

Wahrscheinlichkeitsrechnung



KAD 2018.09.26

Grundbegriffe

Zufallsexperiment

- Vorgang nach einer bestimmten Vorschrift ausgeführt
- (im Prinzip) beliebig oft wiederholbar
- sein Ergebnis ist zufallsabhängig
- bei mehrmaligen Durchführung des Experiments beeinflussen die Ergebnisse einander nicht

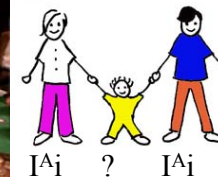
Beispiele:



Würfelspiel



Roulett



Blutgruppenversuch



Fahrtversuch

2

Elementarereignisse

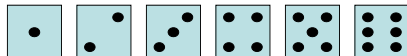
die einzelnen, nicht mehr zerlegbaren und sich gegenseitig ausschließenden Ausgänge oder Ergebnisse eines Zufallsexperimentes

Ereignismenge, Ereignisraum (Ω)

Reihe aller möglichen Elementarereignisse. Z.B:

beim Würfelspiel:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$



beim Münzenexperiment: $\Omega = \{\text{Zahl}, \text{Kopf}\}$



beim „Blutgruppenversuch“: $\Omega = \{I^A I^A, I^A i, iI^A, ii\}$

beim „Fahrtversuch“:

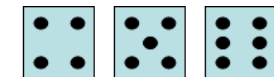
{kein Unfall, Unfall}



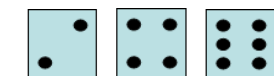
Ereignis

jede beliebige Teilmenge des Ergebnisraumes

zB. Augenzahl des Würfels > 3 : $\{4, 5, 6\}$



Augenzahl des Würfels ist gerade: $\{2, 4, 6\}$



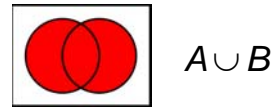
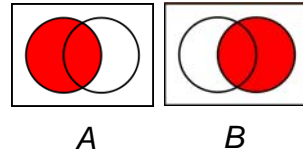
Blutgruppe A: $\{I^A I^A, I^A i, iI^A\}$

Operationen mit Ereignissen

Man kann mit Ereignissen rechnen wie mit Mengen.

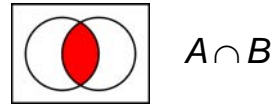
Vereinigung

Menge aller Elementarereignisse, die zu **A oder B** gehören



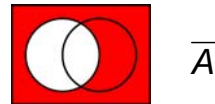
Durchschnitt

Menge aller Elementarereignisse, die zu **A und B** gehören



Komplementbildung

Menge aller Elementarereignisse des Ereignisraumes, die nicht in Ereignis **A** enthalten sind (Gegenereignis von A)



5

Definition der Wahrscheinlichkeit

Bernoulli (1654-1705), Laplace (1749-1827)
(klassische Wahrscheinlichkeit)

Bei einem Zufallsexperiment, was endlich viele Ausgänge hat, die (zB. wegen Symmetriegründen) **gleichwahrscheinlich** sind, die Wahrscheinlichkeit eines Ereignisses (E) ist:

$$p(E) = \frac{\text{Anzahl der für } E \text{ günstigen Elementarereignisse}}{\text{Anzahl aller gleichmöglichen Elementarereignisse}}$$

p =probability, Probabilität

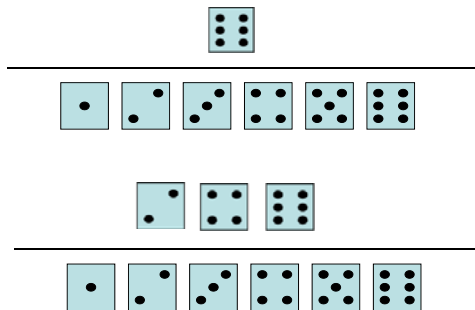
6

Würfelexperiment:

$$p(6) = \frac{1}{6}$$

$$p(E) = \frac{g}{m}$$

Beispiele



$$p(\text{gerade Zahl}) = \frac{3}{6}$$

Münzenexperiment:

$$p(\text{Kopf}) = \frac{1}{2}$$



7

Beispiele

Blutgruppenversuch

$$p(I^A I^A) = \frac{1}{4}$$

$$\frac{\{I^A I^A\}}{\{I^A I^A, I^A i, i I^A, ii\}}$$

$$p(\text{Blutgruppe} = A) = \frac{3}{4}$$

$$\frac{\{I^A I^A, I^A i, i I^A\}}{\{I^A I^A, I^A i, i I^A, ii\}}$$

$$p(\text{Blutgruppe} = 0) = \frac{1}{4}$$

$$\frac{\{ii\}}{\{I^A I^A, I^A i, i I^A, ii\}}$$



$$\Omega = \{I^A I^A, I^A i, i I^A, ii\}$$

8

„Fahrtversuch“ ? keine Symmetrie !

$$p(E) = \frac{g}{m} \neq \frac{1}{2}$$

Beispiel



9

Statistische Wahrscheinlichkeit

Zufallsexperiment \longrightarrow

Ereignis A



Ereignis B



Tritt bei n -maliger Durchführung eines Zufallsexperimentes ein bestimmtes Ereignis **A** k -mal auf, so bezeichnet man die in langen Versuchsreihen zu beobachtende relative Häufigkeit als **Wahrscheinlichkeit, $p(A)$** :

$$p(A) = \frac{k}{n}$$

10

Beispiele



Buffon (1707-1788): $p(A) = \frac{2048}{4040} = 0.5069$

Pearson (1857-1936): $p(A) = \frac{6019}{12000} = 0.5016$ $p(A) = \frac{12012}{24000} = 0.5005$

Häufigkeit der Krankheiten:

“Die Todesursachengruppe der bösartigen Neubildungen war in 2006 für 25,6% der Todesfälle verantwortlich” (<http://www.statistik.at>)

$p(\text{bösa. Neub. als Todesursache}) = 19\,056 / 74\,295 = 25,6\%$

die Wahrscheinlichkeit, dass ein Neugeborener das n -te Lebensjahr erlebt

Jahr	30	40	50	60	70	80	90
M	0.986	0.977	0.955	0.893	0.766	0.513	0.152
W	0.994	0.990	0.978	0.947	0.879	0.703	0.290

Österreich, 2006

Eigenschaften der Wahrscheinlichkeit

$$\rightarrow 0 \leq p(A) \leq 1$$

$$\rightarrow p(\text{sicheres Ereignis}) = 1$$

$$\rightarrow p(\text{unmögliches Ereignis}) = 0$$

$$\text{zB: } p(\text{Augenzahl des Würfels} < 10) = 1$$

$$p(\text{Augenzahl des Würfels} = 10) = 0$$

im unseren Blutgruppenversuch:

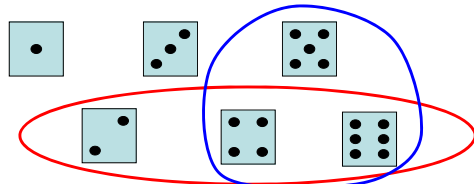
$$p(\text{Blutgruppe} = B) = 0$$

12

Additionsregel

$$p(A \text{ oder } B) = p(A) + p(B) - p(A \text{ und } B)$$

zB: Würfel: $p(\text{gerade oder } >3) = 4/6$
 $p(\text{gerade}) + p(>3) - p(\text{gerade und } >3)$
 $3/6 + 3/6 - 2/6 = 4/6$



für Ereignisse A und B $p(A \text{ oder } B) = p(A) + p(B)$
 ist gültig nur wenn A und B nicht gleichzeitig auftreten können (disjunkte Ereignisse, $p(A \text{ und } B) = 0$)

zB: Würfel: $p(5 \text{ oder } 6) = 1/6 + 1/6 = 1/3$
 Münze: $p(\text{Kopf oder Zahl}) = 1/2 + 1/2 = 1$

13

Unabhängige Ereignisse

Wenn die Ereignisse einander nicht beeinflussen:
 $p(A \text{ und } B) = p(A) \cdot p(B)$

zB: Würfelexperiment mit zwei Würfeln
 $p(\text{beide}=6) = p(\text{erste}=6) \cdot p(\text{zweite}=6) = 1/36$

Die Wahrscheinlichkeit, dass der Mann eines Ehepaares das 80. und die Frau das 70. Lebensjahr erreichen :

$$p(M_{80} \text{ und } W_{70}) = p(M_{80}) \cdot p(W_{70}) = 0.513 \cdot 0.879 = 0.451 = 45.1\%$$

Die Wahrscheinlichkeit, dass ein Neugeborener das n -te Lebensjahr erlebt

Jahr	30	40	50	60	70	80	90
M	0.986	0.977	0.955	0.893	0.766	0.513	0.152
W	0.994	0.990	0.978	0.947	0.879	0.703	0.290

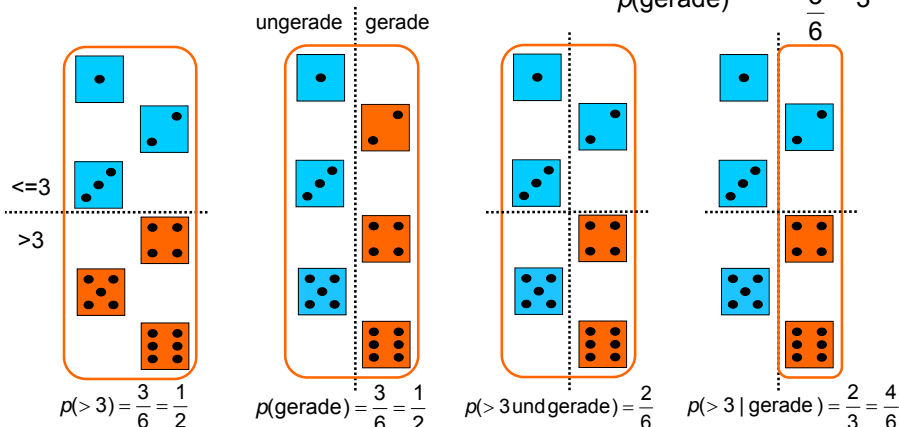
14

Bedingte Wahrscheinlichkeit

Die Wahrscheinlichkeit, dass A zutrifft unter der Voraussetzung, dass B eingetreten ist*.

$$p(A|B) = \frac{p(A \text{ und } B)}{p(B)}$$

zB: bei Würfelexperiment: $p(>3|\text{gerade}) = \frac{p(>3 \text{ und gerade})}{p(\text{gerade})} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$



*oder: $p(A \text{ gegeben } B)$, p von A vorausgesetzt B

Wie groß ist die Wahrscheinlichkeit, dass ein 50 jähriger Mann das 80. Lebensjahr erreicht?

Beispiele

Die Wahrscheinlichkeit, dass ein Neugeborener das n -te Lebensjahr erlebt

Jahr	30	40	50	60	70	80	90
M	0.986	0.977	0.955	0.893	0.766	0.513	0.152
W	0.994	0.990	0.978	0.947	0.879	0.703	0.290

$$p(M_{80}|M_{50}) = 0.513 / 0.955 = 0.537$$

Wie groß ist die Wahrscheinlichkeit, dass ein 60 jähriger Mann das 80. Lebensjahr erreicht?

$$p(M_{80}|M_{60}) = 0.513 / 0.893 = 0.574$$

Wie groß ist die Wahrscheinlichkeit, dass ein 70 jähriger Mann das 80. Lebensjahr erreicht?

$$p(M_{80}|M_{70}) = 0.513 / 0.766 = 0.670$$

16

Beispiel (Daten: USA 1991)

$p(\text{Krankheit} \mid \text{Risikofaktor})$

Risikofaktor: kleines Geburtsgewicht
„Krankheit“: Säuglingsmortalität

		Säuglingsmortalität		
		Krankheit ja	Krankheit nein	Teilsumme
Anzahl von Neugeborenen	Risikofaktor ja	$a=21054$	$b=271269$	$a+b$
	Risikofaktor nein	$c=14442$	$d=3804294$	$c+d$
	Teilsumme	$a+c$	$b+d$	$a+b+c+d$

$p(\text{Säuglingsmortalität} \mid \text{kleines Geburtsgewicht}) =$

$$= p(K_+ | R_+) = \frac{p(K_+ \text{ und } R_+)}{p(R_+)} = \frac{\frac{a}{a+b+c+d}}{\frac{a+b}{a+b+c+d}} = \frac{a}{a+b} = \frac{21054}{21054 + 271269} = 0.0720$$

$p(\text{Säuglingsmortalität} \mid \text{normales Geburtsgewicht}) =$

$$= p(K_+ | R_-) = \frac{p(K_+ \text{ und } R_-)}{p(R_-)} = \frac{\frac{c}{a+b+c+d}}{\frac{c+d}{a+b+c+d}} = \frac{c}{c+d} = \frac{14442}{14442 + 3804294} = 0.0037$$

17

Das relative Risiko

	K_+	K_-
R_+	a	b
R_-	c	d

Erkrankungswahrscheinlichkeit einer Risikogruppe relativ zur Erkrankungswahrscheinlichkeit einer nicht Risikogruppe.

$$RR = \frac{p(K_+ | R_+)}{p(K_+ | R_-)} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

- zB: Relatives Risiko der Säuglingsmortalität mit Risikofaktor von Geburtsgewicht:

$$RR = \frac{p(\text{Säuglingsmortalität} \mid \text{kleines Geburtsgewicht})}{p(\text{Säuglingsmortalität} \mid \text{normales Geburtsgewicht})}$$

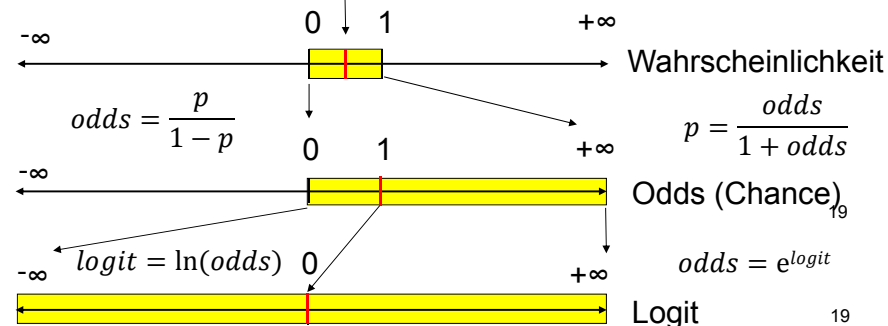
$$RR = \frac{0.0720}{0.0037} = 19.04$$

18

Charakterisierungsmöglichkeiten des Eintretens von Ereignissen

Ereignis A	Wahrscheinlichkeit $p(A)$	odds	logit
unmögliches Ereignis	0	0	$-\infty$
Ereignis A und Gegenereignis von A haben die gleichen Chancen	0.5	1	0
sicheres Ereignis	1	$+\infty$	$+\infty$

Ereignis A und Gegenereignis von A haben die gleiche Chance



19

Das Chancenverhältnis (Odds ratio)

	K_+	K_-
R_+	a	b
R_-	c	d

$$OR = \frac{\frac{p(K_+ | R_+)}{p(K_- | R_+)}}{\frac{p(K_+ | R_-)}{p(K_- | R_-)}} = \frac{\frac{\frac{a}{a+b}}{\frac{b}{a+b}}}{\frac{\frac{c}{c+d}}{\frac{d}{c+d}}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \left(\frac{\frac{a}{c}}{\frac{b}{d}} \right) = \frac{ad}{bc}$$

- zB: Chancenverhältnis der Säuglingsmortalität mit Risikofaktor von Geburtsgewicht:

$$OR = \frac{\frac{0.0720}{1-0.0720}}{\frac{0.00378}{1-0.00378}} = \frac{\frac{0.0720}{0.9280}}{\frac{0.0037}{0.99622}} = 20.4$$

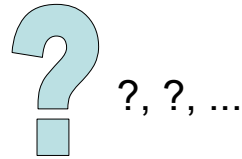
20

Verteilungen

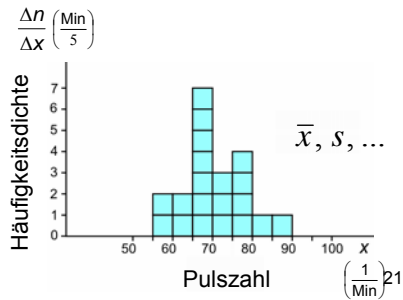
Population



Wahrscheinlichkeitsverteilung



Stichprobe



Verteilungen

Wie kann man die theoretische Verteilung bestimmen?

Vermutung

(nach dem Histogramm)

Modellannahme



Laplace-Prinzip:

wenn nichts dagegen spricht, gehen wir davon aus, dass alle Elementarereignisse gleich wahrscheinlich sind

Laplace-Experiment:

es meint ein Zufalls-Experiment bei dem davon ausgegangen wird, dass jeder Versuchsausgang **gleichwahrscheinlich** ist

➔ Gleichverteilung

22

Klassifizierung der Verteilungen

- **diskrete Verteilungen**
- diskrete Gleichverteilung
- Binomialverteilung
- Poisson Verteilung
- ...
- **kontinuierliche Verteilungen**
- kontinuierliche Gleichverteilung
- Normalverteilung
- Chi-Quadrat Verteilung
- t-Verteilung
- ...

diskrete Zufallsgröße

zB: Anzahl der Kranken, Augenzahl des Würfels

kontinuierliche Zufallsgröße

zB: Blutdruck, Körperhöhe,...

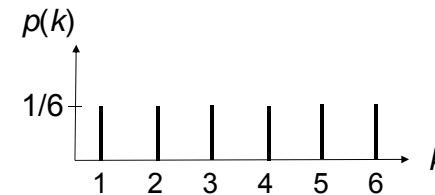
23

Diskrete Gleichverteilung



Beispiel:

Wertebereich	1	2	3	4	5	6
Wahrscheinlichkeit	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



$$p(k) = \frac{1}{6}, \quad k = 1, 2, \dots, 6$$

weitere Beispiele:

Münzenversuch



Würfelexperiment mit einem Ikosaeder



24

Lageparameter der Verteilung

Es sei X eine diskrete Zufallsgröße mit Werten x_1, x_2, \dots dann heisst

$$\mu = \sum_i x_i p(x_i) \quad \text{Erwartungswert von } X.$$

Der Erwartungswert gibt denjenigen Wert an, den man als Mittelwert (durchschnittlichen Wert) über viele Versuchswiederholungen "erwarten" kann.

Dabei ist es durchaus möglich, dass der Erwartungswert bei keinem einzigen Versuch realisiert wird oder sogar überhaupt nicht vorkommen kann.

25

Nützliche Formel des arithmetischen Mittelwertes

(ungewogenes)
arithmetisches Mittel

Berechnung aus
Einzelbeobachtungen

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

gewogenes
arithmetisches Mittel

Berechnung aus
gruppierten Daten
(Merkmalausprägungen)

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_m x_m}{n_1 + n_2 + \dots + n_m} = \frac{\sum_{i=1}^m n_i x_i}{\sum_{i=1}^m n_i} = \frac{\sum_{i=1}^m n_i x_i}{n} = \sum_{i=1}^m h_i x_i = \sum_{i=1}^m x_i h_i$$

n_i : absolute Häufigkeit, h_i : relative Häufigkeit

26

Erwartungswert und Durchschnittswert

$$\mu = \sum_i x_i p(x_i) \quad \bar{x} = \sum_i x_i h_i$$

Beispiel: 100 Würfelexperimente. 2,5,4,3,6,6,1,5,4,2,3...

Insgesamt:

x_i	n_i	h_i
1	15	15/100
2	20	20/100
3	14	14/100
4	16	16/100
5	18	18/100
6	17	17/100

$$\bar{x} = \frac{15 \cdot 1 + 20 \cdot 2 + 14 \cdot 3 + 16 \cdot 4 + 18 \cdot 5 + 17 \cdot 6}{100} =$$

$$= \frac{15}{100} \cdot 1 + \frac{20}{100} \cdot 2 + \frac{14}{100} \cdot 3 + \frac{16}{100} \cdot 4 + \frac{18}{100} \cdot 5 + \frac{17}{100} \cdot 6 = 3.53 =$$

$$= h(1) \cdot 1 + h(2) \cdot 2 + h(3) \cdot 3 + h(4) \cdot 4 + h(5) \cdot 5 + h(6) \cdot 6 \rightarrow$$

$$\xrightarrow{n \rightarrow \infty} P(1) \cdot 1 + P(2) \cdot 2 + P(3) \cdot 3 + P(4) \cdot 4 + P(5) \cdot 5 + P(6) \cdot 6 = \mu$$

x_i : Augenzahl

n_i : absolute Häufigkeit

h_i : relative Häufigkeit

$$\bar{x} \xrightarrow{n \rightarrow \infty} \mu$$

27

Streuung der Verteilung

Es sei X eine diskrete Zufallsgröße mit Werten x_1, x_2, \dots und mit dem Erwartungswert μ . Dann nennt man die Zahl

$$\sigma^2 = \sum_i (x_i - \mu)^2 p(x_i)$$

als Varianz von X , ihre Wurzel als (theoretische) Streuung (σ).

$$s \xrightarrow{n \rightarrow \infty} \sigma$$

empirische
Streuung

theoretische
Streuung

(Standardabweichung)

28

Erwartungswert und Streuung der Gleichverteilung

$$x_i = 1, 2, \dots, n \quad p(x_i) = 1/n$$

$$\mu = \sum_{i=1}^n x_i p(x_i) = \sum_{i=1}^n i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2} = \mu$$

$$\sigma^2 = \frac{n^2 - 1}{12}$$

$$\sigma = \sqrt{\frac{n^2 - 1}{12}}$$

$$\text{zB: Würfel: } n=6 \quad \mu=3,5 \quad \sigma^2=35/12=2,92.. \quad \sigma=1,71\dots$$

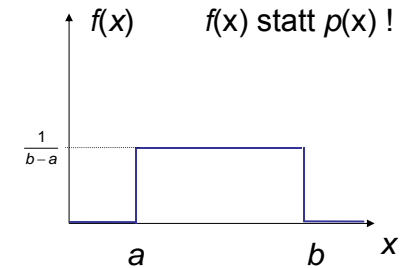
$$\begin{aligned} \sigma^2 &= \sum_i (x_i - \mu)^2 p(x_i) = \sum_i \left(i - \frac{n+1}{2} \right)^2 \frac{1}{n} = \text{Ergänzungsmaterial} \\ &= \frac{1}{n} \sum_i \left(i^2 + \frac{(n+1)^2}{4} - i(n+1) \right) = \frac{1}{n} \sum_i i^2 + \frac{1}{n} \sum_i \frac{(n+1)^2}{4} - \frac{n+1}{n} \sum_i i = \\ &= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} + \frac{1}{n} n \frac{(n+1)^2}{4} - \frac{n+1}{n} \frac{n(n+1)}{2} = \dots = \frac{n^2 - 1}{12} \end{aligned}$$

29

Kontinuierliche Gleichverteilung

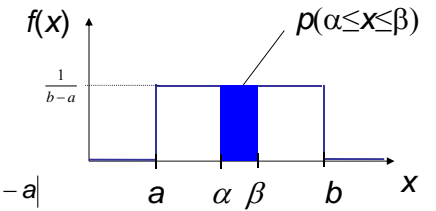
Verteilungsdichtefunktion:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{für } a \leq x \leq b \\ 0, & \text{sonst} \end{cases}$$



Wahrscheinlichkeit entspricht der Fläche

$$\text{Erwartungswert: } \mu = \frac{a+b}{2}$$



$$\text{Streuung: } \sigma = \frac{|b-a|}{\sqrt{12}} = 0.289 |b-a|$$

30

Ergänzungsmaterial

$$\begin{aligned} \mu &= \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b x \frac{1}{b-a} dx = \left[\frac{x^2}{2} \frac{1}{b-a} \right]_a^b = \\ &= \left(\frac{b^2}{2} \frac{1}{b-a} \right) - \left(\frac{a^2}{2} \frac{1}{b-a} \right) = \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = \int_a^b (x^2 - 2\mu x + \mu^2) \frac{1}{b-a} dx = \\ &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{2\mu}{b-a} \int_a^b x dx + \frac{\mu^2}{b-a} \int_a^b 1 dx = \\ &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b - \frac{2\mu}{b-a} \left[\frac{x^2}{2} \right]_a^b + \frac{\mu^2}{b-a} [x]_a^b = \dots = \frac{(b-a)^2}{12} \end{aligned}$$

31