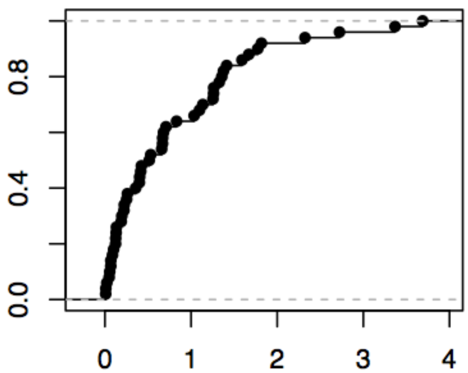
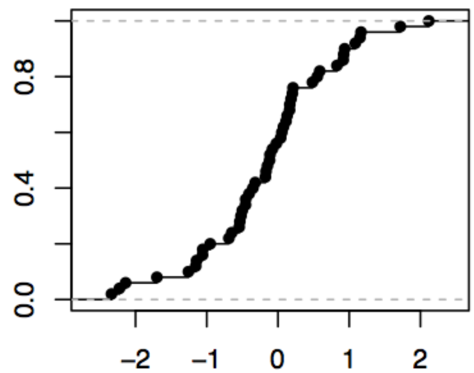
Extra Opgaven over Hoofdstuk 15, 16 en 17

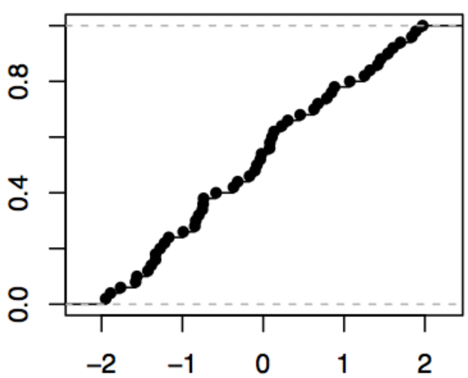
15.1. Van vier datasets ter grootte 50 is de empirische verdelingsfunctie weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(1)$, een $Exp(1/3)$, een $N(0,1)$ en een $U(-2,2)$ verdeling. Welke dataset komt uit de $Exp(1/3)$ en welke dataset komt uit de $U(-2,2)$ verdeling?



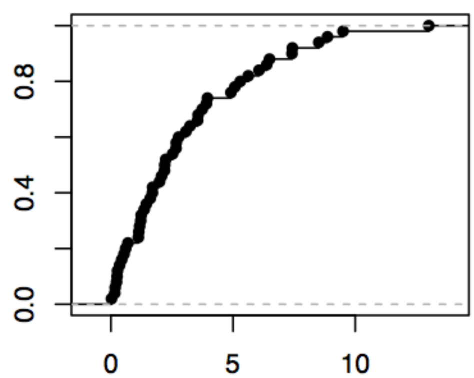
I



II



III



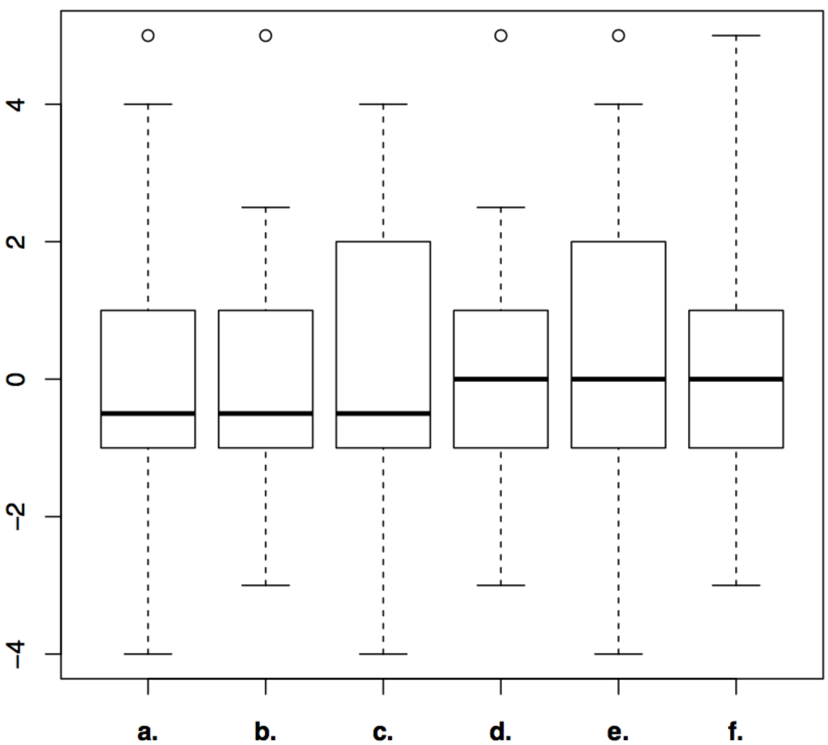
IV

a. $Exp(1/3)$: I; $U(-2,2)$: III
b. $Exp(1/3)$: III; $U(-2,2)$: II
c. $Exp(1/3)$: III; $U(-2,2)$: I
d. $Exp(1/3)$: IV; $U(-2,2)$: III
e. $Exp(1/3)$: IV; $U(-2,2)$: II
f. $Exp(1/3)$: II; $U(-2,2)$: IV

15.2. Gegeven is een dataset met de volgende 11 punten:

$$-3.0 \quad -2.5 \quad -1.0 \quad -0.8 \quad -0.8 \quad -0.5 \quad 0.5 \quad 0.7 \quad 1.0 \quad 2.5 \quad 5.0$$

Welke van de zes onderstaande boxplots hoort bij deze dataset?



15.3. Gegeven is de volgende informatie over een histogram

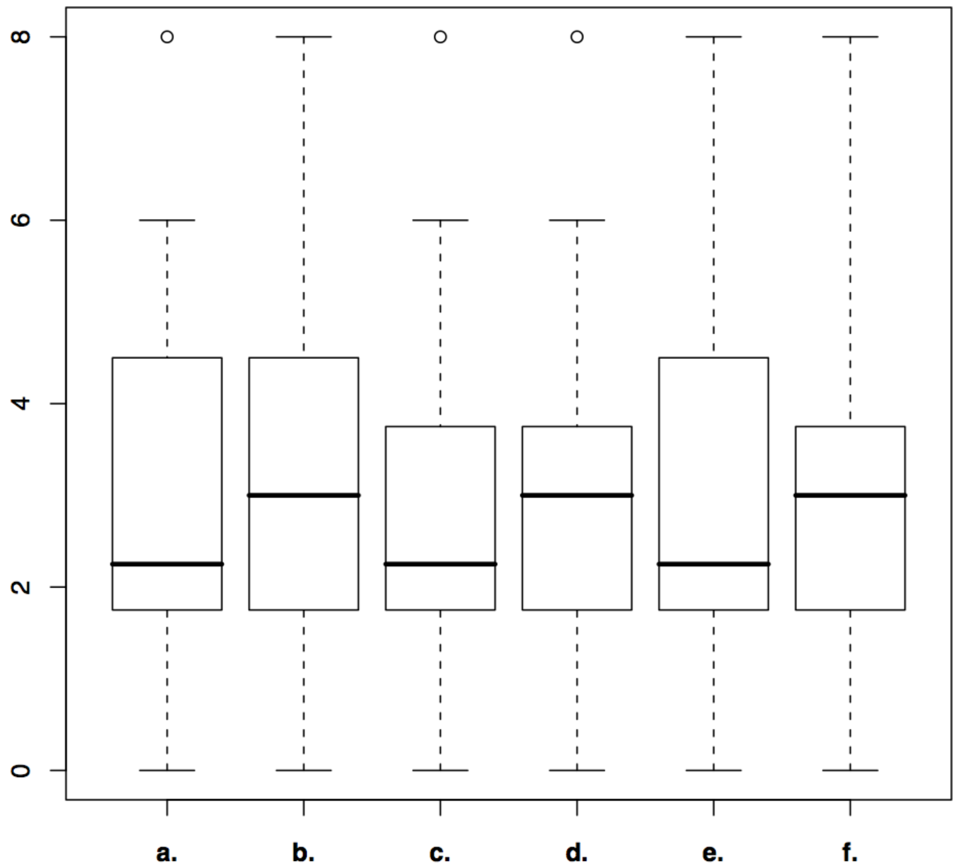
cel	hoogte
(0,1]	0.6583
(1,2]	0.1917
(2,4]	0.0667
(4,7]	0.0056

Bereken de toename van de empirische verdelingsfunctie op het interval $[1, 4]$

15.4. Gegeven is de dataset

$$0.00 \quad 1.5 \quad 1.75 \quad 1.75 \quad 2.00 \quad 2.25 \quad 3.25 \quad 4.5 \quad 4.5 \quad 6.00 \quad 8.00$$

Welke van de zes onderstaande boxplots hoort bij deze dataset?



15.5. Geef aan welke van de volgende steekproefgrootheden informatie geven over de spreiding:

m: de mediaan,	M: de Variantie,
s: de standaarddeviatie,	I: de IQR.

a. M en s **b.** m, M en s **c.** M, s en I
d. M en I **e.** s en I **f.** m, M, s en I

15.6. We hebben de volgende dataset

$$7 \quad 15 \quad 6 \quad 6\frac{1}{2} \quad 5 \quad 10 \quad 5 \quad 15$$

Wat is de empirische verdelingsfunctie in 6 en in $7\frac{1}{2}$?

a. $\frac{2}{8}$ en $\frac{6}{8}$ **b.** $\frac{3}{8}$ en $\frac{6}{8}$ **c.** 0 en 1 **d.** 0 en 4 **e.** $\frac{3}{8}$ en $\frac{5}{8}$ **f.** $\frac{2}{8}$ en $\frac{5}{8}$

15.7. Een geordende dataset van 100 metingen van dagelijkse neerslag in mm in Bogota bevat de volgende gegevens

0.2	0.6	0.9	1.2	1.2	1.5	1.7	1.8	2	2
2.1	2.9	2.2	2.4	2.8	2.9	2.9	3.6	3	3.1
3.3	3.4	3.4	3.8	4.1	4.1	4.1	4.2	4.2	4.2
4.7	5.4	5.6	5.6	5.7	5.8	6	6.2	6.3	6.3
6.4	6.4	6.6	6.9	7	7.2	7.2	7.7	8.2	8.1
8.5	9.7	9.1	10.1	11.1	11.5	11.6	11.7	11.8	12.1
12.2	12.2	12.2	12.8	13.3	13.8	14.6	14.7	14.7	14.8
14.7	15.1	15.1	16.2	16.4	17.1	17.6	18.3	19.8	21.7
22.4	23.1	23.2	24.4	25.5	26.2	26.6	27.5	31.4	32.2
32.7	33.1	36.1	46.8	54.2	57	100.1	117.3	309.4	400.7

De dataset wordt weergegeven via een histogram met bins van breedte 5, als volgt $[0,5], (5,10], \ldots, (395,400]$. Bereken de hoogte van het histogram op bin $(10,15]$.

a. 0.018 **b.** 0.036 **c.** 0.18 **d.** 0.36 **e.** 3.6 **f.** 18

15.8. Gegeven is de volgende dataset

.0018	.1571	.1687	.2006	.2093
.2691	.2774	.3233	.3279	.3922
.4417	.4931	.5704	.5998	.6013
.6221	.6719	.9470	.9531	.9870

Voor welk getal in de dataset is de waarde van de empirische verdelingsfunctie van deze dataset gelijk aan 0.6?

a. 0.2691 **b.** 0.5704 **c.** 0.4931 **d.** 0.9470 **e.** 0.4417 **f.** 0.4991

15.9. We construeren een histogram met cellen [0,1], (1,3], (3,5], (5,8], (8,11], (11,14], en (14,18]. Gegeven zijn de waarden van de empirische verdelingsfunctie in de randen van de cellen:

t	0	1	3	5	8	11	14	18
$F_n(t)$	0	0.225	0.445	0.615	0.735	0.805	0.910	1.000

Dan is de hoogte van het histogram op de cel $(8,11]$ gelijk aan

a. 0.0233 **b.** 0.0035 **c.** 0.0700 **d.** 0.2100 **e.** 0.2450 **f.** 0.2683

15.10. Een onderzoeker bestudeert een data set $x_1, x_2, \ldots, x_n$ met behulp van de empirische verdelingsfunctie F_n. Hij merkt dat hij nog een extra data punt x_{n+1} heeft. Hij vraagt zich af hoe F_n vergelijkt met de empirische verdelingsfunctie F_{n+1} van de data set $x_1, x_2, \ldots, x_n, x_{n+1}$. Van de volgende beweringen is er één juist. Welke?

a. $F_{n+1}(x) = F_n(x) - \frac{1}{n+1}$ voor $-\infty < x < \infty$

b. $F_{n+1}(x) = F_n(x) + \frac{1}{n+1}$ voor $-\infty < x < \infty$

c. $F_{n+1}(x_n) \leq F_n(x_n)$ mits $x_n > x_{n+1}$

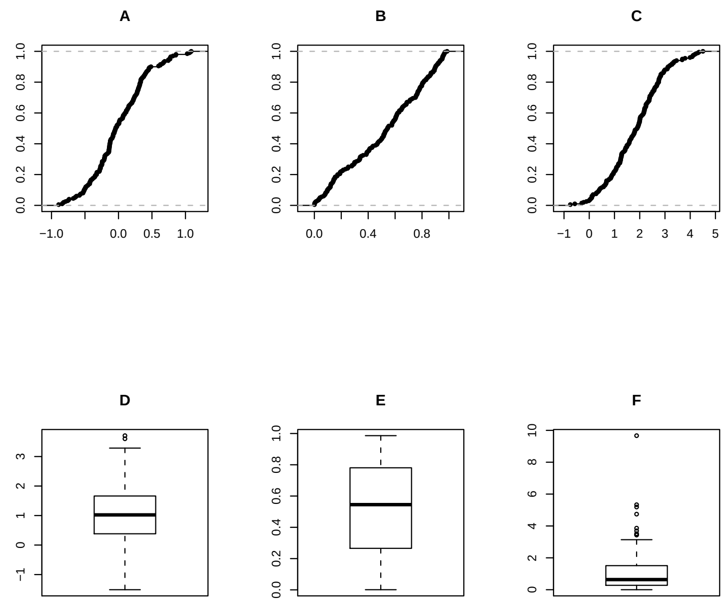
d. $F_{n+1}(x_2) \leq F_n(x_2)$ mits $x_2 < x_{n+1}$

e. $F_{n+1}(x) = F_n(x)$ mits $x > x_{n+1}$

f. $F_{n+1}(x) = F_n(x)$ mits $x < x_{n+1}$

15.11. In onderstaande figuur zijn afbeeldingen A, B en C plots van de empirische verdelingsfunctie van een dataset. Afbeeldingen D, E en F zijn boxplots van een dataset. Alle datasets zijn gebaseerd op 200 waarden, die ofwel uit een normale verdeling, ofwel uit een exponentiële verdeling ofwel uit een uniforme verdeling op $(0, 1)$ getrokken zijn.

Bij welk van deze afbeeldingen is het plausibel dat de data getrokken zijn uit een normale verdeling?



a. A,B,C,D,E **b.** A,C,D,E **c.** B,F
d. B,E **e.** A,C,D **f.** A,C,E

15.12. Voor een dataverzameling construeert men een histogram met cellen $[0, 1]$, $(1, 3]$, $(3, 5]$, $(5, 8]$, $(8, 11]$, $(11, 14]$ and $(14, 18]$. Gegeven zijn de waarden van de empirische verdelingsfunctie op de grenzen van de cellen:

t	$F_n(t)$
0	0
1	0.225
3	0.445
5	0.615
8	0.735
11	0.805
14	0.910
18	1.000

Het 40ste percentiel van de dataverzameling ligt in cel

a. $(1, 3]$ **b.** $(3, 5]$ **c.** $(5, 8]$
d. $(8, 11]$ **e.** $(11, 14]$ **f.** $(14, 18]$

Extra Opgaven over Hoofdstuk 19

19.1. Een dataset wordt gemodelleerd als een realisatie van een steekproef $X_1, X_2, \ldots, X_n$ uit een *Exp*(λ) verdeling, waarbij $\lambda > 0$ onbekend is. Zij μ de verwachting van de modelverdeling (dus $\mu = 1/\lambda$), en M_n het minimum van $X_1, X_2, \ldots, X_n$. Dan heeft M_n een *Exp*(λn) verdeling. De steekproeffunctie $T = c_n M_n$ is een zuivere schatter voor μ, waarbij c_n een constante is die van n afhangt. Dan is c_n gelijk aan

a. 1 **b.** $1/n$ **c.** n **d.** n^2 **e.** $1/n^2$ **f.** $1/(n-1)$

19.2. Stel dat $X_1, X_2, \ldots, X_n$ een steekproef is uit een verdeling met verwachting μ en variantie σ^2. Beschouw de beweringen

A Het steekproefgemiddelde is een zuivere schatter voor μ.

B De steekproefvariantie S_n^2 is een zuivere schatter voor σ^2.

C De steekproefstandaarddeviatie S_n is een zuivere schatter voor σ.

Welke van deze beweringen zijn waar?

a. geen **b.** B en C **c.** A
d. A, B en C **e.** A en C **f.** A en B

19.3. Voor de stochastische variabelen $X_1, X_2, \ldots, X_n$ geldt $\mathrm{E}[X_i] = \mu$ voor alle $i = 1, 2, \ldots, n$, waarbij $\mu \in \mathbb{R}$ onbekend is. Voor welke $a, b \in \mathbb{R}$ is

$$T = a(X_1 + X_2 + \cdots + X_n) + b$$

een zuivere schatter voor μ?

a. $a = 0,\ b = 0$ **b.** $a = 0,\ b = \mu$ **c.** $a = 1/n,\ b = \mu$
d. $a = 1/n,\ b = 0$ **e.** $a = 1,\ b = 0$ **f.** $a = 1,\ b = \mu$

19.4. We beschikken over een dataset $x_1, \ldots, x_n$ die wordt opgevat als realisatie van een steekproef $X_1, \ldots, X_n$ uit een onbekende verdeling. Gegeven is dat

$$\mathrm{E}[X_i] = \theta + \frac{2}{3} \quad \text{en} \quad \mathrm{Var}(X_i) = \frac{1}{18},$$

waarbij θ een onbekende parameter is.

Voor welke waarde van a is de schatter

$$T = \bar{X}_n + a$$

een zuivere schatter voor de parameter θ?

19.5. Van een bepaalde plantensoort komen de vier typen *starchy-green*, *sugary-white*, *starchy-white*, en *sugary-green* voor met kansen

Type	Kans
1: starchy-green	$p_1 = \frac{1}{4}(2+\theta)$
2: sugary-white	$p_2 = \frac{1}{4}\theta$
3: starchy-white	$p_3 = \frac{1}{4}(1-\theta)$
4: sugary-green	$p_4 = \frac{1}{4}(1-\theta)$
Totaal	1

waarbij $0 < \theta < 1$ een onbekende parameter is. Stel dat men beschikt over n bladeren. Het aantal *starchy-green* bladeren modeleert men met een $Bin(n, p_1)$ variabele N_1 en het aantal *sugary-white* bladeren met een $Bin(n, p_2)$ variabele N_2. We beschouwen twee schatters voor θ:

$$T_1 = \frac{4}{n}N_1 - 2 \quad \text{en} \quad T_2 = \frac{4}{n}N_2.$$

Toon aan dat beide schatters zuiver zijn voor θ.

19.6. Gegeven zijn twee steekproefgrootheden (Engels: sample statistics) T en S. We willen een parameter θ schatten, die uit het interval $(1, \infty)$ komt. Bekend is dat $\mathrm{E}[T] = 2\theta$, en dat $\mathrm{E}[S] = \sqrt{\theta}$. Welke van de volgende schatters is een zuivere schatter voor θ?

a. $2T$ **b.** T^2 **c.** $T - 2$
d. $T + S$ **e.** $\sqrt{S}$ **f.** i.h.a. géén van deze vijf

19.7. We beschikken over een dataset van omvang 20 die we opvatten als realisatie van onafhankelijke stochasten $X_1, X_2, \ldots, X_{20}$. Gegeven is dat elke X_i een continue verdeling heeft met kansdichtheid $f(x) = \frac{1}{2\theta}$, als $-\theta \leq x \leq \theta$, en $f(x) = 0$ voor alle andere waarden van x. Verder is $\theta > 0$ onbekend. Definieer de volgende schatter voor θ:

$$T = a(|X_1| + |X_2| + \cdots + |X_{20}|).$$

Voor welke a is T een zuivere schatter voor θ?

a. $\frac{1}{40}$ **b.** $\frac{1}{20}$ **c.** $\frac{3}{20}$ **d.** $\frac{1}{10}$ **e.** $\frac{5}{40}$ **f.** $\frac{3}{40}$

19.8. Gegeven onafhankelijke stochasten $X_1, \ldots, X_n$ uit een geometrische verdeling met parameter p. We hebben twee schatters voor p, namelijk

$$S = \frac{1}{\bar{X}_n} \quad \text{en} \quad T = \frac{\#\{i : X_i = 1\}}{n}.$$

Hierbij gebruiken we de notatie $\#A$ voor het aantal elementen in de verzameling A. Dan geldt

a. S en T zijn beide zuiver voor p
b. S is zuiver en T heeft een negatieve bias
c. T is zuiver en S heeft een negatieve bias
d. S is zuiver en T heeft een positiev bias
e. T is zuiver en S heeft een positieve bias
f. S en T zijn beide onzuiver

Extra Opgaven over Hoofdstuk 20

20.1. Stel dat $X_1, \ldots, X_n$ onafhankelijke stochasten zijn met een exponentiële verdeling met parameter λ. Beschouw de schatter $T = nM_n$ voor $1/\lambda$, waarbij $M_n = \min\{X_1, \ldots, X_n\}$. We kunnen laten zien dat de verdeling van M_n exponentieel met parameter $n\lambda$ is. De mean squared error van T (als schatter voor $1/\lambda$) is gelijk aan

a. $\frac{1}{\lambda}$ **b.** $\frac{n}{\lambda}$ **c.** $\frac{1}{n\lambda}$ **d.** $\frac{1}{n^2\lambda^2}$ **e.** $\frac{1}{\lambda^2}$ **f.** $\frac{1}{n\lambda^2}$

20.2. Laat T een schatter zijn voor een parameter θ. Gegeven is dat $\mathrm{E}[T] = \theta - 2$ en dat $\mathrm{Var}(T) = 2$. Dan is de *mean squared error* van T gelijk aan:

a. -2 **b.** 0 **c.** 1 **d.** 2 **e.** 4 **f.** 6

20.3. Zij $X_1, X_2, \ldots, X_n$ onafhankelijke stochastische variabelen, allemaal met dezelfde verwachting μ en variantie σ^2. Een zuivere schatter voor μ is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Zij Y een stochastische variabele onafhankelijk van $X_1, X_2, \ldots, X_n$ wederom met verwachting μ en variantie σ^2. Dan is voor elke $r \in [0,1]$, de combinatie

$$T = r\bar{X}_n + (1-r)Y$$

een zuivere schatter voor μ. Welke waarde voor $r \in [0,1]$ leidt tot de beste (in de zin van MSE) schatter voor μ?

a. $r = 0$ **b.** $r = 1/n$ **c.** $r = 1/2$
d. $r = n/(n+1)$ **e.** $r = (n-1)/n$ **f.** $r = 1$

20.4. Zij $\bar{X}_n$ and $\bar{Y}_m$ de steekproefgemiddelden van twee onafhankelijke steekproeven van omvang n respektievelijk m uit dezelfde kansverdeling met verwachting μ en variantie σ^2. We combineren de twee schatters tot een nieuwe schatter

$$T = r\bar{X}_n + (1-r)\bar{Y}_m,$$

waarbij r is een getal is tussen 0 en 1. Dan geldt

a. T is zuiver voor μ met de kleinste MSE als $r = n/(m+n)$
b. T is zuiver voor μ met de kleinste MSE als $r = m/(m+n)$
c. T is zuiver voor μ met de kleinste MSE als $r = 1/2$
d. T is onzuiver voor μ met de kleinste MSE als $r = n/(m+n)$
e. T is onzuiver voor μ met de kleinste MSE als $r = m/(m+n)$
f. T is onzuiver voor μ met de kleinste MSE als $r = 1/2$

20.5. Laat T een schatter zijn voor een parameter θ. Gegeven is dat $\mathrm{E}[T] = \theta - 2$ en dat $\mathrm{Var}(T) = 2$. Bereken de *mean squared error* van T

Antwoorden:

15.1. d.

15.2. b.

15.3. 0.3251

15.4. e.

15.5. c.

15.6. e.

15.7. b.

15.8. c.

15.9. a.

15.10. d.

15.11. e.

15.12. a.

19.1. c.

19.2. f.

19.3. d.

19.4. $-2/3$.

19.6. f.

19.7. d.

19.8. e.

20.1. e.

20.2. f.

20.3. d.

20.4. a.

20.5. 6.