Mittelwert

Sinn und Unsinn eines Alltagsbegriffs

Friedrich Barth

Der klassische Mittelwert

Arithmetisches Mittel - Erwartungswert

ist **gut**

- für Verteilungen

  eingipflig und **symmetrisch**

- stabiler als Einzelwerte:

  $\sqrt{n} - \text{Gesetz}$

- neue Werte drängen zur Mitte:
  Regression Galton (1877)

problematisch bei **Ausreißern**

Korrektur:

gestutztes, winsorisiertes Mittel:

Ein 10 % winsorisiertes Mittel erhält man, wenn man 5 % aller Werte am unteren und 5 % am oberen Ende auslässt.

Anwendung z.B. im Sport

Probleme

- mehrgipflige
- unsymmetrische Verteilungen
- Informationsverlust

  durch Verdichtung
- verschiedene Grundmengen

Seltsames

- Eine deutsche Frau hat 2005 im Mittel 1,4 Kinder.
- Alle Personen im Saal haben überdurchschnittlich viele Arme und Beine.
- Im Mittel sind die Teilnehmer 40 Jahre alt.
  (Zehn 70-jährige mit je einem Enkelkind)
- Die meisten Deutschen sind überdurchschnittlich gute Autofahrer
- Eine Aktie von 1000€ steigt im ersten Jahr um 60% auf 1600€ und fällt im zweiten um 50% auf 800€.
  Die Aktie steigt also im Mittel pro Jahr um
  (60% – 50%):2 = 5% (?) >>> geometr. Mittel

- Bei der Hinfahrt fährt ein Autofahrer mit 100 km/h, bei der Rückfahrt mit 160 km/h, also im Mittel mit 130 km/h. Wenn er dort bleibt (Rückfahrt mit 0 km/h), fährt er dann im Mittel mit 50 km/h ?

  >>> harmonisches Mittel

Fazit:

Das Arithmetische Mittel macht meist nur Sinn als „Etwa-Wert“ einer eingipfligen (unimodularen) und annähernd symmetrischen Verteilung.

Alternativen sind z. B. Median und Modus oder die anderen Mittel.

Die klassischen Mittelwerte

Arithmetisches Mittel $\quad A = \frac{a+b}{2}$

Geometrisches Mittel $\quad G = \sqrt{ab}$

Harmonisches Mittel $\quad H = \frac{2ab}{a+b}$

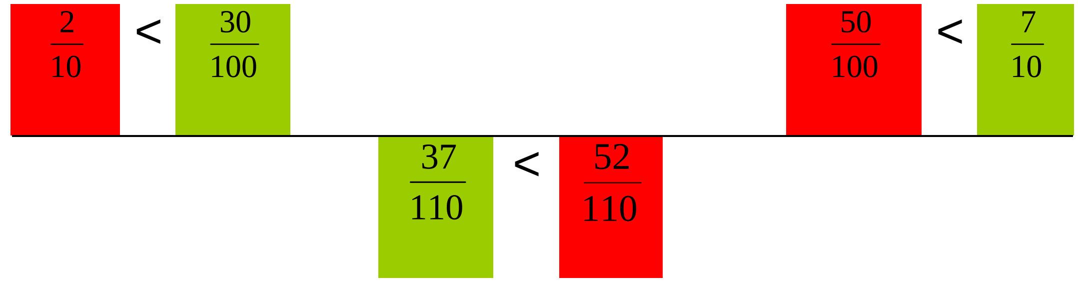
Es gilt: $A \cdot H = G^2$

Daneben gibt es auch kuriosere Mittel, wie das **Bruchmittel** zweier Brüche, das beim Simpsonparadox eine entscheidende Rolle spielt:

$$u = \frac{a}{c} \qquad\qquad v = \frac{b}{d}$$

$$m_b(u,v) = \frac{a+b}{c+d}$$

Beispiel zu Simpson



Ein Medikament A wirkt

sowohl bei Männern wie auch bei Frauen

besser als das Medikament B.

Bei Menschen wirkt aber B besser als A.

$$A = \frac{a+b}{2} \qquad G = \sqrt{ab} \qquad H = \frac{2ab}{a+b}$$

Allgemeines Mittel m: $\quad f(m) = \dfrac{f(a) + f(b)}{2}$

$$f(x) = kx + t \quad \Rightarrow \quad m = A$$

$$f(x) = \ln(x) \quad \Rightarrow \quad m = G$$

$$f(x) = \frac{k}{x} \quad \Rightarrow \quad m = H$$

Mittlerer Wert einer Funktion über einem Intervall [a,b] = Höhe des flächenbilanzgleichen Rechtecks über [a,b]

$$m = \frac{1}{b-a} \cdot \int_a^b f(x)dx$$

S. R. Wassell (1998)

nach Archytas von Tarent (430 - 345 v. Chr.):

$$\frac{m - a}{b - m} = \frac{a}{x} \quad \Rightarrow \quad m(x) = a \cdot \frac{b + x}{a + x}$$

$m(a) = A \qquad m(b) = H \qquad m(G) = G$

Also ist G Attraktor der Iteration $x_{k+1} = m(x_k)$

S. R. Wassell (1998)

nach Archytas von Tarent (430 - 345 v. Chr.):

$$\frac{m - a}{b - m} = \frac{a}{x} \quad \Rightarrow \quad m(x) = a \cdot \frac{b + x}{a + x}$$

m(a) = A   m(b) = H   m(G) = G

Also ist G Attraktor der Iteration $x_{k+1} = m(x_k)$

Cauchy-Hölder-Minkowski : $m(p) = \left(\frac{a^p + b^p}{2}\right)^{1/p}$

p = 1; 0; -1 liefert A; G; H

S. R. Wassell (1998)

nach Archytas von Tarent (430 - 345 v. Chr.):

$$\frac{m - a}{b - m} = \frac{a}{x} \quad \Rightarrow \quad m(x) = a \cdot \frac{b + x}{a + x}$$

$m(a) = A \qquad m(b) = H \qquad m(G) = G$

Also ist G Attraktor der Iteration $x_{k+1} = m(x_k)$

Cauchy-Hölder-Minkowski : $m(p) = \left(\dfrac{a^p + b^p}{2}\right)^{1/p}$

p = 1; 0; -1 liefert A; G; H

Polya-Beckenbach-Lehmer : $m(p) = \dfrac{a^{p+1} + b^{p+1}}{a^p + b^p}$

p = 0; -0,5; -1 liefert A; G; H

Erwartungswert

- mit Wahrscheinlichkeiten gewichtetes arithmetisches Mittel
- Mittelwert einer Zufallsgröße auf lange Sicht
- Durchschnitt
- „normaler Wert“ der Zufallsgröße
- Wert eines Angebots

$$E(X) = \mu = \sum_{k=1}^{\infty} x_k \cdot p_k = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

Eigenschaften des Erwartungswerts

1. Gebräuchlichster Mittelwert
2. Gibt in etwa die Lage der Werte der Zufallsgröße an
3. Leicht zu berechnen
4. Das gewichtete Mittel der Abweichungen vom Erwartungswert ist null
5. Das gewichtete Mittel der quadratischen Abweichungen vom Erwartungswert ist minimal
6. Der Erwartungswert wird merklich beeinflusst durch Ausreißer

Der erste schwarze König

Ein Kartenspiel (52 Karten) wird gemischt, die Karten werden der Reihe nach abgehoben.
Gewonnen hat, wer die Stelle S des ersten schwarzen Königs angibt.

P(k) = P(Der 1. schw. König liegt an der k-ten Stelle)

$$= \frac{52 - k}{26 \cdot 51}$$

$$E(S) = 17\frac{2}{3}$$

Ist E(S) eine gute Wahl?

Bei schiefen Verteilungen ist der Erwartungswert zwar das Mittel auf lange Sicht, der Modus aber bei einer Einzelentscheidung sinnvoller.

Rencontreproblem

Montmort 1708, De Moivre 1718
Euler 1751
Die Zahlen 1,2,…,n werden permutiert.
Ein Rencontre bei k liegt vor,
wenn die Zahl k an k-ter Stelle liegt.

Die Wahrscheinlichkeit für ein Rencontre
an jeder Stelle k ist 1/n.
Auf welche Stelle sollte man setzen?
Die mittlere Stelle ist (n+1)/2,
die wahrscheinlichste Stelle fürs erste Rencontre ist
aber ist die erste!

Der Erwartungswert der Zufallsgröße
„Anzahl der Rencontres“ ist unabhängig von n gleich 1!.

Das Petersburger Problem

Nikolaus und Daniel Bernoulli, Petersburger Commentarien (1738)

A wirft eine Münze und B verpflichtet sich, ihm 2^{k-1} Taler zu geben, wenn Wappen beim k-ten Wurf zum ersten Mal fällt. Damit endet das Spiel.

Was wäre ein fairer Einsatz für A?

Zufallsgröße: Auszahlung Z $\quad W(2^{k-1}) = \frac{1}{2^k}$

Der Erwartungswert für k Würfe ist $k \cdot \frac{1}{2}$

Er wird unendlich für k $\to \infty$!

Gabriel Cramer (1704 - 1752): ... die Rechnung liefert für den Einsatz einen unendlich großen Wert, aber kein halbwegs vernünftiger Mensch wird mehr als 20 Taler einsetzen.

Einschätzung von Wahrscheinlichkeiten nach Kitaigorodski und Borel.

p	Bedeutung	Wurfzahl	Auszahlung
$< 0{,}05$	signifkante Abw.	≥ 5	2,5 Taler
$< 0{,}01$	hochsignifkante Abw.	≥ 7	3,5 Taler
$< 10^{-6}$	persönlich unmöglich	≥ 20	10 Taler
$< 10^{-15}$	praktisch unmöglich	≥ 50	25 Taler
$< 10^{-50}$	„unmöglich“	≥ 167	83,5 Taler

Der Erwartungswert täuscht die Erwartung

A

a	1	3
W(a)	0,2	0,8

E(A) = 0,2 + 2,4 = 2,6

B

b	2	4
W(b)	0,8	0,2

E(B) = 1,6 + 0,8 = 2,4

A hat den größeren Erwartungswert,
aber B ist an jeder Stelle größer.

A

a	1	3
W(a)	0,2	0,8

E(A) = 0,2 + 2,4 = 2,6

B

b	2	4
W(b)	0,8	0,2

E(B) = 1,6 + 0,8 = 2,4

Händler A und B bieten Waren zu verschiedenen Preisen an.
Unterschiedliche Verkaufszahlen führen zu unterschiedlichen „mittleren Preisen“.

A

a	1	3
W(a)	0,2	0,8

E(A) = 0,2 + 2,4 = 2,6

B

b	2	4
W(b)	0,8	0,2

E(B) = 1,6 + 0,8 = 2,4

Zwei Lotterien A und B.
In B liegen die höheren Gewinne,
aber die Verteilung in A ist günstiger.

A

a	1	3
W(a)	0,2	0,8

E(A) = 0,2 + 2,4 = 2,6

B

b	2	4
W(b)	0,8	0,2

E(B) = 1,6 + 0,8 = 2,4

Zwei Urnen A(1,3,3,3,3) und B(2,2,2,2,4).
Man zieht einmal, die größere Zahl hat gewonnen.
P(A siegt) = 0,8·0,8 = 0,64 > 50%,
obwohl in A die kleineren Zahlen liegen.
Der Erwartungswert spricht auch für A!

Der Erwartungswert als unzuverlässiger Helfer

Efrons nicht-transitive Würfel (1970)

A: 111555 B: 004444 C: 333333 D: 222266

Wähle einen Würfel, der Gegner nimmt einen anderen.
Wer die höhere Zahl wirft, hat gewonnen.

E(A) = E(C) = 3 $E(B) = 2\frac{2}{3}$ $E(D) = 3\frac{1}{3}$

A schlägt B B schlägt C C schlägt D D schlägt A

jeweils mit $p = \frac{2}{3}$

E ignoriert Details der Verteilung und bringt daher nichts!

Probleme beim Ändern der Bezugsmenge

Ist Heiraten für Männer gesund ?

Westergard: „Die Lehre von der Mortalität und Morbidität".
(Jena, 1901)

Sterblichkeit von Männern pro Jahr in Promille:

Alter	Ledig	Verheiratet	Gesamt
22 – 26	6,70	3,80	6,13
27 – 31	7,80	4,19	5,89
32 – 36	8,63	4,86	5,88

Die Sterblichkeit ist bei Junggesellen immer höher als bei Ehemännern; sie nimmt bei beiden Gruppen mit zunehmendem Alter zu, insgesamt aber ab.

Umgang mit Sterbetafeln

x = Lebensjahr

l_x = Anzahl der Lebenden Anfang des Jahres x

d_x = Anzahl der Todesfälle im Jahre x,

„er wurde x Jahre alt“

$$q_x = \frac{d_x}{l_x} = \text{Sterbewahrscheinlichkeit im Jahr x}$$

$1 – q_x$ = Überlebenswahrscheinlichkeit im Jahr x

$$l_x = l_{x-1} \cdot (1 - q_{x-1})$$

$$r_x = \frac{d_x}{l_0} = \frac{q_x \cdot l_x}{l_0} = q_x \cdot (1 - q_0) \cdot (1 - q_1) \cdot \ldots \cdot (1 - q_{x-1}) =$$

= Anteil derer, die im x-ten Lebensjahr sterben; $r_0 = q_0$

x	**l_x**	**d_x**	**q_x**	**$1 - q_x$**	**r_x**
0	100	36	0,36	0,64	0,36
1	64	24	0,38	0,62	0,24
2	40	15	0,38	0,62	0,15
3	25	9	0,36	0,64	0,09
4	16	6	0,38	0,62	0,06
5	10	4	0,40	0,60	0,04
6	6	3	0,50	0,50	0,03
7	3	2	0,67	0,33	0,02
8	1	1	1,00	0,00	0,01
9	0	0	nd	nd	nd

$E_x(A)$ = Lebenserwartung im Jahre x = Erwartungswert der Zufallsgröße A = „Sterbejahr“

$$E_0(A) = \sum_{k=0}^{\infty} k \cdot r_k = 1{,}65 \qquad E_5(A) = \sum_{k=5}^{\infty} k \cdot r_k = 6{,}00$$

Wunderbare Geldvermehrung

Das Geldbeutel-Paradoxon

Maurice Kraitchik, Mathematical Recreations (1953)
The Paradox of the Neckties

A und B legen ihre unterschiedlich gefüllten Geldbeutel auf den Tisch. Wer weniger hat erhält das Geld des anderen.

A überlegt:
Wenn ich verliere, dann verliere ich das, was ich habe.
Wenn ich gewinne, dann gewinne ich mehr als ich habe.
Das Spiel ist also günstig für mich !

B überlegt genau so !

Wunderbare Geldvermehrung ! ?

Analyse:

A habe a, B habe b

a < b und b < a mit gleicher Wahrscheinlichkeit

GA:= Gewinn des A GB:= Gewinn des B

GA	b	–a
W(GA)	0,5	0,5

$E(GA) = 0{,}5(b - a)$ $E(GB) = 0{,}5(a - b)$

Beide Erwartungswerte sind mit gleicher Wahrscheinlichkeit positiv oder negativ und bis aufs Vorzeichen gleich.

Beachte:
Wenn A verliert, dann verliert er relativ viel,
Wenn er gewinnt, dann gewinnt er relativ wenig.

Das Paradoxon der zwei Umschläge

Linzer, E. (1994): The two envelope paradox
Chalmers, D. (1996): The Petersburg two envelope paradox

(1) Auf dem Tisch liegen zwei Umschläge A und B.
Einer enthält einen Geldbetrag x, der andere das Doppelte 2x.
Man darf einen Umschlag wählen und das Geld behalten.

Der Spieler wählt A und überlegt:
In B liegt mit gleicher Wahrscheinlichkeit entweder die Hälfte oder das Doppelte.

Liegt in A der Betrag a, dann ist der Erwartungswert für B

$$\frac{1}{2}\cdot\frac{1}{2}a+\frac{1}{2}\cdot 2a = \frac{5}{4}a$$

also 25% mehr als in A

Man sollte (immer wieder !) tauschen ! ?

(2) Petersburger Variante:
Die Umschläge werden nach folgender Regel gefüllt:
Man wirft für jeden Umschlag eine Münze.

Kommt Adler beim i-ten Wurf zum ersten Mal, dann
Legt man 2^i € hinein.

Der Erwartungswert für den Inhalt eines Umschlags ist

$$E(A) = E(B) = \sum_{i=1}^{\infty} 2^{-i} \cdot 2^{i} = \infty$$

Liegt in A irgendein Betrag a, dann sollte man tauschen,
weil der Erwartungswert für B unendlich viel größer ist.

Bemerkung zu (1): $$\frac{1}{2} \cdot x + \frac{1}{2} \cdot 2x = \frac{3}{2} x$$

Der Erwartungswert ist gleich für A und B.
Tauschen bringt nichts !

Die genannte Überlegung beschreibt ein anderes Zufallsexperiment:
Man betrachtet den Inhalt von A und entscheidet durch Münzenwurf, ob man in B $\frac{1}{2}y$ oder $2y$ legt.

Bemerkung zu (2): $$a = 2^m$$

Tausche, wenn $E(B) > a$
$p < 2^{-20}$ wird ignoriert !

Erwartungstreue Schätzer

Warum hat der Taschenrechner zwei unterschiedliche Tasten für die Varianz ?

Eine Schätzgröße T für den Parameter t heißt erwartungstreu, wenn gilt: E(t) = t

Die relative Haufigkeit ist ein erwartungstreuer Schätzer für die Wahrscheinlichkeit: X_i sei die Anzahl der Treffer beim i-ten Versuch

$H = \frac{1}{n} \cdot (X_1 + X_2 + ... + X_n)$ ist die relative Häufigkeit mit E(H) = p

$\boxed{2 \cdot \overline{X} - 1}$ ist ein erwartungstreuer Schätzer der Anzahl einer von 1 bis N nummerierten Menge.

$$E(2 \cdot \overline{X} - 1) = 2 \cdot \frac{N+1}{2} - 1 = N$$

Für eine binomial verteilte Zufallsgröße X gilt:

Ist X_k der Wert der Zufallsgröße X bei der k-ten unabhängigen Wiederholung, dann sind

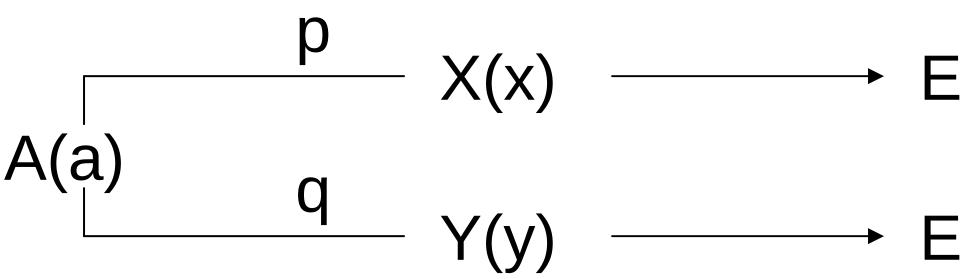
$$\boxed{\overline{X} = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k} \quad \text{und} \quad \boxed{S^2 = \frac{1}{n-1} \cdot \sum_{k-1}^{n} (X_k - \overline{X})^2}$$

erwartungstreue Schätzer des Erwartungswerts μ bzw. der Varianz σ^2 .

Mittelwertsregeln

Pierre Rémond de Montmort (1678 - 1719)
Essay d‘Analyse sur les Jeux de Hazard (1708)

Baum:



Die erste und die zweite Pfadregel liefern die
Mittelwertsregel für Wahrscheinlichkeiten

$$\boxed{a = px + qy}$$

mit $a = P_A(E)$ $\quad x = P_X(E)$ $\quad y = P_Y(E)$

$p = P_A(X)$ $\quad q = P_A(Y)$

a ist das mit p und q gewichtete Mittel von x und y.

Analog bestimmt man den bedingten Erwartungswert $E_A(S)$ der Anzahl der Schritte S_A bis zum Eintreten von E, ausgehend von A

S_A	$1 + E(S_X)$	$1 + E(S_Y)$
	p	q

$$E(S_A) = (1 + E(S_X))\cdot p + (1 + E(S_Y))\cdot q = 1 + E(S_X)\cdot p + E(S_Y)\cdot q$$

Mittelwertsregel für die Schrittzahl

$$\boxed{a = 1 + px + qy}$$

mit $a = E(S_A)$ $x = E(S_X)$ $y = E(S_Y)$

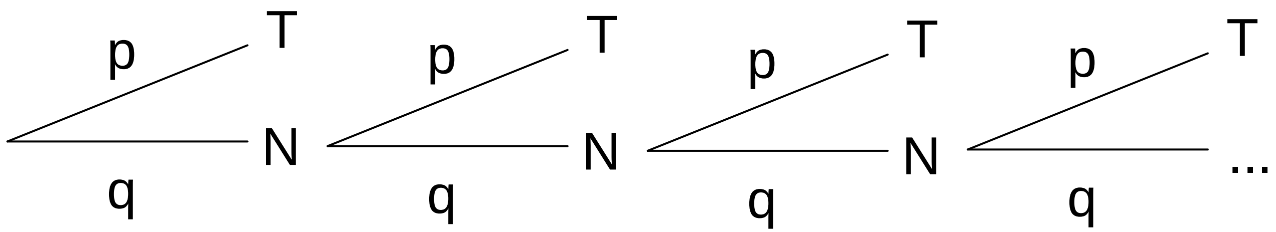
$p = P_A(X)$ $q = P_A(Y)$

Beispiele zu den Mittelwertsregeln

Warten auf den ersten Treffer

$p = P(T)$ S = Anzahl der Versuche bis zum ersten Treffer

S	1	2	...	k	...
W	p	qp	...	$q^{k-1}p$	...

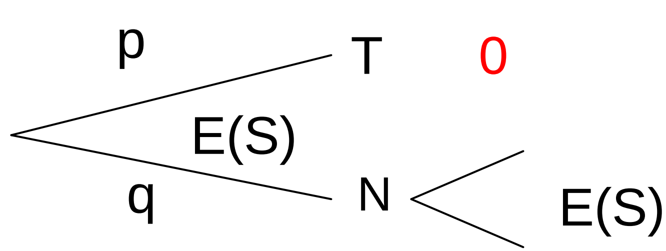


$$E(S) = p\cdot(1 + 2q + 3q^2 + ... + kq^{k-1} + ...) = ... = \frac{1}{p}$$

Mittelwertsregeln:

$$E(S) = 1 + p\cdot 0 + q\cdot E(S)$$

$$\Rightarrow E(S) = \frac{1}{p}$$



Das Penney-Paradoxon

Walter Penney (1969), Journal of Recreational Mathematics

Eine L-Münze (0|1) wird so lange geworfen bis ein bestimmtes Muster (z.B. 110) erscheint.
Jeder Spieler wählt ein n-stelliges Muster.
Gewonnen hat der, dessen Muster zuerst kommt.

Für n = 2 ist die Welt in Ordnung:

Die schlechten sind erwartungsgemäß 00 und 11, die andern beiden sind gleich gut.
Zu jedem schlechten gibt es ein gutes, das es mit 3 : 1 schlägt.
Die Wartezeiten sind entsprechend 6 für xx und 4 für xy.

n = 3:

Naive Vorstellung: P({abc}) = 12,5% ; alle sind gleich gut.

Aber: Zu jedem Muster gibt es ein besseres !

100 vs 000 : $\frac{7}{8}$ 100 vs 001 : $\frac{3}{4}$

001 vs 010 : $\frac{2}{3}$ 001 vs 011 : $\frac{2}{3}$ 110 vs 100 : $\frac{2}{3}$

110 vs 101 : $\frac{2}{3}$

Wartezeit für 110 ist 8,
für 101 aber 10, fürs Spielende 6

011 vs 110 : $\frac{3}{4}$ 011 vs 111 : $\frac{7}{8}$

Erfreulicherweise stimmt hier wenigstens der Satz:
Das Muster mit der längeren Wartezeit ist schlechter.

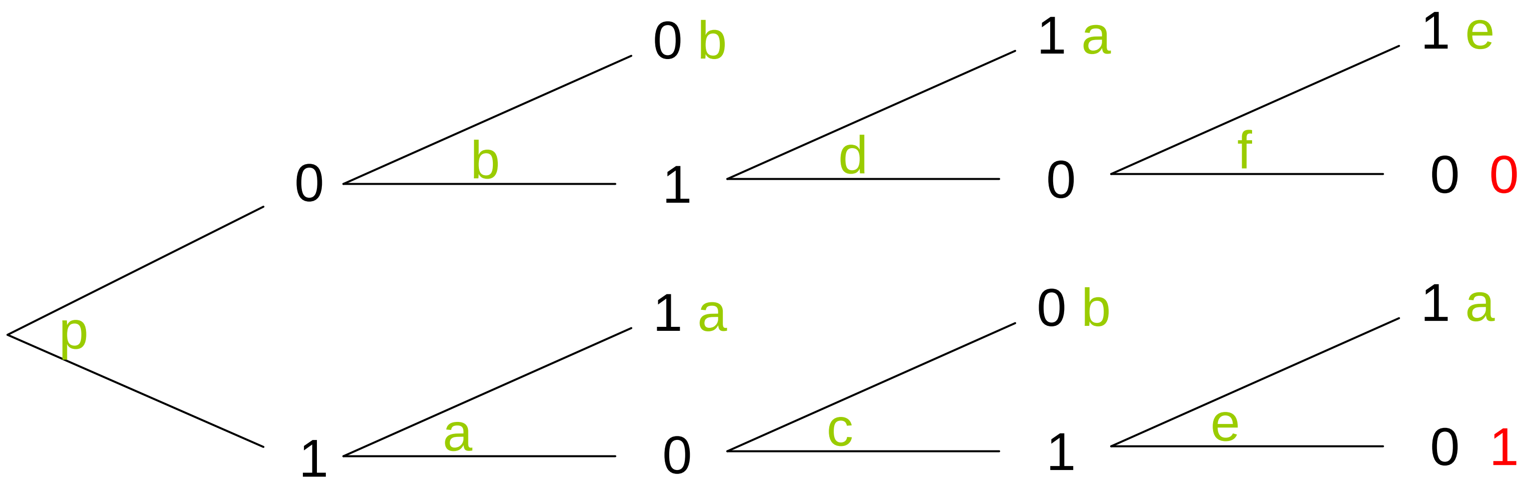
n = 4:

Wartezeit für 0100 ist 18, Wartezeit für 1010 ist 20,

Aber 1010 schlägt 0100 deutlich mit $\frac{9}{14}$.

0100 kommt offenbar insgesamt öfter,
aber seltener vor 1010.

1010 gegen 0100



$$c = \frac{1}{2}b + \frac{1}{2}e \qquad b = \frac{1}{2}d + \frac{1}{2}b \qquad a = \frac{1}{2}c + \frac{1}{2}a \qquad p = \frac{1}{2}a + \frac{1}{2}b$$

$$f = 0 + \frac{1}{2}e \qquad d = \frac{1}{2}a + \frac{1}{2}f \qquad e = \frac{1}{2}a + \frac{1}{2}$$

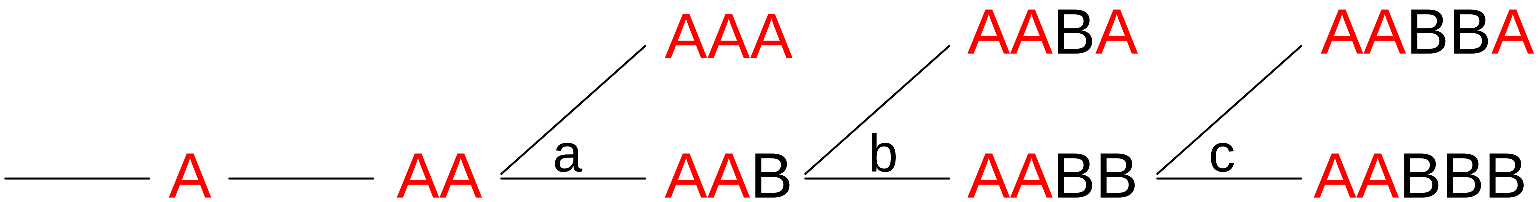
$$a = \frac{5}{7} \quad b = \frac{4}{7} \quad c = \frac{5}{7} \quad d = \frac{4}{7} \quad e = \frac{3}{7} \quad f = \frac{3}{7} \qquad p = \frac{9}{14}$$

Lotto-Show

ARD-Sendung (1998-2000)

6 Spieler, 2 davon spielen schließlich um 1 Million:
In einer Urne liegen 6 Kugeln, dreimal A und dreimal B.
Man zieht ohne Zurücklegen so lange, bis alle drei
Kugeln eines der Kandidaten gezogen sind.
Er gewinnt die Million.

Zwei Kugeln sind gezogen: AA
Welchen Wert hat das Spiel jetzt für A?
Um welchen Betrag fällt der Wert, wenn der dritte
Zug eine B-Kugel ist ?



Mittelwertsregeln: $a = \frac{1}{4} + \frac{3}{4} \cdot b$ $\quad b = \frac{1}{3} + \frac{2}{3} \cdot c$

$\boxed{c = \frac{1}{2}}$ $\quad \boxed{b = \frac{2}{3}}$ $\quad \boxed{a = \frac{3}{4}}$

Beim Stand AA hat das Spiel den Wert 750 000,
beim Stand AAB hat das Spiel den Wert $666\,666\frac{2}{3}$.

Pascals Wette

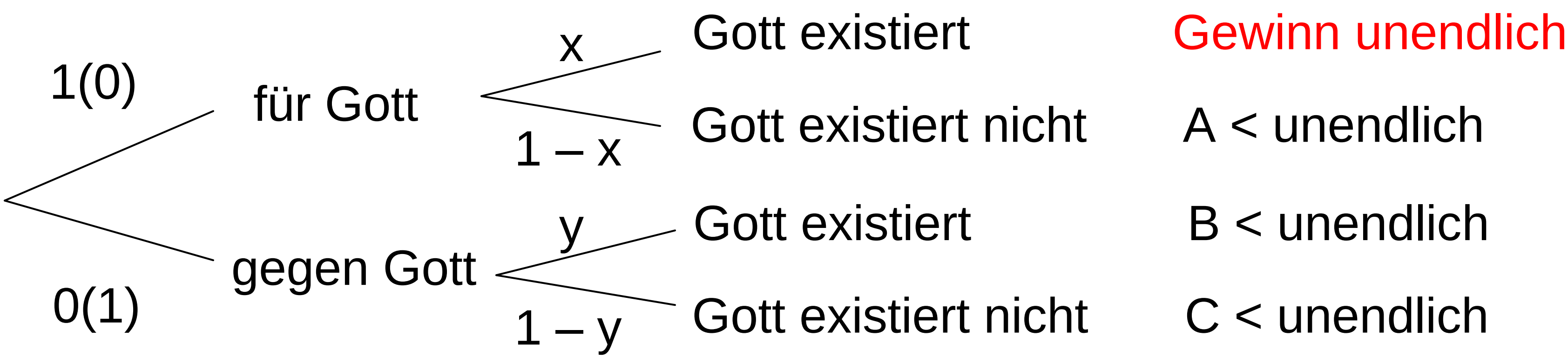
Le Pari: Infini - Rien (Pensées, Section III, 1670)

1) Die Annahme:
   „Es gibt eine persönliche Existenz nach dem Tod.“
   ist vernünftig.
   Sie kann nur bestätigt, aber nicht widerlegt werden.
2) Nach dem Tod ist Unendlichkeit zu erwarten.
   Unendliche Zeit oder gar keine (?)
3) Pascal: Dieu est, ou il n‘est pas

Entscheidung zwischen den zwei Möglichkeiten mit allen Konsequenzen:

- Es gibt einen Gott, der das Leben nach dem Tod auf Grund des Lebens vor dem Tod einrichtet.
- Es gibt keinen Gott; mit dem Tod ist alles aus.

Pesons le gain et la perte



Bedingte Erwartungswerte:

$E_{\text{Entscheidung für Gott}}(\text{Gewinn}) = x \cdot \infty + (1 - x) \cdot A = \infty$

$E_{\text{Entscheidung gegen Gott}}(\text{Gewinn}) = y \cdot B + (1 - y) \cdot C < \infty$