Bayes-statistik -- Einführung

Klassisch: „Frequenzbestimmt” (frequentist)

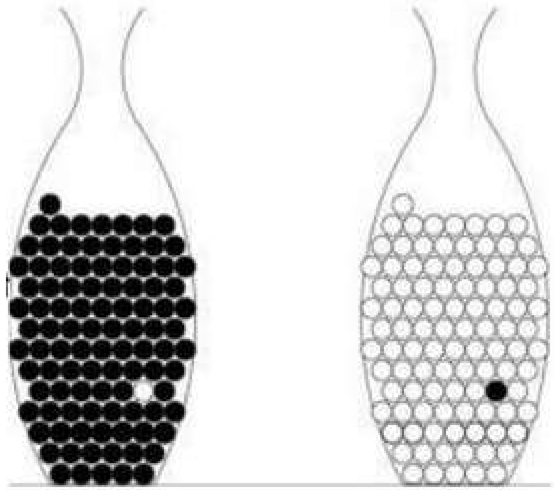
P ~ rel.Häufigkeit

Oder:
rel.Häufigkeit -> P **wenn** N-> ∞

Beispiel:

Kopf oder Zahl: Münze werfen

Wahrscheinlichkeit: das Urnenmodell.



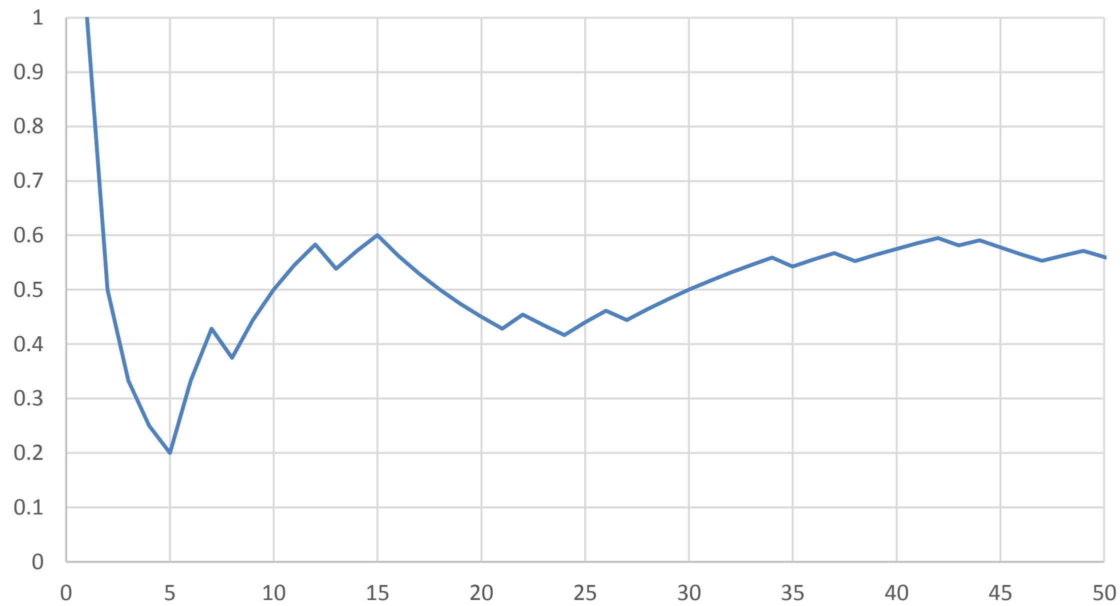
Beliebige Wahrscheinlichkeiten können wir so beschreiben.

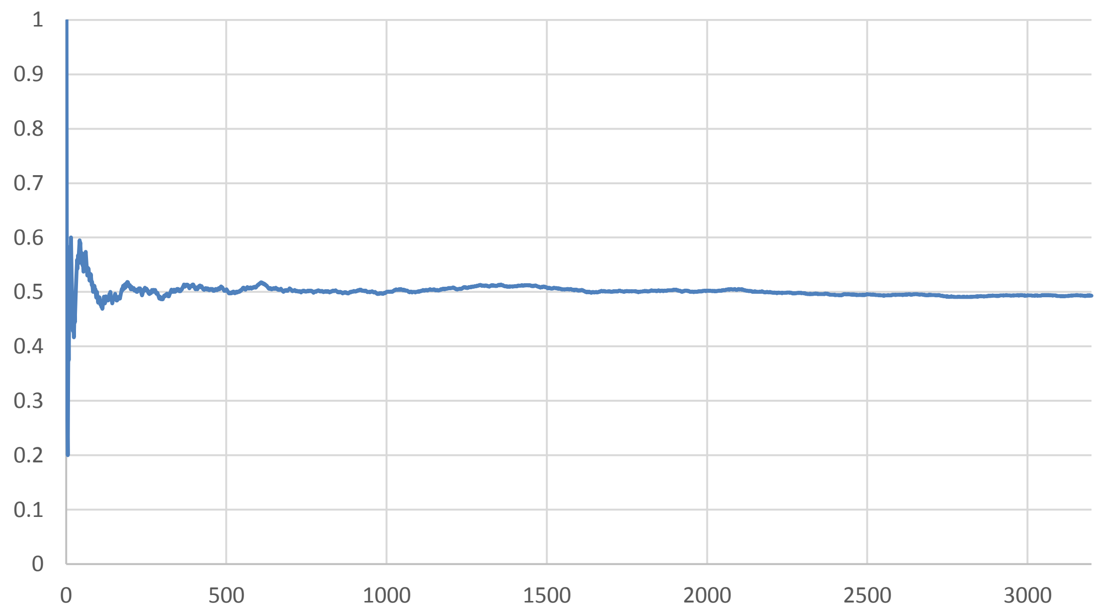
Alle reelle Probleme können mit einem gewissen Urnenmodell gleich gesetzt werden, deshalb is das Urnenmodell der Standard.

Am einfachsten: 1Weiss + 1Grün

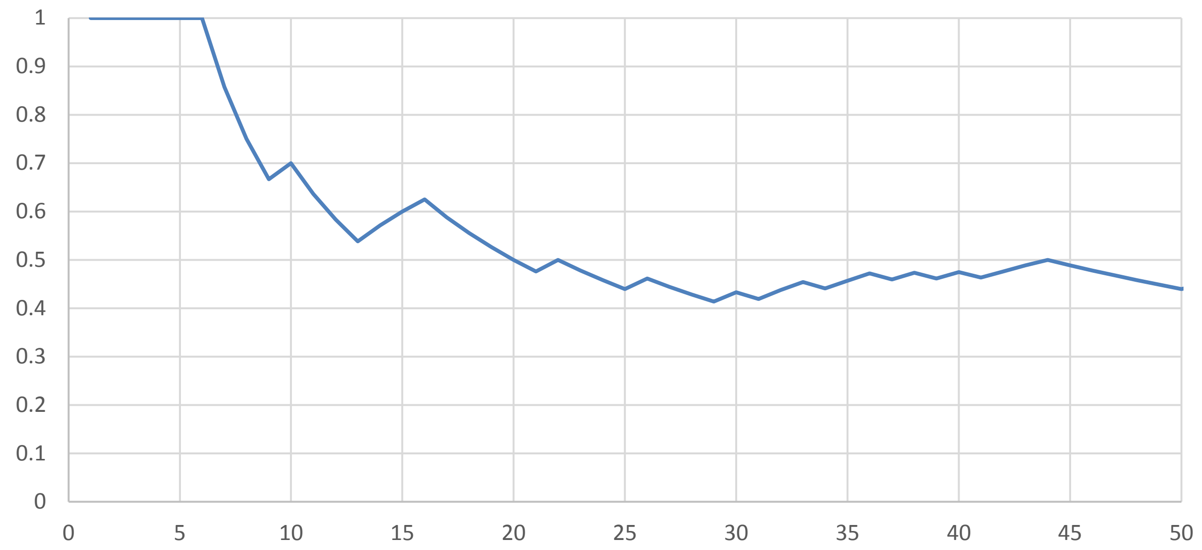
Equivalent mit Münzenwerfen.

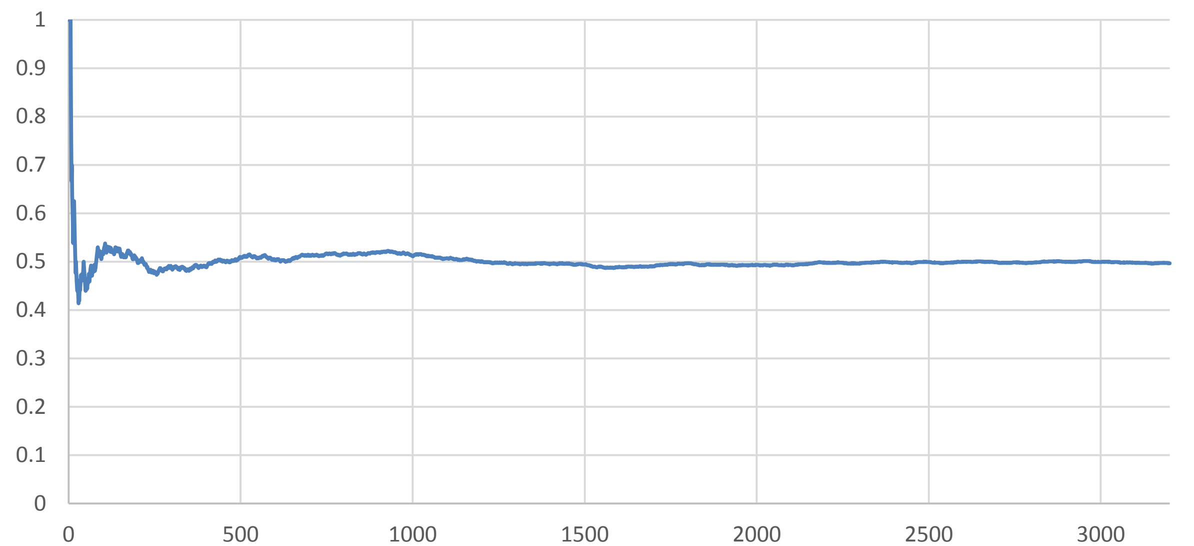
K	1	1
Z	2	0.5
Z	3	0.333333
Z	4	0.25
Z	5	0.2
K	6	0.333333
K	7	0.428571
Z	8	0.375
K	9	0.444444
K	10	0.5
K	11	0.545455
K	12	0.583333
Z	13	0.538462
K	14	0.571429
K	15	0.6
Z	16	0.5625
Z	17	0.529412
Z	18	0.5
Z	19	0.473684
Z	20	0.45
Z	21	0.428571
K	22	0.454545
Z	23	0.434783
Z	24	0.416667
K	25	0.44
K	26	0.461538
Z	27	0.444444
K	28	0.464286
K	29	0.482759
K	30	0.5
K	31	0.516129
K	32	0.53125
K	33	0.545455
K	34	0.558824
Z	35	0.542857
K	36	0.555556
K	37	0.567568
Z	38	0.552632
K	39	0.564103
K	40	0.575
K	41	0.585366
K	42	0.595238
Z	43	0.581395
K	44	0.590909





K	1	1
K	2	1
K	3	1
K	4	1
K	5	1
K	6	1
Z	7	0.857143
Z	8	0.75
Z	9	0.666667
K	10	0.7
Z	11	0.636364
Z	12	0.583333
Z	13	0.538462
K	14	0.571429
K	15	0.6
K	16	0.625
Z	17	0.588235
Z	18	0.555556
Z	19	0.526316
Z	20	0.5
Z	21	0.47619
K	22	0.5
Z	23	0.478261
Z	24	0.458333
Z	25	0.44
K	26	0.461538
Z	27	0.444444
Z	28	0.428571
Z	29	0.413793





Rel.Häufigkeit geht zu 0.5, aber ziemlich langsam…

G
W
W
W
W
W
W
W
W
W
G
W
G
W
W
W
W
G
W
W
W
W
W
W
W
G
G
W

Wir haben ein Sack, und darin Kugeln: **W**eiss oder **G**rün.

Wir ziehen einen Kugel raus, und schreiben die Farbe auf,
dann geht der Kugel zurück.

Was können wir über die Zusammensetzung sagen?

G
W
W
W
W
W
W
W
W
W
G
W
G
W
W
W
W
G
W
W
W
W
W
W
W
G
G
W

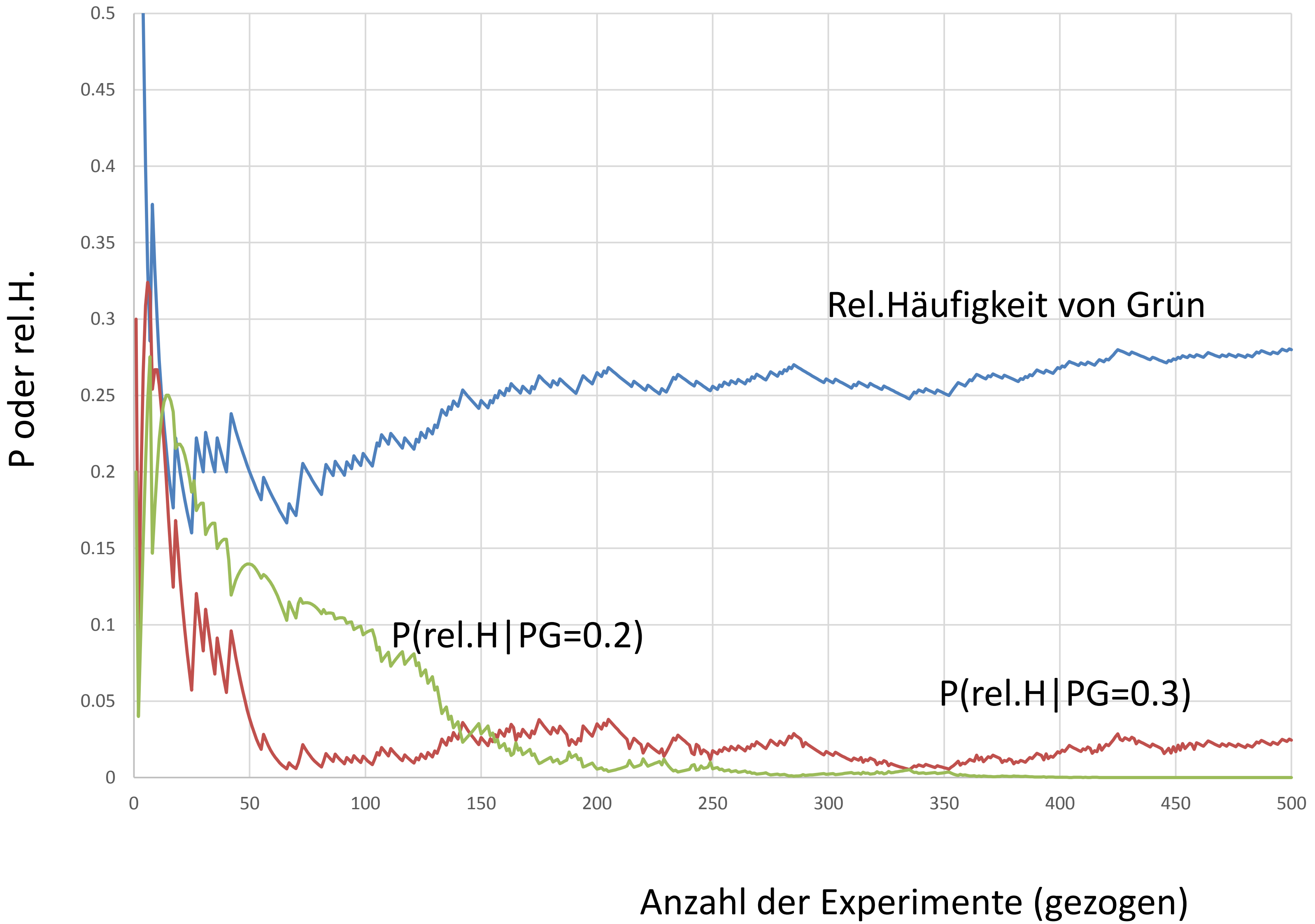
a.) wir können die relative Häufigkeiten darstellen.

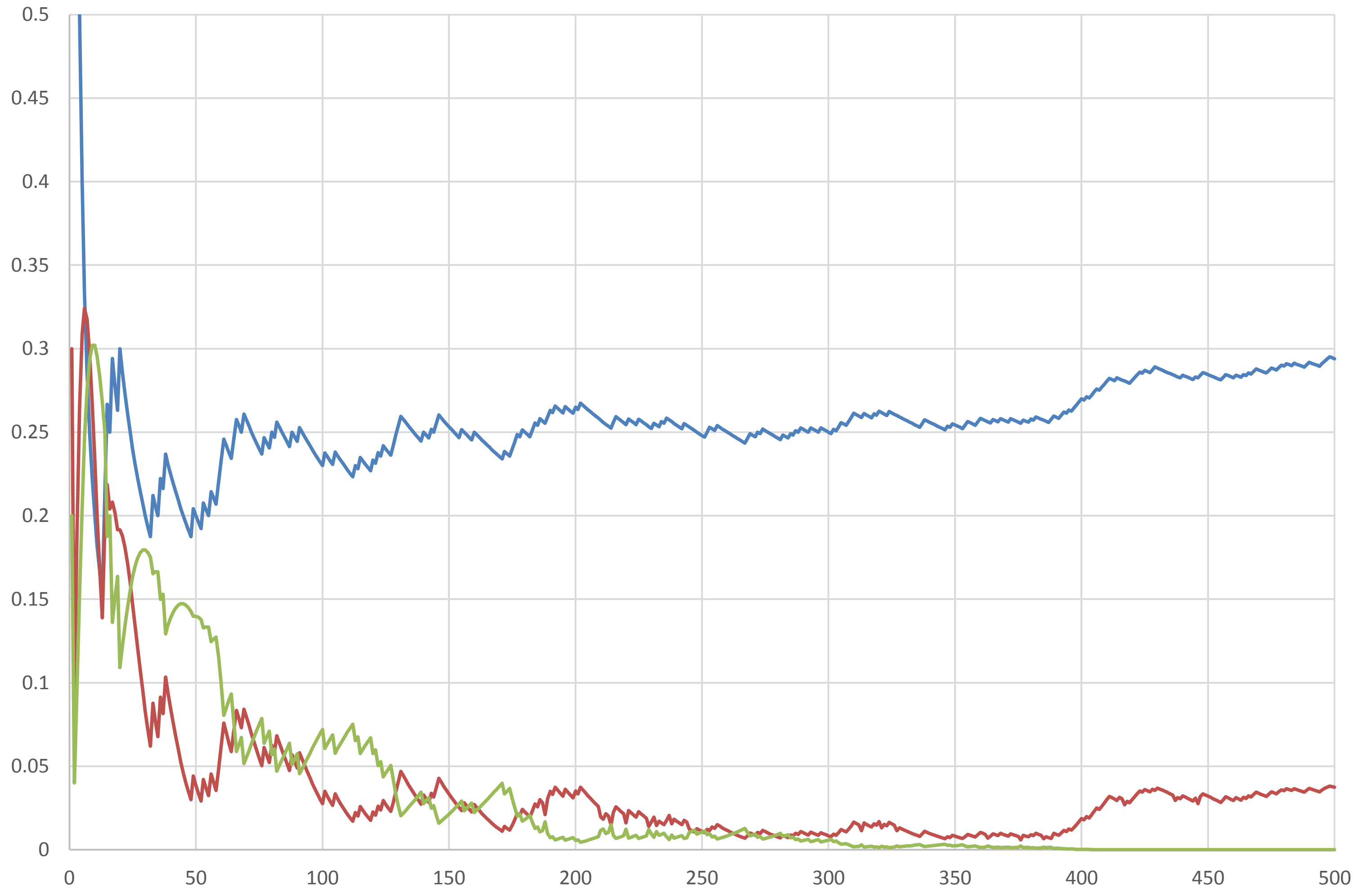
b.) wir können Hypothesen haben, und Wahrscheinlichkeiten aufgrund dessen ausrechnen

c.) intuitiv: was für eine Zusammensetzung könnte das sein, wenn wir das hier sehen…? Gibt es hier Wahrscheinlichkeiten?

Wie können wir Entscheiden?

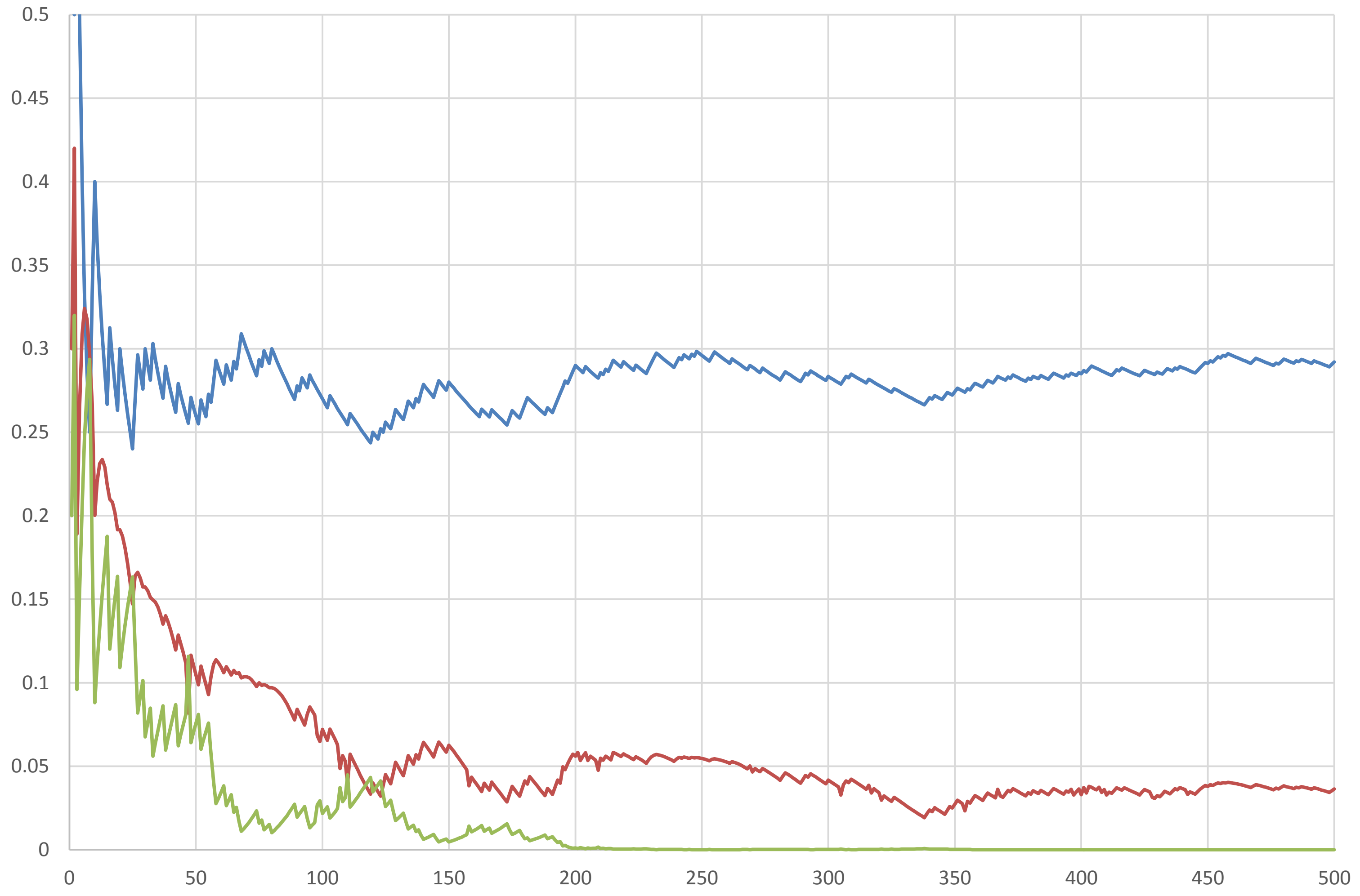
Wenn wir nichts wissen, dann können wir sogar an P=0.2 glauben, statt 0.3, echt lange.





0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
50
100
150
200
250
300
350
400
450
500

Glück gehabt: hier könnten wir schnell entscheiden



Probieren wir aus!

R, Excel,…

Rel.H -> P ist nicht so einfach…

Theoretisch gut, aber praktisch die Schätzungen sind oft ungenau.

Noch dazu:
P(Krank) = ?
P(Gesund) = ?

Hypothesen haben keinen Frequenz!

Bayes – Statistik

“Wahrscheinlichkeit” als “Glaubensgrad”
aber doch rechnerisch benutzbar!

Hypothesenprüfung?

2 Beispiele:

Alles richtig aus 10 -> Entscheindung laut Hypothesentest <-> Vernunft?

Arzt: 10 richtige diagnosen

„Übermeschalkoholiker”: 10 mal richtig geraten.

2 1 M A R C H 2 0 1 9 | VO L 5 6 7 | N AT U R E | 3 0 5
Retire statistical significance

H0: “nur geraten, keine richtige Kentnisse”

$P(10 \text{ mal richtig} \mid H0) = (1/2)^{10}$
$=1/1024 = 0.0009765 < 5\%$ also H0 ablehnen.

H0: nur geraten, keine richtige Kentnisse

P(10 mal richtig | H0) = $(1/2)^{10}$
=1/1024 = 0.0009765 < 5% also H0 ablehnen.

Aber….
Arzt : Ja, das ist was wir erwarten
Übermensch: im Kino wäre OK, aber in der Wirklichkeit sind wir eher immer noch nicht überzeugt, würden weitere Versuche empfehlen.

Wobei rechentechnisch müssten wir die selbe Entscheidung treffen!

Ewas ist hier nicht ganz in Ordnung.

Vorkenntnisse!

Wir *erwarten* das ein Arzt sein Fach gut kennt.

Und

Wir *denken* das Übermenschen ins Kino gehören,
und uns über den Gegenteil zu überzeugen braucht man echt „*harte*" Beweise.

-> Wir Entscheiden oft *nicht nur* mit Hypothesenprüfungsergebissen.

P(Arzt ist gut) = ?

Rel.Häufigkeit ist hier nicht benutzbar…

Wir brauchen eine neue Definition für die Wahrscheinlichkeit.

Intuitiv: *wie „stark" wir an etwas glauben.*

Probleme damit:

1. Skala?

2. ABER: sind dann alle P-Werte ganz zufällig und individualistisch? Wie können wir dann
vergelichen?

Zurück zu der Frage:

Für den Bayesianer (Subjektivist) geht es bei Wahrscheinlichkeiten um den **Grad unseres Glaubens oder unserer Überzeugung** im Lichte einer bestimmten Beweislage – es spielt dabei keine Rolle, *worauf* sich die Überzeugung bezieht.

Wie kann man sowas benutzen? Wie können wir dann Wahrscheinlichkeiten vergelichen?

1.) Regeln für die subjektive „Wahrscheinlichkeiten"
2.) „**The Principal Principle"**: die subjektiven Wahrscheinlichkeiten sollten mit den objektiven Wahrscheinlichkeiten übereinstimmen, falls objektive (Frequenzbestimmte) Wahrscheinlichkeiten existieren.
3.) Updating (rationale Aktualisieren von Glaubensgraden)

„reasonable expectation” (sinnvolle, vernünftige Erwartungen) = P = Wahrscheinlichkeit

Wir können die Regeln auch ganz ohne Häufigkeit (Kugeln, etc) definieren, nur mit Logik.

3 wichtigsten Regeln:

Sei „a” eine Aussage (z.B. „es regnet”)

dann

a = a (identity)

~ ~a = a (non-contradiction) oder ~a v a = TRUE

~a·a = FALSE (excluded middle)

~ = Nicht

. = und

v = oder

Wenn wir rechnen wollen, dann brauchen wir eine Algebra: **Boole-algebra**

Frequenzbestimmt oder generell?

„reasonable expectation” (sinnvolle, vernünftige Erwartungen) = P = Wahrscheinlichkeit

Wir können die Regeln auch ganz ohne Häufigkeit (Kugeln, etc) definieren,
nur mit **Boole-algebra.**

~ = Nicht
. = und
v = oder

a,b,c : Aussagen
(und nicht unbedingt Ereignissen!)

$$\sim\sim a = a, \qquad (1)$$

$$a \cdot b = b \cdot a, \qquad (2) \qquad a \vee b = b \vee a, \qquad (2')$$

$$a \cdot a = a, \qquad (3) \qquad a \vee a = a, \qquad (3')$$

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c = a \cdot b \cdot c, \qquad (4)$$

$$a \vee (b \vee c) = (a \vee b) \vee c = a \vee b \vee c, \qquad (4')$$

$$\sim(a \cdot b) = \sim a \vee \sim b, \qquad (5)$$

$$\sim(a \vee b) = \sim a \cdot \sim b, \qquad (5')$$

$$a \cdot (a \vee b) = a, \qquad (6) \qquad a \vee (a \cdot b) = a. \qquad (6')$$

Nur 1 und noch5 Regeln sind unabhängig.

R. T. Cox, AmJ.Phys

b|a : wie stark glaube ich an der Wahrheit von b, falls a ist wahr.
(Likelihood)
b|a existiert nur wenn es ein logischer zusammenhang gibt.

I. (cb)|a = F(c|ba , b|a) F(x,y), also sollte eine Kombination sein.
II. ~b|a = S(b|a) S(x), auch eine Funktion zu bestimmen.

Ad I. : fahren auf Urlaub (a), und auch zurück. b|a: hingefahren, c: auch zurück. cb|a: hin und zurück, c|ba : zurück falls schon hingefahren. Hier ist I. einfach Vernunft.

Es gibt aber die Regeln der Logik (z.B. 4) -> F(F(x,y),z)=F(x,F(y,z)) (manipulieren wir d.c.b|a ein weing… und d|c.b.a=x, c|b.a = y, b|a=z), sollte stimmen für alle x,y,z Werte.

Mathematisch: C*f(F(p,q)) = f(p)*f(q), wo f beliebige Funktion ist (aber kontinuierlich), und C konst.
Dann, um einfacher zu sein: f(p):=p.
So aber C* c.b|a = c|b.a * b|a. Jetzt aber, wenn c=b, und so b.b=b, wir haben:
C=b|b.a das ist aber die **sicherheit**! (Glaubensgrad von b, falls b und a sind beide wahr)
-> Glaubensgrad von sicherheit können wir frei wählen.

Was bedeutet das?

Ich bin sicher das…

1

100%

365

Etc…

Ad II.

$\sim\sim b|a = b|a \to S(S(x))=x$

Und auch $S(cvb|a)=\sim(cvb)|a = \sim c.\sim b|a$

Wovon folgt

$S(cvb|a)=S(c|\sim b.a)*S(b|a)$.

Nun müssen wir noch ~b weg bekommen (etwas lang)

Und dann haben wir

$S(cb|a / c|a)*c|a = S(S(cvb|a)/S(b|a))/S(b|a)$

Das muss für alle Aussagen gelten.

Davon kann man mit noch mehr Analyse bekommen:

$S(x)=(1-x^m)^{(1/m)}$

Wo m ist ein beliebiger Zahl.

Sei $x=b|a$, dann, weil $S(x)^m = 1-x^m$, aber $S(b|a)=\sim b|a$

$(b|a)^m + (\sim b|a)^m = 1$

Lassen wir m:=1 sein, dann

$b|a + \sim b|a = 1$.

Und

$a|a + \sim a|a = 1$. Sicherheit und Unmöglichkeit.

Wenn wir Glaubensgrad der Sicherheit den Wert 1 geben, dann müssen wir für Unmöglichkeit den Wert von 0 geben.

Es folgen noch:

$c|b.a + {\sim}c|b.a = 1$, wovon aber: $cb|a + {\sim}cb|a = b|a$
Und
$cvb|a = c|a + b|a - c.b|a$

Davon folgt aber, dass „Glaubensgrad" oder „Likelihood" ist eingentlich Wahrscheinlichkeit!

Kolmogorov-Axiome

(1) Für jede Aussage *p*, $0 \leq P(p) \leq 1$
(2) Wenn *p* mit Sicherheit wahr ist, $P(p) = 1$
    Wenn p ist sicher nicht wahr, dann $P(p)=0$
(3) Wenn *p* und *q* inkompatibel sind, $P(p \text{ oder } q) = P(p) + P(q)$
    inkompatibel : sie schliessen sich gegenseitug aus. Also: entwerder p oder q.

Wovon aber folgt: (mit 2+3)

$P(\textit{nicht-p}) = 1 - P(p)$

Standardisierung von Glaubensgraden

Alles ist eine Wette!

"Glaubensgrad" als Glückspiel?

Sei P ϵ[0…1] ~ Glaubensgrad.
Sei Z ein Zustand (oder Aussage)
Je grösser ist P , desto eher wollen wir so handeln, das im Falle Z ist wahr, der Konsequenz der Handlung für uns angenehm sein wird.

Wette ist so etwas: Wie viel wollen wir maximum wetten, so das wenn Z ist wahr, bekommen wir 1 EUR?

Mehr als 1 wäre sicherer Verlust, 0 wäre kein Spiel.
Genau so viel würden wir maximum reskieren, damit *im langen Lauf* die Kosten abgedeckt sind.
(Darunter erwarten wir gewinn, oberhalb aber Verlust)

Wette:
Es gibt 2 Leute, die um ein Ereigniss wetten: jeder gibt eine Menge ein, und wer gewinnt (hat auf das Auftreten des Ereignisses gewettet) bekommt alles.

Probieren wir aus:

Wie viel würden wir reskieren falls:

$P(Z) = 0.5$

$P(Z) = 0.1$

$P(Z) = 0$

$P(Z) = 1$

Die Wette is maximal so viel Wert wie der Erwartungswert.

z.B. $P(Z)=0.1$
Dann 9 Mal aus 10 Spiele gehen wir leer aus, und 1 Mal mit 1 EUR.
Erw=$0.1*1 + 0.9*0 = 0.1$, mehr dürften wir *vernünftigerweise* nicht bezahlen.

P-Werte können nur dann verglichen werden, wenn aus der selben Wissensmenge alle Leute zu der selben „glaubensgrad" kommen.

Intersubjektive Vereinbarung

Faire Wette (Pferderennen)
Gewinn für ein Spiel: G
Wahrscheinlichkeit von P_{ferd}: Pg.
Wetten: ***1:x*** , also Pg=1/(1+x)
Man muss B bezahlen, bekommt G wenn der Pferd gewinnt, und nichts wenn der Pferd nicht gewinnt.

Sinnvolles, vernünftiges Benehmen sollte *durchschnittlich* in kein Verlust enden!
Wenn B ≠ Pg*G dann *durchschnittlich* gibt es entweder Gewinn oder Verlust.

Erwartungswert: E=Pg*G – B*(1-Pg) = 0 wenn B=Pg*G, in diesem Fall ist der Spiel Fair.

also die Wahrscheinlichkeiten addieren zu 1, genau wie klassisch.

Noch ein Beispiel:
Kopf oder Zahl: A wettet für Kopf, B für Zahl.
A bekommt Ga Geld von B für Kopf, und B bekommt Gb für Zahl.

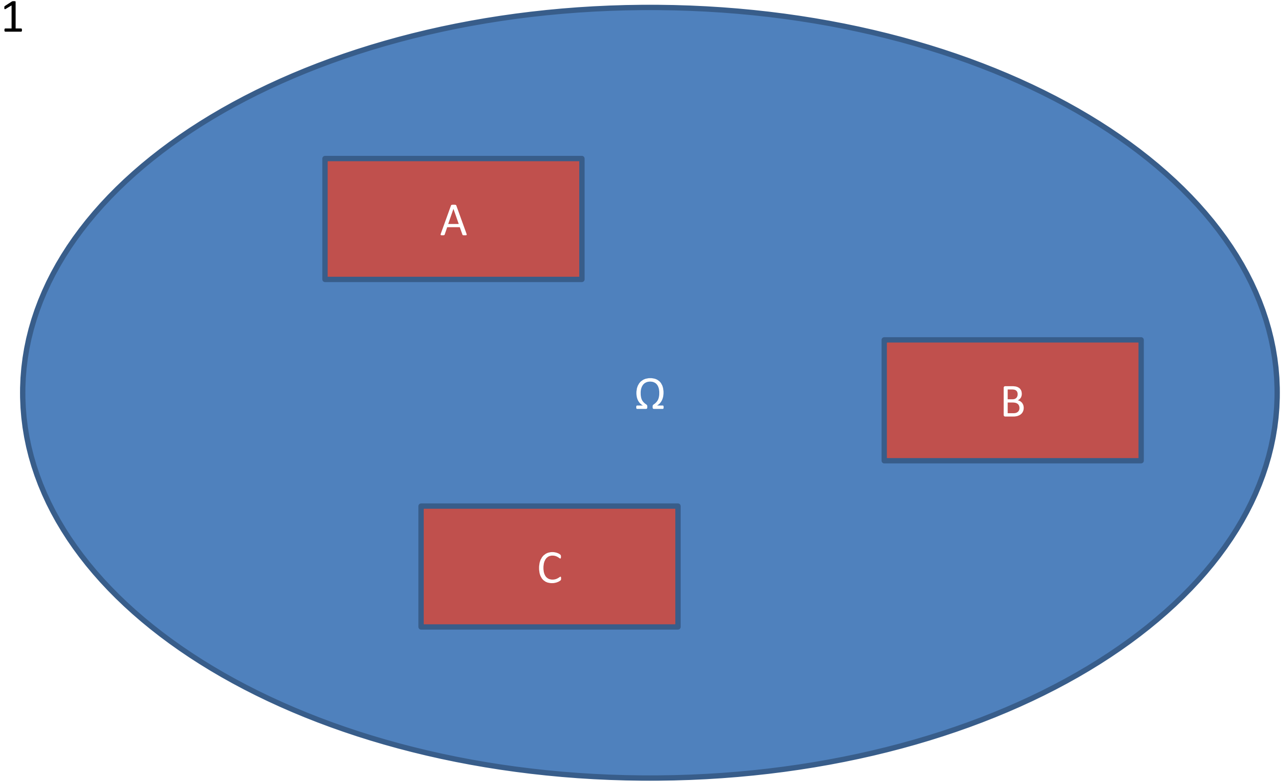
In einem fairen Spiel, *durchschnittlich* gewinnen beide das selbe.
Sonst würde es unsinn zu spielen, nämlich eine Seite würde sicher Verlust realisieren.
(Lotto ist nicht fair)

Also muss $P_{kopf}*Ga - P_{zahl}*Gb = 0$ (Standpunkt A)
Aber: wenn $P_{kopf}=1$ dann gewinnt nur A, sofern müsste B für den Spiel nicht bezahlen, A aber am Anfang Ga eingeben

Es ist äußerst wichtig, das **Glaubensgrade**, mit denen ein vollkommen rationaler Denker verschiedene Aussagen für wahr hält, keine Widersprüche erzeugen, also **kohärent** sind.

$P(A)+P(B)+P(C) <= 1$

$P(\Omega)=1$



Dutch-book

Kombination von Wetten, die garantiert gewinnt – egal, was passiert.
(historisch: niederländische Händler galten im 17. Jh. als besonders gerissen)

Nur möglich, wenn die Glaubensgrade nicht kohärent sind.

Beispiel: 2 Pferde – Rennen, der Wettanbieter (bookmaker) sagt:

100 EUR auf „Sonne” mit Quote 1:1
120 EUR auf „Mond” mit Quote 2:1

Das ist ein sicherer Gewinn für den Wettanbieter!

Warum?

n.B: Quote 1:1 bedeutet: man setzt 1 EUR um +1 EUR zu gewinnen, oder den originellen 1EUR zu verlieren.

a.) „Sonne” gewinnt:
Einkommen: 100+120 EUR = 220 EUR.
Auszahlen: 100+100 EUR = 200 EUR.
Gewinn: 20 EUR.

b.) „Mond” gewinnt:
Einkommen: 100+120 EUR = 220 EUR.
Auszahlen: 120 für „Mond” + 120/2=60 für „Mond”
Gewinn: 220-180 = 40 EUR

P(„Sonne”) = 1/(1+1) = 0.5
P(„Mond”) = 1/(1+2) = 0.3
Aber Ω = entweder „Sonne” oder „Mond”.

Das gilt auch für $|\Omega| > 2$

Wette								
a zu b fuer	a	b		Kosten				
	3	2		a				
				Einkommen				
Erwartet	p*(a+b)+(1-p)*0-a=0			p*(a+b)+(1-p)*0				
	p*a+p*b-a=0							
	p*(a+b)=a							
	p=a/(a+b)							
	p	0.6		fairer Preis ist = N*p, N kann beliebig sein				
					p glauben	Einkommen	**Preis**	**ausgezahlt**
Schones Wetter		1	4		**0.2**	1	**20**	**100**
Regenschauer		1	3		**0.25**	1	**25**	**100**
Bewolkt		1	1		**0.5**	1	**50**	**100**
				summe P	**0.95**		**die Preise sind Fair in allen Fallen**	
total eingesetzt>			95					
Bekommen			100					
Balanz			5					

Balanz schlecht für Spieler

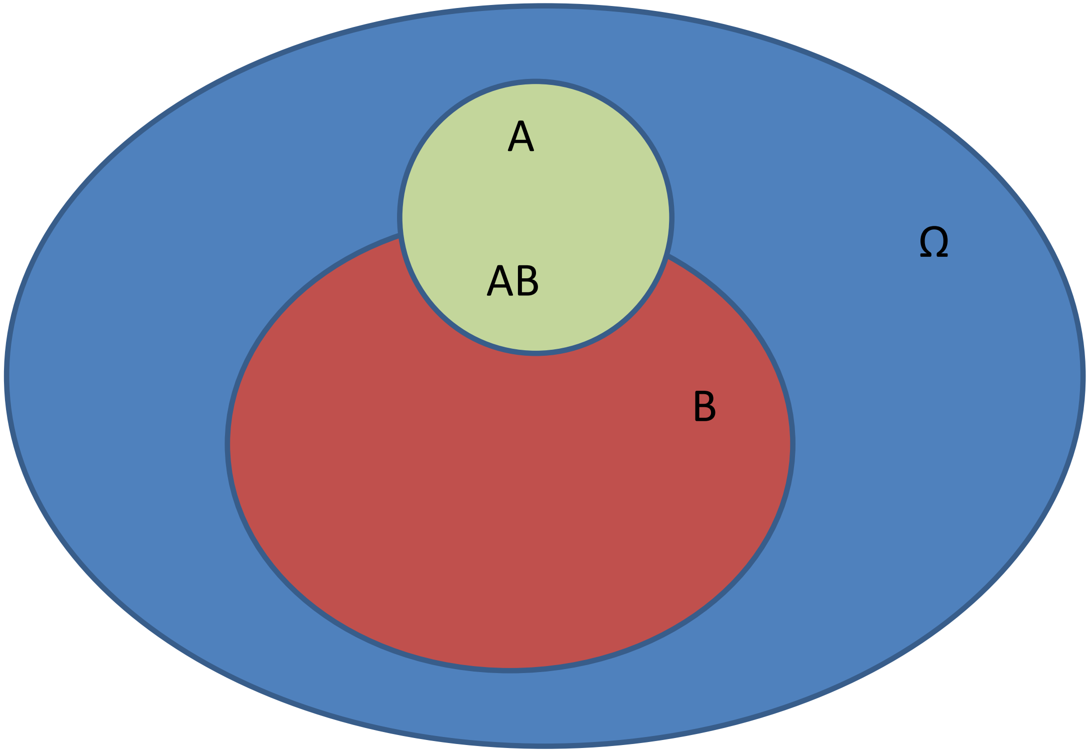
Das „dutch-book”-Argument unterstützt auch, das die subjektive Wahrscheinlichkeiten den Kolmogorov-Regeln folgen müssen.

Updating : Das rationale (vernünftige) Aktualisieren von Glaubensgraden (Wahrscheinlichkeiten).

Bedingte Wahrscheinlichkeiten:

P(A|B) = P(A wenn B ist wahr)

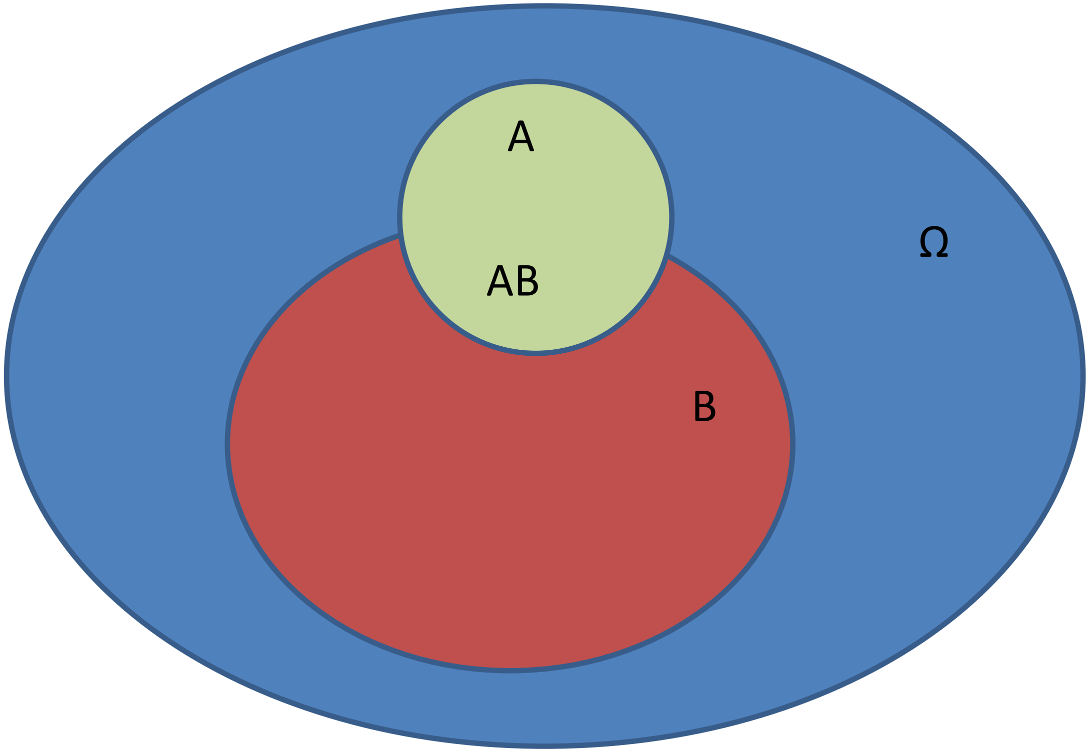
Standard Urne:
Ω: alle Kugeln,
A,B spezifische Kugeln.
(z.B. Farbe, Muster)



Bedingte Wahrscheinlichkeiten:

P(A|B) = P(A wenn B ist wahr)

Beispiel:
B: Grün
A: mit Streifen

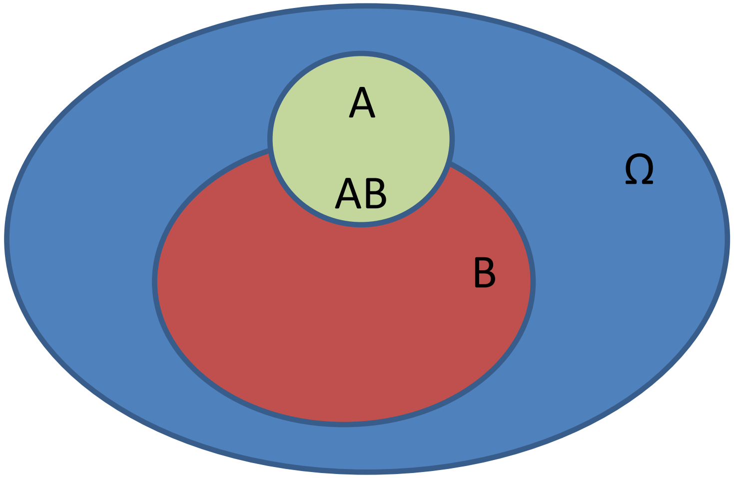


Urne:
100 Kugeln, davon 60 Grün, wovon 12 sind mint Streifen, die anderen ohne Muster. Aus den Weißen 15 sind mit Streifen.

Bedingte Wahrscheinlichkeiten:

P(A|B) = P(A wenn B ist wahr)

Beispiel:
B: Grün
A: mit Streifen



Urne:
100 Kugeln, davon 60 Grün, wovon 12 sind mint Streifen, die anderen ohne Muster. Aus den Weißen 15 sind mit Streifen.

|Ω|=100 , |B|=60 , |A|=12+15=27 , |w|=40, |AB|=12 , |Aw|=15

Und so P(B) = |B|/|Ω|, P(AB)=|AB|/|Ω|, etc.

Wenn B wahr ist, dann können wir ein neues Urnenmodell aufbauen, weil alles mit w kommt nicht mehr ins Frage.
Ωneu = B, dann P(A|B) = |AB|/| Ωneu |=|AB|/|B| = 12/60 = 0.2

Aber |AB| = P(AB)*|Ω| und |B|=P(B)*|Ω|, und so |AB|/|B|=P(AB)/P(B).

$$\boxed{P(A|B) = P(AB)/P(B), \text{ oder } P(AB)=P(A|B)*P(B)}$$

$P(\text{Streifen}|\text{Blau})=20/(20+30)=2/5=0{,}4$
$P(SB) = 20/100 = 0{,}2$
$P(S)=(20+20)/100=0{,}4$
$P(B)=(20+30)/100=0{,}5$

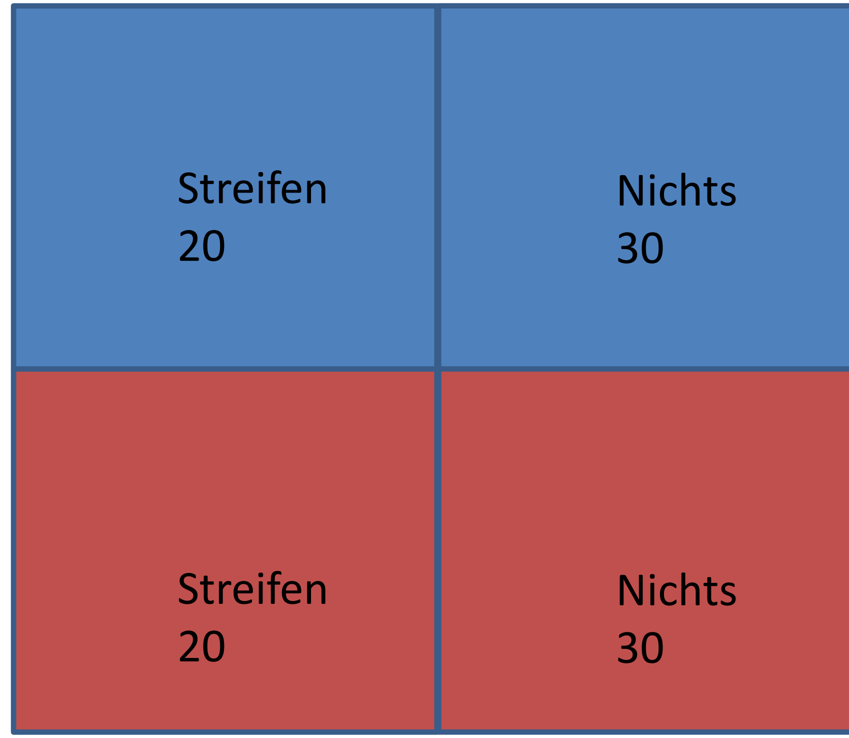
$P(SB)=P(S|B)*P(B)=0{,}4*0{,}5=0{,}2$

Aber $P(S|B) = P(S)$, noch dazu
$P(S|\sim B) = 20/(20+30)=2/5=0{,}4$, also
$P(S|B)=P(S|\sim B)=P(S)$.

Wovon:

$P(SB)=P(S)*P(B)$

stochastische Unabhängigkeit.



Updating

Spiel: 3 Säcke mit Kugeln, jeweils 10 Kugeln, aber die Zusammensetzung ist unterschiedlich:

A: 1G 9W
B: 3G 7W
C: 7G 3W

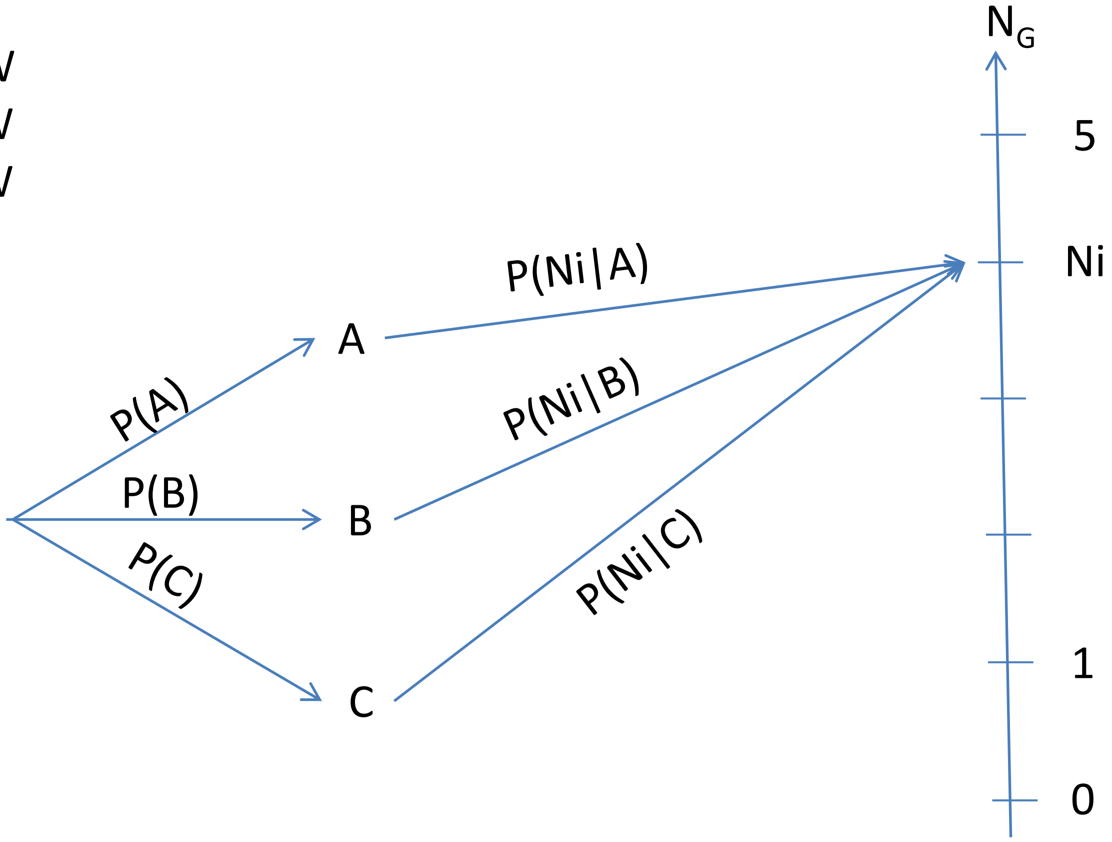
Falls wir 5 Mal ziehen (mit Rückgabe), und N=Ni Mal Grün gezogen haben, was können wir dann darüber sagen, welcher Sack in dem Spiel war?

Spiel: 3 Säcke mit Kugeln, jeweils 10 Kugeln, aber die Zusammensetzung ist unterschiedlich:

A: 1G 9W

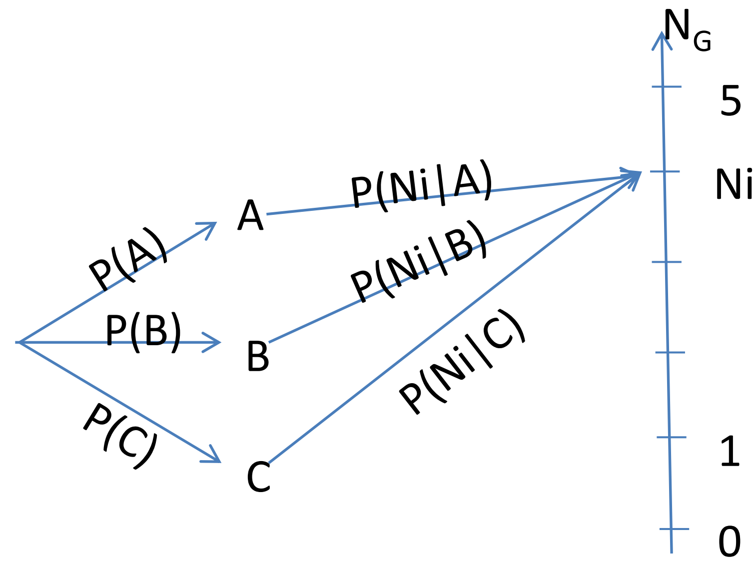
B: 3G 7W

C: 7G 3W



$P(Ni) = P(\text{Weg } A)+P(\text{Weg } B)+P(\text{Weg } C)$
(disjunkte Wege)

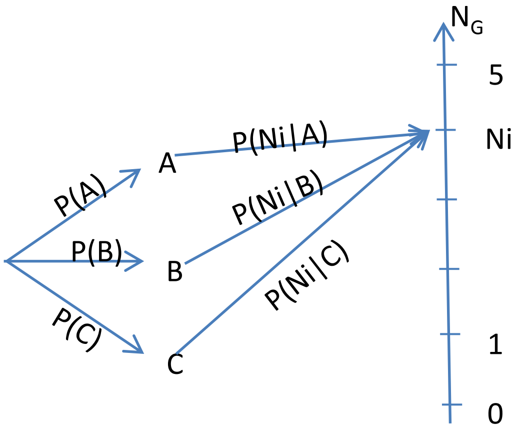
Wir wissen aber, dass $N=Ni$ ist passiert, davon wollen wir etwas über $P(A)$,$P(B)$,$P(C)$ sagen.



$P(\text{Weg } A|N=Ni) = P(\text{Weg } A \text{ und } Ni)/P(Ni)$
Wo aber
$P(\text{Weg } A \text{ und } Ni) = P(Ni|A)*P(A)$
Und
$P(Ni)=P(\text{Weg } A \text{ und } Ni)+P(\text{Weg } B \text{ und } Ni)+P(\text{Weg } C \text{ und } Ni)$

$\boxed{P(A|B) = P(AB)/P(B), \text{ oder } P(AB)=P(A|B)*P(B)}$

P(Weg A|N=Ni) = P(Weg A und Ni)/P(Ni)
Hier alles einsetzen:



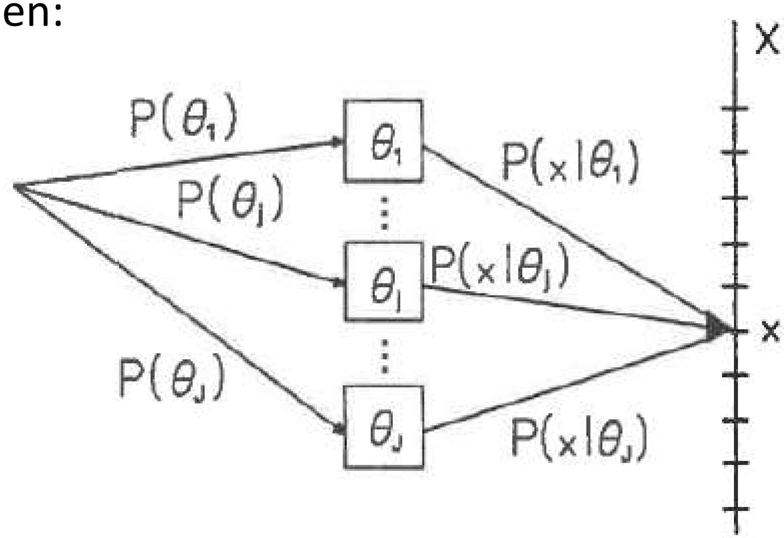
$$\boxed{P(A|N = Ni) = \frac{P(A) * P(Ni|A)}{\sum_{k=A,B,C} P(k) * P(Ni|k)}}$$

P(A) : Wahrscheinlichkeit (Glaubensgrad!) von „wir haben Sack A in der Hand”.
Priori Wahrscheinlichkeit (vorausgehenden Wahrscheinlichkeit)

P(Ni|A): Likelihood

P(A|Ni): Posteriori Wahrscheinlichkeit, die neue Werte, ***nach*** der Beobachtung.

Ganz allgemein, wenn wir viele Möglichkeiten haben:



$$P(\theta_j|x) = \frac{P(\theta_j)P(x|\theta_j)}{P(x)} = \frac{P(\theta_j)P(x|\theta_j)}{\sum_{j=1}^{J} P(\theta_j)P(x|\theta_j)}$$

Konditionalisierungsregel
(Bayes Theorem)

Diskrete Möglichkeiten (θ_j)

$$P(\theta_j|x) = \frac{P(\theta_j)P(x|\theta_j)}{P(x)} = \frac{P(\theta_j)P(x|\theta_j)}{\sum_{j=1}^{J} P(\theta_j)P(x|\theta_j)}$$

Kontinuierliche Möglichkeiten:
Hier rechnen wir die
Dichtefunktion aus

$$f(\theta|x) = \frac{f(\theta)P(x|\theta)}{\int_\Theta f(\theta)P(x|\theta)\,d\theta}$$

Rechnen wir aus!

ABER:

$P(A)=?$

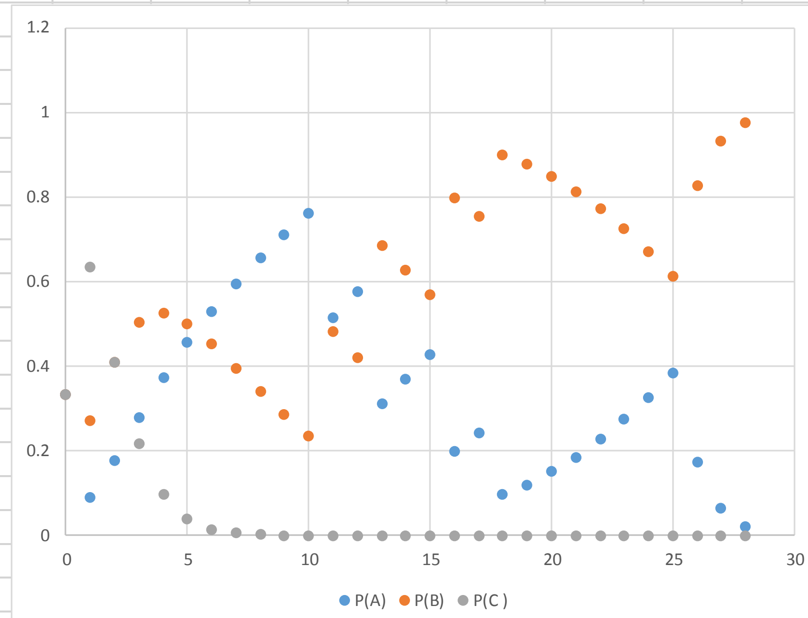
Wenn wir nichts wissen, dann können wir $P(A)=P(B)=P(C)$ als Ansatz nehmen.

Rechnen wir aus!

Spiel:
Ich habe 3 Säcke 1G9W, 3G7W,7G3W

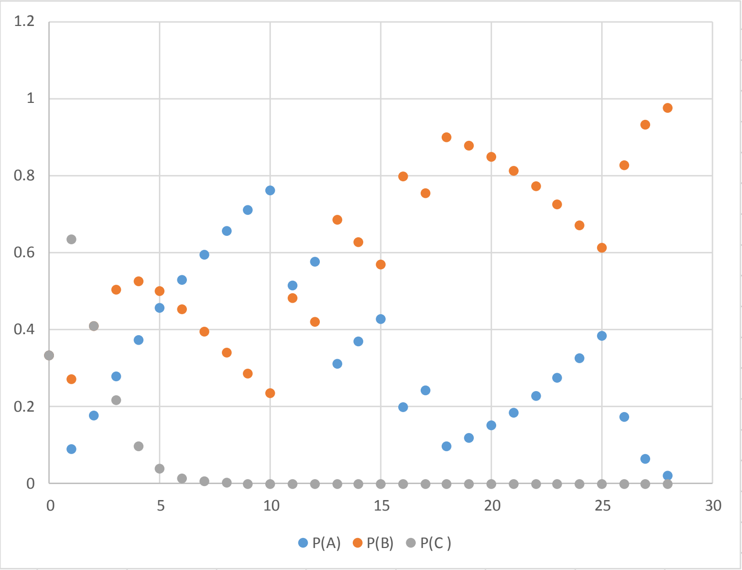
Aus welchem ziehen wir?

	count	N_G		P(A)	P(B)	P(C)		L(A)	L(B)	L(C)		P(G|A)	P(G|B)	P(G|C)
START	0	0		0.333333	0.333333	0.333333						0.1	0.3	0.7
G	1	1		0.090909	0.272727	0.636364		0.1	0.3	0.7		P(W|A)	P(W|B)	P(W|C)
W	2	1		0.176471	0.411765	0.411765		0.9	0.7	0.3		0.9	0.7	0.3
W	3	1		0.278351	0.505155	0.216495		0.9	0.7	0.3				
W	4	1		0.374422	0.528505	0.097072		0.9	0.7	0.3				
W	5	1		0.457819	0.502617	0.039565		0.9	0.7	0.3				
W	6	1		0.531155	0.453545	0.015301		0.9	0.7	0.3				
W	7	1		0.597466	0.396797	0.005737		0.9	0.7	0.3				
W	8	1		0.658004	0.33989	0.002106		0.9	0.7	0.3				
W	9	1		0.712847	0.286393	0.000761		0.9	0.7	0.3				
W	10	1		0.76171	0.238019	0.000271		0.9	0.7	0.3				
G	11	2		0.515483	0.483233	0.001283		0.1	0.3	0.7				
W	12	2		0.578052	0.421468	0.00048		0.9	0.7	0.3				
G	13	3		0.313169	0.685012	0.001819		0.1	0.3	0.7				
W	14	3		0.36993	0.629353	0.000716		0.9	0.7	0.3				
W	15	3		0.430319	0.569404	0.000278		0.9	0.7	0.3				
G	16	4		0.201039	0.798053	0.000908		0.1	0.3	0.7				
W	17	4		0.244558	0.755073	0.000368		0.9	0.7	0.3				
G	18	5		0.097342	0.901632	0.001026		0.1	0.3	0.7				
W	19	5		0.121837	0.877735	0.000428		0.9	0.7	0.3				
W	20	5		0.151414	0.848409	0.000177		0.9	0.7	0.3				
W	21	5		0.186621	0.813307	7.29E-05		0.9	0.7	0.3				
W	22	5		0.227804	0.772167	2.96E-05		0.9	0.7	0.3				
W	23	5		0.274996	0.724992	1.19E-05		0.9	0.7	0.3				
W	24	5		0.327813	0.672183	4.74E-06		0.9	0.7	0.3				
W	25	5		0.38538	0.614619	1.86E-06		0.9	0.7	0.3				
G	26	6		0.172874	0.82712	5.83E-06		0.1	0.3	0.7				
G	27	7		0.065131	0.934854	1.54E-05		0.1	0.3	0.7				
G	28	8		0.022695	0.977267	3.75E-05		0.1	0.3	0.7				

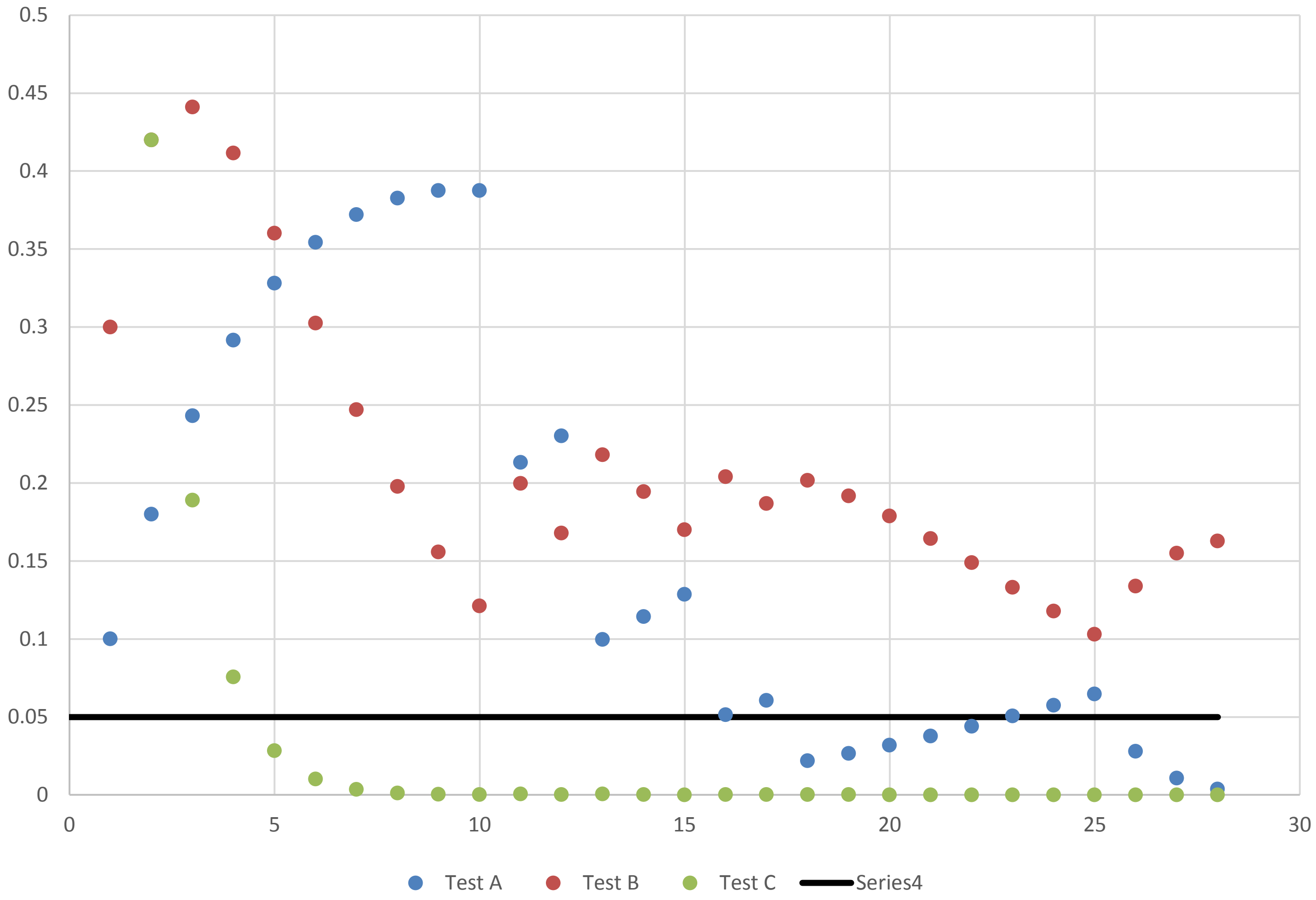


Mehrere Versuche können als 1 Kombi-Experiment genommen werden

	count	N_G		P(A)	P(B)	P(C)	P(A) ein	P(B) ein	P(C) ein	L(A)	L(B)	L(C)		P(G|A)	P(G|B)	P(G|C)
START	0	0		0.333333	0.333333	0.333333	0.333333	0.333333	0.333333					0.1	0.3	0.7
G	1	1		0.090909	0.272727	0.636364	0.090909	0.272727	0.636364	0.1	0.3	0.7		P(W|A)	P(W|B)	P(W|C)
W	2	1		0.176471	0.411765	0.411765	0.176471	0.411765	0.411765	0.9	0.7	0.3		0.9	0.7	0.3
W	3	1		0.278351	0.505155	0.216495	0.278351	0.505155	0.216495	0.9	0.7	0.3				
W	4	1		0.374422	0.528505	0.097072	0.374422	0.528505	0.097072	0.9	0.7	0.3				
W	5	1		0.457819	0.502617	0.039565	0.457819	0.502617	0.039565	0.9	0.7	0.3				
W	6	1		0.531155	0.453545	0.015301	0.531155	0.453545	0.015301	0.9	0.7	0.3				
W	7	1		0.597466	0.396797	0.005737	0.597466	0.396797	0.005737	0.9	0.7	0.3				
W	8	1		0.658004	0.33989	0.002106	0.658004	0.33989	0.002106	0.9	0.7	0.3				
W	9	1		0.712847	0.286393	0.000761	0.712847	0.286393	0.000761	0.9	0.7	0.3				
W	10	1		0.76171	0.238019	0.000271	0.76171	0.238019	0.000271	0.9	0.7	0.3				
G	11	2		0.515483	0.483233	0.001283	0.515483	0.483233	0.001283	0.1	0.3	0.7				
W	12	2		0.578052	0.421468	0.00048	0.578052	0.421468	0.00048	0.9	0.7	0.3				
G	13	3		0.313169	0.685012	0.001819	0.313169	0.685012	0.001819	0.1	0.3	0.7				
W	14	3		0.36993	0.629353	0.000716	0.36993	0.629353	0.000716	0.9	0.7	0.3				
W	15	3		0.430319	0.569404	0.000278	0.430319	0.569404	0.000278	0.9	0.7	0.3				
G	16	4		0.201039	0.798053	0.000908	0.201039	0.798053	0.000908	0.1	0.3	0.7				
W	17	4		0.244558	0.755073	0.000368	0.244558	0.755073	0.000368	0.9	0.7	0.3				
G	18	5		0.097342	0.901632	0.001026	0.097342	0.901632	0.001026	0.1	0.3	0.7				
W	19	5		0.121837	0.877735	0.000428	0.121837	0.877735	0.000428	0.9	0.7	0.3				
W	20	5		0.151414	0.848409	0.000177	0.151414	0.848409	0.000177	0.9	0.7	0.3				
W	21	5		0.186621	0.813307	7.29E-05	0.186621	0.813307	7.29E-05	0.9	0.7	0.3				
W	22	5		0.227804	0.772167	2.96E-05	0.227804	0.772167	2.96E-05	0.9	0.7	0.3				
W	23	5		0.274996	0.724992	1.19E-05	0.274996	0.724992	1.19E-05	0.9	0.7	0.3				
W	24	5		0.327813	0.672183	4.74E-06	0.327813	0.672183	4.74E-06	0.9	0.7	0.3				
W	25	5		0.38538	0.614619	1.86E-06	0.38538	0.614619	1.86E-06	0.9	0.7	0.3				
G	26	6		0.172874	0.82712	5.83E-06	0.172874	0.82712	5.83E-06	0.1	0.3	0.7				
G	27	7		0.065131	0.934854	1.54E-05	0.065131	0.934854	1.54E-05	0.1	0.3	0.7				
G	28	8		0.022695	0.977267	3.75E-05	0.022695	0.977267	3.75E-05	0.1	0.3	0.7				



Hypothesentest



Oft müssten wir mehr als 1 H0 annehmen…

Wie viel lernen wir von einem Versuch?

n.B. Informationsmenge:

Informationsübertragung – Informationsmaß

Wie können wir den Informationsmaß (H) matematisch definieren? (Shannon 1948)

1.: H muss kontinuierlich sein in p_i (kleine Veränderung von p_i → keine Sprünge in H)

2.: seltene Ereignisse haben großen Informationsgehalt:

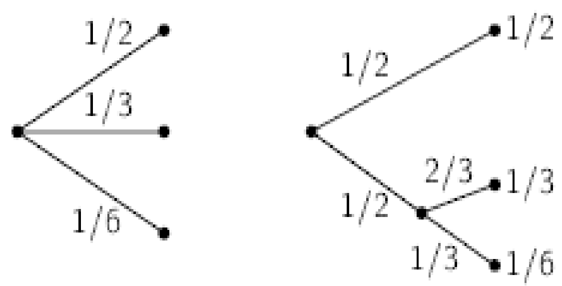
H ist indirekt proportional zu p

Wenn alle Ereignisse haben die gleiche Wahrscheinlichkeit ($p_i = 1/n$)
dann muss H monotonisch steigend sein in n.

3.: Entscheidungsverzweigungen:

Wenn ein Ereignis mit mehreren Entscheidungen hervorgerufen werden kann,
dann die möglichen Wege sollen in H additiv sein,
mit Wichtungsfaktoren gleich den Wa

$$H\left(\frac{1}{2},\frac{1}{3},\frac{1}{6}\right)=H\left(\frac{1}{2},\frac{1}{2}\right)+\frac{1}{2}\cdot H\left(\frac{2}{3},\frac{1}{3}\right)$$



Und noch: $H(p_1,p_2,p_3)=H(p_2,p_1,p_3)$, also ist H symmetrisch

Informationsübertragung – Informationsmaß

Man kann den Informationsgehalt eines einzigen Ereignisses ausrechnen:

$$I = \log_2\left(\frac{1}{p}\right)$$

Maß für Überraschung !

Shannon : definieren wir H mit $H = p \cdot \log_2\left(\frac{1}{p}\right)$

Also H= p*I, somit ist H ein Erwartungswert, oder Mittelwert.
(wenn wir über alle Möglichkeiten summieren)

$\log_2$: 2-Basis logaritmus

Beispiele:

$\log_2 (2) = 1$
$\log_2 (4) = 2$
$\log_2 (8) = 3$

Informationsübertragung – Informationsmaß

Shannon $$H = p\ log_2\left(\frac{1}{p}\right)$$ [bit]

Wenn wir mehrere Möglichkeiten haben, dann ist H eine Summe:
(Erwartungswert der Informationsmenge)

$$H = \sum_i p_i\ log_2\left(\frac{1}{p_i}\right)$$

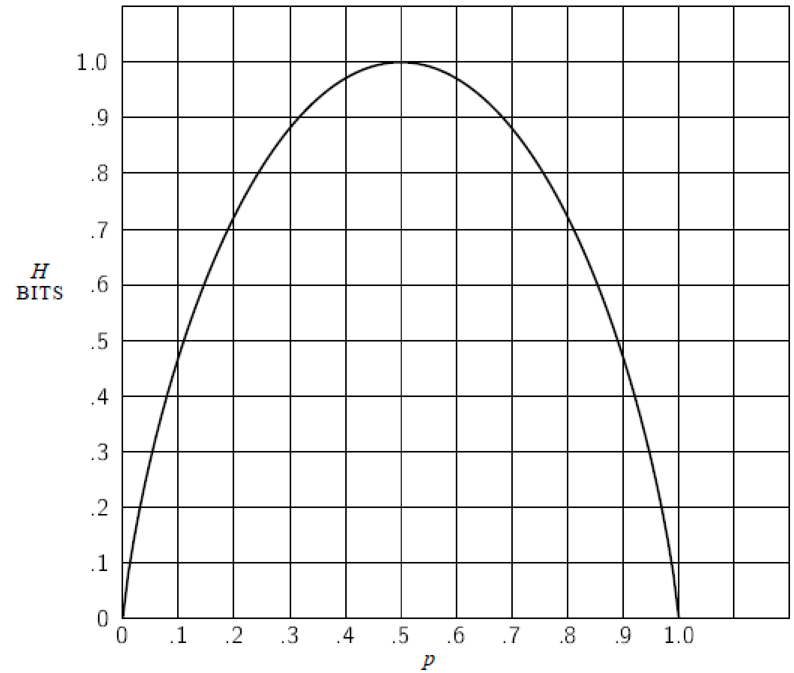
andere log-basis Möglichkeiten:
$\log_e$ (ln) : [nat]
$\log_{10}$ (lg) : [ban]

Informationsmaß - Entropie

Kopf oder Zahl?

p

q = 1-p



$$H=\sum_i -p_i \log_2 p_i = -p\cdot\log_2 p - q\cdot\log_2 q = -p\cdot\log_2 p - (1-p)\cdot\log_2(1-p)$$

Informationsmaß - Entropie

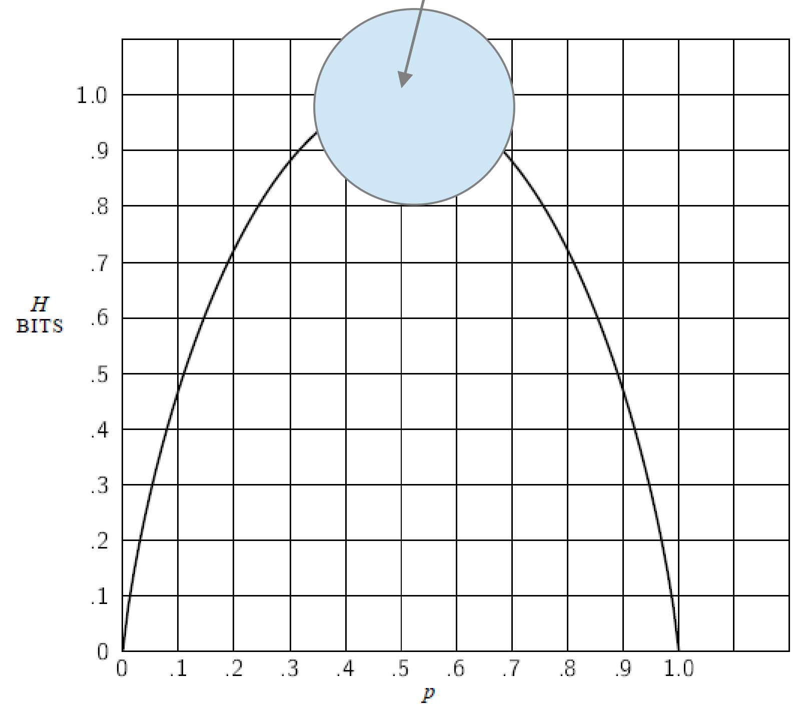
faire Münze: p = ½

Keine Erwartungen
maximale Unsicherheit

Kopf oder Zahl?

p

q = 1-p



H deshalb wird auch benannt als:
Shannon-entropy

Wenn wir nichts im voraus wissen, und so alle Ereignisse sind gleich möglich, dann ist Anzahl der erwarteten Ereignisse maximal.

⇕

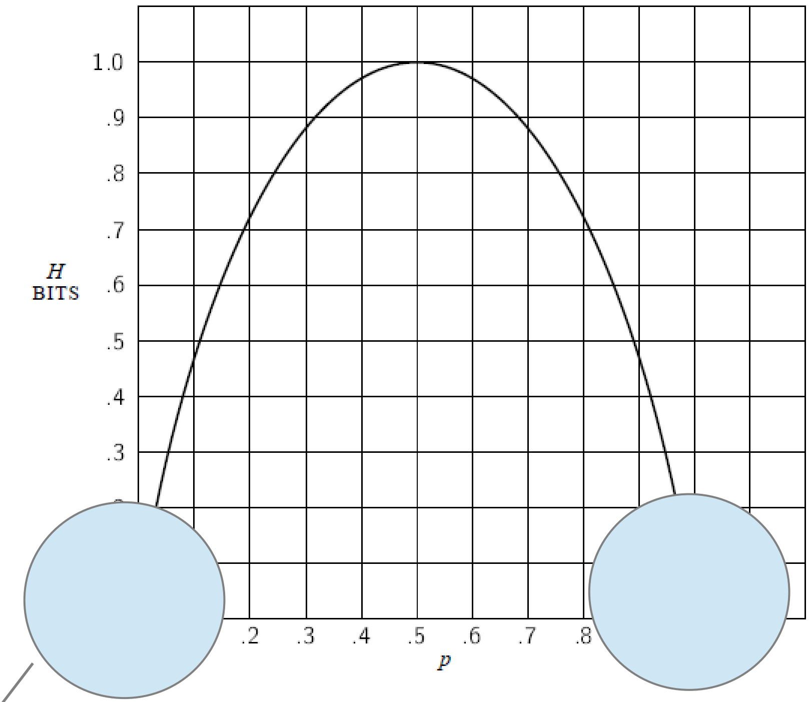
In der Physik ist Entropie dann maximal, wenn der Anzahl der Mikrozustände maximal ist.

Informationsmaß - Entropie

Kopf oder Zahl?

p

q = 1-p



H deshalb wird auch benannt als:

Shannon-entropy

H ist nur dann 0, wenn es nur eine Möglichkeit gibt: p=0 oder p=1

Physikalische Entropie (S) ist nur dann 0 wenn nur 1 Mikrozustand möglich ist.

Kullback–Leibler Divergenz

$$D_{KL}(P|Q) = \sum_{i} p_i \cdot log\left(\frac{p_i}{q_i}\right)$$

$$D_{KL}(P|Q) = \int_{-\infty}^{+\infty} p(x) \cdot log\left(\frac{p(x)}{q(x)}\right) dx$$

Zusammenhang mit Shannon-Entropie:

$$H = \log(N) - D_{KL}(p(x) | PG_N)$$

Die gemessene Verteilung

PG_N ist die Gleichverteilung (N Möglichkeiten)

Um die gemessene Verteilung übertragen zu können Verhältnissmässig zu der Gleichverteilung braucht man H Bits *weniger* bei der Übertragung.

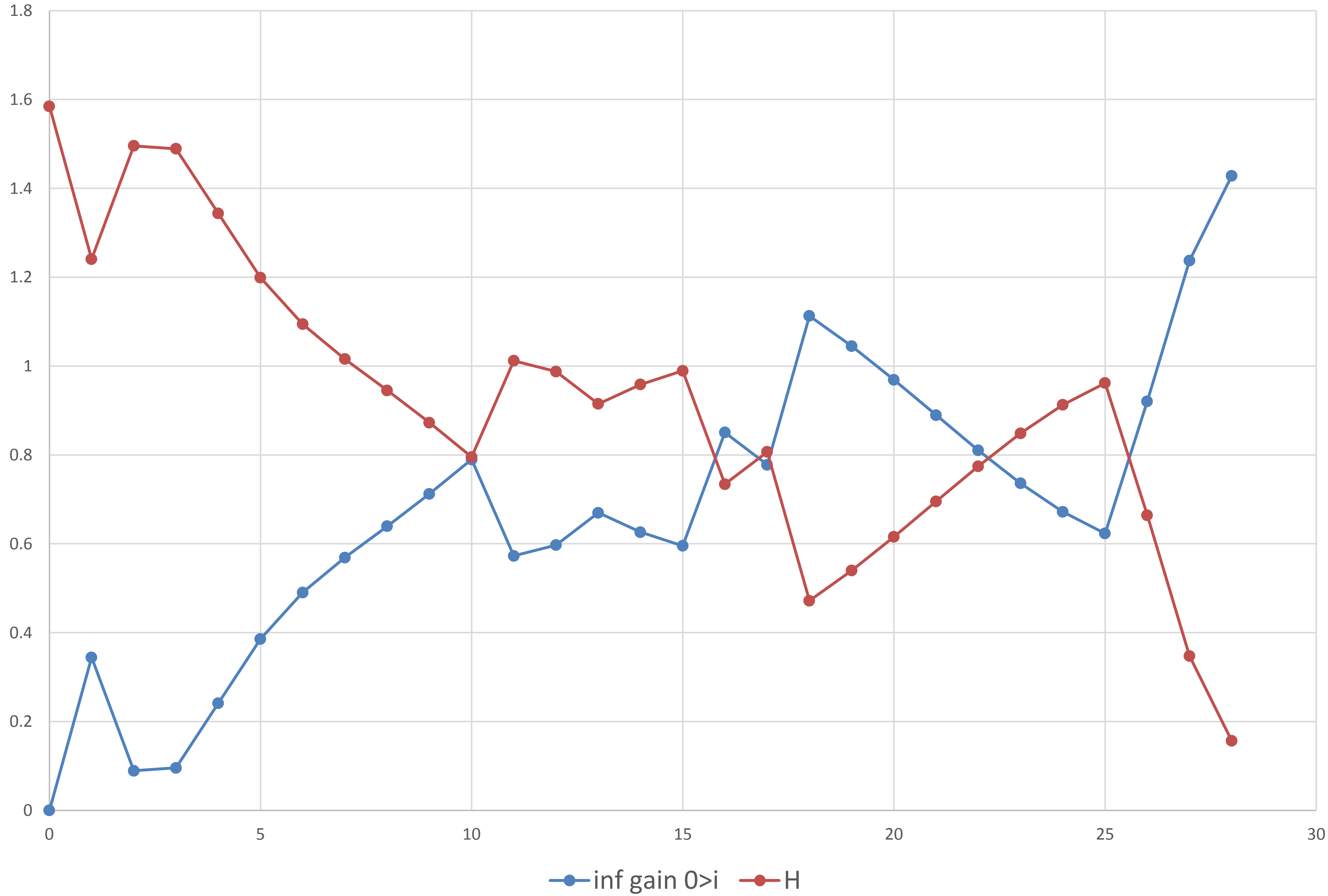
Dreieckregel gilt aber nicht!

$$D_{KL}(P(x|1,2)|Q) \quad <> D_{KL}(P(x|2)|P(x|1))+ D_{KL}(P(x|1)|Q)$$

Noch dazu Info 2 bekommen

Beide Infos gleichzeitig bekommen

Info 1 bekommen


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
inf gain 0>i
H

Utility function: Gewinnfunktion, Nützlichkeitsfunktion, etc.

U(Zustand der Welt) = ein reeller Zahl.

z.B. U(Arbeitstag) = 0,5
U(Urlaubstag) = 30

Subjektiv, ABER nicht ganz zufällig!

- Orderability: $(A \succ B) \vee (A \prec B) \vee (A \sim B)$
- Transitivity: $(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$
- Continuity: $A \succ B \succ C \Rightarrow \exists p [p, A; 1-p, C] \sim B$
- Substitutability: $A \sim B \Rightarrow [p, A; 1-p, C] \sim [p, B; 1-p, C]$
- Monotonicity: $A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A; 1-p, B] \geq [q, A; 1-q, B])$
- Decomposability:

$$[p, A; (1-p), [q, B; (1-q), C]] \sim [p, A; (1-p)q, B; (1-p)(1-q), C]$$

Skaleninvarianz

Schiebungsinvarianz (0-Punkt kann geschoben werden)

Erwartungswert:

$E(U) = Sum(p_i * U(i))$

Wir wollen das maximalisieren.

Wenn wir etwas unternehmen (Therapie-möglichkeiten!) damit verändern wir die Zustände
-> E(U) wird auch verändert.

Nun suchen wir die optimale Handlung, so das E(U)=max.

Suppose you want to be famous

You can be either (N,M,C)

- Nobody
- Modestly Famous
- Celebrity

Your utility function:

- U(N) = 20
- U(M) = 50
- U(C) = 100

You have to decide between going to grad school and becoming a professor (G) or going to Hollywood and becoming an actor (A)

P(N|G)=0.5, P(M|G)=0.4, P(C|G)=0.1

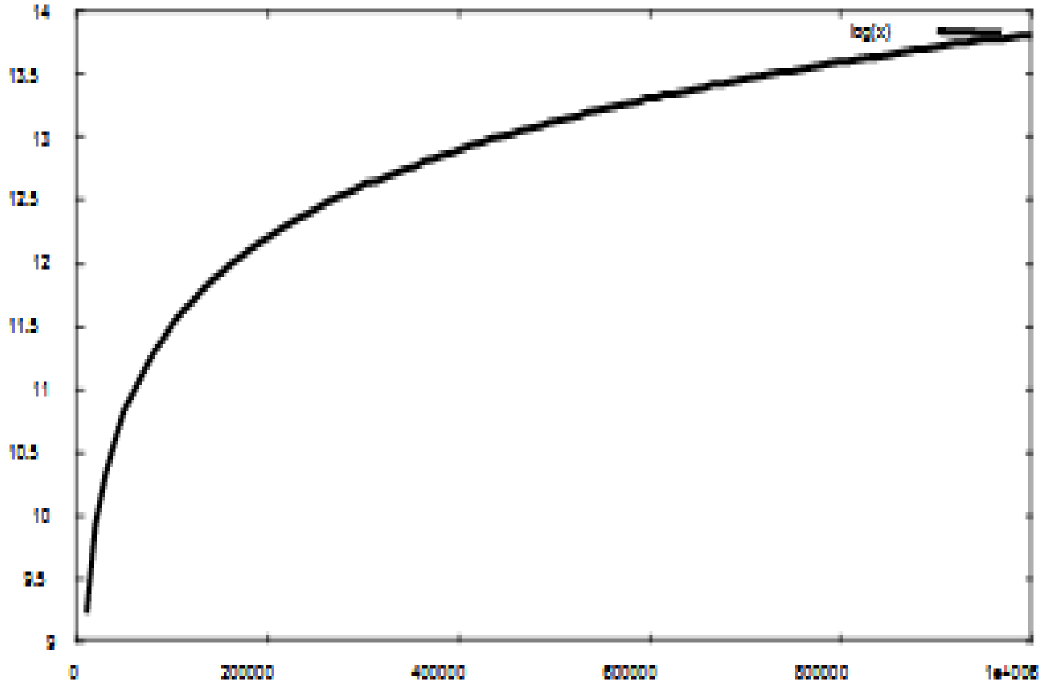
P(N|H)=0.6, P(M|H)=0.2, P(C|H)=0.2

Maximize expected utility:

- U(N) = 20, U(M) = 50, U(C) = 100

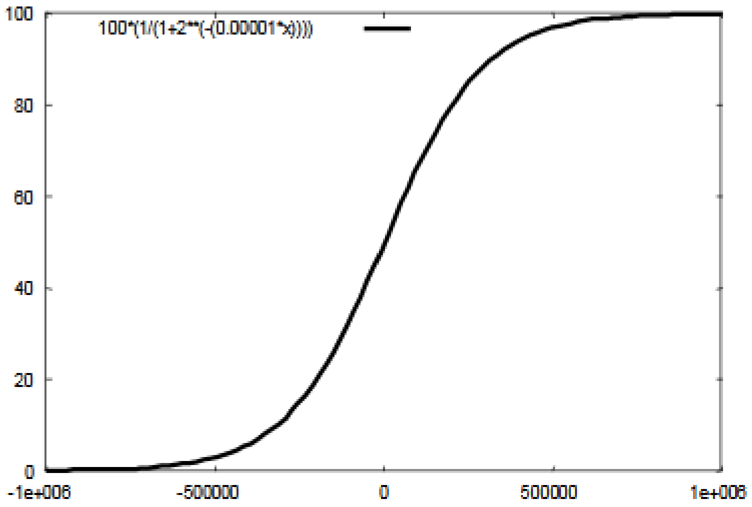
$$EU_G = 0.5(20) + 0.4(50) + 0.1(100) = 40$$
$$EU_H = 0.6(20) + 0.2(50) + 0.2(100) = 42$$

Mit Geld ist die Funktion nicht linear.



(wenn ich viel habe, dann noch 100 EUR ergibt weniger Gewinnsgefühl)

Das gilt auch für Schulden, also:



Wert von Information:

Wie viel verbesasrt sich E(U) wenn
ich die P-werte aktualisieren kann?

Braucht man ein Modell

Oft ist das eine Suche innerhalb von
Omega, das kostet aber auch Geld