A

Summary of distributions

Discrete distributions

1. **Bernoulli distribution**: $Ber(p)$, where $0 \leq p \leq 1$.
   $\mathrm{P}(X = 1) = p \quad \text{and} \quad \mathrm{P}(X = 0) = 1 - p.$
   $\mathrm{E}[X] = p \quad \text{and} \quad \mathrm{Var}(X) = p(1 - p).$
2. **Binomial distribution**: $Bin(n, p)$, where $0 \leq p \leq 1$.
   $$\mathrm{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$
   $\mathrm{E}[X] = np \quad \text{and} \quad \mathrm{Var}(X) = np(1 - p).$
3. **Geometric distribution**: $Geo(p)$, where $0 < p \leq 1$.
   $\mathrm{P}(X = k) = p(1 - p)^{k-1} \quad \text{for } k = 1, 2, \ldots .$
   $\mathrm{E}[X] = 1/p \quad \text{and} \quad \mathrm{Var}(X) = (1 - p)/p^2.$
4. **Poisson distribution**: $Pois(\mu)$, where $\mu > 0$.
   $$\mathrm{P}(X = k) = \frac{\mu^k}{k!} \mathrm{e}^{-\mu} \quad \text{for } k = 0, 1, \ldots .$$
   $\mathrm{E}[X] = \mu \quad \text{and} \quad \mathrm{Var}(X) = \mu.$

Continuous distributions

1. **Cauchy distribution**: $Cau(\alpha, \beta)$, where $-\infty < \alpha < \infty$ and $\beta > 0$.
   $$f(x) = \frac{\beta}{\pi\left(\beta^2 + (x - \alpha)^2\right)} \quad \text{for } -\infty < x < \infty.$$
   $$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - \alpha}{\beta}\right) \quad \text{for } -\infty < x < \infty.$$
   $\mathrm{E}[X]$ and $\mathrm{Var}(X)$ do not exist.

2. **Exponential distribution**: $Exp(\lambda)$, where $\lambda > 0$.

   $f(x) = \lambda e^{-\lambda x} \quad$ for $x \geq 0$.

   $F(x) = 1 - \mathrm{e}^{-\lambda x} \quad$ for $x \geq 0$.

   $\mathrm{E}[X] = 1/\lambda \quad$ and $\quad \mathrm{Var}(X) = 1/\lambda^2$.

3. **Gamma distribution**: $Gam(\alpha, \lambda)$, where $\alpha > 0$ and $\lambda > 0$.

   $$f(x) = \frac{\lambda\,(\lambda x)^{\alpha-1}\,\mathrm{e}^{-\lambda x}}{\Gamma(\alpha)} \quad \text{for } x \geq 0.$$

   $$F(x) = \int_0^x \frac{\lambda\,(\lambda t)^{\alpha-1}\,\mathrm{e}^{-\lambda t}}{\Gamma(\alpha)}\,\mathrm{d}t \quad \text{for } x \geq 0.$$

   $\mathrm{E}[X] = \alpha/\lambda \quad$ and $\quad \mathrm{Var}(X) = \alpha/\lambda^2$.

4. **Normal distribution**: $N(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma > 0$.

   $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty.$$

   $$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\,\mathrm{d}t \quad \text{for } -\infty < x < \infty.$$

   $\mathrm{E}[X] = \mu \quad$ and $\quad \mathrm{Var}(X) = \sigma^2$.

5. **Pareto distribution**: $Par(\alpha)$, where $\alpha > 0$.

   $$f(x) = \frac{\alpha}{x^{\alpha+1}} \quad \text{for } x \geq 1.$$

   $F(x) = 1 - x^{-\alpha} \quad$ for $x \geq 1$.

   $\mathrm{E}[X] = \alpha/(\alpha - 1)$ for $\alpha > 1$ and ∞ for $0 < \alpha \leq 1$.

   $\mathrm{Var}(X) = \alpha/((\alpha-1)^2(\alpha-2))$ for $\alpha > 2$ and ∞ for $0 < \alpha \leq 1$.

6. **Uniform distribution**: $U(a, b)$, where $a < b$.

   $$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b.$$

   $$F(x) = \frac{x-a}{b-a} \quad \text{for } a \leq x \leq b.$$

   $\mathrm{E}[X] = (a+b)/2 \quad$ and $\quad \mathrm{Var}(X) = (b-a)^2/12$.

B

Tables of the normal and t-distributions

Table B.1. Right tail probabilities $1 - \Phi(a) = \mathrm{P}(Z \geq a)$ for an $N(0,1)$ distributed random variable Z.

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

Table B.2. Right critical values $t_{m,p}$ of the t-distribution with m degrees of freedom corresponding to right tail probability p: $\mathrm{P}(T_m \geq t_{m,p}) = p$. The last row in the table contains right critical values of the $N(0,1)$ distribution: $t_{\infty,p} = z_p$.

	Right tail probability p							
m	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

C

Answers to selected exercises

2.1 $P(A \cup B) = 13/18$.

2.4 Yes.

2.7 0.7.

2.8 $P(D_1 \cup D_2) \le 2 \cdot 10^{-6}$ and $P(D_1 \cap D_2) \le 10^{-6}$.

2.11 $p = (-1 + \sqrt{5})/2$.

2.12 a 1/10!

2.12 b $5! \cdot 5!$

2.12 c $8/63 = 12.7$ percent.

2.14 a

	a	b	c
a	0	1/6	1/6
b	0	0	1/3
c	0	1/3	0

2.14 b $P(\{(a,b),(a,c)\}) = 1/3$.

2.14 c $P(\{(b,c),(c,b)\}) = 2/3$.

2.16 $P(E) = 2/3$.

2.19 a $\Omega = \{2, 3, 4, \ldots\}$.

2.19 b $4p^2(1-p)^3$.

3.1 7/36.

3.2 a $P(A \mid B) = 2/11$.

3.2 b No.

3.3 a $P(S_1) = 13/52 = 1/4$, $P(S_2 \mid S_1) = 12/51$, and $P(S_2 \mid S_1^c) = 13/51$.

3.3 b $P(S_2) = 1/4$.

3.4 $P(B \mid T) = 9.1 \cdot 10^{-5}$ and $P(B \mid T^c) = 4.3 \cdot 10^{-6}$.

3.7 a $P(A \cup B) = 1/2$.

3.7 b $P(B) = 1/3$.

3.8 a $P(W) = 0.117$.

3.8 b $P(F \mid W) = 0.846$.

3.9 $P(B \mid A) = 7/15$.

3.14 a $P(W \mid R) = 0$ and $P(W \mid R^c) = 1$.

3.14 b $P(W) = 2/3$.

3.16 a $P(D \mid T) = 0.165$.

3.16 b 0.795.

4.1 a

a	0	1	2
$p_Z(a)$	25/36	10/36	1/36

Z has a $Bin(2, 1/6)$ distribution.

4.1 b $\{M = 2, Z = 0\} = \{(2,1), (1,2), (2,2)\}$, $\{S = 5, Z = 1\} = \emptyset$, and $\{S = 8, Z = 1\} = \{(6,2), (2,6)\}$. $P(M = 2, Z = 0) = 1/12$, $P(S = 5, Z = 1) = 0$, and $P(S = 8, Z = 1) = 1/18$.

4.1 c The events are dependent.

4.3

a	0	1/2	3/4
$p(a)$	1/3	1/6	1/2

4.6 a $p_{\bar{X}}(1) = p_{\bar{X}}(3) = 1/27$, $p_{\bar{X}}(4/3) = p_{\bar{X}}(8/3) = 3/27$, $p_{\bar{X}}(5/3) = p_{\bar{X}}(7/3) = 6/27$, and $p_{\bar{X}}(2) = 7/27$.

4.6 b 6/27.

4.7 a $Bin(1000, 0.001)$.

4.7 b $\mathrm{P}(X = 0) = 0.3677$, $\mathrm{P}(X = 1) = 0.3681$, and $\mathrm{P}(X > 2) = 0.0802$.

4.8 a $Bin(6, 0.8178)$.

4.8 b 0.9999634.

4.10 a Determine $\mathrm{P}(R_i = 0)$ first.

4.10 b No!

4.10 c See the birthday problem in Section 3.2.

4.12 No!

4.13 a $Geo(1/N)$.

4.13 b Let D_i be the event that the marked bolt was drawn (for the first time) in the ith draw, and use conditional probabilities in
$\mathrm{P}(Y = k) = \mathrm{P}(D_1^c \cap \cdots \cap D_{k-1}^c \cap D_k)$.

4.13 c Count the number of ways the event $\{Z = k\}$ can occur, and divide this by the number of ways $\binom{N}{r}$ we can select r objects from N objects.

5.2 $\mathrm{P}(1/2 < X \le 3/4) = 5/16$.

5.4 a $\mathrm{P}(X < 4\,1/2) = 1/4$.

5.4 b $\mathrm{P}(X = 5)$=1/2.

5.4 c X is neither discrete nor continuous!

5.5 a $c = 1$.

5.5 b $F(x) = 0$ for $x \le -3$;
$F(x) = (x + 3)^2/2$ for $-3 \le x \le -2$;
$F(x) = 1/2$ for $-2 \le x \le 2$;
$F(x) = 1 - (3 - x)^2/2$ for $2 \le x \le 3$;
$F(x) = 1$ for $x \ge 3$.

5.8 a $g(y) = 1/(2\sqrt{ry})$.

5.8 b Yes.

5.8 c Consider $F(r/10)$.

5.9 a 1/2 and $\{(x, y) : 2 \le x \le 3, 1 \le y \le 3/2\}$.

5.9 b $F(x) = 0$ for $x < 0$;
$F(x) = 2x$ for $0 \le x \le 1/2$;
$F(x) = 1$ for $x > 1/2$.

5.9 c $f(x) = 2$ for $0 \le x \le 1/2$;
$f(x) = 0$ elsewhere.

5.12 2.

5.13 a Change variables from x to $-x$.

5.13 b $\mathrm{P}(Z \le -2) = 0.0228$.

6.2 a $1 + 2\sqrt{0.378\cdots} = 2.2300795$.

6.2 b Smaller.

6.2 c 0.3782739.

6.5 Show, for $a \ge 0$, that $X \le a$ is equivalent with $U \ge \mathrm{e}^{-a}$.

6.6 $U = \mathrm{e}^{-2X}$.

6.7 $Z = \sqrt{-\ln(1 - U)/5}$, or
$Z = \sqrt{-\ln U/5}$.

6.9 a 6/8.

6.9 b $Geo(6/8)$.

6.10 a Define $B_i = 1$ if $U_i \le p$ and $B_i = 0$ if $U_i > p$, and N as the position in the sequence of B_i where the first 1 occurs.

6.10 b $\mathrm{P}(Z > n) = (1 - p)^n$, for $n = 0, 1, \ldots$; Z has a $Geo(p)$ distribution.

7.1 a Outcomes: 1, 2, 3, 4, 5, and 6. Each has probability 1/6.

7.1 b $\mathrm{E}[T] = 7/2$, $\mathrm{Var}(T) = 35/12$.

7.2 a $\mathrm{E}[X] = 1/5$.

7.2 b

y	0	1
$\mathrm{P}(Y = y)$	2/5	3/5

and $\mathrm{E}[Y] = 3/5$.

7.2 c $\mathrm{E}\left[X^2\right] = 3/5$.

7.2 d $\mathrm{Var}(X) = 14/25$.

7.5 $\mathrm{E}[X] = p$ and $\mathrm{Var}(X) = p(1 - p)$.

7.6 195/76.

7.8 $\mathrm{E}[X] = 1/3$.

7.10 a $\mathrm{E}[X] = 1/\lambda$ and $\mathrm{E}\left[X^2\right] = 2/\lambda^2$.

7.10 b $\mathrm{Var}(X) = 1/\lambda^2$.

7.11 a 2.

7.11 b The expectation is infinite!

7.11 c $\mathrm{E}[X] = \int_1^\infty x \cdot \alpha x^{-\alpha-1}\,\mathrm{d}x$.

7.15 a Start with
$\mathrm{Var}(rX) = \mathrm{E}\left[(rX - \mathrm{E}[rX])^2\right]$.

7.15 b Start with $\mathrm{Var}(X + s) =$
$\mathrm{E}\left[((X + s) - \mathrm{E}[X + s])^2\right]$.

7.15 c Apply **b** with rX instead of X.

7.16 $\mathrm{E}[X] = 4/9$.

7.17 a If positive terms add to zero, they must all be zero.

7.17 b Note that $\mathrm{E}\left[(V - \mathrm{E}[V])^2\right] = \mathrm{Var}(V)$.

8.1

y	0	10	20
$\mathrm{P}(Y = y)$	0.2	0.4	0.4

8.2 a

y	-1	0	1
$\mathrm{P}(Y = y)$	1/6	1/2	1/3

8.2 b

z	-1	0	1
$\mathrm{P}(Z = z)$	1/3	1/2	1/6

8.2 c $\mathrm{P}(W = 1) = 1$.

8.3 a V has a $U(7, 9)$ distribution.

8.3 b $rU + s$ has a $U(s, s + r)$ distribution if $r > 0$ and a $U(s+r, s)$ distribution if $r < 0$.

8.5 a $x^2(3 - x)/4$ for $0 \leq x \leq 2$.

8.5 b $F_Y(y) = (3/4)y^4 - (1/4)y^6$ for $0 \leq y \leq \sqrt{2}$.

8.5 c $3y^3 - (3/2)y^5$ for $0 \leq y \leq \sqrt{2}$, 0 elsewhere.

8.8 $F_W(w) = 1 - \mathrm{e}^{-\gamma w^\alpha}$, with $\gamma = \lambda^\alpha$.

8.10 0.1587.

8.11 Apply Jensen with $-g$.

8.12 a

y	0	1	10	100
$\mathrm{P}(Y = y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

8.12 b $\sqrt{\mathrm{E}[X]} \geq \mathrm{E}\left[\sqrt{X}\right]$.

8.12 c $\sqrt{\mathrm{E}[X]} = 50.25$, but $\mathrm{E}\left[\sqrt{X}\right] = 27.75$.

8.18 V has an exponential distribution with parameter $n\lambda$.

8.19 a The upper right quarter of the circle.

8.19 b $F_Z(t) = 1/2 + \arctan(t)/\pi$.

8.19 c $1/[\pi(1 + z^2)]$.

9.2 a $\mathrm{P}(X = 0, Y = -1) = 1/6$,
$\mathrm{P}(X = 0, Y = 1) = 0$,
$\mathrm{P}(X = 1, Y = -1) = 1/6$,
$\mathrm{P}(X = 2, Y = -1) = 1/6$,
and $\mathrm{P}(X = 2, Y = 1) = 0$.

9.2 b Dependent.

9.5 a $1/16 \leq \eta \leq 1/4$.

9.5 b No.

9.6 a

v \ u	0	1	2	
0	1/4	0	1/4	1/2
1	0	1/2	0	1/2
	1/4	1/2	1/4	1

9.6 b Dependent.

9.8 a

z	0	1	2	3
$p_Z(z)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

9.8 b

z	-2	-1	0	1	2	3
$p_{\tilde{X}}(z)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

9.9 a $F_X(x) = 1 - \mathrm{e}^{-2x}$ for $x > 0$ and $F_Y(y) = 1 - \mathrm{e}^{-y}$ for $y > 0$.

9.9 b $f(x, y) = 2\mathrm{e}^{-(2x+y)}$ for $x > 0$ and $y > 0$.

9.9 c $f_X(x) = 2\mathrm{e}^{-2x}$ $x > 0$ and $f_Y(y) = \mathrm{e}^{-y}$ for $y > 0$.

9.9 d Independent.

9.10 a 41/720.

9.10 b $F(a, b) = \frac{3}{5}a^2b^2 + \frac{2}{5}a^2b^3$.

9.10 c $F_X(a) = a^2$.

9.10 d $f_X(x) = 2x$ for $0 \leq x \leq 1$.

9.10 e Independent.

9.11 27/50.

9.13 a $1/\pi$.

9.13 b $F_R(r) = r^2$ for $0 \leq r \leq 1$.

9.13 c $f_X(x) = \frac{2}{\pi}\sqrt{1 - x^2} = f_Y(x)$ for x between -1 and 1.

9.15 a Since $F(a, b) = \frac{\text{area } (\Delta \cap \square(a,b))}{\text{area of } \Delta}$, where $\square(a, b)$ is the set of points (x, y), for which $x \leq a$ and $y \leq b$, one needs to calculate the areas for the various cases.

9.15 b $f(x, y) = 2$ for $(x, y) \in \Delta$, and $f(x, y) = 0$ otherwise.

9.15 c Use the rule on page 122.

9.19 a $a = 5\sqrt{2}$, $b = 4\sqrt{2}$, and $c = 18$.

9.19 b Use that $\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$ is the probability density function of an $N(\mu, \sigma^2)$ distributed random variable.

9.19 c $N(0, 1/36)$.

10.1 a $\mathrm{Cov}(X, Y) = 0.142$. Positively correlated.

10.1 b $\rho(X, Y) = 0.0503$.

10.2 a $\mathrm{E}[XY] = 0$.

10.2 b $\mathrm{Cov}(X, Y) = 0$.

10.2 c $\mathrm{Var}(X + Y) = 4/3$.

10.2 d $\mathrm{Var}(X - Y) = 4/3$.

10.5 a

	a			
b	0	1	2	
0	8/72	6/72	10/72	1/3
1	12/72	9/72	15/72	1/2
2	4/72	3/72	5/72	1/6
	1/3	1/4	5/12	1

10.5 b $\mathrm{E}[X] = 13/12$, $\mathrm{E}[Y] = 5/6$, and $\mathrm{Cov}(X, Y) = 0$.

10.5 c Yes.

10.6 a $\mathrm{E}[X] = \mathrm{E}[Y] = 0$ and $\mathrm{Cov}(X, Y) = 0$.

10.6 b $\mathrm{E}[X] = \mathrm{E}[Y] = c$; $\mathrm{E}[XY] = c^2$.

10.6 c No.

10.7 a $\mathrm{Cov}(X, Y) = -1/8$.

10.7 b $\rho(X, Y) = -1/2$.

10.7 c For ε equal to 1/4, 0 or $-1/4$.

10.9 a $\mathrm{P}(X_i = 1) = (1-0.001)^{40} = 0.96$ and $\mathrm{P}(X_i = 41) = 0.04$.

10.9 b $\mathrm{E}[X_i] = 2.6$ and $\mathrm{E}[X_1 + \cdots + X_{25}] = 65$.

10.10 a $\mathrm{E}[X] = 109/50$, $\mathrm{E}[Y] = 157/100$, and $\mathrm{E}[X + Y] = 15/4$.

10.10 b $\mathrm{E}\left[X^2\right] = 1287/250$, $\mathrm{E}\left[Y^2\right] = 318/125$, and $\mathrm{E}[X + Y] = 3633/250$.

10.10 c $\mathrm{Var}(X) = 989/2500$, $\mathrm{Var}(Y) = 791/10\,000$, and $\mathrm{Var}(X + Y) = 4747/10\,000$.

10.14 a Use the alternative expression for the covariance.

10.14 b Use the alternative expression for the covariance.

10.14 c Combine parts **a** and **b**.

10.16 a $\mathrm{Var}(X) + \mathrm{Cov}(X, Y)$.

10.16 b Anything can happen.

10.16 c X and $X + Y$ are positively correlated.

10.18 Solve $0 = N(N-1)(N+1)/12 + N(N-1)\mathrm{Cov}(X_1, X_2)$.

11.1 a Check that for k between 2 and 6, the summation runs over $\ell = 1, \ldots, k-1$, whereas for k between 7 and 12 it runs over $\ell = k-6, \ldots, 12$.

11.1 b Check that for $2 \leq k \leq N$, the summation runs over $\ell = 1, \ldots, k-1$, whereas for k between $N+1$ and $2N$ it runs over $\ell = k-N, \ldots, 2N$.

11.2 a Check that the summation runs over $\ell = 0, 1, \ldots, k$.

11.2 b Use that $\lambda^{k-\ell}\mu^{\ell}/(\lambda+\mu)^k$ is equal to $p^{\ell}(1-p)^{k-\ell}$, with $p = \mu/(\lambda + \mu)$.

11.4 a $\mathrm{E}[Z] = -3$ and $\mathrm{Var}(Z) = 81$.

11.4 b Z has an $N(-3, 81)$ distribution.

11.4 c $\mathrm{P}(Z \leq 6) = 0.8413$.

11.5 Check that for $0 \leq z < 1$, the integral runs over $0 \leq y \leq z$, whereas for $1 \leq z \leq 2$, it runs over $z - 1 \leq y \leq 1$.

11.6 Check that the integral runs over $0 \leq y \leq z$.

11.7 Recall that a $Gam(k, \lambda)$ random variable can be represented as the sum of k independent $Exp(\lambda)$ random variables.

11.9 a $f_Z(z) = \dfrac{3}{2}\left(\dfrac{1}{z^2} - \dfrac{1}{z^4}\right)$, for $z \geq 1$.

11.9 b $f_Z(z) = \dfrac{\alpha\beta}{\beta - \alpha}\left(\dfrac{1}{z^{\beta+1}} - \dfrac{1}{z^{\alpha+1}}\right)$, for $z \geq 1$.

12.1 e 1: no, 2: no, 3: okay, 4: okay, 5: okay.

12.5 a 0.00049.

12.5 b 1 (correct to 8281 decimals).

12.6 0.256.

12.7 a $\lambda \approx 0.192$.

12.7 b 0.1583 is close to 0.147.

12.7 c $2.71 \cdot 10^{-5}$.

12.8 a $\mathrm{E}[X(X-1)] = \mu^2$.

12.8 b $\mathrm{Var}(X) = \mu$.

12.11 The probability of the event in the hint equals $(\lambda s)^n \mathrm{e}^{-\lambda 2s}/(k!(n-k)!)$.

12.14 a Note: $1-1/n \to 1$ and $1/n \to 0$.

12.14 b $\mathrm{E}[X_n] = (1-1/n) \cdot 0 + (1/n) \cdot 7n = 7$.

13.2 a $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1/12$.

13.2 b 1/12.

13.4 a $n \geq 63$.

13.4 b $n \geq 250$.

13.4 c $n \geq 125$.

13.4 d $n \geq 240$.

13.6 Expected income per game € 1/37; per year: € 9865.

13.8 a $\mathrm{Var}\big(\bar{Y}_n/2h\big) = 0.171/h\sqrt{n}$.

13.8 b $n \geq 801$.

13.9 a T_n is the average of a sequence of independent and identically distributed random variables.

13.9 b $a = \mathrm{E}\left[X_i^2\right] = 1/3$.

13.10 a $\mathrm{P}(|M_n - 1| > \varepsilon) = (1-\varepsilon)^n$ for $0 \leq \varepsilon \leq 1$.

13.10 b No.

14.2 0.9977.

14.3 17.

14.4 1/2.

14.5 Use that X has the same probability distribution as $X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent $Ber(p)$ distributed random variables.

14.6 a $\mathrm{P}(X \leq 25) \approx 0.5$, $\mathrm{P}(X < 26) \approx 0.6141$.

14.6 b $\mathrm{P}(X \leq 2) \approx 0$.

14.9 a 5.71%.

14.9 b Yes!

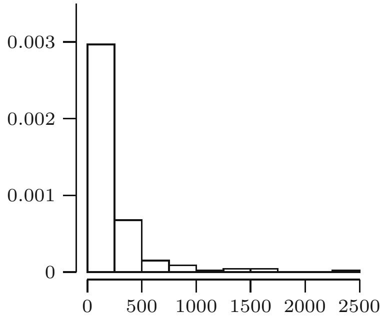
14.10 a 91.

14.10 b Use that $(\bar{M}_n - c)/\sigma$ has an $N(0,1)$ distribution.

15.3 a

Bin	Height
(0,250]	0.00297
(250,500]	0.00067
(500,750]	0.00015
(750,1000]	0.00008
(1000,1250]	0.00002
(1250,1500]	0.00004
(1500,1750]	0.00004
(1750,2000]	0
(2250,2500]	0
(2250,2500]	0.00002

15.3 b Skewed.



15.4 a

Bin	Height
[0,500]	0.0012741
(500,1000]	0.0003556
(1000,1500]	0.0001778
(1500,2000]	0.0000741
(2000,2500]	0.0000148
(2500,3000]	0.0000148
(3000,3500]	0.0000296
(3500,4000]	0
(4000,4500]	0.0000148
(4500,5000]	0
(5000,5500]	0.0000148
(5500,6000]	0.0000148
(6000,6500]	0.0000148

15.4 b

t	$F_n(t)$	t	$F_n(t)$
0	0	3500	0.9704
500	0.6370	4000	0.9704
1000	0.8148	4500	0.9778
1500	0.9037	5000	0.9778
2000	0.9407	5500	0.9852
2500	0.9481	6000	0.9926
3000	0.9556	6500	1

15.4 c Both are equal to 0.0889.

15.5

Bin	Height
$(0, 1]$	0.2250
$(1, 3]$	0.1100
$(3, 5]$	0.0850
$(5, 8]$	0.0400
$(8, 11]$	0.0230
$(11, 14]$	0.0350
$(14, 18]$	0.0225

15.6 $F_n(7) = 0.9$.

15.11 Use that the number of x_i in $(a, b]$ equals the number of $x_i \le b$ minus the number of $x_i \le a$.

15.12 a Bring the integral into the sum, change the integration variable to $u = (t - x_i)/h$, and use the properties of kernel functions.

15.12 b Similar to **a**.

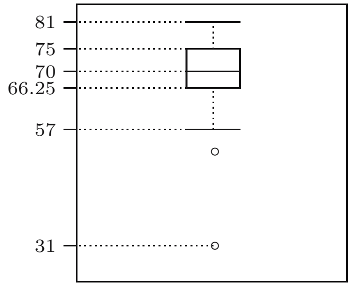
16.1 a Median: 290.

16.1 b Lower quartile: 81; upper quartile: 843; IQR: 762.

16.1 c 144.6.

16.3 a Median: 70; lower quartile: 66.25; upper quartile: 75.

16.3 b



16.3 c Note the position of 31 in the boxplot.

16.4 a Yes, they both equal 7.056.

16.4 b Yes.

16.4 c Yes.

16.6 a Yes.

16.6 b In general this will not be true.

16.6 c Yes.

16.8 MAD is 3.

16.10 a The sample mean goes to infinity, whereas the sample median changes to 4.6.

16.10 b At least three elements need to be replaced.

16.10 c For the sample mean only one; for the sample median at least $\lfloor (n+1)/2 \rfloor$ elements.

16.12 $\bar{x}_n = (N + 1)/2$; $\text{Med}_n = (N + 1)/2$.

16.15 Write $(x_i - \bar{x}_n)^2 = x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2$.

17.1

$N(3, 1)$	$N(0, 1)$	$N(0, 1)$
$N(3, 1)$	$Exp\,(1/3)$	$Exp\,(1)$
$N(0, 1)$	$N(0, 9)$	$Exp\,(1)$
$N(3, 1)$	$N(0, 9)$	$Exp\,(1/3)$
$N(0, 9)$	$Exp\,(1/3)$	$Exp\,(1)$

17.2

$Exp(1/3)$	$N(0,9)$	$Exp(1/3)$
$N(0,1)$	$N(3,1)$	$Exp(1)$
$N(0,9)$	$N(0,9)$	$N(3,1)$
$Exp(1)$	$N(3,1)$	$Exp(1)$
$N(0,1)$	$N(0,1)$	$Exp(1/3)$

17.3 a $Bin(10, p)$.

17.3 b $p = 0.435$.

17.5 a One possibility is $p = 93/331$; another is $p = 29/93$.

17.5 b $p = 474/1285$ or $p = 198/474$.

17.5 c 0.6281 or 0.6741 for smokers and 0.7486 or 0.8026 for nonsmokers.

17.7 a An exponential distribution.

17.7 b One possibility is $\lambda = 0.00469$.

17.9 a Recall the formula for the volume of a cylinder with diameter d (at the base) and height h.

17.9 b $\bar{z}_n = 0.3022$; $\bar{y}/\bar{x} = 0.3028$; least squares: 0.3035.

18.1 $5^6 = 15625$. Not equally likely.

18.3 a 0.0574.

18.3 b 0.0547.

18.3 c 0.000029.

18.4 a 0.3487.

18.4 b $(1 - 1/n)^n$.

18.5 values 0, ± 1, ± 2, and ± 3 with probabilities 7/27, 6/27, 3/27, and 1/27.

18.7 Determine from which parametric distribution you generate the bootstrap datasets and what the bootstrapped version is of $\bar{X}_n - \mu$.

18.8 a Determine from which $\hat{F}$ you generate the bootstrap datasets and what the bootstrapped version is of $\bar{X}_n - \mu$.

18.8 b Similar to **a**.

18.8 c Similar to **a** and **b**.

18.9 Determine which normal distribution corresponds to $X_1^*, X_2^*, \ldots, X_n^*$ and use this to compute $\mathrm{P}\left(|\bar{X}_n^* - \mu^*| > 1\right)$.

19.1 a First show that $\mathrm{E}\left[X_1^2\right] = \theta^2/3$, and use linearity of expectations.

19.1 b $\sqrt{T}$ has negative bias.

19.3 $a = 1/n$, $b = 0$.

19.5 $c = n$.

19.6 a Use linearity of expectations and plug in the expressions for $\mathrm{E}[M_n]$ and $\mathrm{E}\left[\bar{X}_n\right]$.

19.6 b $(nM_n - \bar{X}_n)/(n-1)$.

19.6 c Estimate for δ: 2073.5.

19.8 Check that $\mathrm{E}[Y_i] = \beta x_i$ and use linearity of expectations.

20.2 a We prefer T.

20.2 b If $a < 6$ we prefer T; if $a \geq 6$ we prefer S.

20.3 T_1.

20.4 a $\mathrm{E}[3L-1] = 3\mathrm{E}[N+1-M]-1 = N$.

20.4 b $(N+1)(N-2)/2$.

20.4 c 4 times.

20.7 $\mathrm{Var}(T_1) = (4 - \theta^2)/n$ and $\mathrm{Var}(T_2) = \theta(4-\theta)/n$. We prefer T_2.

20.8 a Use linearity of expectations.

20.8 b Differentiate with respect to r.

20.11 $\mathrm{MSE}(T_1) = \sigma^2/(\sum_{i=1}^n x_i^2)$, $\mathrm{MSE}(T_2) = (\sigma^2/n^2)\cdot\sum_{i=1}^n(1/x_i^2)$, $\mathrm{MSE}(T_3) = \sigma^2\, n/(\sum_{i=1}^n x_i)^2$.

21.1 D_2.

21.2 $\hat{p} = 1/4$.

21.4 a Use that $X_1, \ldots, X_n$ are independent $Pois(\mu)$ distributed random variables.

21.4 b $\ell(\mu) = \left(\sum_{i=1}^n x_i\right)\ln(\mu) - \ln\left(x_1!\cdot x_2!\cdots x_n!\right) - n\mu$, $\hat{\mu} = \bar{x}_n$.

21.4 c $\mathrm{e}^{-\bar{x}_n}$.

21.5 a $\bar{x}_n$.

21.5 b $\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}$.

21.7 $\sqrt{\frac{1}{2n}\sum_{i=1}^n x_i^2}$.

21.8 a $L(\theta) = \frac{C}{4^{3839}} \cdot (2+\theta)^{1997} \cdot \theta^{32} \cdot (1-\theta)^{1810}$; $\ell(\theta) = \ln(C) - 3839\ln(4) + 1997\ln(2+\theta) + 32\ln(\theta) + 1810\ln(1-\theta)$.

21.8 b 0.0357.

21.8 c $(-b+\sqrt{D})/(2n)$, with $b = -n_1 + n_2 + 2n_3 + 2n_4$, and $D = (n_1 - n_2 - 2n_3 - 2n_4)^2 + 8nn_2$.

21.9 $\hat{\alpha} = x_{(1)}$ and $\hat{\beta} = x_{(n)}$.

21.11 a $1/\bar{x}_n$.

21.11 b $y_{(n)}$.

22.1 a $\hat{\alpha} = 2.35$, $\hat{\beta} = -0.25$.

22.1 b $r_1 = -0.1$, $r_2 = 0.2$, $r_3 = -0.1$.

22.1 c The estimated regression line goes through $(0, 2.35)$ and $(3, 1.6)$.

22.5 Minimize $\sum_{i=1}^{n} (y_i - \beta x_i)^2$.

22.6 2218.45.

22.8 The model with no intercept.

22.10 a $\hat{\alpha} = 7/3$, $\hat{\beta} = -1$, $A(\hat{\alpha}, \hat{\beta}) = 4/3$.

22.10 b $17/9 < \alpha < 7/3$, $\alpha = 2$.

22.10 c $\alpha = 2$, $\beta = -1$.

22.12 a Use that the denominator of $\hat{\beta}$ and that $\sum x_i$ are numbers, *not* random variables.

22.12 b Use that $\mathrm{E}[Y_i] = \alpha + \beta x_i$.

22.12 c Simplify the expression in **b**.

22.12 d Combine **a** and **c**.

23.1 $(740.55, 745.45)$.

23.2 $(3.486, 3.594)$.

23.5 a $(0.050, 1.590)$.

23.5 b See Section 23.3.

23.5 c (0.045, 1.600).

23.6 a Rewrite the probability in terms of L_n and U_n.

23.6 b $(3l_n + 7, 3u_n + 7)$.

23.6 c $\tilde{L}_n = 1 - U_n$ and $\tilde{U}_n = 1 - L_n$. The confidence interval: $(-4, 3)$.

23.6 d $(0, 25)$ is a conservative 95% confidence interval for θ.

23.7 $\left(\mathrm{e}^{-3}, \mathrm{e}^{-2}\right) = (0.050, 0.135)$.

23.11 a Yes.

23.11 b Not necessarily.

23.11 c Not necessarily.

24.1 $(0.620, 0.769)$.

24.4 a 609.

24.4 b No.

24.6 a $(1.68, \infty)$.

24.6 b $[0, 2.80)$.

24.8 a $(0.449, 0.812)$.

24.8 b $(0.481, 1]$.

24.9 a See Section 8.4.

24.9 b $c_l = 0.779$, $c_u = 0.996$.

24.9 c $(3.013, 3.851)$.

24.9 d $(m/(1-\alpha/2)^{1/n}, m/(\alpha/2)^{1/n})$.

25.2 $H_1 : \mu > 1472$.

25.4 a The difference or the ratio of the average numbers of cycles for the two groups.

25.4 b The difference or the ratio of the maximum likelihood estimators $\hat{p}_1$ and $\hat{p}_2$.

25.4 c $H_1 : p_1 < p_2$.

25.5 a Relevant values of T_1 are in $[0, 5]$; those close to 0, or close to 5, are in favor of H_1.

25.5 b Relevant values of T_2 are in $[0, 5]$; only those close to 0 are in favor of H_1.

25.6 a The p-value is 0.23. Do not reject.

25.6 b The p-value is 0.77. Do not reject.

25.6 c The p-value is 0.968. Do not reject.

25.6 d The p-value is 0.019. Reject.

25.6 e The p-value is 0.99. Do not reject.

25.6 f The p-value is smaller than 0.019. Reject.

25.6 g The p-value is smaller than 0.200. We cannot say anything about rejection of H_0.

25.10 a $H_1 : \mu > 23.75$.

25.10 b The p-value is 0.0344.

25.11 0.0456.

26.3 a 0.1.

26.3 b 0.72.

26.5 a The p-value is 0.1050. Do not reject H_0; this agrees with Exercise 24.8 **b**.

26.5 b $K = \{16, 17, \ldots, 23\}$.

26.5 c 0.0466.

26.5 d 0.6950.

26.6 a Right critical value.

26.6 b Right critical value $c = 1535.1$; critical region $[1536, \infty)$.

26.8 a For T we find $K = (0, c_l]$ and for T' we find $K' = [c_u, 1)$.

26.8 b For T we find $K = (0, c_l] \cup [c_u, \infty)$ and for T' we find $K' = (0, c_l'] \cup [c_u', 1)$.

26.9 a For T we find $K = [c_u, \infty)$ and for T' we find $K' = [c_l', 0) \cup (0, c_u']$.

26.9 b For T we find $K = [c_u, \infty)$ and for T' we find $K' = (0, c_u']$.

27.2 a $H_0 : \mu = 2550$ and $H_1 : \mu \neq 2550$.

27.2 b $t = 1.2096$. Do not reject H_0.

27.5 a $H_0 : \mu = 0$; $H_1 : \mu > 0$; $t = 0.70$.

27.5 b p-value: 0.2420. Do not reject H_0.

27.7 a $H_0 : \beta = 0$ and $H_1 : \beta < 0$; $t_b = -20.06$. Reject H_0.

27.7 b Same testing problem; $t_b = -11.03$. Reject H_0.

28.1 a $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$; $t_p = -2.130$. Reject H_0.

28.1 b $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$; $t_d = -2.130$. Reject H_0.

28.1 c Reject H_0. The salaries differ significantly.

28.3 a $t_p = 2.492$. Reject H_0.

28.3 b Reject H_0.

28.3 c $t_d = 2.463$. Reject H_0.

28.3 d Reject H_0.

28.5 a Determine $\mathrm{E}\left[aS_X^2 + bS_Y^2\right]$, using that S_X^2 and S_Y^2 are both unbiased for σ^2.

28.5 b Determine $\mathrm{E}\left[aS_X^2 + (1-a)S_Y^2\right]$, using that S_X^2 and S_Y^2 are independent, and minimize over a.

D

Full solutions to selected exercises

2.8 From the rule for the probability of a union we obtain $P(D_1 \cup D_2) \leq P(D_1) + P(D_2) = 2 \cdot 10^{-6}$. Since $D_1 \cap D_2$ is contained in both D_1 and D_2, we obtain $P(D_1 \cap D_2) \leq \min\{P(D_1), P(D_2)\} = 10^{-6}$. Equality may hold in both cases: for the union, take D_1 and D_2 disjoint, for the intersection, take D_1 and D_2 equal to each other.

2.12 a This is the same situation as with the three envelopes on the doormat, but now with ten possibilities. Hence an outcome has probability $1/10!$ to occur.

2.12 b For the five envelopes labeled $1, 2, 3, 4, 5$ there are $5!$ possible orders, and for each of these there are $5!$ possible orders for the envelopes labeled $6, 7, 8, 9, 10$. Hence in total there are $5! \cdot 5!$ outcomes.

2.12 c There are $32{\cdot}5!{\cdot}5!$ outcomes in the event "dream draw." Hence the probability is $32 \cdot 5!5!/10! = 32 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5/(6 \cdot 7 \cdot 8 \cdot 9 \cdot 10) = 8/63 =$12.7 percent.

2.14 a Since door a is never opened, $P((a,a)) = P((b,a)) = P((c,a)) = 0$. If the candidate chooses a (which happens with probability $1/3$), then the quizmaster chooses without preference from doors b and c. This yields that $P((a,b)) = P((a,c)) = 1/6$. If the candidate chooses b (which happens with probability $1/3$), then the quizmaster can only open door c. Hence $P((b,c)) = 1/3$. Similarly, $P((c,b)) = 1/3$. Clearly, $P((b,b)) = P((c,c)) = 0$.

2.14 b If the candidate chooses a then she or he wins; hence the corresponding event is $\{(a,a), (a,b), (a,c)\}$, and its probability is $1/3$.

2.14 c To end with a the candidate should have chosen b or c. So the event is $\{(b,c), (c,b)\}$ and $P(\{(b,c), (c,b)\}) = 2/3$.

2.16 Since $E \cap F \cap G = \emptyset$, the three sets $E \cap F$, $F \cap G$, and $E \cap G$ are disjoint. Since each has probability $1/3$, they have probability 1 together. From these two facts one deduces $P(E) = P(E \cap F) + P(E \cap G) = 2/3$ (make a diagram or use that $E = E \cap (E \cap F) \cup E \cap (F \cap G) \cup E \cap (E \cap G)$).

3.1 Define the following events: B is the event "point B is reached on the second step," C is the event "the path to C is chosen on the first step," and similarly we define D and E. Note that the events C, D, and E are mutually exclusive and that one of them must occur. Furthermore, that we can only reach B by first going to C

or D. For the computation we use the law of total probability, by conditioning on the result of the first step:

$$\begin{aligned}\mathrm{P}(B) &= \mathrm{P}(B \cap C) + \mathrm{P}(B \cap D) + \mathrm{P}(B \cap E)\\ &= \mathrm{P}(B \mid C)\,\mathrm{P}(C) + \mathrm{P}(B \mid D)\,\mathrm{P}(D) + \mathrm{P}(B \mid E)\,\mathrm{P}(E)\\ &= \frac{1}{3}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{3}\cdot 0 = \frac{7}{36}.\end{aligned}$$

3.2 a Event A has three outcomes, event B has 11 outcomes, and $A \cap B = \{(1,3),(3,1)\}$. Hence we find $\mathrm{P}(B) = 11/36$ and $\mathrm{P}(A \cap B) = 2/36$ so that

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

3.2 b Because $\mathrm{P}(A) = 3/36 = 1/12$ and this is not equal to $2/11 = \mathrm{P}(A \mid B)$ the events A and B are *dependent*.

3.3 a There are 13 spades in the deck and each has probability $1/52$ of being chosen, hence $\mathrm{P}(S_1) = 13/52 = 1/4$. Given that the first card is a spade there are $13-1 = 12$ spades left in the deck with $52 - 1 = 51$ remaining cards, so $\mathrm{P}(S_2 \mid S_1) = 12/51$. If the first card is not a spade there are 13 spades left in the deck of 51, so $\mathrm{P}(S_2 \mid S_1^c) = 13/51$.

3.3 b We use the law of total probability (based on $\Omega = S_1 \cup S_1^c$):

$$\begin{aligned}\mathrm{P}(S_2) &= \mathrm{P}(S_2 \cap S_1) + \mathrm{P}(S_2 \cap S_1^c) = \mathrm{P}(S_2 \mid S_1)\,\mathrm{P}(S_1) + \mathrm{P}(S_2 \mid S_1^c)\,\mathrm{P}(S_1^c)\\ &= \frac{12}{51}\cdot\frac{1}{4} + \frac{13}{51}\cdot\frac{3}{4} = \frac{12+39}{51\cdot 4} = \frac{1}{4}.\end{aligned}$$

3.7 a The best approach to a problem like this one is to write out the conditional probability and then see if we can somehow combine this with $\mathrm{P}(A) = 1/3$ to solve the puzzle. Note that $\mathrm{P}(B \cap A^c) = \mathrm{P}(B \mid A^c)\,\mathrm{P}(A^c)$ and that $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B \cap A^c)$. So

$$\mathrm{P}(A \cup B) = \frac{1}{3} + \frac{1}{4}\cdot\left(1 - \frac{1}{3}\right) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

3.7 b From the conditional probability we find $\mathrm{P}(A^c \cap B^c) = \mathrm{P}(A^c \mid B^c)\,\mathrm{P}(B^c) = \frac{1}{2}\,(1 - \mathrm{P}(B))$. Recalling DeMorgan's law we know $\mathrm{P}(A^c \cap B^c) = \mathrm{P}((A \cup B)^c) = 1 - \mathrm{P}(A \cup B) = 1/3$. Combined this yields an equation for $\mathrm{P}(B)$: $\frac{1}{2}\,(1 - \mathrm{P}(B)) = 1/3$ from which we find $\mathrm{P}(B) = 1/3$.

3.8 a This asks for $\mathrm{P}(W)$. We use the law of total probability, decomposing $\Omega = F \cup F^c$. Note that $\mathrm{P}(W \mid F) = 0.99$.

$$\begin{aligned}\mathrm{P}(W) &= \mathrm{P}(W \cap F) + \mathrm{P}(W \cap F^c) = \mathrm{P}(W \mid F)\,\mathrm{P}(F) + \mathrm{P}(W \mid F^c)\,\mathrm{P}(F^c)\\ &= 0.99 \cdot 0.1 + 0.02 \cdot 0.9 = 0.099 + 0.018 = 0.117.\end{aligned}$$

3.8 b We need to determine $\mathrm{P}(F \mid W)$, and this can be done using Bayes' rule. Some of the necessary computations have already been done in **a**, we can copy $\mathrm{P}(W \cap F)$ and $\mathrm{P}(W)$ and get:

$$\mathrm{P}(F \mid W) = \frac{\mathrm{P}(F \cap W)}{\mathrm{P}(W)} = \frac{0.099}{0.117} = 0.846.$$

4.1 a In two independent throws of a die there are 36 possible outcomes, each occurring with probability 1/36. Since there are 25 ways to have no 6's, 10 ways to have one 6, and one way to have two 6's, we find that $p_Z(0) = 25/36$, $p_Z(1) = 10/36$, and $p_Z(2) = 1/36$. So the probability mass function p_Z of Z is given by the following table:

z	0	1	2
$p_Z(z)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

The distribution function F_Z is given by

$$F_Z(a) = \begin{cases} 0 & \text{for } a < 0 \\ \frac{25}{36} & \text{for } 0 \leq a < 1 \\ \frac{25}{36} + \frac{10}{36} = \frac{35}{36} & \text{for } 1 \leq a < 2 \\ \frac{25}{36} + \frac{10}{36} + \frac{1}{36} = 1 & \text{for } a \geq 2. \end{cases}$$

Z is the sum of two independent $Ber(1/6)$ distributed random variables, so Z has a $Bin(2, 1/6)$ distribution.

4.1 b If we denote the outcome of the two throws by (i, j), where i is the outcome of the first throw and j the outcome of the second, then $\{M = 2, Z = 0\} = \{(2,1), (1,2), (2,2)\}$, $\{S = 5, Z = 1\} = \emptyset$, $\{S = 8, Z = 1\} = \{(6,2), (2,6)\}$. Furthermore, $\mathrm{P}(M = 2, Z = 0) = 3/36$, $\mathrm{P}(S = 5, Z = 1) = 0$, and $\mathrm{P}(S = 8, Z = 1) = 2/36$.

4.1 c The events are dependent, because, e.g., $\mathrm{P}(M = 2, Z = 0) = \frac{3}{36}$ differs from $\mathrm{P}(M = 2) \cdot \mathrm{P}(Z = 0) = \frac{3}{36} \cdot \frac{253}{36}$.

4.10 a Each R_i has a Bernoulli distribution, because it can only attain the values 0 and 1. The parameter is $p = \mathrm{P}(R_i = 1)$. It is not easy to determine $\mathrm{P}(R_i = 1)$, but it is fairly easy to determine $\mathrm{P}(R_i = 0)$. The event $\{R_i = 0\}$ occurs when none of the m people has chosen the ith floor. Since they make their choices independently of each other, and each floor is selected by each of these m people with probability 1/21, it follows that

$$\mathrm{P}(R_i = 0) = \left(\frac{20}{21}\right)^m.$$

Now use that $p = \mathrm{P}(R_i = 1) = 1 - \mathrm{P}(R_i = 0)$ to find the desired answer.

4.10 b If $\{R_1 = 0\}, \ldots, \{R_{20} = 0\}$, we must have that $\{R_{21} = 1\}$, so we cannot conclude that the events $\{R_1 = a_1\}, \ldots, \{R_{21} = a_{21}\}$, where a_i is 0 or 1, are independent. Consequently, we cannot use the argument from Section 4.3 to conclude that S_m is $Bin(21, p)$. In fact, S_m is not $Bin(21, p)$ distributed, as the following shows. The elevator will stop at least once, so $\mathrm{P}(S_m = 0) = 0$. However, if S_m would have a $Bin(21, p)$ distribution, then $\mathrm{P}(S_m = 0) = (1 - p)^{21} > 0$, which is a contradiction.

4.10 c This exercise is a variation on finding the probability of no coincident birthdays from Section 3.2. For $m = 2$, $S_2 = 1$ occurs precisely if the two persons entering the elevator select the same floor. The first person selects any of the 21 floors, the second selects the same floor with probability 1/21, so $\mathrm{P}(S_2 = 1) = 1/21$. For $m = 3$, $S_3 = 1$ occurs if the second and third persons entering the elevator both select the same floor as was selected by the first person, so $\mathrm{P}(S_3 = 1) = (1/21)^2 = 1/441$. Furthermore, $S_3 = 3$ occurs precisely when all three persons choose a different floor. Since there are $21 \cdot 20 \cdot 19$ ways to do this out of a total of 21^3 possible ways, we find that $P(S_3 = 3) = 380/441$. Since S_3 can only attain the values $1, 2, 3$, it follows that $P(S_3 = 2) = 1 - P(S_3 = 1) - P(S_3 = 3) = 60/441$.

4.13 a Since we wait for the first time we draw the marked bolt in independent draws, each with a $Ber(p)$ distribution, where p is the probability to draw the bolt (so $p = 1/N$), we find, using a reasoning as in Section 4.4, that X has a $Geo(1/N)$ distribution.

4.13 b Clearly, $P(Y = 1) = 1/N$. Let D_i be the event that the marked bolt was drawn (for the first time) in the ith draw. For $k = 2, \ldots, N$ we have that

$$\begin{aligned} P(Y = k) &= P(D_1^c \cap \cdots \cap D_{k-1}^c \cap D_k) \\ &= P(D_k \mid D_1^c \cap \cdots \cap D_{k-1}^c) \cdot P(D_1^c \cap \cdots \cap D_{k-1}^c). \end{aligned}$$

Now $P(D_k \mid D_1^c \cap \cdots \cap D_{k-1}^c) = \frac{1}{N-k+1}$,

$$P(D_1^c \cap \cdots \cap D_{k-1}^c) = P(D_{k-1}^c \mid D_1^c \cap \cdots \cap D_{k-2}^c) \cdot P(D_1^c \cap \cdots \cap D_{k-2}^c),$$

and

$$P(D_{k-1}^c \mid D_1^c \cap \cdots \cap D_{k-1}^c) = 1 - P(D_{k-1} \mid D_1^c \cap \cdots \cap D_{k-1}^c) = 1 - \frac{1}{N-k+2}.$$

Continuing in this way, we find after k steps that

$$P(Y = k) = \frac{1}{N-k+1} \cdot \frac{N-k+1}{N-k+2} \cdot \frac{N-k+2}{N-k+3} \cdots \frac{N-2}{N-1} \cdot \frac{N-1}{N} = \frac{1}{N}.$$

See also Section 9.3, where the distribution of Y is derived in a different way.

4.13 c For $k = 0, 1, \ldots, r$, the probability $P(Z = k)$ is equal to the number of ways the event $\{Z = k\}$ can occur, divided by the number of ways $\binom{N}{r}$ we can select r objects from N objects, see also Section 4.3. Since one can select k marked bolts from m marked ones in $\binom{m}{k}$ ways, and $r-k$ nonmarked bolts from $N-m$ nonmarked ones in $\binom{N-m}{r-k}$ ways, it follows that

$$P(Z = k) = \frac{\binom{m}{k}\binom{N-m}{r-k}}{\binom{N}{r}}, \quad \text{for } k = 0, 1, 2, \ldots, r.$$

5.4 a Let T be the time until the next arrival of a bus. Then T has $U(4, 6)$ distribution. Hence $P(X \leq 4.5) = P(T \leq 4.5) = \int_4^{4.5} 1/2 \,dx = 1/4$.

5.4 b Since Jensen leaves when the next bus arrives after more than 5 minutes, $P(X = 5) = P(T > 5) = \int_5^6 \frac{1}{2} \,dx = 1/2$.

5.4 c Since $P(X = 5) = 0.5 > 0$, X cannot be continuous. Since X can take any of the uncountable values in $[4, 5]$, it can also not be discrete.

5.8 a The probability density $g(y) = 1/(2\sqrt{ry})$ has an asymptote in 0 and decreases to $1/2r$ in the point r. Outside $[0, r]$ the function is 0.

5.8 b The second darter is better: for each $0 < b < r$ one has $(b/r)^2 < \sqrt{b/r}$ so the second darter always has a larger probability to get closer to the center.

5.8 c Any function F that is 0 left from 0, increasing on $[0, r]$, takes the value 0.9 in $r/10$, and takes the value 1 in r and to the right of r is a correct answer to this question.

5.13 a This follows with a change of variable transformation $x \mapsto -x$ in the integral: $\Phi(-a) = \int_{-\infty}^{-a} \phi(x)\,dx = \int_a^{\infty} \phi(-x)\,dx = \int_a^{\infty} \phi(x)\,dx = 1 - \Phi(a)$.

5.13 b This is straightforward: $P(Z \leq -2) = \Phi(-2) = 1 - \Phi(2) = 0.0228$.

6.5 We see that

$$X \leq a \quad \Leftrightarrow \quad -\ln U \leq a \quad \Leftrightarrow \quad \ln U \geq -a \quad \Leftrightarrow \quad U \geq e^{-a},$$

and so $P(X \leq a) = P\big(U \geq e^{-a}\big) = 1 - P\big(U \leq e^{-a}\big) = 1 - e^{-a}$, where we use $P(U \leq p) = p$ for $0 \leq p \leq 1$ applied to $p = e^{-a}$ (remember that $a \geq 0$).

6.7 We need to obtain F^{inv}, and do this by solving $F(x) = u$, for $0 \leq u \leq 1$:

$$\begin{aligned} 1 - e^{-5x^2} = u \quad &\Leftrightarrow \quad e^{-5x^2} = 1 - u \quad \Leftrightarrow \quad -5x^2 = \ln(1-u) \\ &\Leftrightarrow \quad x^2 = -0.2\ln(1-u) \quad \Leftrightarrow \quad x = \sqrt{-0.2\ln(1-u)}. \end{aligned}$$

The solution is $Z = \sqrt{-0.2 \ln U}$ (replacing $1-U$ by U, see Exercise 6.3). Note that Z^2 has an $Exp(5)$ distribution.

6.10 a Define random variables $B_i = 1$ if $U_i \leq p$ and $B_i = 0$ if $U_i > p$. Then $P(B_i = 1) = p$ and $P(B_i = 0) = 1 - p$: each B_i has a $Ber(p)$ distribution. If $B_1 = B_2 = \cdots = B_{k-1} = 0$ and $B_k = 1$, then $N = k$, i.e., N is the position in the sequence of Bernoulli random variables, where the first 1 occurs. This is a $Geo(p)$ distribution. This can be verified by computing the probability mass function: for $k \geq 1$,

$$\begin{aligned} P(N = k) &= P(B_1 = B_2 = \cdots = B_{k-1} = 0, B_k = 1) \\ &= P(B_1 = 0)\ P(B_2 = 0) \cdots P(B_{k-1} = 0)\ P(B_k = 1) \\ &= (1-p)^{k-1}\,p. \end{aligned}$$

6.10 b If Y is (a real number!) greater than n, then rounding *upwards* means we obtain $n+1$ or higher, so $\{Y > n\} = \{Z \geq n+1\} = \{Z > n\}$. Therefore, $P(Z > n) = P(Y > n) = e^{-\lambda n} = \big(e^{-\lambda}\big)^n$. From $\lambda = -\ln(1-p)$ we see: $e^{-\lambda} = 1 - p$, so the last probability is $(1-p)^n$. From $P(Z > n-1) = P(Z = n) + P(Z > n)$ we find: $P(Z = n) = P(Z > n-1) - P(Z > n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1}\,p$. Z has a $Geo(p)$ distribution.

6.12 We need to generate stock prices for the next five years, or 60 months. So we need sixty $U(0,1)$ random variables $U_1, \ldots, U_{60}$. Let S_i denote the stock price in month i, and set $S_0 = 100$, the initial stock price. From the U_i we obtain the stock movement, as follows, for $i = 1, 2, \ldots$:

$$S_i = \begin{cases} 0.95\,S_{i-1} & \text{if } U_i < 0.25, \\ S_{i-1} & \text{if } 0.25 \leq U_i \leq 0.75, \\ 1.05\,S_{i-1} & \text{if } U_i > 0.75. \end{cases}$$

We have carried this out, using the realizations below:

1–10:	0.72	0.03	0.01	0.81	0.97	0.31	0.76	0.70	0.71	0.25
11–20:	0.88	0.25	0.89	0.95	0.82	0.52	0.37	0.40	0.82	0.04
21–30:	0.38	0.88	0.81	0.09	0.36	0.93	0.00	0.14	0.74	0.48
31–40:	0.34	0.34	0.37	0.30	0.74	0.03	0.16	0.92	0.25	0.20
41–50:	0.37	0.24	0.09	0.69	0.91	0.04	0.81	0.95	0.29	0.47
51–60:	0.19	0.76	0.98	0.31	0.70	0.36	0.56	0.22	0.78	0.41

We do not list all the stock prices, just the ones that matter for our investment strategy (you can verify this). We first wait until the price drops below € 95, which happens at $S_4 = 94.76$. Our money has been in the bank for four months, so we own € $1000 \cdot 1.005^4 =$ € 1020.15, for which we can buy $1020.15/94.76 = 10.77$ shares. Next we wait until the price hits € 110, this happens at $S_{15} = 114.61$. We sell the our shares for € $10.77 \cdot 114.61 =$ € 1233.85, and put the money in the bank. At $S_{42} = 92.19$ we buy stock again, for the € $1233.85 \cdot 1.005^{27} =$ € 1411.71 that has accrued in the bank. We can buy 15.31 shares. For the rest of the five year period nothing happens, the final price is $S_{60} = 100.63$, which puts the value of our portfolio at € 1540.65.
For a real simulation the above should be repeated, say, one thousand times. The one thousand net results then give us an impression of the probability distribution that corresponds to this model and strategy.

7.6 Since f is increasing on the interval $[2, 3]$ we know from the interpretation of expectation as center of gravity that the expectation should lie closer to 3 than to 2. The computation: $\mathrm{E}[Z] = \int_2^3 \frac{3}{19} z^3 \,\mathrm{d}z = \left[\frac{3}{76} z^4\right]_2^3 = 2\frac{43}{76}$.

7.15 a We use the change-of-units rule for the expectation twice:

$$\begin{aligned}\mathrm{Var}(rX) &= \mathrm{E}\left[(rX - \mathrm{E}[rX]^2)\right] = \mathrm{E}\left[(rX - r\mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[r^2(X - \mathrm{E}[X])^2\right] = r^2\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = r^2\mathrm{Var}(X)\,.\end{aligned}$$

7.15 b Now we use the change-of-units rule for the expectation once:

$$\begin{aligned}\mathrm{Var}(X + s) &= \mathrm{E}\left[((X + s) - \mathrm{E}[X + s])^2\right] \\ &= \mathrm{E}\left[((X + s) - \mathrm{E}[X] + s)^2\right] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{Var}(X)\,.\end{aligned}$$

7.15 c With first **b**, and then **a**: $\mathrm{Var}(rX + s) = \mathrm{Var}(rX) = r^2\mathrm{Var}(X)$.

7.17 a Since $a_i \geq 0$ and $p_i \geq 0$ it must follow that $a_1p_1 + \cdots + a_rp_r \geq 0$. So $0 = \mathrm{E}[U] = a_1p_1 + \cdots + a_rp_r \geq 0$. As we may assume that all $p_i > 0$, it follows that $a_1 = a_2 = \cdots = a_r = 0$.

7.17 b Let $m = \mathrm{E}[V] = p_1b_1 + \cdots + p_rb_r$. Then the random variable $U = (V - \mathrm{E}[V])^2$ takes the values $a_1 = (b_1 - m)^2, \ldots, a_r = (b_r - m)^2$. Since $\mathrm{E}[U] = \mathrm{Var}(V) = 0$, part **a** tells us that $0 = a_1 = (b_1 - m)^2, \ldots, 0 = a_r = (b_r - m)^2$. But this is only possible if $b_1 = m, \ldots, b_r = m$. Since $m = \mathrm{E}[V]$, this is the same as saying that $\mathrm{P}(V = \mathrm{E}[V]) = 1$.

8.2 a First we determine the possible values that Y can take. Here these are $-1, 0$, and 1. Then we investigate which x-values lead to these y-values and sum the probabilities of the x-values to obtain the probability of the y-value. For instance,

$$\mathrm{P}(Y = 0) = \mathrm{P}(X = 2) + \mathrm{P}(X = 4) + \mathrm{P}(X = 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Similarly, we obtain for the two other values

$$\mathrm{P}(Y = -1) = \mathrm{P}(X = 3) = \frac{1}{6}, \qquad \mathrm{P}(Y = 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = 5) = \frac{1}{3}.$$

8.2 b The values taken by Z are -1, 0, and 1. Furthermore

$$\mathrm{P}(Z = 0) = \mathrm{P}(X = 1) + \mathrm{P}(X = 3) + \mathrm{P}(X = 5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

and similarly $\mathrm{P}(Z = -1) = 1/3$ and $\mathrm{P}(Z = 1) = 1/6$.

8.2 c Since for any α one has $\sin^2(\alpha)+\cos^2(\alpha)=1$, W can only take the value 1, so $\mathrm{P}(W=1)=1$.

8.10 Because of symmetry: $\mathrm{P}(X\geq 3)=0.500$. Furthermore: $\sigma^2=4$, so $\sigma=2$. Then $Z=(X-3)/2$ is an $N(0,1)$ distributed random variable, so that $\mathrm{P}(X\leq 1)=\mathrm{P}((X-3)/2)\leq(1-3)/2=\mathrm{P}(Z\leq -1)=\mathrm{P}(Z\geq 1)=0.1587$.

8.11 Since $-g$ is a convex function, Jensen's inequality yields that $-g(\mathrm{E}[X])\leq \mathrm{E}[-g(X)]$. Since $\mathrm{E}[-g(X)]=-\mathrm{E}[g(X)]$, the inequality follows by multiplying both sides by -1.

8.12 a The possible values Y can take are $\sqrt{0}=0$, $\sqrt{1}=1$, $\sqrt{100}=10$, and $\sqrt{10\,000}=100$. Hence the probability mass function is given by

y	0	1	10	100
$\mathrm{P}(Y=y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

8.12 b Compute the second derivative: $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sqrt{x}=-\frac{1}{4}x^{-3/2}<0$. Hence $g(x)=-\sqrt{x}$ is a convex function. Jensen's inequality yields that $\sqrt{\mathrm{E}[X]}\geq \mathrm{E}\left[\sqrt{X}\right]$.

8.12 c We obtain $\sqrt{\mathrm{E}[X]}=\sqrt{(0+1+100+10\,000)/4}=50.25$, but

$$\mathrm{E}\left[\sqrt{X}\right]=\mathrm{E}[Y]=(0+1+10+100)/4=27.75.$$

8.19 a This happens for all φ in the interval $[\pi/4,\pi/2]$, which corresponds to the upper right quarter of the circle.

8.19 b Since $\{Z\leq t\}=\{X\leq \arctan(t)\}$, we obtain

$$F_Z(t)=\mathrm{P}(Z\leq t)=\mathrm{P}(X\leq\arctan(t))=\frac{1}{2}+\frac{1}{\pi}\arctan(t).$$

8.19 c Differentiating F_Z we obtain that the probability density function of Z is

$$f_Z(z)=\frac{\mathrm{d}}{\mathrm{d}z}F_Z(z)=\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{2}+\frac{1}{\pi}\arctan(z)\right)=\frac{1}{\pi(1+z^2)}\quad\text{for } -\infty<z<\infty.$$

9.2 a From $\mathrm{P}(X=1,Y=1)=1/2$, $\mathrm{P}(X=1)=2/3$, and the fact that $\mathrm{P}(X=1)=\mathrm{P}(X=1,Y=1)+\mathrm{P}(X=1,Y=-1)$, it follows that $\mathrm{P}(X=1,Y=-1)=1/6$. Since $\mathrm{P}(Y=1)=1/2$ and $\mathrm{P}(X=1,Y=1)=1/2$, we must have: $\mathrm{P}(X=0,Y=1)$ and $\mathrm{P}(X=2,Y=1)$ are both zero. From this and the fact that $\mathrm{P}(X=0)=1/6=\mathrm{P}(X=2)$ one finds that $\mathrm{P}(X=0,Y=-1)=1/6=\mathrm{P}(X=2,Y=-1)$.

9.2 b Since, e.g., $\mathrm{P}(X=2,Y=1)=0$ is different from $\mathrm{P}(X=2)\,\mathrm{P}(Y=1)=\frac{1}{6}\cdot\frac{1}{2}$, one finds that X and Y are dependent.

9.8 a Since X can attain the values 0 and 1 and Y the values 0 and 2, Z can attain the values 0, 1, 2, and 3 with probabilities: $\mathrm{P}(Z=0)=\mathrm{P}(X=0,Y=0)=1/4$, $\mathrm{P}(Z=1)=\mathrm{P}(X=1,Y=0)=1/4$, $\mathrm{P}(Z=2)=\mathrm{P}(X=0,Y=2)=1/4$, and $\mathrm{P}(Z=3)=\mathrm{P}(X=1,Y=2)=1/4$.

9.8 b Since $\tilde{X}=\tilde{Z}-\tilde{Y}$, $\tilde{X}$ can attain the values -2, -1, 0, 1, 2, and 3 with probabilities

$$\begin{aligned}
\mathrm{P}\left(\tilde{X}=-2\right) &= \mathrm{P}\left(\tilde{Z}=0,\tilde{Y}=2\right)=1/8,\\
\mathrm{P}\left(\tilde{X}=-1\right) &= \mathrm{P}\left(\tilde{Z}=1,\tilde{Y}=2\right)=1/8,\\
\mathrm{P}\left(\tilde{X}=0\right) &= \mathrm{P}\left(\tilde{Z}=0,\tilde{Y}=0\right)+\mathrm{P}\left(\tilde{Z}=2,\tilde{Y}=2\right)=1/4,\\
\mathrm{P}\left(\tilde{X}=1\right) &= \mathrm{P}\left(\tilde{Z}=1,\tilde{Y}=0\right)+\mathrm{P}\left(\tilde{Z}=3,\tilde{Y}=2\right)=1/4,\\
\mathrm{P}\left(\tilde{X}=2\right) &= \mathrm{P}\left(\tilde{Z}=2,\tilde{Y}=0\right)=1/8,\\
\mathrm{P}\left(\tilde{X}=3\right) &= \mathrm{P}\left(\tilde{Z}=3,\tilde{Y}=0\right)=1/8.
\end{aligned}$$

We have the following table:

z	-2	-1	0	1	2	3
$p_{\tilde{X}}(z)$	$1/8$	$1/8$	$1/4$	$1/4$	$1/8$	$1/8$

9.9 a One has that $F_X(x) = \lim_{y\to\infty} F(x,y)$. So for $x \le 0$: $F_X(x) = 0$, and for $x > 0$: $F_X(x) = F(x,\infty) = 1 - \mathrm{e}^{-2x}$. Similarly, $F_Y(y) = 0$ for $y \le 0$, and for $y > 0$: $F_Y(y) = F(\infty, y) = 1 - \mathrm{e}^{-y}$.

9.9 b For $x > 0$ and $y > 0$: $f(x,y) = \frac{\partial^2}{\partial x\,\partial y}F(x,y) = \frac{\partial}{\partial x}\left(\mathrm{e}^{-y} - \mathrm{e}^{-(2x+y)}\right) = 2\mathrm{e}^{-(2x+y)}$.

9.9 c There are two ways to determine $f_X(x)$:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^{\infty} \mathrm{e}^{-(2x+y)}\mathrm{d}y = 2\mathrm{e}^{-2x} \quad \text{for } x > 0$$

and

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}F_X(x) = 2\mathrm{e}^{-2x} \quad \text{for } x > 0.$$

Using either way one finds that $f_Y(y) = \mathrm{e}^{-y}$ for $y > 0$.

9.9 d Since $F(x,y) = F_X(x)F_Y(y)$ for all x, y, we find that X and Y are independent.

9.11 To determine $\mathrm{P}(X < Y)$ we must integrate $f(x,y)$ over the region G of points (x,y) in $\mathbb{R}^2$ for which x is smaller than y:

$$\begin{aligned}
\mathrm{P}(X<Y) &= \iint_{\{(x,y)\in\mathbb{R}^2;\, x<y\}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y = \int_0^1\left(\int_0^y \frac{12}{5}xy(1+y)\,\mathrm{d}x\right)\mathrm{d}y\\
&= \frac{12}{5}\int_0^1 y(1+y)\left(\int_0^y x\,\mathrm{d}x\right)\mathrm{d}y = \frac{12}{10}\int_0^1 y^3(1+y)\,\mathrm{d}y = \frac{27}{50}.
\end{aligned}$$

Here we used that $f(x,y) = 0$ for (x,y) outside the unit square.

9.15 a Setting $\square(a,b)$ as the set of points (x,y), for which $x \le a$ and $y \le b$, we have that

$$F(a,b) = \frac{\text{area}\,(\Delta \cap \square(a,b))}{\text{area of}\,\Delta}.$$

- If $a < 0$ or if $b < 0$ (or both), then $\text{area}\,(\Delta \cap \square(a,b)) = \emptyset$, so $F(a,b) = 0$,

- If $(a,b) \in \Delta$, then area $(\Delta \cap \square(a,b)) = a(b - \frac{1}{2}a)$, so $F(a,b) = a(2b-a)$,
- If $0 \le b \le 1$, and $a > b$, then area $(\Delta \cap \square(a,b)) = \frac{1}{2}b^2$, so $F(a,b) = b^2$,
- If $0 \le a \le 1$, and $b > 1$, then area $(\Delta \cap \square(a,b)) = a - \frac{1}{2}a^2$, so $F(a,b) = 2a - a^2$,
- If both $a > 1$ and $b > 1$, then area $(\Delta \cap \square(a,b)) = \frac{1}{2}$, so $F(a,b) = 1$.

9.15 b Since $f(x,y) = \frac{\partial^2}{\partial x\,\partial y}F(x,y)$, we find for $(x,y) \in \Delta$ that $f(x,y) = 2$. Furthermore, $f(x,y) = 0$ for (x,y) outside the triangle Δ.

9.15 c For x between 0 and 1,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_x^1 2\,\mathrm{d}y = 2(1-x).$$

For y between 0 and 1,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_0^y 2\,\mathrm{d}x = 2y.$$

10.6 a When $c = 0$, the joint distribution becomes

	a			
b	-1	0	1	$\mathrm{P}(Y=b)$
-1	2/45	9/45	4/45	1/3
0	7/45	5/45	3/45	1/3
1	6/45	1/45	8/45	1/3
$\mathrm{P}(X=a)$	1/3	1/3	1/3	1

We find $\mathrm{E}[X] = (-1)\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0$, and similarly $\mathrm{E}[Y] = 0$. By leaving out terms where either $X = 0$ or $Y = 0$, we find

$$\mathrm{E}[XY] = (-1)\cdot(-1)\cdot\frac{2}{45} + (-1)\cdot 1\cdot\frac{4}{45} + 1\cdot(-1)\cdot\frac{6}{45} + 1\cdot 1\cdot\frac{8}{45} = 0,$$

which implies that $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = 0$.

10.6 b Note that the variables X and Y in part **b** are equal to the ones from part **a**, shifted by c. If we write U and V for the variables from **a**, then $X = U + c$ and $Y = V + c$. According to the rule on the covariance under change of units, we then immediately find $\mathrm{Cov}(X,Y) = \mathrm{Cov}(U+c, V+c) = \mathrm{Cov}(U,V) = 0$.
Alternatively, one could also compute the covariance from $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$. We find $\mathrm{E}[X] = (c-1)\cdot\frac{1}{3} + c\cdot\frac{1}{3} + (c+1)\cdot\frac{1}{3} = c$, and similarly $\mathrm{E}[Y] = c$. Since

$$\begin{aligned}\mathrm{E}[XY] = {} & (c-1)\cdot(c-1)\cdot\frac{2}{45} + (c-1)\cdot c\cdot\frac{9}{45} + (c+1)\cdot(c+1)\cdot\frac{4}{45}\\ & + c\cdot(c-1)\cdot\frac{7}{45} + c\cdot c\cdot\frac{5}{45} + c\cdot(c+1)\cdot\frac{3}{45}\\ & + (c+1)\cdot(c-1)\cdot\frac{6}{45} + (c+1)\cdot c\cdot\frac{1}{45} + (c+1)\cdot(c+1)\cdot\frac{8}{45} = c^2,\end{aligned}$$

we find $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] = c^2 - c\cdot c = 0$.

10.6 c No, X and Y are not independent. For instance, $\mathrm{P}(X = c, Y = c+1) = 1/45$, which differs from $\mathrm{P}(X = c)\ \mathrm{P}(Y = c+1) = 1/9$.

10.9 a If the aggregated blood sample tests negative, we do not have to perform additional tests, so that X_i takes on the value 1. If the aggregated blood sample tests positive, we have to perform 40 additional tests for the blood sample of each person in the group, so that X_i takes on the value 41. We first find that $\mathrm{P}(X_i = 1) = \mathrm{P}(\text{no infections in group of } 40) = (1 - 0.001)^{40} = 0.96$, and therefore $\mathrm{P}(X_i = 41) = 1 - \mathrm{P}(X_i = 1) = 0.04$.

10.9 b First compute $\mathrm{E}[X_i] = 1\cdot 0.96 + 41\cdot 0.04 = 2.6$. The expected total number of tests is $\mathrm{E}[X_1 + X_2 + \cdots + X_{25}] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_{25}] = 25\cdot 2.6 = 65$. With the original procedure of blood testing, the total number of tests is $25\cdot 40 = 1000$. On average the alternative procedure would only require 65 tests. Only with very small probability one would end up with doing more than 1000 tests, so the alternative procedure is better.

10.10 a We find

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_0^3 \frac{2}{225}\left(9x^3 + 7x^2\right)\mathrm{d}x = \frac{2}{225}\left[\frac{9}{4}x^4 + \frac{7}{3}x^3\right]_0^3 = \frac{109}{50},$$

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y = \int_1^2 \frac{1}{25}(3y^3 + 12y^2)\,\mathrm{d}y = \frac{1}{25}\left[\frac{3}{4}y^4 + 4y^3\right]_1^2 = \frac{157}{100},$$

so that $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 15/4$.

10.10 b We find

$$\mathrm{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)\,\mathrm{d}x = \int_0^3 \frac{2}{225}\left(9x^4 + 7x^3\right)\mathrm{d}x = \frac{2}{225}\left[\frac{9}{5}x^5 + \frac{7}{4}x^4\right]_0^3 = \frac{1287}{250},$$

$$\mathrm{E}\left[Y^2\right] = \int_{-\infty}^{\infty} y^2 f_Y(y)\,\mathrm{d}y = \int_1^2 \frac{1}{25}(3y^4 + 12y^3)\,\mathrm{d}y = \frac{1}{25}\left[\frac{3}{5}y^5 + 3y^4\right]_1^2 = \frac{318}{125},$$

$$\begin{aligned}
\mathrm{E}[XY] &= \int_0^3\int_1^2 xy f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^3\int_1^2 \frac{2}{75}\left(2x^3y^2 + x^2y^3\right)\mathrm{d}y\,\mathrm{d}x \\
&= \frac{4}{75}\int_0^3 x^3\left(\int_1^2 y^2\,\mathrm{d}y\right)\mathrm{d}x + \frac{2}{75}\int_0^3 x^2\left(\int_1^2 y^3\,\mathrm{d}y\right)\mathrm{d}x \\
&= \frac{4}{75}\frac{7}{3}\int_0^3 x^3\,\mathrm{d}x + \frac{2}{75}\frac{15}{4}\int_0^3 x^2\,\mathrm{d}x = \frac{171}{50},
\end{aligned}$$

so that $\mathrm{E}\left[(X+Y)^2\right] = \mathrm{E}\left[X^2\right] + \mathrm{E}\left[Y^2\right] + 2\mathrm{E}[XY] = 3633/250$.

10.10 c We find

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \frac{1287}{250} - \left(\frac{109}{50}\right)^2 = \frac{989}{2500},$$

$$\mathrm{Var}(Y) = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = \frac{318}{125} - \left(\frac{157}{100}\right)^2 = \frac{791}{10\,000},$$

$$\mathrm{Var}(X+Y) = \mathrm{E}\left[(X+Y)^2\right] - (\mathrm{E}[X+Y])^2 = \frac{3633}{250} - \left(\frac{15}{4}\right)^2 = \frac{939}{2000}.$$

Hence, $\mathrm{Var}(X) + \mathrm{Var}(Y) = 0.4747$, which differs from $\mathrm{Var}(X+Y) = 0.4695$.

10.14 a By using the alternative expression for the covariance and linearity of expectations, we find

$$\begin{aligned}
&\mathrm{Cov}(X+s, Y+u) \\
&\quad = \mathrm{E}\left[(X+s)(Y+u)\right] - \mathrm{E}\left[X+s\right]\mathrm{E}\left[Y+u\right] \\
&\quad = \mathrm{E}\left[XY + sY + uX + su\right] - (\mathrm{E}\left[X\right] + s)(\mathrm{E}\left[Y\right] + u) \\
&\quad = (\mathrm{E}\left[XY\right] + s\mathrm{E}\left[Y\right] + u\mathrm{E}\left[X\right] + su) - (\mathrm{E}\left[X\right]\mathrm{E}\left[Y\right] + s\mathrm{E}\left[Y\right] + u\mathrm{E}\left[X\right] + su) \\
&\quad = \mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right] \\
&\quad = \mathrm{Cov}(X, Y)\,.
\end{aligned}$$

10.14 b By using the alternative expression for the covariance and the rule on expectations under change of units, we find

$$\begin{aligned}
\mathrm{Cov}(rX, tY) &= \mathrm{E}\left[(rX)(tY)\right] - \mathrm{E}\left[rX\right]\mathrm{E}\left[tY\right] \\
&= \mathrm{E}\left[rtXY\right] - (r\mathrm{E}\left[X\right])(t\mathrm{E}\left[Y\right]) \\
&= rt\mathrm{E}\left[XY\right] - rt\mathrm{E}\left[X\right]\mathrm{E}\left[Y\right] \\
&= rt\left(\mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]\right) \\
&= rt\mathrm{Cov}(X, Y)\,.
\end{aligned}$$

10.14 c First applying part **a** and then part **b** yields

$$\mathrm{Cov}(rX+s, tY+u) = \mathrm{Cov}(rX, tY) = rt\mathrm{Cov}(X, Y)\,.$$

10.18 First note that $X_1 + X_2 + \cdots + X_N$ is the sum of all numbers, which is a nonrandom constant. Therefore, $\mathrm{Var}(X_1 + X_2 + \cdots + X_N) = 0$. In Section 9.3 we argued that, although we draw without replacement, each X_i has the same distribution. By the same reasoning, we find that each pair (X_i, X_j), with $i \neq j$, has the same joint distribution, so that $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_1, X_2)$ for all pairs with $i \neq j$. Direct application of Exercise 10.17 with $\sigma^2 = (N-1)(N+1)$ and $\gamma = \mathrm{Cov}(X_1, X_2)$ gives

$$0 = \mathrm{Var}(X_1 + X_2 + \cdots + X_N) = N \cdot \frac{(N-1)(N+1)}{12} + N(N-1)\mathrm{Cov}(X_1, X_2)\,.$$

Solving this identity gives $\mathrm{Cov}(X_1, X_2) = -(N+1)/12$.

11.2 a By using the rule on addition of two independent discrete random variables, we have

$$\mathrm{P}(X+Y=k) = p_Z(k) = \sum_{\ell=0}^{\infty} p_X(k-\ell)p_Y(\ell).$$

Because $p_X(a) = 0$ for $a \leq -1$, all terms with $\ell \geq k+1$ vanish, so that

$$\mathrm{P}(X+Y=k) = \sum_{\ell=0}^{k} \frac{1^{k-\ell}}{(k-\ell)!}\mathrm{e}^{-1} \cdot \frac{1^{\ell}}{\ell!}\mathrm{e}^{-1} = \frac{\mathrm{e}^{-2}}{k!}\sum_{\ell=0}^{k}\binom{k}{\ell} = \frac{2^k}{k!}\mathrm{e}^{-2},$$

also using $\sum_{\ell=0}^{k}\binom{k}{\ell} = 2^k$ in the last equality.

11.2 b Similar to part **a**, by using the rule on addition of two independent discrete random variables and leaving out terms for which $p_X(a) = 0$, we have

$$\mathrm{P}(X+Y=k) = \sum_{\ell=0}^{k} \frac{\lambda^{k-\ell}}{(k-\ell)!}\mathrm{e}^{-\lambda} \cdot \frac{\mu^\ell}{\ell!}\mathrm{e}^{-\mu} = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)} \sum_{\ell=0}^{k} \binom{k}{\ell} \frac{\lambda^{k-\ell}\mu^\ell}{(\lambda+\mu)^k}.$$

Next, write

$$\frac{\lambda^{k-\ell}\mu^\ell}{(\lambda+\mu)^k} = \left(\frac{\mu}{\lambda+\mu}\right)^\ell \left(\frac{\lambda}{\lambda+\mu}\right)^{k-\ell} = \left(\frac{\mu}{\lambda+\mu}\right)^\ell \left(1-\frac{\mu}{\lambda+\mu}\right)^{k-\ell} = p^\ell(1-p)^{k-\ell}$$

with $p = \mu/(\lambda+\mu)$. This means that

$$\mathrm{P}(X+Y=k) = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)} \sum_{\ell=0}^{k} \binom{k}{\ell} p^\ell(1-p)^{k-\ell} = \frac{(\lambda+\mu)^k}{k!}\mathrm{e}^{-(\lambda+\mu)},$$

using that $\sum_{\ell=0}^{k}\binom{k}{\ell}p^\ell(1-p)^{k-\ell} = 1$.

11.4 a From the fact that X has an $N(2,5)$ distribution, it follows that $\mathrm{E}[X] = 2$ and $\mathrm{Var}(X) = 5$. Similarly, $\mathrm{E}[Y] = 5$ and $\mathrm{Var}(Y) = 9$. Hence by linearity of expectations,

$$\mathrm{E}[Z] = \mathrm{E}[3X-2Y+1] = 3\mathrm{E}[X] - 2\mathrm{E}[Y] + 1 = 3\cdot 2 - 2\cdot 5 + 1 = -3.$$

By the rules for the variance and covariance,

$$\mathrm{Var}(Z) = 9\mathrm{Var}(X) + 4\mathrm{Var}(Y) - 12\mathrm{Cov}(X,Y) = 9\cdot 5 + 4\cdot 9 - 12\cdot 0 = 81,$$

using that $\mathrm{Cov}(X,Y) = 0$, due to independence of X and Y.

11.4 b The random variables $3X$ and $-2Y+1$ are independent and, according to the rule for the normal distribution under a change of units (page 106), it follows that they both have a normal distribution. Next, the sum rule for independent normal random variables then yields that $Z = (3X) + (-2Y+1)$ also has a normal distribution. Its parameters are the expectation and variance of Z. From **a** it follows that Z has an $N(-3,81)$ distribution.

11.4 c From **b** we know that Z has an $N(-3,81)$ distribution, so that $(Z+3)/9$ has a standard normal distribution. Therefore

$$\mathrm{P}(Z \le 6) = \mathrm{P}\left(\frac{Z+3}{9} \le \frac{6+3}{9}\right) = \Phi(1),$$

where Φ is the standard normal distribution function. From Table B.1 we find that $\Phi(1) = 1 - 0.1587 = 0.8413$.

11.9 a According to the product rule on page 160,

$$\begin{aligned} f_Z(z) &= \int_1^z f_Y\left(\frac{z}{x}\right) f_X(x) \frac{1}{x}\,\mathrm{d}x = \int_1^z \frac{1}{\left(\frac{z}{x}\right)^2} \frac{3}{x^4} \frac{1}{x}\,\mathrm{d}x \\ &= \frac{3}{z^2}\int_1^z \frac{1}{x^3}\,\mathrm{d}x = \frac{3}{z^2}\left[-\frac{1}{2}x^{-2}\right]_1^z = \frac{3}{2}\frac{1}{z^2}\left(1-\frac{1}{z^2}\right) \\ &= \frac{3}{2}\left(\frac{1}{z^2} - \frac{1}{z^4}\right). \end{aligned}$$

11.9 b According to the product rule,

$$f_Z(z) = \int_1^z f_Y\left(\frac{z}{x}\right) f_X(x) \frac{1}{x}\,\mathrm{d}x = \int_1^z \frac{\beta}{\left(\frac{z}{x}\right)^{\beta+1}} \frac{\alpha}{x^{\alpha+1}} \frac{1}{x}\,\mathrm{d}x$$
$$= \frac{\alpha\beta}{z^{\beta+1}} \int_1^z x^{\beta-\alpha-1}\,\mathrm{d}x = \frac{\alpha\beta}{z^{\beta+1}} \left[\frac{x^{\beta-\alpha}}{\beta-\alpha}\right]_1^z = \frac{\alpha\beta}{\alpha-\beta} \frac{1}{z^{\beta+1}} \left(1 - z^{\beta-\alpha}\right)$$
$$= \frac{\alpha\beta}{\beta-\alpha} \left(\frac{1}{z^{\beta+1}} - \frac{1}{z^{\alpha+1}}\right).$$

12.1 e This is certainly open to discussion. Bankruptcies: no (they come in clusters, don't they?). Eggs: no (I suppose after one egg it takes the chicken some time to produce another). Examples 3 and 4 are the best candidates. Example 5 could be modeled by the Poisson process if the crossing is not a dangerous one; otherwise authorities might take measures and destroy the homogeneity.

12.6 The expected numbers of flaws in 1 meter is $100/40 = 2.5$, and hence the number of flaws X has a $Pois(2.5)$ distribution. The answer is $\mathrm{P}(X = 2) = \frac{1}{2!}(2.5)^2\mathrm{e}^{-2.5} = 0.256$.

12.7 a It is reasonable to estimate λ with (nr. of cars)/(total time in sec.) = 0.192.

12.7 b $19/120 = 0.1583$, and if $\lambda = 0.192$ then $\mathrm{P}(N(10) = 0) = \mathrm{e}^{-0.192\cdot 10} = 0.147$.

12.7 c $\mathrm{P}(N(10) = 10)$ with λ from **a** seems a reasonable approximation of this probability. It equals $\mathrm{e}^{-1.92} \cdot (0.192 \cdot 10)^{10}/10! = 2.71 \cdot 10^{-5}$.

12.11 Following the hint, we obtain:

$$\begin{aligned} \mathrm{P}(N([0,s] = k, N([0,2s]) = n) &= \mathrm{P}(N([0,s]) = k, N((s,2s]) = n-k) \\ &= \mathrm{P}(N([0,s]) = k) \cdot \mathrm{P}(N((s,2s]) = n-k) \\ &= (\lambda s)^k \mathrm{e}^{-\lambda s}/(k!) \cdot (\lambda s)^{n-k} \mathrm{e}^{-\lambda s}/((n-k)!) \\ &= (\lambda s)^n \mathrm{e}^{-\lambda 2s}/(k!(n-k)!). \end{aligned}$$

So

$$\begin{aligned} \mathrm{P}(N([0,s]) = k \,|\, N([0,2s]) = n) &= \frac{\mathrm{P}(N([0,s]) = k, N([0,2s]) = n)}{\mathrm{P}(N([0,2s]) = n)} \\ &= n!/(k!(n-k)!) \cdot (\lambda s)^n/(2\lambda s)^n \\ &= n!/(k!(n-k)!) \cdot (1/2)^n. \end{aligned}$$

This holds for $k = 0, \ldots, n$, so we find the $Bin(n, \frac{1}{2})$ distribution.

13.2 a From the formulas for the $U(a, b)$ distribution, substituting $a = -1/2$ and $b = 1/2$, we derive that $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = 1/12$.

13.2 b We write $S = X_1 + X_2 + \cdots + X_{100}$, for which we find $\mathrm{E}[S] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_{100}] = 0$ and, by independence, $\mathrm{Var}(S) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{100}) = 100 \cdot \frac{1}{12} = 100/12$. We find from Chebyshev's inequality:

$$\mathrm{P}(|S| > 10) = \mathrm{P}(|S - 0| > 10) \le \frac{\mathrm{Var}(S)}{10^2} = \frac{1}{12}.$$

13.4 a Because X_i has a $Ber(p)$ distribution, $\mathrm{E}[X_i] = p$ and $\mathrm{Var}(X_i) = p(1-p)$, and so $\mathrm{E}\left[\bar{X}_n\right] = p$ and $\mathrm{Var}\left(\bar{X}_n\right) = \mathrm{Var}(X_i)/n = p(1-p)/n$. By Chebyshev's inequality:

$$\mathrm{P}\left(|\bar{X}_n - p| \geq 0.2\right) \leq \frac{p(1-p)/n}{(0.2)^2} = \frac{25p(1-p)}{n}.$$

The right-hand side should be at most 0.1 (note that we switched to the complement). If $p = 1/2$ we therefore require $25/(4n) \leq 0.1$, or $n \geq 25/(4 \cdot 0.1) = 62.5$, i.e., $n \geq 63$. Now, suppose $p \neq 1/2$, using $n = 63$ and $p(1-p) \leq 1/4$ we conclude that $25p(1-p)/n \leq 25 \cdot (1/4)/63 = 0.0992 < 0.1$, so (because of the inequality) the computed value satisfies for other values of p as well.

13.4 b For arbitrary $a > 0$ we conclude from Chebyshev's inequality:

$$\mathrm{P}\left(|\bar{X}_n - p| \geq a\right) \leq \frac{p(1-p)/n}{a^2} = \frac{p(1-p)}{na^2} \leq \frac{1}{4na^2},$$

where we used $p(1-p) \leq 1/4$ again. The question now becomes: when $a = 0.1$, for what n is $1/(4na^2) \leq 0.1$? We find: $n \geq 1/(4 \cdot 0.1 \cdot (0.1)^2) = 250$, so $n = 250$ is large enough.

13.4 c From part **a** we know that an error of size 0.2 or occur with a probability of at most $25/4n$, regardless of the values of p. So, we need $25/(4n) \leq 0.05$, i.e., $n \geq 25/(4 \cdot 0.05) = 125$.

13.4 d We compute $\mathrm{P}\left(\bar{X}_n \leq 0.5\right)$ for the case that $p = 0.6$. Then $\mathrm{E}\left[\bar{X}_n\right] = 0.6$ and $\mathrm{Var}\left(\bar{X}_n\right) = 0.6 \cdot 0.4/n$. Chebyshev's inequality cannot be used directly, we need an intermediate step: the probability that $\bar{X}_n \leq 0.5$ is contained in the event "the prediction is off by at least 0.1, in either direction." So

$$\mathrm{P}\left(\bar{X}_n \leq 0.5\right) \leq \mathrm{P}\left(|\bar{X}_n - 0.6| \geq 0.1\right) \leq \frac{0.6 \cdot 0.4/n}{(0.1)^2} = \frac{24}{n}$$

For $n \geq 240$ this probability is 0.1 or smaller.

13.9 a The statement looks like the law of large numbers, and indeed, if we look more closely, we see that T_n is the average of an i.i.d. sequence: define $Y_i = X_i^2$, then $T_n = \bar{Y}_n$. The law of large numbers now states: if $\bar{Y}_n$ is the average of n independent random variables with expectation μ and variance σ^2, then for any $\varepsilon > 0$: $\lim_{n\to\infty} \mathrm{P}\left(|\bar{Y}_n - \mu| > \varepsilon\right) = 0$. So, if $a = \mu$ and the variance σ^2 is finite, then it is true.

13.9 b We compute expectation and variance of Y_i: $\mathrm{E}[Y_i] = \mathrm{E}\left[X_i^2\right] = \int_{-1}^{1} \frac{1}{2}x^2 \,\mathrm{d}x = 1/3$. And: $\mathrm{E}\left[Y_i^2\right] = \mathrm{E}\left[X_i^4\right] = \int_{-1}^{1} \frac{1}{2}x^4 \,\mathrm{d}x = 1/5$, so $\mathrm{Var}(Y_i) = 1/5 - (1/3)^2 = 4/45$. The variance is finite, so indeed, the law of large numbers applies, and the statement is true if $a = \mathrm{E}\left[X_i^2\right] = 1/3$.

14.3 First note that $\mathrm{P}\left(|\bar{X}_n - p| < 0.2\right) = 1 - \mathrm{P}\left(\bar{X}_n - p \geq 0.2\right) - \mathrm{P}\left(\bar{X}_n - p \leq -0.2\right)$. Because $\mu = p$ and $\sigma^2 = p(1-p)$, we find, using the central limit theorem:

$$\begin{aligned}
\mathrm{P}\left(\bar{X}_n - p \geq 0.2\right) &= \mathrm{P}\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \\
&= \mathrm{P}\left(Z_n \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \approx \mathrm{P}\left(Z \geq \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right),
\end{aligned}$$

where Z has an $N(0,1)$ distribution. Similarly,

$$\mathrm{P}\big(\bar{X}_n - p \le -0.2\big) \approx \mathrm{P}\left(Z \ge \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right),$$

so we are looking for the smallest positive integer n such that

$$1 - 2\mathrm{P}\left(Z \ge \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \ge 0.9,$$

i.e., the smallest positive integer n such that

$$\mathrm{P}\left(Z \ge \sqrt{n}\frac{0.2}{\sqrt{p(1-p)}}\right) \le 0.05.$$

From Table B.1 it follows that

$$\sqrt{n}\frac{0.2}{\sqrt{p(1-p)}} \ge 1.645.$$

Since $p(1-p) \le 1/4$ for all p between 0 and 1, we see that n should be at least 17.

14.5 In Section 4.3 we have seen that X has the same probability distribution as $X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent $Ber(p)$ distributed random variables. Recall that $\mathrm{E}[X_i] = p$, and $\mathrm{Var}(X_i) = p(1-p)$. But then we have for any real number a that

$$\mathrm{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \le a\right) = \mathrm{P}\left(\frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} \le a\right) = \mathrm{P}(Z_n \le a)\,;$$

see also (14.1). It follows from the central limit theorem that

$$\mathrm{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \le a\right) \approx \Phi(a),$$

i.e., the random variable $\frac{X-np}{\sqrt{np(1-p)}}$ has a distribution that is approximately standard normal.

14.9 a The probability that for a chain of at least 50 meters more than 1002 links are needed is the same as the probability that a chain of 1002 chains is shorter than 50 meters. Assuming that the random variables $X_1, X_2, \ldots, X_{1002}$ are independent, and using the central limit theorem, we have that

$$\mathrm{P}(X_1 + X_2 + \cdots + X_{1002} < 5000) \approx \mathrm{P}\left(Z < \sqrt{1002} \cdot \frac{\frac{5000}{1002} - 5}{\sqrt{0.04}}\right) = 0.0571,$$

where Z has an $N(0,1)$ distribution. So about 6% of the customers will receive a free chain.

14.9 b We now have that

$$\mathrm{P}(X_1 + X_2 + \cdots + X_{1002} < 5000) \approx \mathrm{P}(Z < 0.0032)\,,$$

which is slightly larger than 1/2. So about half of the customers will receive a free chain. Clearly something has to be done: a seemingly minor change of expected value has major consequences!

15.6 Because $(2-0)\cdot 0.245+(4-2)\cdot 0.130+(7-4)\cdot 0.050+(11-7)\cdot 0.020+(15-11)\cdot 0.005=1$, there are no data points outside the listed bins. Hence

$$\begin{aligned}F_n(7) &= \frac{\text{number of } x_i \leq 7}{n}\\ &= \frac{\text{number of } x_i \text{ in bins } (0,2],\ (2,4] \text{ and } (4,7]}{n}\\ &= \frac{n\cdot(2-0)\cdot 0.245+n\cdot(4-2)\cdot 0.130+n\cdot(7-4)\cdot 0.050}{n}\\ &= 0.490+0.260+0.150=0.9.\end{aligned}$$

15.11 The height of the histogram on a bin $(a,b]$ is

$$\begin{aligned}\frac{\text{number of } x_i \text{ in } (a,b]}{n(b-a)} &= \frac{(\text{number of } x_i \leq b)-(\text{number of } x_i \leq a)}{n(b-a)}\\ &= \frac{F_n(b)-F_n(a)}{b-a}.\end{aligned}$$

15.12 a By inserting the expression for $f_{n,h}(t)$, we get

$$\begin{aligned}\int_{-\infty}^{\infty} t\cdot f_{n,h}(t)\,\mathrm{d}t &= \int_{-\infty}^{\infty} t\cdot \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t.\end{aligned}$$

For each i fixed we find with change of integration variables $u=(t-x_i)/h$,

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t &= \int_{-\infty}^{\infty}(x_i+hu)K(u)\,\mathrm{d}u\\ &= x_i\int_{-\infty}^{\infty}K(u)\,\mathrm{d}u+h\int_{-\infty}^{\infty}uK(u)\,\mathrm{d}u=x_i,\end{aligned}$$

using that K integrates to one and that $\int_{-\infty}^{\infty} uK(u)\,\mathrm{d}u=0$, because K is symmetric. Hence

$$\int_{-\infty}^{\infty} t\cdot f_{n,h}(t)\,\mathrm{d}t=\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

15.12 b By means of similar reasoning

$$\begin{aligned}\int_{-\infty}^{\infty} t^2\cdot f_{n,h}(t)\,\mathrm{d}t &= \int_{-\infty}^{\infty} t^2\cdot \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{t^2}{h}K\left(\frac{t-x_i}{h}\right)\,\mathrm{d}t.\end{aligned}$$

For each i:

$$\int_{-\infty}^{\infty} \frac{t^2}{h} K\left(\frac{t-x_i}{h}\right) \mathrm{d}t$$
$$= \int_{-\infty}^{\infty} (x_i + hu)^2 K(u)\, \mathrm{d}u = \int_{-\infty}^{\infty} (x_i^2 + 2x_i hu + h^2u^2) K(u)\, \mathrm{d}u$$
$$= x_i^2 \int_{-\infty}^{\infty} K(u)\, \mathrm{d}u + 2x_i h \int_{-\infty}^{\infty} uK(u)\, \mathrm{d}u + h^2 \int_{-\infty}^{\infty} u^2 K(u)\, \mathrm{d}u$$
$$= x_i^2 + h^2 \int_{-\infty}^{\infty} u^2 K(u)\, \mathrm{d}u,$$

again using that K integrates to one and that K is symmetric.

16.3 a Because $n = 24$, the sample median is the average of the 12th and 13th elements. Since these are both equal to 70, the sample median is also 70. The lower quartile is the pth empirical quantile for $p = 1/4$. We get $k = \lfloor p(n+1) \rfloor = 6$, so that

$$q_n(0.25) = x_{(6)} + 0.25 \cdot (x_{(7)} - x_{(6)}) = 66 + 0.25 \cdot (67 - 66) = 66.25.$$

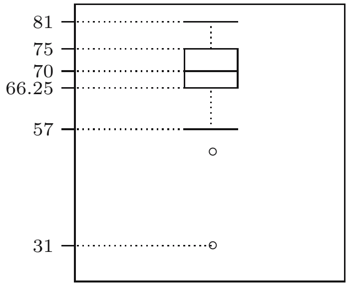
Similarly, the upper quartile is the pth empirical quantile for $p = 3/4$:

$$q_n(0.75) = x_{(18)} + 0.75 \cdot (x_{(19)} - x_{(18)}) = 75 + 0.75 \cdot (75 - 75) = 75.$$

16.3 b In part **a** we found the sample median and the two quartiles. From this we compute the IQR: $q_n(0.75) - q_n(0.25) = 75 - 66.25 = 8.75$. This means that

$$q_n(0.25) - 1.5 \cdot \text{IQR} = 66.25 - 1.5 \cdot 8.75 = 53.125,$$
$$q_n(0.75) + 1.5 \cdot \text{IQR} = 75 + 1.5 \cdot 8.75 = 88.125.$$

Hence, the last element below 88.125 is 88, and the first element above 53.125 is 57. Therefore, the upper whisker runs until 88 and the lower whisker until 57, with two elements 53 and 31 below. This leads to the following boxplot:



16.3 c The values 53 and 31 are outliers. Value 31 is far away from the bulk of the data and appears to be an *extreme* outlier.

16.6 a Yes, we find $\bar{x} = (1+5+9)/3 = 15/3 = 5$, $\bar{y} = (2+4+6+8)/4 = 20/4 = 5$, so that $(\bar{x} + \bar{y})/2 = 5$. The average for the combined dataset is also equal to 5: $(15 + 20)/7 = 5$.

16.6 b The mean of $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m$ equals

$$\frac{x_1 + \cdots + x_n + y_1 + \cdots + y_m}{n+m} = \frac{n\bar{x}_n + m\bar{y}_m}{n+m} = \frac{n}{n+m}\,\bar{x}_n + \frac{m}{n+m}\,\bar{y}_m.$$

In general, this is not equal to $(\bar{x}_n+\bar{y}_m)/2$. For instance, replace 1 in the first dataset by 4. Then $\bar{x}_n = 6$ and $\bar{y}_m = 5$, so that $(\bar{x}_n + \bar{y}_m)/2 = 5\frac{1}{2}$. However, the average of the combined dataset is $38/7 = 5\frac{2}{7}$.

16.6 c Yes, $m = n$ implies $n/(n+m) = m/(n+m) = 1/2$. From the expressions found in part **b** we see that the sample mean of the combined dataset equals $(\bar{x}_n + \bar{y}_m)/2$.

16.8 The ordered combined dataset is 1, 2, 4, 5, 6, 8, 9, so that the sample median equals 5. The absolute deviations from 5 are: 4, 3, 1, 0, 1, 3, 4, and if we put them in order: 0, 1, 1, 3, 3, 4, 4. The MAD is the sample median of the absolute deviations, which is 3.

16.15 First write

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2\right) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}_n^2.$$

Next, by inserting

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}_n \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\bar{x}_n^2 = \frac{1}{n}\cdot n \cdot \bar{x}_n^2 = \bar{x}_n^2,$$

we find

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n^2 + \bar{x}_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}_n^2.$$

17.3 a The model distribution corresponds to the number of women in a queue. A queue has 10 positions. The occurrence of a woman in any position is independent of the occurrence of a woman in other positions. At each position a woman occurs with probability p. Counting the occurrence of a woman as a "success," the number of women in a queue corresponds to the number of successes in 10 independent experiments with probability p of success and is therefore modeled by a $Bin(10, p)$ distribution.

17.3 b We have 100 queues and the number of women x_i in the ith queue is a realization of a $Bin(10, p)$ random variable. Hence, according to Table 17.2, the average number of women $\bar{x}_{100}$ resembles the expectation $10p$ of the $Bin(10, p)$ distribution. We find $\bar{x}_{100} = 435/100 = 4.35$, so an estimate for p is $4.35/10 = 0.435$.

17.7 a If we model the series of disasters by a Poisson process, then as a property of the Poisson process, the interdisaster times should follow an exponential distribution (see Section 12.3). This is indeed confirmed by the histogram and empirical distribution of the observed interdisaster times; they resemble the probability density and distribution function of an exponential distribution.

17.7 b The average length of a time interval is $40\,549/190 = 213.4$ days. Following Table 17.2 this should resemble the expectation of the $Exp(\lambda)$ distribution, which is $1/\lambda$. Hence, as an estimate for λ we could take $190/40\,549 = 0.00469$.

17.9 a A (perfect) cylindrical cone with diameter d (at the base) and height h has volume $\pi d^2 h/12$, or about $0.26 d^2 h$. The effective wood of a tree is the trunk without the branches. Since the trunk is similar to a cylindrical cone, one can expect a linear relation between the effective wood and $d^2 h$.

17.9 b We find

$$\bar{z}_n = \frac{\sum y_i/x_i}{n} = \frac{9.369}{31} = 0.3022$$
$$\bar{y}/\bar{x} = \frac{(\sum y_i)/n}{(\sum x_i)/n} = \frac{26.486/31}{87.456/31} = 0.3028$$
$$\text{least squares} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{95.498}{314.644} = 0.3035.$$

18.3 a Note that generating from the empirical distribution function is the same as choosing one of the elements of the original dataset with equal probability. Hence, an element in the bootstrap dataset equals 0.35 with probability 0.1. The number of ways to have exactly three out of ten elements equal to 0.35 is $\binom{10}{3}$, and each has probability $(0.1)^3(0.9)^7$. Therefore, the probability that the bootstrap dataset has exactly three elements equal to 0.35 is equal to $\binom{10}{3}(0.1)^3(0.9)^7 = 0.0574$.

18.3 b Having at most two elements less than or equal to 0.38 means that 0, 1, or 2 elements are less than or equal to 0.38. Five elements of the original dataset are smaller than or equal to 0.38, so that an element in the bootstrap dataset is less than or equal to 0.38 with probability 0.5. Hence, the probability that the bootstrap dataset has at most two elements less than or equal to 0.38 is equal to $(0.5)^{10} + \binom{10}{1}(0.5)^{10} + \binom{10}{2}(0.5)^{10} = 0.0547$.

18.3 c Five elements of the dataset are smaller than or equal to 0.38 and two are greater than 0.42. Therefore, obtaining a bootstrap dataset with two elements less than or equal to 0.38, and the other elements greater than 0.42 has probability $(0.5)^2\,(0.2)^8$. The number of such bootstrap datasets is $\binom{10}{2}$. So the answer is $\binom{10}{2}\,(0.5)^2\,(0.2)^8 = 0.000029$.

18.7 For the parametric bootstrap, we must estimate the parameter θ by $\hat{\theta} = (n+1)m_n/n$, and generate bootstrap samples from the $U(0, \hat{\theta})$ distribution. This distribution has expectation $\mu_{\hat{\theta}} = \hat{\theta}/2 = (n+1)m_n/(2n)$. Hence, for each bootstrap sample $x_1^*, x_2^*, \ldots, x_n^*$ compute $\bar{x}_n^* - \mu_{\hat{\theta}} = \bar{x}_n^* - (n+1)m_n/(2n)$.
Note that this is different from the *empirical* bootstrap simulation, where one would estimate μ by $\bar{x}_n$ and compute $\bar{x}_n^* - \bar{x}_n$.

18.8 a Since we know nothing about the distribution of the interfailure times, we estimate F by the empirical distribution function F_n of the software data and we estimate the expectation μ of F by the expectation $\mu^* = \bar{x}_n = 656.8815$ of F_n. The bootstrapped centered sample mean is the random variable $\bar{X}_n^* - 656.8815$. The corresponding empirical bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from F_n, i.e., draw with replacement 135 numbers from the software data.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815$$

   where $\bar{x}_n$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times.

18.8 b Because the interfailure times are now assumed to have an $Exp(\lambda)$ distribution, we must estimate λ by $\hat{\lambda} = 1/\bar{x}_n = 0.0015$ and estimate F by the distribution function of the *Exp*(0.0015) distribution. Estimate the expectation $\mu = 1/\lambda$ of the *Exp*(λ) distribution by $\mu^* = 1/\hat{\lambda} = \bar{x}_n = 656.8815$. Also now, the bootstrapped centered sample mean is the random variable $\bar{X}_n^* - 656.8815$. The corresponding parametric bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the *Exp*(0.0015) distribution.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 656.8815,$$

   where $\bar{x}_n$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times. We see that in this simulation the bootstrapped centered sample mean is the *same* in both cases: $\bar{X}_n^* - \bar{x}_n$, but the corresponding simulation procedures differ in step 1.

18.8 c Estimate λ by $\hat{\lambda} = \ln 2/m_n = 0.0024$ and estimate F by the distribution function of the *Exp*(0.0024) distribution. Estimate the expectation $\mu = 1/\lambda$ of the *Exp*(λ) distribution by $\mu^* = 1/\hat{\lambda} = 418.3816$. The corresponding parametric bootstrap simulation is described as follows:

1. Generate a bootstrap dataset $x_1^*, x_2^*, \ldots, x_n^*$ from the *Exp*(0.0024) distribution.
2. Compute the centered sample mean for the bootstrap dataset:

$$\bar{x}_n^* - 418.3816,$$

   where $\bar{x}_n$ is the sample mean of $x_1^*, x_2^*, \ldots, x_n^*$.

Repeat steps 1 and 2 one thousand times. We see that in this parametric bootstrap simulation the bootstrapped centered sample mean is *different* from the one in the empirical bootstrap simulation: $\bar{X}_n^* - (\ln 2)/m_n$ instead of $\bar{X}_n^* - \bar{x}_n$.

19.1 a From the formulas for the expectation and variance of uniform random variables we know that $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = (2\theta)^2/12 = \theta^2/3$. Hence $\mathrm{E}\left[X_i^2\right] = \mathrm{Var}(X_i) + (\mathrm{E}[X_i])^2 = \theta^2/3$. Therefore, by linearity of expectations

$$\mathrm{E}[T] = \frac{3}{n}\left(\frac{\theta^2}{3} + \cdots + \frac{\theta^2}{3}\right) = \frac{3}{n} \cdot n \cdot \frac{\theta^2}{3} = \theta^2.$$

Since $\mathrm{E}[T] = \theta^2$, the random variable T is an unbiased estimator for θ^2.

19.1 b The function $g(x) = -\sqrt{x}$ is a strictly convex function, because $g''(x) = (x^{-3/4})/4 > 0$. Therefore, by Jensen's inequality, $-\sqrt{\mathrm{E}[T]} < -\mathrm{E}\left[\sqrt{T}\right]$. Since, from part **a** we know that $\mathrm{E}[T] = \theta^2$, this means that $\mathrm{E}\left[\sqrt{T}\right] < \theta$. In other words, $\sqrt{T}$ is a biased estimator for θ, with negative bias.

19.8 From the model assumptions it follows that $\mathrm{E}[Y_i] = \beta x_i$ for each i. Using linearity of expectations, this implies that

$$\mathrm{E}[B_1] = \frac{1}{n}\left(\frac{\mathrm{E}[Y_1]}{x_1} + \cdots + \frac{\mathrm{E}[Y_n]}{x_n}\right) = \frac{1}{n}\left(\frac{\beta x_1}{x_1} + \cdots + \frac{\beta x_n}{x_n}\right) = \beta,$$
$$\mathrm{E}[B_2] = \frac{\mathrm{E}[Y_1] + \cdots + \mathrm{E}[Y_n]}{x_1 + \cdots + x_n} = \frac{\beta x_1 + \cdots + \beta x_n}{x_1 + \cdots + x_n} = \beta,$$
$$\mathrm{E}[B_3] = \frac{x_1\mathrm{E}[Y_1] + \cdots + x_n\mathrm{E}[Y_n]}{x_1^2 + \cdots + x_n^2} = \frac{\beta x_1^2 + \cdots + \beta x_n^2}{x_1^2 + \cdots + x_n^2} = \beta.$$

20.2 a Compute the mean squared errors of S and T: $\text{MSE}(S) = \text{Var}(S) + [\text{bias}(S)]^2 = 40 + 0 = 40$; $\text{MSE}(T) = \text{Var}(T) + [\text{bias}(T)]^2 = 4 + 9 = 13$. We prefer T, because it has a smaller MSE.

20.2 b Compute the mean squared errors of S and T: $\text{MSE}(S) = 40$, as in **a**; $\text{MSE}(T) = \text{Var}(T) + [\text{bias}(T)]^2 = 4 + a^2$. So, if $a < 6$: prefer T. If $a \geq 6$: prefer S. The preferences are based on the MSE criterion.

20.3 $\text{Var}(T_1) = 1/(n\lambda^2), \text{Var}(T_2) = 1/\lambda^2$; hence we prefer T_1, because of its smaller variance.

20.8 a This follows directly from linearity of expectations:

$$\text{E}[T] = \text{E}\left[r\bar{X}_n + (1-r)\bar{Y}_m\right] = r\text{E}\left[\bar{X}_n\right] + (1-r)\text{E}\left[\bar{Y}_m\right] = r\mu + (1-r)\mu = \mu.$$

20.8 b Using that $\bar{X}_n$ and $\bar{Y}_m$ are independent, we find $\text{MSE}(T)=\text{Var}(T) = r^2\text{Var}\left(\bar{X}_n\right) + (1-r)^2\text{Var}\left(\bar{Y}_m\right) = r^2 \cdot \sigma^2/n + (1-r)^2 \cdot \sigma^2/m$.
To find the minimum of this parabola we differentiate with respect to r and equate the result to 0: $2r/n - 2(1-r)/m = 0$. This gives the minimum value: $2rm - 2n(1-r) = 0$ or $r = n/(n+m)$.

21.1 Setting $X_i = j$ if *red* appears in the ith experiment for the first time on the jth throw, we have that X_1, X_2, and X_3 are independent $Geo(p)$ distributed random variables, where p is the probability that *red* appears when throwing the selected die. The likelihood function is

$$\begin{aligned} L(p) &= \text{P}(X_1 = 3, X_2 = 5, X_3 = 4) = (1-p)^2 p \cdot (1-p)^4 p \cdot (1-p)^3 p \\ &= p^3 (1-p)^9, \end{aligned}$$

so for D_1 one has that $L(p) = L(\frac{5}{6}) = \left(\frac{5}{6}\right)^3 \left(1 - \frac{5}{6}\right)^9$, whereas for D_2 one has that $L(p) = L(\frac{1}{6}) = \left(\frac{1}{6}\right)^3 \left(1 - \frac{1}{6}\right)^9 = 5^6 \cdot L(\frac{5}{6})$. It is very likely that we picked D_2.

21.4 a The likelihood $L(\mu)$ is given by

$$\begin{aligned} L(\mu) &= \text{P}(X_1 = x_1, \ldots, X_n = x_n) = \text{P}(X_1 = x_1) \cdots \text{P}(X_n = x_n) \\ &= \frac{\mu^{x_1}}{x_1!} \cdot \text{e}^{-\mu} \cdots \frac{\mu^{x_n}}{x_n!} \cdot \text{e}^{-\mu} = \frac{\text{e}^{-n\mu}}{x_1! \cdots x_n!} \mu^{x_1 + x_2 + \cdots + x_n}. \end{aligned}$$

21.4 b We find that the loglikelihood $\ell(\mu)$ is given by

$$\ell(\mu) = \left(\sum_{i=1}^{n} x_i\right) \ln(\mu) - \ln\left(x_1! \cdots x_n!\right) - n\mu.$$

Hence

$$\frac{\text{d}\ell}{\text{d}\mu} = \frac{\sum x_i}{\mu} - n,$$

and we find—after checking that we indeed have a maximum!—that $\bar{x}_n$ is the maximum likelihood estimate for μ.

21.4 c In **b** we have seen that $\bar{x}_n$ is the maximum likelihood estimate for μ. Due to the invariance principle from Section 21.4 we thus find that $\text{e}^{-\bar{x}_n}$ is the maximum likelihood estimate for $\text{e}^{-\mu}$.

21.8 a The likelihood $L(\theta)$ is given by

$$L(\theta) = C \cdot \left(\frac{1}{4}(2+\theta)\right)^{1997} \cdot \left(\frac{1}{4}\theta\right)^{32} \cdot \left(\frac{1}{4}(1-\theta)\right)^{906} \cdot \left(\frac{1}{4}(1-\theta)\right)^{904}$$
$$= \frac{C}{4^{3839}} \cdot (2+\theta)^{1997} \cdot \theta^{32} \cdot (1-\theta)^{1810},$$

where C is the number of ways we can assign 1997 starchy-greens, 32 sugary-whites, 906 starchy-whites, and 904 sugary-greens to 3839 plants. Hence the loglikelihood $\ell(\theta)$ is given by

$$\ell(\theta) = \ln(C) - 3839\ln(4) + 1997\ln(2+\theta) + 32\ln(\theta) + 1810\ln(1-\theta).$$

21.8 b A short calculation shows that

$$\frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} = 0 \qquad \Leftrightarrow \qquad 3810\theta^2 - 1655\theta - 64 = 0,$$

so the maximum likelihood estimate of θ is (after checking that $L(\theta)$ indeed attains a maximum for this value of θ):

$$\frac{-1655 + \sqrt{3714385}}{7620} = 0.0357.$$

21.8 c In this general case the likelihood $L(\theta)$ is given by

$$L(\theta) = C \cdot \left(\frac{1}{4}(2+\theta)\right)^{n_1} \cdot \left(\frac{1}{4}\theta\right)^{n_2} \cdot \left(\frac{1}{4}(1-\theta)\right)^{n_3} \cdot \left(\frac{1}{4}(1-\theta)\right)^{n_4} \cdot$$
$$= \frac{C}{4^n} \cdot (2+\theta)^{n_1} \cdot \theta^{n_2} \cdot (1-\theta)^{n_3+n_4},$$

where C is the number of ways we can assign n_1 starchy-greens, n_2 sugary-whites, n_3 starchy-whites, and n_4 sugary-greens to n plants. Hence the loglikelihood $\ell(\theta)$ is given by

$$\ell(\theta) = \ln(C) - n\ln(4) + n_1\ln(2+\theta) + n_2\ln(\theta) + (n_3+n_4)\ln(1-\theta).$$

A short calculation shows that

$$\frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} = 0 \qquad \Leftrightarrow \qquad n\theta^2 - (n_1 - n_2 - 2n_3 - 2n_4)\theta - 2n_2 = 0,$$

so the maximum likelihood estimate of θ is (after checking that $L(\theta)$ indeed attains a maximum for this value of θ):

$$\frac{n_1 - n_2 - 2n_3 - 2n_4 + \sqrt{(n_1 - n_2 - 2n_3 - 2n_4)^2 + 8nn_2}}{2n}.$$

21.11 a Since the dataset is a realization of a random sample from a $Geo(1/N)$ distribution, the likelihood is $L(N) = \mathrm{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, where each X_i has a $Geo(1/N)$ distribution. So

$$L(N) = \left(1 - \frac{1}{N}\right)^{x_1 - 1}\frac{1}{N}\left(1 - \frac{1}{N}\right)^{x_2-1}\frac{1}{N}\cdots\left(1 - \frac{1}{N}\right)^{x_n-1}\frac{1}{N}$$
$$= \left(1 - \frac{1}{N}\right)^{\left(-n + \sum_{i=1}^n x_i\right)}\left(\frac{1}{N}\right)^n.$$

But then the loglikelihood is equal to

$$\ell(N) = -n \ln N + \Big(-n + \sum_{i=1}^{n} x_i \Big) \ln \Big(1 - \frac{1}{N} \Big).$$

Differentiating to N yields

$$\frac{\mathrm{d}}{\mathrm{d}N}\big(\ell(N)\big) = \frac{-n}{N} + \Big(-n + \sum_{i=1}^{n} x_i \Big) \frac{1}{N(N-1)},$$

Now $\frac{\mathrm{d}}{\mathrm{d}N}\big(\ell(N)\big) = 0$ if and only if $N = \bar{x}_n$. Because $\ell(N)$ attains its maximum at $\bar{x}_n$, we find that the maximum likelihood estimate of N is $\hat{N} = \bar{x}_n$.

21.11 b Since $\mathrm{P}(Y = k) = 1/N$ for $k = 1, 2, \ldots, N$, the likelihood is given by

$$L(N) = \Big(\frac{1}{N}\Big)^n \quad \text{for } N \geq y_{(n)},$$

and $L(N) = 0$ for $N < y_{(n)}$. So $L(N)$ attains its maximum at $y_{(n)}$; the maximum likelihood estimate of N is $\hat{N} = y_{(n)}$.

22.1 a Since $\sum x_i y_i = 12.4$, $\sum x_i = 9$, $\sum y_i = 4.8$, $\sum x_i^2 = 35$, and $n = 3$, we find (c.f. (22.1) and (22.2)), that

$$\hat{\beta} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} = \frac{3 \cdot 12.4 - 9 \cdot 4.8}{3 \cdot 35 - 9^2} = -\frac{1}{4},$$

and $\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n = 2.35$.

22.1 b Since $r_i = y_i - \hat{\alpha} - \hat{\beta} x_i$, for $i = 1, \ldots, n$, we find $r_1 = 2 - 2.35 + 0.25 = -0.1$, $r_2 = 1.8 - 2.35 + 0.75 = 0.2$, $r_3 = 1 - 2.35 + 1.25 = -0.1$, and $r_1 + r_2 + r_3 = -0.1 + 0.2 - 0.1 = 0$.

22.1 c See Figure D.1.

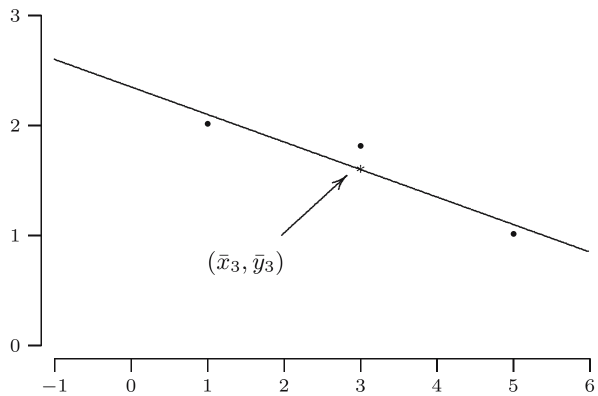


Fig. D.1. Solution of Exercise 22.1 **c**.

22.5 With the assumption that $\alpha = 0$, the method of least squares tells us now to minimize

$$S(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2 .$$

Now

$$\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta} = -2\sum_{i=1}^{n}(y_i - \beta x_i)x_i = -2\left(\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2\right),$$

so

$$\frac{\mathrm{d}S(\beta)}{\mathrm{d}\beta} = 0 \quad \Leftrightarrow \quad \beta = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Because $S(\beta)$ has a minimum for this last value of β, we see that the least squares estimator $\hat{\beta}$ of β is given by

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

22.12 a Since the denominator of $\hat{\beta}$ is a number, *not* a random variable, one has that

$$\mathrm{E}\left[\hat{\beta}\right] = \frac{\mathrm{E}\left[n(\sum x_i Y_i) - (\sum x_i)(\sum Y_i)\right]}{x\sum x_i^2 - (\sum x_i)^2}.$$

Furthermore, the numerator of this last fraction can be written as

$$\mathrm{E}\left[n\sum x_i Y_i\right] - \mathrm{E}\left[(\sum x_i)(\sum Y_i)\right],$$

which is equal to

$$n\sum(x_i \mathrm{E}\left[Y_i\right]) - (\sum x_i)\sum \mathrm{E}\left[Y_i\right].$$

22.12 b Substituting $\mathrm{E}\left[Y_i\right] = \alpha + \beta x_i$ in the last expression, we find that

$$\mathrm{E}\left[\hat{\beta}\right] = \frac{n\sum(x_i(\alpha + \beta x_i)) - (\sum x_i)\left[\sum(\alpha + \beta x_i)\right]}{x\sum x_i^2 - (\sum x_i)^2}.$$

22.12 c The numerator of the previous expression for $\mathrm{E}\left[\hat{\beta}\right]$ can be simplified to

$$\frac{n\alpha\sum x_i + n\beta\sum x_i^2 - n\alpha\sum x_i - \beta(\sum x_i)(\sum x_i)}{n\sum x_i^2 - (\sum x_i)^2},$$

which is equal to

$$\frac{\beta(n\sum x_i^2 - (\sum x_i)^2)}{n\sum x_i^2 - (\sum x_i)^2}.$$

22.12 d From **c** it now follows that $\mathrm{E}\left[\hat{\beta}\right] = \beta$, i.e., $\hat{\beta}$ is an unbiased estimator for β.

23.5 a The standard confidence interval for the mean of a normal sample with unknown variance applies, with $n = 23$, $\bar{x} = 0.82$ and $s = 1.78$, so:

$$\left(\bar{x} - t_{22,0.025} \cdot \frac{s}{\sqrt{23}},\ \bar{x} + t_{22,0.025} \cdot \frac{s}{\sqrt{23}}\right).$$

The critical values come from the $t(22)$ distribution: $t_{22,0.025} = 2.074$. The actual interval becomes:

$$\left(0.82 - 2.074 \cdot \frac{1.78}{\sqrt{23}},\ 0.82 + 2.074 \cdot \frac{1.78}{\sqrt{23}}\right) = (0.050,\ 1.590).$$

23.5 b Generate one thousand samples of size 23, by drawing with replacement from the 23 numbers

$$1.06, \quad 1.04, \quad 2.62, \quad \ldots, \quad 2.01.$$

For each sample $x_1^*, x_2^*, \ldots, x_{23}^*$ compute: $t^* = \bar{x}_{23}^* - 0.82/(s_{23}^*/\sqrt{23})$, where $s_{23}^* = \sqrt{\frac{1}{22}\sum(x_i^* - \bar{x}_{23}^*)^2}$.

23.5 c We need to estimate the critical value c_l^* such that $\mathrm{P}(T^* \leq c_l^*) \approx 0.025$. We take $c_l^* = -2.101$, the 25th of the ordered values, an estimate for the $25/1000 = 0.025$ quantile. Similarly, c_l^* is estimated by the 976th, which is 2.088.
The bootstrap confidence interval uses the c^* values instead of the t-distribution values $\pm t_{n-1,\alpha/2}$, but beware: c_l^* is from the *left tail* and appears on the *right-hand side* of the interval and c_u^* on the left-hand side:

$$\left(\bar{x}_n - c_u^* \frac{s_n}{\sqrt{n}}, \, \bar{x}_n - c_l^* \frac{s_n}{\sqrt{n}}\right).$$

Substituting $c_l^* = -2.101$ and $c_u^* = 2.088$, the confidence interval becomes:

$$\left(0.82 - 2.088 \cdot \frac{1.78}{\sqrt{23}}, \, 0.82 + 2.101 \cdot \frac{1.78}{\sqrt{23}}\right) = (0.045, \, 1.600).$$

23.6 a Because events described by inequalities do not change when we multiply the inqualities by a positive constant or add or subtract a constant, the following equalities hold: $\mathrm{P}\left(\tilde{L}_n < \theta < \tilde{U}_n\right) = \mathrm{P}(3L_n + 7 < 3\mu + 7 < 3U_n + 7) = \mathrm{P}(3L_n < 3\mu < 3U_n) = \mathrm{P}(L_n < \mu < U_n)$, and this equals 0.95, as is given.

23.6 b The confidence interval for θ is obtained as the realization of $(\tilde{L}_n, \tilde{U}_n)$, that is: $(\tilde{l}_n, \tilde{u}_n) = (3l_n + 7, 3u_n + 7)$. This is obtained by transforming the confidence interval for μ (using the transformation that is applied to μ to get θ).

23.6 c We start with $\mathrm{P}(L_n < \mu < U_n) = 0.95$ and try to get $1 - \mu$ in the middle: $\mathrm{P}(L_n < \mu < U_n) = \mathrm{P}(-L_n > -\mu > -U_n) = \mathrm{P}(1 - L_n > 1 - \mu > 1 - U_n) = \mathrm{P}(1 - U_n < 1 - \mu < 1 - L_n)$, where we see that the minus sign causes an interchange: $\tilde{L}_n = 1 - U_n$ and $\tilde{U}_n = 1 - L_n$. The confidence interval: $(1 - 5, 1 - (-2)) = (-4, 3)$.

23.6 d If we knew that L_n and U_n were always positive, then we could conclude: $\mathrm{P}(L_n < \mu < U_n) = \mathrm{P}\big(L_n^2 < \mu^2 < U_n^2\big)$ and we could just square the numbers in the confidence interval for μ to get the one for θ. Without the positivity assumption, the sharpest conclusion you can draw from $L_n < \mu < U_n$ is that μ^2 is smaller than the maximum of L_n^2 and U_n^2. So, $0.95 = \mathrm{P}(L_n < \mu < U_n) \leq \mathrm{P}\big(0 \leq \mu^2 < \max\{L_n^2, U_n^2\}\big)$ and the confidence interval $[0, \max\{l_n^2, u_n^2\}) = [0, 25)$ has a confidence of *at least* 95%. This kind of problem may occur when the transformation is not one-to-one (both -1 and 1 are mapped to 1 by squaring).

23.11 a For the 98% confidence interval the same formula is used as for the 95% interval, replacing the critical values by larger ones. This is the case, no matter whether the critical values are from the normal or t-distribution, or from a bootstrap experiment. Therefore, the 98% interval *contains* the 95%, and so must also contain the number 0.

23.11 b From a new bootstrap experiment we would obtain new and, most probably, different values c_u^* and c_l^*. It therefore could be, if the number 0 is close to the edge of the first bootstrap confidence interval, that it is just outside the new interval.

23.11 c The new dataset will resemble the old one in many ways, but things like the sample mean would most likely differ from the old one, and so there is no guarantee that the number 0 will again be in the confidence interval.

24.6 a The environmentalists are interested in a *lower* confidence bound, because they would like to make a statement like "We are 97.5% confidence that the concentration exceeds 1.68 ppm [and that is much too high.]" We have normal data, with σ unknown so we use $s_{16} = \sqrt{1.12} = 1.058$ as an estimate and use the critical value corresponding to 2.5% from the $t(15)$ distribution: $t_{15,0.025} = 2.131$. The lower confidence bound is $2.24 - 2.131 \cdot 1.058/\sqrt{16} = 2.24 - 0.56 = 1.68$, the interval: $(1.68, \infty)$.

24.6 b For similar reasons, the plant management constructs an *upper* confidence bound ("We are 97.5% confident pollution does not exceed 2.80 [and this is acceptable.]"). The computation is the same except for a minus sign: $2.24 + 2.131 \cdot 1.058/\sqrt{16} = 2.24 + 0.56 = 2.80$, so the interval is $[0, 2.80)$. Note that the computed upper and lower bounds are in fact the endpoints of the 95% two-sided confidence interval.

24.9 a From Section 8.4 we know: $\mathrm{P}(M \leq a) = [F_X(a)]^{12}$, so $\mathrm{P}(M/\theta \leq t) = \mathrm{P}(M \leq \theta t) = [F_X(\theta t)]^{12}$. Since X_i has a $U(0,\theta)$ distribution, $F_X(\theta t) = t$, for $0 \leq t \leq 1$. Substituting this shows the result.

24.9 b For c_l we need to solve $(c_l)^{12} = \alpha/2$, or $c_l = (\alpha/2)^{1/12} = (0.05)^{1/12} = 0.7791$. For c_u we need to solve $(c_u)^{12} = 1-\alpha/2$, or $c_u = (1-\alpha/2)^{1/12} = (0.95)^{1/12} = 0.9958$.

24.9 c From **b** we know that $\mathrm{P}(c_l < M/\theta < c_u) = \mathrm{P}(0.7790 < M/\theta < 0.9958) = 0.90$. Rewriting this equation, we get: $\mathrm{P}(0.7790\,\theta < M < 0.9958\,\theta) = 0.90$ and $\mathrm{P}(M/0.9958 < \theta < M/0.7790) = 0.90$. This means that $(m/0.9958, m/0.7790) = (3.013, 3.851)$ is a 90% confidence interval for θ.

24.9 d From **b** we derive the general formula:

$$\mathrm{P}\left((\alpha/2)^{1/n} < \frac{M}{\theta} < (1-\alpha/2)^{1/n}\right) = 1 - \alpha.$$

The left hand inequality can be rewritten as $\theta < M/(\alpha/2)^{1/n}$ and the right hand one as $M/(1-\alpha/2)^{1/n} < \theta$. So, the statement above can be rewritten as:

$$\mathrm{P}\left(\frac{M}{(1-\alpha/2)^{1/n}} < \theta < \frac{M}{(\alpha/2)^{1/n}}\right) = 1 - \alpha,$$

so that the general formula for the confidence interval becomes:

$$\left(\frac{m}{(1-\alpha/2)^{1/n}}, \frac{m}{(\alpha/2)^{1/n}}\right).$$

25.4 a Denote the observed numbers of cycles for the smokers by $X_1, X_2, \ldots, X_{n_1}$ and similarly $Y_1, Y_2, \ldots, Y_{n_2}$ for the nonsmokers. A test statistic should compare estimators for p_1 and p_2. Since the geometric distributions have expectations $1/p_1$

and $1/p_2$, we could compare the estimator $1/\bar{X}_{n_1}$ for p_1 with the estimator $1/\bar{Y}_{n_2}$ for p_2, or simply compare $\bar{X}_{n_1}$ with $\bar{Y}_{n_2}$. For instance, take test statistic $T = \bar{X}_{n_1} - \bar{Y}_{n_2}$. Values of T close to zero are in favor of H_0, and values far away from zero are in favor of H_1. Another possibility is $T = \bar{X}_{n_1}/\bar{Y}_{n_2}$.

25.4 b In this case, the maximum likelihood estimators $\hat{p}_1$ and $\hat{p}_2$ give better indications about p_1 and p_2. They can be compared in the same way as the estimators in **a**.

25.4 c The probability of getting pregnant during a cycle is p_1 for the smoking women and p_2 for the nonsmokers. The alternative hypothesis should express the belief that smoking women are *less likely* to get pregnant than nonsmoking women. Therefore take $H_1 : p_1 < p_2$.

25.10 a The alternative hypothesis should express the belief that the gross calorific *exceeds* 23.75 MJ/kg. Therefore take $H_1 : \mu > 23.75$.

25.10 b The p-value is the probability $\mathrm{P}(\bar{X}_n \geq 23.788)$ under the null hypothesis. We can compute this probability by using that under the null hypothesis $\bar{X}_n$ has an $N(23.75, (0.1)^2/23)$ distribution:

$$\mathrm{P}(\bar{X}_n \geq 23.788) = \mathrm{P}\left(\frac{\bar{X}_n - 23.75}{0.1/\sqrt{23}} \geq \frac{23.788 - 23.75}{0.1/\sqrt{23}}\right) = \mathrm{P}(Z \geq 1.82)\,,$$

where Z has an $N(0,1)$ distribution. From Table B.1 we find $\mathrm{P}(Z \geq 1.82) = 0.0344$.

25.11 A type I error occurs when $\mu = 0$ and $|t| \geq 2$. When $\mu = 0$, then T has an $N(0,1)$ distribution. Hence, by symmetry of the $N(0,1)$ distribution and Table B.1, we find that the probability of committing a type I error is

$$\mathrm{P}(|T| \geq 2) = \mathrm{P}(T \leq -2) + \mathrm{P}(T \geq 2) = 2 \cdot \mathrm{P}(T \geq 2) = 2 \cdot 0.0228 = 0.0456.$$

26.5 a The p-value is $\mathrm{P}(X \geq 15)$ under the null hypothesis $H_0 : p = 1/2$. Using Table 26.3 we find $\mathrm{P}(X \geq 15) = 1 - \mathrm{P}(X \leq 14) = 1 - 0.8950 = 0.1050$.

26.5 b Only values close to 23 are in favor of $H_1 : p > 1/2$, so the critical region is of the form $K = \{c, c+1, \ldots, 23\}$. The critical value c is the smallest value, such that $\mathrm{P}(X \geq c) \leq 0.05$ under $H_0 : p = 1/2$, or equivalently, $1 - \mathrm{P}(X \leq c-1) \leq 0.05$, which means $\mathrm{P}(X \leq c-1) \geq 0.95$. From Table 26.3 we conclude that $c - 1 = 15$, so that $K = \{16, 17, \ldots, 23\}$.

26.5 c A type I error occurs if $p = 1/2$ and $X \geq 16$. The probability that this happens is $\mathrm{P}(X \geq 16 \mid p = 1/2) = 1 - \mathrm{P}(X \leq 15 \mid p = 1/2) = 1 - 0.9534 = 0.0466$, where we have used Table 26.3 once more.

26.5 d In this case, a type II error occurs if $p = 0.6$ and $X \leq 15$. To approximate $\mathrm{P}(X \leq 15 \mid p = 0.6)$, we use the same reasoning as in Section 14.2, but now with $n = 23$ and $p = 0.6$. Write X as the sum of independent Bernoulli random variables: $X = R_1 + \cdots + R_n$, and apply the central limit theorem with $\mu = p = 0.6$ and $\sigma^2 = p(1-p) = 0.24$. Then

$$\begin{aligned}
\mathrm{P}(X \leq 15) &= \mathrm{P}(R_1 + \cdots + R_n \leq 15) \\
&= \mathrm{P}\left(\frac{R_1 + \cdots + R_n - n\mu}{\sigma\sqrt{n}} \leq \frac{15 - n\mu}{\sigma\sqrt{n}}\right) \\
&= \mathrm{P}\left(Z_{23} \geq \frac{15 - 13.8}{\sqrt{0.24}\sqrt{23}}\right) \approx \Phi(0.51) = 0.6950.
\end{aligned}$$

26.8 a Test statistic $T = \bar{X}_n$ takes values in $(0, \infty)$. Recall that the $Exp(\lambda)$ distribution has expectation $1/\lambda$, and that according to the law of large numbers $\bar{X}_n$ will be close to $1/\lambda$. Hence, values of $\bar{X}_n$ close to 1 are in favor of $H_0 : \lambda = 1$, and *only* values of $\bar{X}_n$ close to zero are in favor $H_1 : \lambda > 1$. Large values of $\bar{X}_n$ also provide evidence against $H_0 : \lambda = 1$, but even stronger evidence against $H_1 : \lambda > 1$. We conclude that $T = \bar{X}_n$ has critical region $K = (0, c_l]$. This is an example in which the alternative hypothesis and the test statistic deviate from the null hypothesis in *opposite* directions.
Test statistic $T' = e^{-\bar{X}_n}$ takes values in $(0, 1)$. Values of $\bar{X}_n$ close to zero correspond to values of T' close to 1, and large values of $\bar{X}_n$ correspond to values of T' close to 0. Hence, *only* values of T' close to 1 are in favor $H_1 : \lambda > 1$. We conclude that T' has critical region $K' = [c_u, 1)$. Here the alternative hypothesis and the test statistic deviate from the null hypothesis in the *same* direction.

26.8 b Again, values of $\bar{X}_n$ close to 1 are in favor of $H_0 : \lambda = 1$. Values of $\bar{X}_n$ close to zero suggest $\lambda > 1$, whereas large values of $\bar{X}_n$ suggest $\lambda < 1$. Hence, both small and large values of $\bar{X}_n$ are in favor of $H_1 : \lambda \neq 1$. We conclude that $T = \bar{X}_n$ has critical region $K = (0, c_l] \cup [c_u, \infty)$.
Small and large values of $\bar{X}_n$ correspond to values of T' close to 1 and 0. Hence, values of T' both close to 0 and close 1 are in favor of $H_1 : \lambda \neq 1$. We conclude that T' has critical region $K' = (0, c'_l] \cup [c'_u, 1)$. Both test statistics deviate from the null hypothesis in the same directions as the alternative hypothesis.

26.9 a Test statistic $T = (\bar{X}_n)^2$ takes values in $[0, \infty)$. Since μ is the expectation of the $N(\mu, 1)$ distribution, according to the law of large numbers, $\bar{X}_n$ is close to μ. Hence, values of $\bar{X}_n$ close to zero are in favor of $H_0 : \mu = 0$. Large negative values of $\bar{X}_n$ suggest $\mu < 0$, and large positive values of $\bar{X}_n$ suggest $\mu > 0$. Therefore, both large negative and large positive values of $\bar{X}_n$ are in favor of $H_1 : \mu \neq 0$. These values correspond to large positive values of T, so T has critical region $K = [c_u, \infty)$. This is an example in which the test statistic deviates from the null hypothesis in *one* direction, whereas the alternative hypothesis deviates in *two* directions.
Test statistic T' takes values in $(-\infty, 0) \cup (0, \infty)$. Large negative values and large positive values of $\bar{X}_n$ correspond to values of T' close to zero. Therefore, T' has critical region $K' = [c'_l, 0) \cup (0, c'_u]$. This is an example in which the test statistic deviates from the null hypothesis for small values, whereas the alternative hypothesis deviates for large values.

26.9 b Only large positive values of $\bar{X}_n$ are in favor of $\mu > 0$, which correspond to large values of T. Hence, T has critical region $K = [c_u, \infty)$. This is an example where the test statistic has the *same type* of critical region with a one-sided or two-sided alternative. Of course, the critical value c_u in part **b** is different from the one in part **a**.
Large positive values of $\bar{X}_n$ correspond to small positive values of T'. Hence, T' has critical region $K' = (0, c'_u]$. This is another example where the test statistic deviates from the null hypothesis for small values, whereas the alternative hypothesis deviates for large values.

27.5 a The interest is whether the inbreeding coefficient exceeds 0. Let μ represent this coefficient for the species of wasps. The value 0 is the a priori specified value of the parameter, so test null hypothesis $H_0 : \mu = 0$. The alternative hypothesis should express the belief that the inbreeding coefficient *exceeds* 0. Hence, we take alternative hypothesis $H_1 : \mu > 0$. The value of the test statistic is

$$t = \frac{0.044}{0.884/\sqrt{197}} = 0.70.$$

27.5 b Because $n = 197$ is large, we approximate the distribution of T under the null hypothesis by an $N(0,1)$ distribution. The value $t = 0.70$ lies to the right of zero, so the p-value is the right tail probability $\mathrm{P}(T \geq 0.70)$. By means of the normal approximation we find from Table B.1 that the right tail probability

$$\mathrm{P}(T \geq 0.70) \approx 1 - \Phi(0.70) = 0.2420.$$

This means that the value of the test statistic is not very far in the (right) tail of the distribution and is therefore not to be considered exceptionally large. We do not reject the null hypothesis.

27.7 a The data are modeled by a simple linear regression model: $Y_i = \alpha + \beta x_i$, where Y_i is the gas consumption and x_i is the average outside temperature in the ith week. Higher gas consumption as a consequence of smaller temperatures corresponds to $\beta < 0$. It is natural to consider the value 0 as the a priori specified value of the parameter (it corresponds to no change of gas consumption). Therefore, we take null hypothesis $H_0 : \beta = 0$. The alternative hypothesis should express the belief that the gas consumption increases as a consequence of smaller temperatures. Hence, we take alternative hypothesis $H_1 : \beta < 0$. The value of the test statistic is

$$t_b = \frac{\hat{\beta}}{s_b} = \frac{-0.3932}{0.0196} = -20.06.$$

The test statistic T_b has a t-distribution with $n - 2 = 24$ degrees of freedom. The value -20.06 is smaller than the left critical value $t_{24,0.05} = -1.711$, so we reject.

27.7 b For the data after insulation, the value of the test statistic is

$$t_b = \frac{-0.2779}{0.0252} = -11.03,$$

and T_b has a $t(28)$ distribution. The value -11.03 is smaller than the left critical value $t_{28,0.05} = -1.701$, so we reject.

28.5 a When $aS_X^2 + bS_Y^2$ is unbiased for σ^2, we should have $\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = \sigma^2$. Using that S_X^2 and S_Y^2 are both unbiased for σ^2, i.e., $\mathrm{E}\left[S_X^2\right] = \sigma^2$ and $\mathrm{E}\left[S_Y^2\right] = \sigma^2$, we get

$$\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = a\mathrm{E}\left[S_X^2\right] + b\mathrm{E}\left[S_Y^2\right] = (a+b)\sigma^2.$$

Hence, $\mathrm{E}\left[aS_X^2 + bS_Y^2\right] = \sigma^2$ for all $\sigma > 0$ if and only if $a + b = 1$.

28.5 b By independence of S_X^2 and S_Y^2 write

$$\begin{aligned}\mathrm{Var}\big(aS_X^2 + (1-a)S_Y^2\big) &= a^2\mathrm{Var}\big(S_X^2\big) + (1-a)^2\mathrm{Var}\big(S_Y^2\big)\\ &= \left(\frac{a^2}{n-1} + \frac{(1-a)^2}{m-1}\right)2\sigma^4.\end{aligned}$$

To find the value of a that minimizes this, differentiate with respect to a and put the derivative equal to zero. This leads to

$$\frac{2a}{n-1} - \frac{2(1-a)}{m-1} = 0.$$

Solving for a yields $a = (n-1)/(n+m-2)$. Note that the second derivative of $\mathrm{Var}\big(aS_X^2 + (1-a)S_Y^2\big)$ is positive so that this is indeed a minimum.

References

1. J. Bernoulli. *Ars Conjectandi*. Basel, 1713.
2. J. Bernoulli. The most probable choice between several discrepant observations and the formation therefrom of the most likely induction. ( ):3–33, 1778. With a comment by Euler.
3. P. Billingsley. *Probability and measure*. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
4. L.D. Brown, T.T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Stat. Science*, 16(2):101–133, 2001.
5. S.R. Dalal, E.B. Fowlkes, and B. Hoadley. Risk analysis of the space shuttle: pre-*Challenger* prediction of failure. *J. Am. Stat. Assoc.*, 84:945–957, 1989.
6. J. Daugman. Wavelet demodulation codes, statistical independence, and pattern recognition. In *Institute of Mathematics and its Applications, Proc. 2nd IMA-IP: Mathematical Methods, Algorithms, and Applications (Blackledge and Turner, Eds)*, pages 244–260. Horwood, London, 2000.
7. B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.
8. W. Feller. *An introduction to probability theory and its applications, Vol. II.* John Wiley & Sons Inc., New York, 1971.
9. R.A. Fisher. On an absolute criterion for fitting frequency curves. *Mess. Math.*, 41:155–160, 1912.
10. R.A. Fisher. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1(4):3–32, 1921.
11. H.S. Fogler. *Elements of chemical reaction engineering.* Prentice-Hall, Upper Saddle River, 1999.
12. D. Freedman and P. Diaconis. On the histogram as a density estimator: L_2 theory. *Z. Wahrsch. Verw. Gebiete*, 57(4):453–476, 1981.
13. C.F. Gauss. *Theoria motus corporum coelestium in sectionis conicis solem ambientum. In: Werke. Band VII.* Georg Olms Verlag, Hildesheim, 1973. Reprint of the 1906 original.
14. P. Hall. *The bootstrap and Edgeworth expansion.* Springer-Verlag, New York, 1992.
15. R. Herz, H.G. Schlichter, and W. Siegener. *Angewandte Statistik für Verkehrs- und Regionalplaner.* Werner-Ingenieur-Texte 42, Werner-Verlag, Düsseldorf, 1992.

16. J.L. Lagrange. *Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations.* Paris, 1770–73. Œvres 2, 1886.
17. J.H. Lambert. *Photometria.* Augustae Vindelicorum, 1760.
18. R.J. MacKay and R.W. Oldford. Scientific method, statistical method and the speed of light. *Stat. Science*, 15(3):254–278, 2000.
19. J. Moynagh, H. Schimmel, and G.N. Kramer. The evaluation of tests for the diagnosis of transmissible spongiform encephalopathy in bovines. Technical report, European Commission, Directorate General XXIV, Brussels, 1999.
20. V. Pareto. *Cours d'economie politique.* Rouge, Lausanne et Paris, 1897.
21. E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1076, 1962.
22. K. Pearson. *Philos. Trans.*, 186:343–414, 1895.
23. R. Penner and D.G. Watts. Mining information. *The Amer. Stat.*, 45:4–9, 1991.
24. Commission Rogers. Report on the space shuttle *Challenger* accident. Technical report, Presidential commission on the Space Shuttle Challenger Accident, Washington, DC, 1986.
25. M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27:832–837, 1956.
26. S.M. Ross. *A first course in probability.* Prentice-Hall, Inc., New Jersey, sixth edition, 1984.
27. R. Ruggles and H. Brodie. An empirical approach to economic intelligence in World War II. *Journal of the American Statistical Association*, 42:72–91, 1947.
28. E. Rutherford and H. Geiger (with a note by H. Bateman). The probability variations in the distribution of α particles. *Phil.Mag.*, 6:698–704, 1910.
29. D.W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979.
30. S. Siegel and N.J. Castellan. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, New York, second edition, 1988.
31. B.W. Silverman. *Density estimation for statistics and data analysis.* Chapman & Hall, London, 1986.
32. K. Singh. On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, 9:1187–1195, 1981.
33. S.M. Stigler. *The history of statistics — the measurement of uncertainty before 1900.* Cambridge, Massachusetts, 1986.
34. H.A. Sturges. *J. Amer. Statist. Ass.*, 21, 1926.
35. J.W. Tukey. *Exploratory data analysis.* Addison-Wesley, Reading, 1977.
36. S.A. van de Geer. *Applications of empirical process theory.* Cambridge University Press, Cambridge, 2000.
37. J.G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers*, 1, 1952.
38. C.R. Weinberg and B.C. Gladen. The beta-geometric distribution applied to comparative fecundability studies. *Biometrics*, 42(3):547–560, 1986.
39. H. Westergaard. *Contributions to the history of statistics.* Agathon, New York, 1968.
40. E.B. Wilson. Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc.*, 22:209–212, 1927.
41. D.R. Witte et al. Cardiovascular mortality in Dutch men during 1996 European foolball championship: longitudinal population study. *British Medical Journal*, 321:1552–1554, 2000.

List of symbols

$\emptyset$	empty set, page 14
α	significance level, page 384
A^c	complement of the event A, page 14
$A \cap B$	intersection of A and B, page 14
$A \subset B$	A subset of B, page 15
$A \cup B$	union of A and B, page 14
$Ber(p)$	Bernoulli distribution with parameter p, page 45
$Bin(n, p)$	binomial distribution with parameters n and p, page 48
c_l, c_u	left and right critical values, page 388
$Cau(\alpha, \beta)$	Cauchy distribution with parameters α en β, page 161
$\mathrm{Cov}(X, Y)$	covariance between X and Y, page 139
$\mathrm{E}[X]$	expectation of the random variable X, page 90, 91
$Exp(\lambda)$	exponential distribution with parameter λ, page 62
Φ	distribution function of the standard normal distribution, page 65
ϕ	probability density of the standard normal distribution, page 65
f	probability density function, page 57
f	joint probability density function, page 119
F	distribution function, page 44
F	joint distribution function, page 118
F^{inv}	inverse function of distribution function F, page 73
F_n	empirical distribution function, page 219
$f_{n,h}$	kernel density estimate, page 213
$Gam(\alpha, \lambda)$	gamma distribution with parameters α en λ, page 157
$Geo(p)$	geometric distribution with parameter p, page 49
H_0, H_1	null hypothesis and alternative hypothesis, page 374

$L(\theta)$	likelihood function, page 317
$\ell(\theta)$	loglikelihood function, page 319
Med_n	sample median of a dataset, page 231
$n!$	n factorial, page 14
$N(\mu, \sigma^2)$	normal distribution with parameters μ and σ^2, page 64
Ω	sample space, page 13
$Par(\alpha)$	Pareto distribution with parameter α, page 63
$Pois(\mu)$	Poisson distribution with parameter μ, page 170
$\mathrm{P}(A \mid C)$	conditional probability of A given C, page 26
$\mathrm{P}(A)$	probability of the event A, page 16
$q_n(p)$	pth empirical quantile, page 234
q_p	pth quantile or 100pth percentile, page 66
$\rho(X, Y)$	correlation coefficient between X and Y, page 142
s_n^2	sample variance of a dataset, page 233
S_n^2	sample variance of random sample, page 292
$t(m)$	t-distribution with m degrees of freedom, page 348
$t_{m,p}$	critical value of the $t(m)$ distribution, page 348
$U(\alpha, \beta)$	uniform distribution with parameters α and β, page 60
$\mathrm{Var}(X)$	variance of the random variable X, page 96
$\bar{x}_n$	sample mean of a dataset, page 231
$\bar{X}_n$	average of the random variables $X_1, \ldots, X_n$, page 182
z_p	critical value of the $N(0, 1)$ distribution, page 345

Index