &= \mathbb{E} \left[\frac{a}{n} \sum_{i=1}^n X_i \right] = \frac{a}{n} \sum_{i=1}^n \mathbb{E}[X_i] = a \mathbb{E}[X_1] = \frac{a\lambda}{3}. \end{aligned}$$

Voor $a = 3$ geldt $\mathbb{E}[T_a] = \lambda$, en dus is T_a zuiver voor $a = 3$. In het algemeen geldt $\text{Bias}(T_a) = \frac{a\lambda}{3} - \lambda$.

2b We rekenen achtereenvolgens $\mathbb{E}[X_1^2]$, $\text{Var}(X_1)$, $\text{Var}(T_a)$ en $\text{MSE}(T_a)$ uit.

$$\mathbb{E}[X_1^2] = \int_{-\infty}^{\infty} x^2 f_\lambda(x) dx = \frac{2}{\lambda^2} \int_0^\lambda \lambda x^2 - x^3 dx = \frac{\lambda^2}{6}.$$

$$\text{Var}(X_1) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = \frac{\lambda^2}{6} - \frac{\lambda^2}{9} = \frac{\lambda^2}{18}.$$

$$\text{Var}(T_a) = \text{Var}\left[\frac{a}{n} \sum_{i=1}^n X_i\right] = \frac{a^2}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{a^2}{n} \text{Var}[X_1] = \frac{a^2 \lambda^2}{18n}.$$

$$\begin{aligned} \text{MSE}(T_a) &= \text{Var}(T_a) + (\text{Bias}(T_a))^2 = \frac{a^2 \lambda^2}{18n} + \left(\frac{a\lambda}{3} - \lambda\right)^2 \\ &= \frac{(2n+1)a^2 - 12na + 18n}{18n} \lambda^2. \end{aligned}$$

De teller van de breuk is een dalparabool in a , dus heeft een minimum als de afgeleide gelijk is aan 0. Dit geeft

$$2(2n+1)a - 12n = 0, \quad \text{dus} \quad a = \frac{12n}{4n+2} = 3 - \frac{3}{2n+1}.$$

3a Onder H_0 heeft T_1 een $\text{Bin}(100, \frac{1}{6})$ verdeling. Het kritieke gebied is van de vorm

$$K_{T_1} = \{0, \dots, k_1\} \cup \{k_2, \dots, 100\}$$

met $P(T_1 \leq k_1) \leq 0.025$ en $P(T_1 \geq k_2) \leq 0.025$. Onder deze voorwaarden kiezen we K_{T_1} zo groot mogelijk. Onder de nulhypothese geldt

$$\begin{aligned} P(T_1 \leq k) &\approx P\left(N(0, 1) \leq \frac{k + \frac{1}{2} - \frac{n}{6}}{\sqrt{n \frac{1}{6} (1 - \frac{1}{6})}}\right) \\ &= P\left(N(0, 1) \leq \frac{6k + 3 - n}{\sqrt{5n}}\right) \end{aligned}$$

Gelijkstellen aan 0.025 en oplossen voor k geeft

$$k = \frac{1}{6}(\xi_{0.025} \cdot \sqrt{5n} + n - 3) \approx \frac{1}{6}(-1.96 \cdot \sqrt{500} + 97) \approx 8.86.$$

We nemen dus $k_1 = 8$ (want voor 9 is de kans op onterecht verwerpen te groot). Analoog vinden we

$$k_2 = \left\lceil \frac{1}{6}(1.96 \cdot \sqrt{500} + 103) \right\rceil = 25.$$

dus

$$K_{T_1} = \{0, \dots, 8\} \cup \{25, \dots, 100\}.$$

De waarneming X geeft $T_1 = 13$ en dat ligt niet in het kritieke gebied. Dus H_0 wordt niet verworpen.

3b Onder H_0 heeft T_2 een $\text{Bin}(100, \frac{1}{2})$ verdeling. De waarneming is $T_2 = 39$. De p -waarde is dan (kansen onder H_0)

$$\begin{aligned} p\text{-waarde} &= 2 \min \{P(T_2 \leq 39), P(T_2 \geq 39)\} \\ &= 2P(T_2 \leq 39) \\ &\approx 2P\left(N(0, 1) \leq \frac{39 + \frac{1}{2} - \frac{n}{2}}{\sqrt{n\frac{1}{2}(1 - \frac{1}{2})}}\right) \\ &= 2P(N(0, 1) \leq -2.1) \approx 0.036 < 0.05. \end{aligned}$$

Nu wordt de nulhypothese wel verworpen.

3c Neem bijvoorbeeld $\theta = (0, 0, \frac{1}{2}, \frac{1}{2}, 0, 0)$. Dan geldt

$$\pi(\theta, T_1) = P_\theta(X \in K_{T_1}) = 1,$$

omdat T_1 met kans 1 gelijk is aan 0.

Verder geldt voor deze θ dat de verdeling van T_2 gelijk is aan de verdeling van T_2 onder de nulhypothese. Dus moet gelden

$$\pi(\theta, T_2) = P_\theta(T_2 \in K_{T_2}) = P_{\theta_0}(T_2 \in K_{T_2}) \leq 0.05.$$

4a De likelihoodfunctie is

$$L(\theta, X) = \pi^{-n/2} \theta^{n/2} e^{-\theta \sum X_i^2}.$$

De loglikelihood is dus

$$l(\theta, X) = -\frac{n}{2} \log(\pi) + \frac{n}{2} \log(\theta) - \theta \sum X_i^2.$$

Merk op dat $\lim_{\theta \rightarrow 0} l(\theta, X) = -\infty$ and $\lim_{\theta \rightarrow \infty} l(\theta, X) = -\infty$, dus de functie heeft een maximum. De afgeleide is

$$\frac{d}{d\theta} l(\theta, X) = \frac{n}{2\theta} - \sum X_i^2.$$

Gelijkstellen aan 0 geeft de maximum likelihoodschatter:

$$\hat{\theta} = \frac{n}{2 \sum X_i^2}.$$

De maximum likelihoodschatter voor $\text{Var}(X)$ is dus

$$\frac{1}{2\hat{\theta}} = \frac{1}{n} \sum X_i^2.$$

4b De dichtheid van de a posteriori verdeling voldoet aan

$$\begin{aligned} p_{\Theta|X}(\theta) &\propto \pi(\theta)L(\theta, X) \\ &\propto \theta^{n/2} e^{-\theta(\beta + \sum X_i^2)}. \end{aligned}$$

Dit is de dichtheid van een $\Gamma_{\alpha, \lambda}$ -verdeling met

$$\alpha = \frac{n}{2} + 1, \quad \lambda = \beta + \sum X_i^2.$$

De Bayes schatter voor $\text{Var}(X) = \frac{1}{2\theta}$ is nu

$$\begin{aligned}\int_0^\infty \frac{1}{2\theta} p_{\Theta|X}(\theta) d\theta &= \int_0^\infty \frac{1}{2\theta} \frac{\theta^{\alpha-1} \lambda^\alpha e^{-\lambda\theta}}{\Gamma(\alpha)} d\theta \\ &= \frac{\lambda \Gamma(\alpha-1)}{2\Gamma(\alpha)} \int_0^\infty \frac{\theta^{\alpha-2} \lambda^{\alpha-1} e^{-\lambda\theta}}{\Gamma(\alpha-1)} d\theta \\ &= \frac{\lambda}{2(\alpha-1)} = \frac{\beta + \sum X_i^2}{n}.\end{aligned}$$

Hierbij is gebruikt dat de integraal over een dichtheid van de $\Gamma_{\alpha-1, \lambda}$ -verdeling gaat en is dus gelijk aan 1.