Die bedingte Wahrscheinlichkeit in der Spracherkennung

Florian Schiel

Beispiele

Münchner Winterwetter

Automatische Spracherkennung

Was ist die bedingte Wahrscheinlichkeit?

»Einfache« Wahrscheinlichkeiten (1)

R : »Es schneit«

!R : »Es schneit nicht«

} <u>Zufallsereignisse</u>

Anruf beim deutschen Wetterdienst:

$P(R) = 0.05 \qquad P(!R) = 0.95$

Beachte: »Gross P!« => <u>Diskrete Wahrscheinl.</u>

»Einfache« Wahrscheinlichkeiten (2)

/a:/ : »Der Sprachlaut /a:/ wird gesprochen«

/i:/ : »Der Sprachlaut /i:/ wird gesprochen«

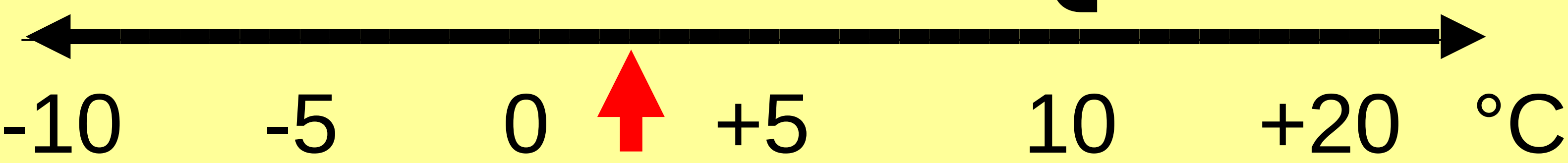
Phonetische Auszählung eines Korpus:

$P(/a:/) = 0.03287 \qquad P(/i:/) = 0.02634$

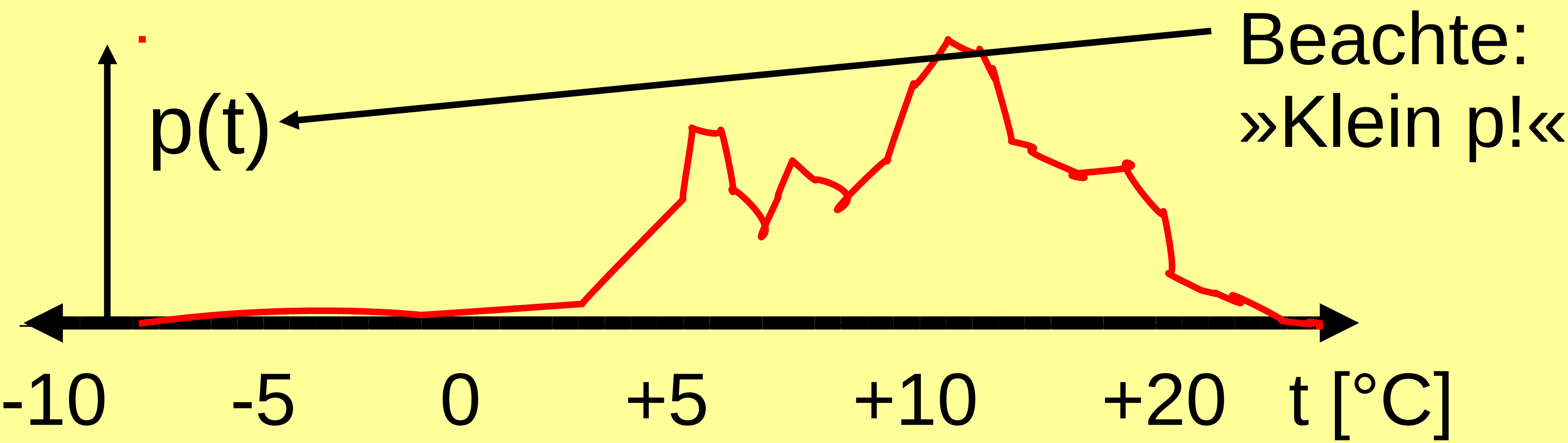
Beachte: »Gross P!« => <u>Diskrete Wahrscheinl.</u>

»Einfache« Wahrscheinlichkeiten (3)

t : »Temperatur um 12 Uhr« { <u>Kontinuierliche Zufallsgrösse</u>

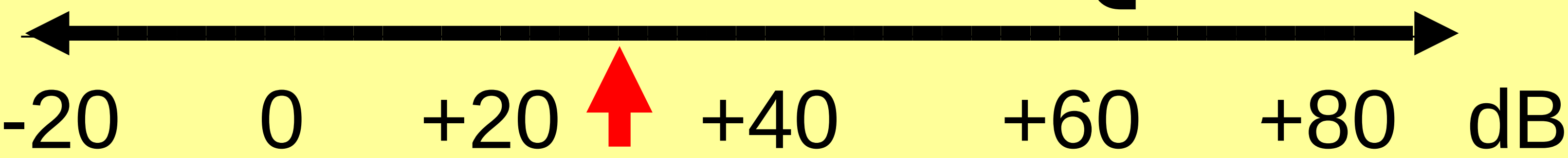


Anruf beim deutschen Wetterdienst:

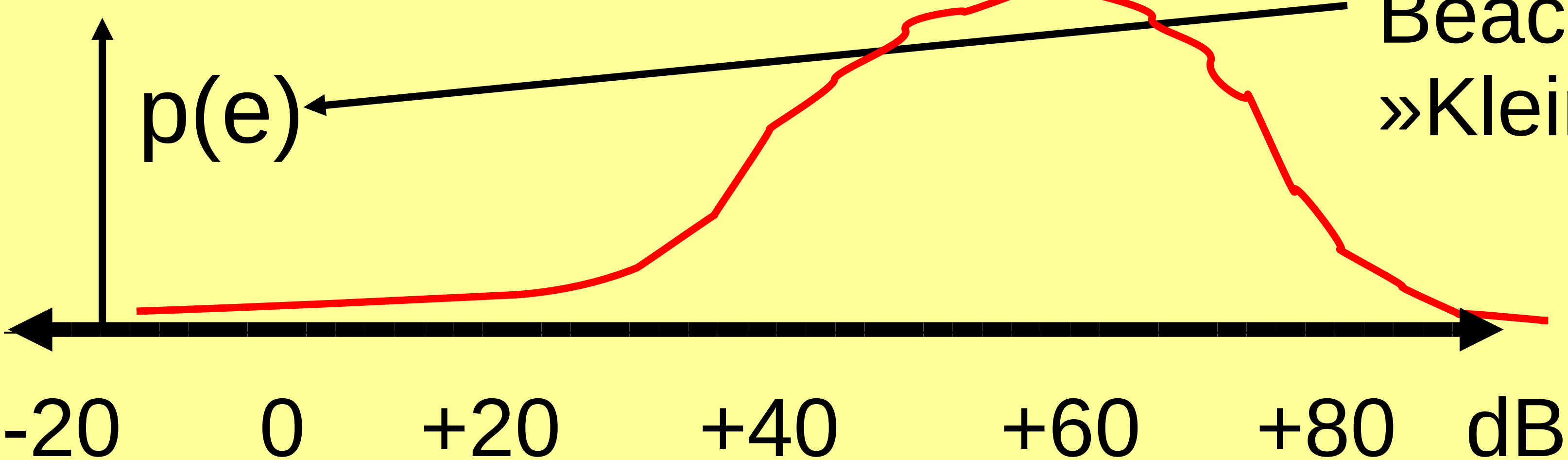


»Einfache« Wahrscheinlichkeiten (4)

e : »Energie von Sprache« { <u>Kontinuierliche Zufallsgrösse</u>



Auswertung von Rundfunksendungen:



Beachte:
»Klein p!«

»Taxonomie« der Wahrscheinlichkeit			
Diskrete		Kontinuierliche	
Einfache		Einfache	
P(R)		p(t)	

Zwischenfragen:

»Würfelwurf?« => diskret P('1')=1/6

»Blutdruck b?« => kontinuierlich p(b)

»Amerikanische Präsidentenwahl?«

=> diskret P(Clinton), P(Trump)

Bedingte Wahrscheinlichkeiten (1)

R : »Es schneit«

G : »Strasse ist glatt«

Anruf beim deutschen Wetterdienst:

$P(R) = 0.05$ $P(G) = ?$

»Mit welcher Wk ist es glatt, wenn es schneit?«

=> $P(G|R) = 0.92$ <u>Diskrete bedingte WK</u>

Bedingte Wahrscheinlichkeiten (2)

$$P(G|R)$$

Ereignis
Zufallsgrösse

Bedingung:
Anderes Ereignis
Andere Zufallsgrösse

G ist diskret => »Gross P!«

Bedingte Wahrscheinlichkeiten (3)

'der' : »Das Wort 'der' wird gesprochen«

'das' : »Das Wort 'das' wird gesprochen«

'Korpus' : »Das Wort 'Korpus' wird gesprochen«

$P(W_2|W_1)$: Wk, dass W_2 <u>nach</u> W_1 gesprochen

Auswertung Schiel'scher Vortragssprache:

$$P('Korpus' \mid 'das') = 0.00267$$

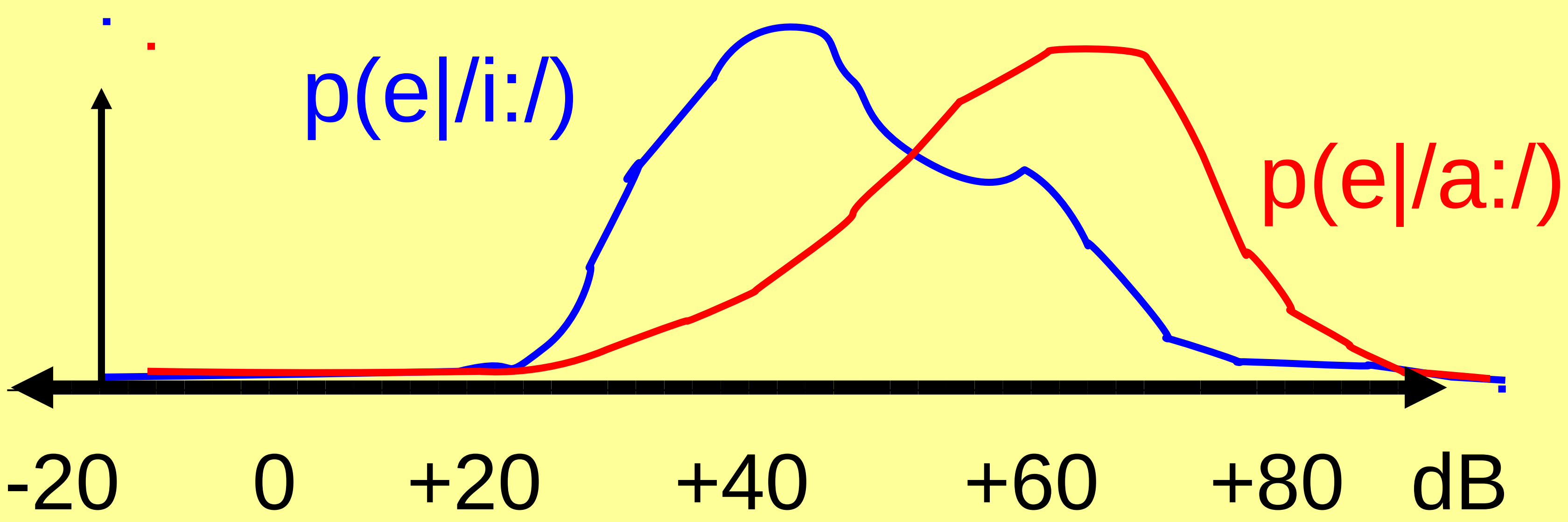
$$P('Korpus' \mid 'der') = 0.00225$$

Bedingte Wahrscheinlichkeiten (4)

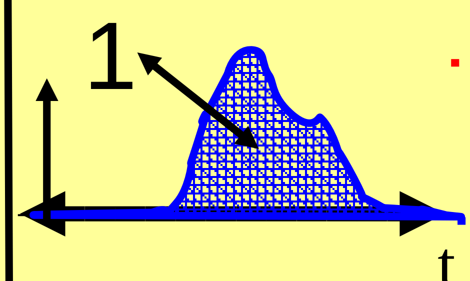
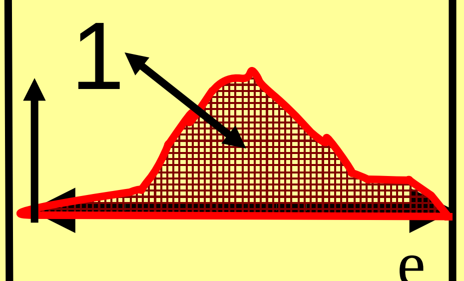
e : »Energie im Sprachsignal«

/a:/ : »Der Laut /a:/ wird gesprochen«

/i:/ : »Der Laut /i:/ wird gesprochen«



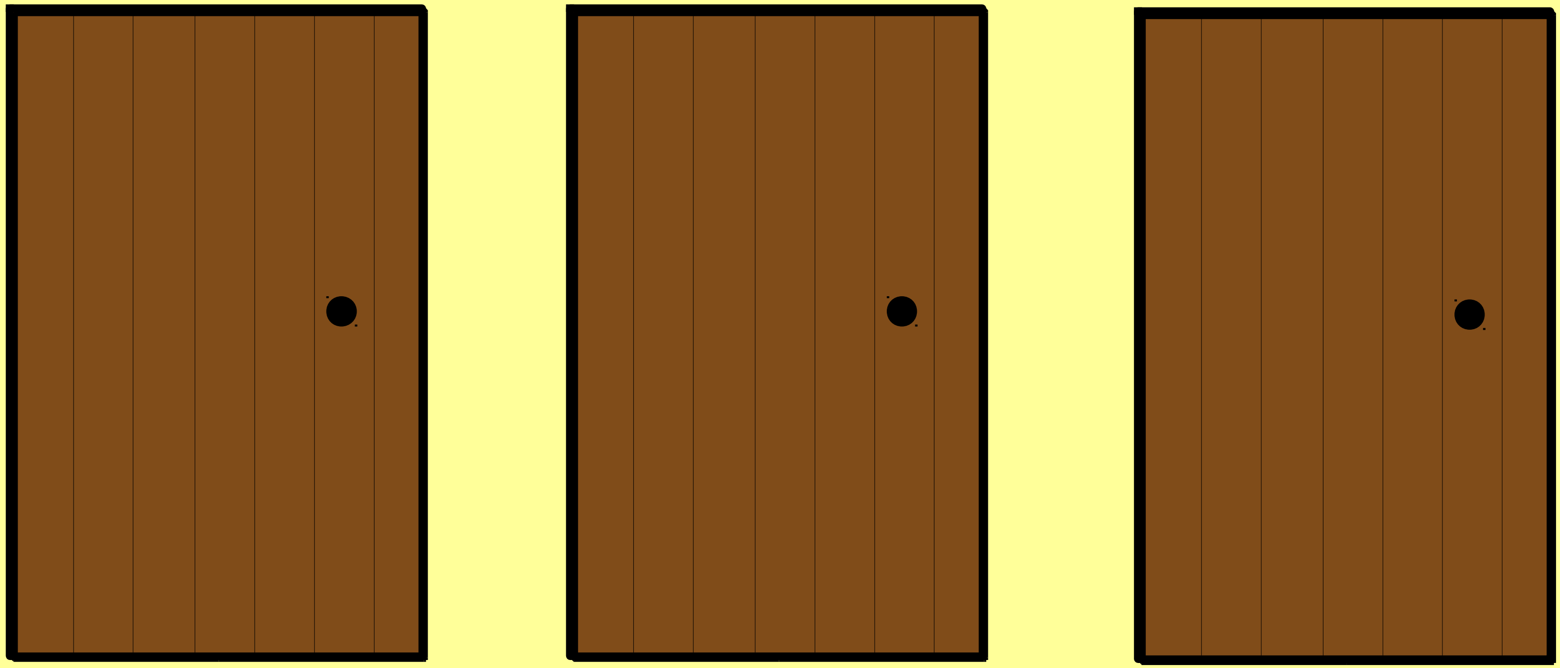
»Taxonomie« der Wahrscheinlichkeit

Diskrete		Kontinuierliche	
Einfache	Bedingte	Einfache	Bedingte
P(R)	P(G|R)	p(t)	p(e|/a:/)
$\sum_{R} P(R)=1$	$\sum_{G} P(G\|R) =1$		

WARNUNG:

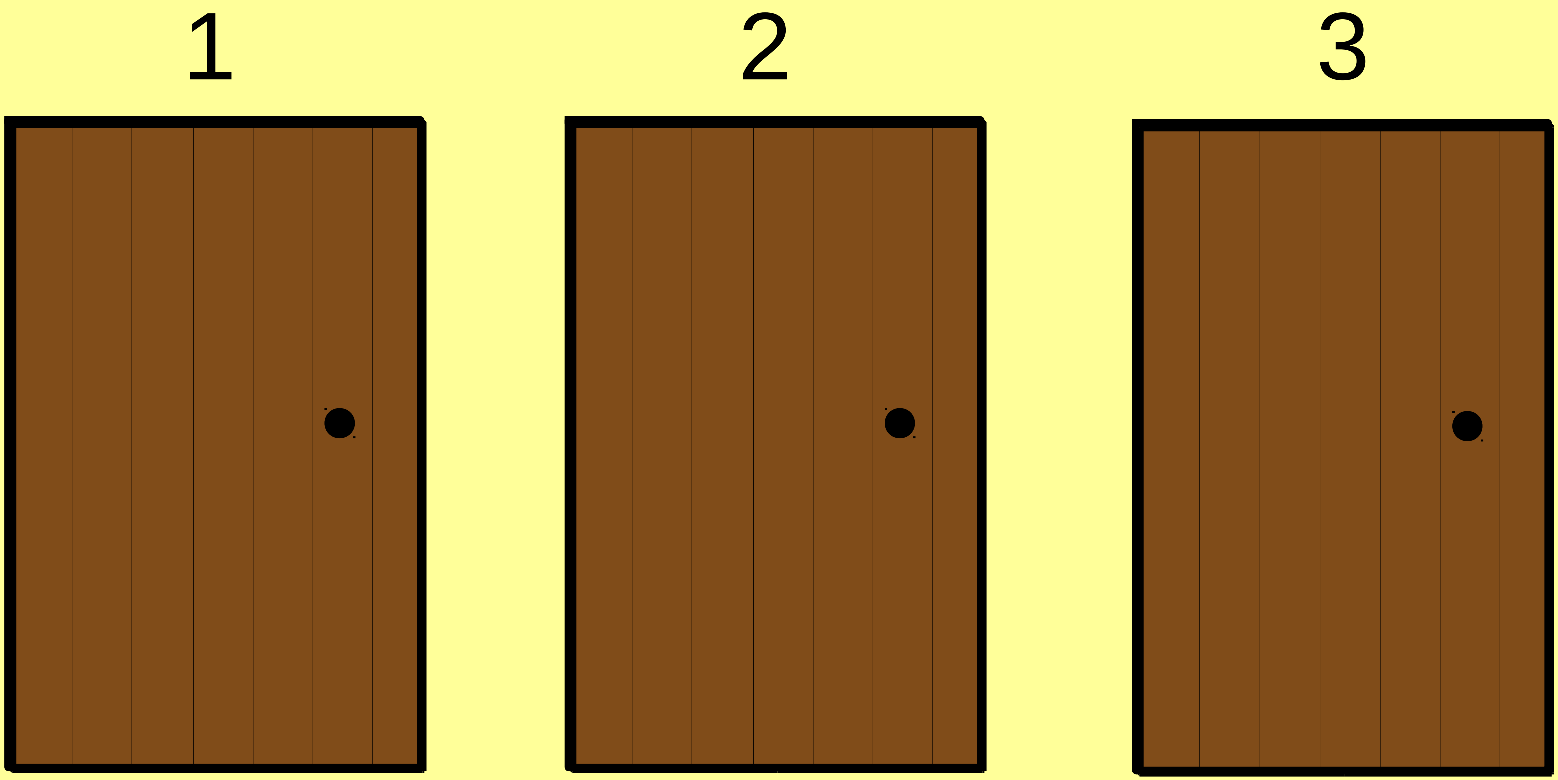
Bedingte Wahrscheinlichkeiten sind oft kontra-intuitiv!

1 2 3

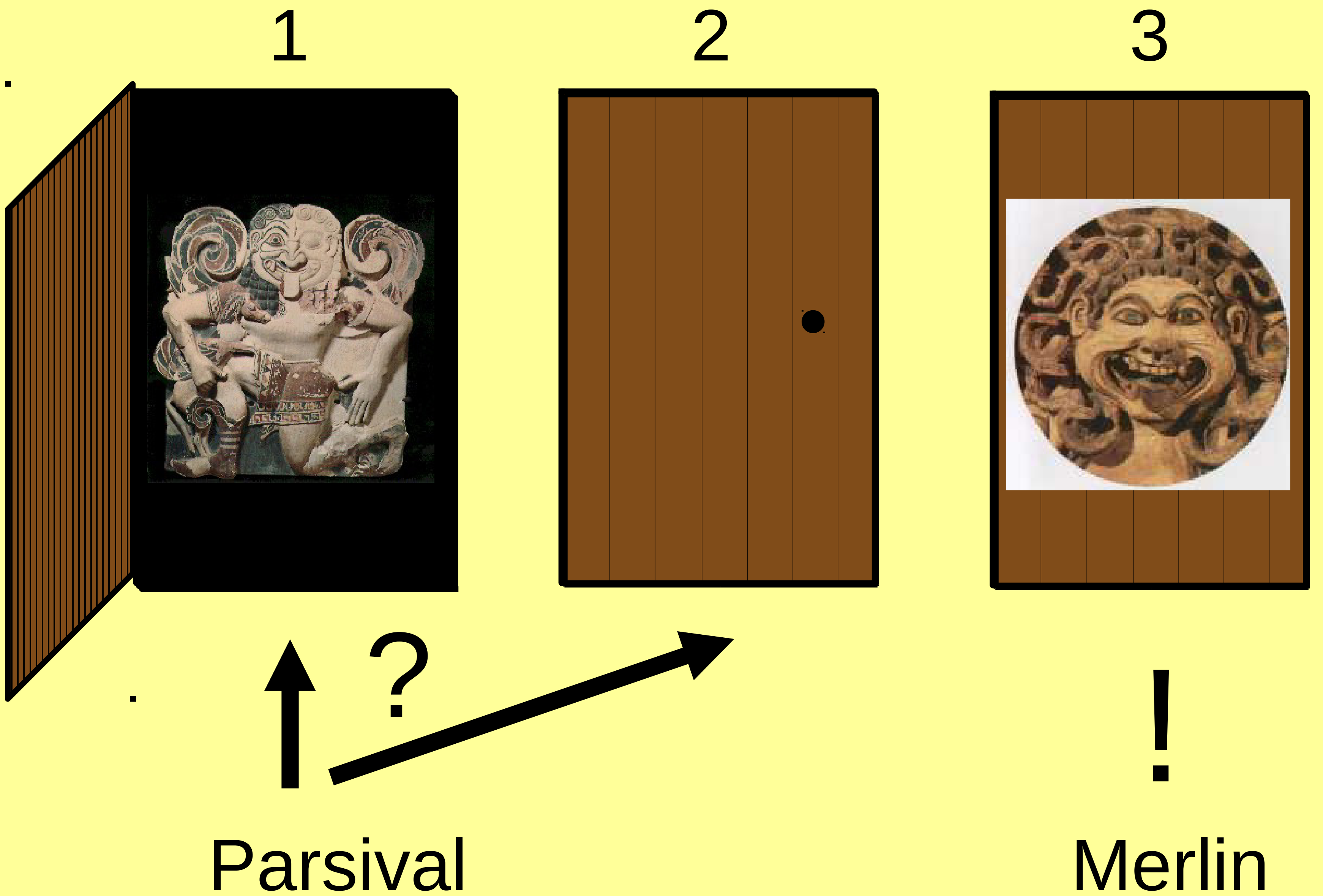


Stheno und Euryale

Gral



P(Gral) = 1/3 P(Gorgone) = 2/3


1
2
3
?
!
Parsival
Merlin

Frage:

Hat Parsival richtig gehandelt?

Ist die Wahrscheinlichkeit tatsächlich gleich oder wäre Parsival (statistisch) besser dran gewesen, sich für die andere Türe zu entscheiden?

Experiment:
Ritter Parsival bleibt immer bei der ersten Tür.
Ritter Mordred entscheidet sich immer anders.

Spiele	Parsival	Mordred
1	gewinnt 0	gewinnt 1
10	gewinnt 4	gewinnt 6
20	gewinnt 7	gewinnt 13
50	gewinnt 18	gewinnt 32
100	gewinnt 34	gewinnt 66
150	gewinnt 50	gewinnt 100

=> Mordred gewinnt doppelt so häufig!

Anschauliche Erklärung:

X P(Gral)=1/3 → X → X <u>P(Gral)=1/3</u>

Parsival wählt das erste Mal

Merlin zeigt Gorgone

Aber

Parsival bleibt dabei

X P(Gorg)=2/3 → X → X <u>P(Gorg)=2/3</u>

Anschauliche Erklärung:

P(Gral)=1/3

P(Gorg)=1/3

Mordred wählt das erste Mal

Merlin zeigt Gorgone

Und

Mordred entscheidet sich anders

P(Gorg)=2/3

P(Gral)=2/3

Erklärung mit bedingter Wahrscheinlichkeit

1. Entscheidung

$\underline{P(Gral=1) = 1/3}$

$P(Gral=2) = 1/3$

$P(Gral=3) = 1/3$

2. Entscheidung

$P(Gral=1|Gorg=2) = 1/3$

$P(Gral=2|Gorg=2) = 0$

$P(Gral=3|Gorg=2) = 2/3$

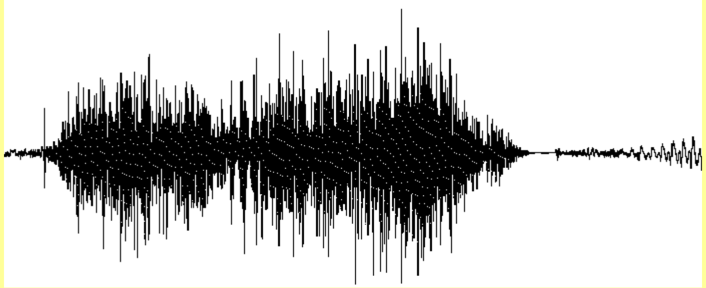
$$\sum_{Gral} P(Gral|Gorg=2) = 1$$

Wozu braucht man die bedingte Wahrscheinlichkeit?

SPRACHERKENNUNG

KATEGORIEN

SIGNALE



Laute
Silben
Wörter
Bedeutungen

Energie
Grundfrequenz
Formanten
Spektren

?

DISKRET

KONTINUIERLICH

Kontinuierliche bedingte Wahrscheinlichkeit

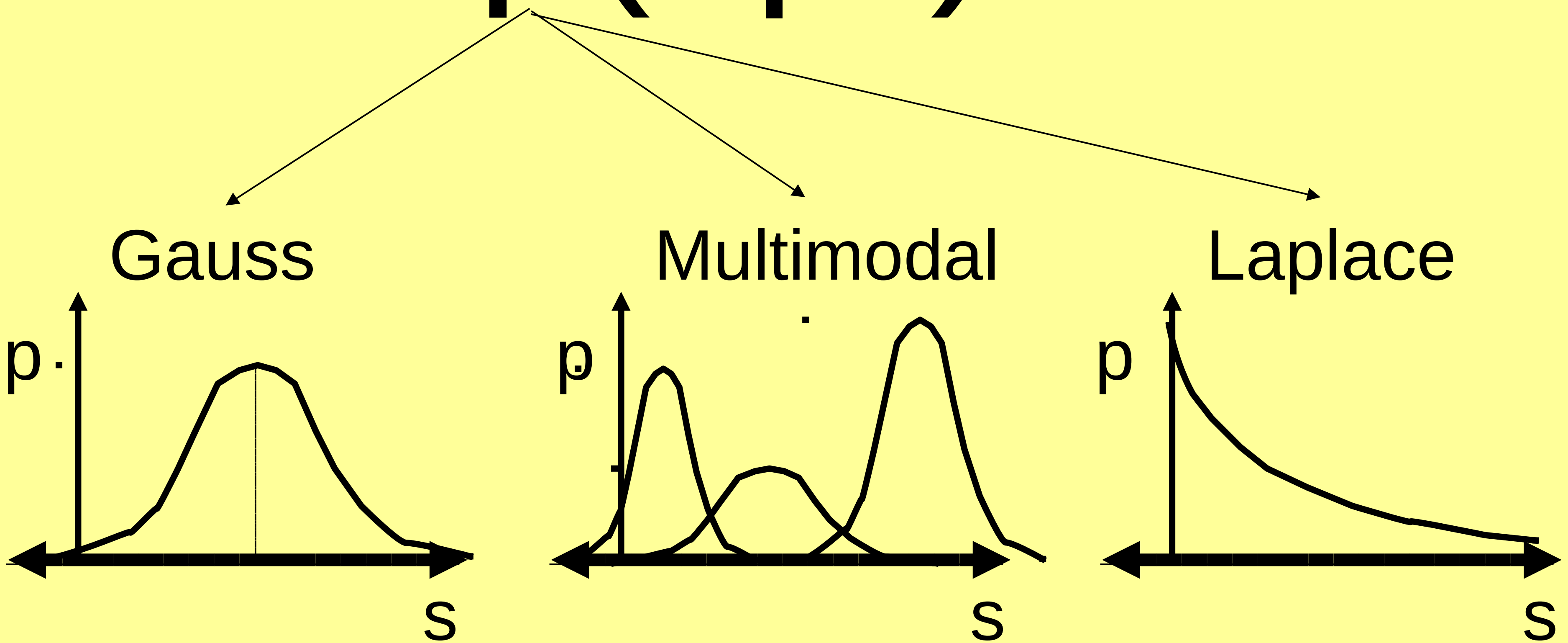
$$p(s|K)$$

SPRACHSIGNAL
Kontinuierliche
Zufallsgrösse

SPRACHKATEGORIE
Diskrete
Zufallsgrösse

Kontinuierliche bedingte Wahrscheinlichkeit

$$p(s|K)$$



SPRACHERKENNUNG

+ Phonotaktik $P(P_2|P_1)$

'Wk, dass Phonem P_2 nach P_1 kommt'

+ Aussprache-Varianten $P(V|W)$

'Wk, dass Wort W als V ausgesprochen wird'

+ Wortfolgen-Statistik $P(W_2|W_1)$

Syntax $P(W_3|W_1,W_2)$

'Wk, dass Wort W_3 nach Wörtern W_1,W_2 auftritt'

Wie rechnet man mit der bedingten Wahrscheinlichkeit?

Normierungen

Diskret

$$\sum_{R} P(R)=1$$

$$\sum_{G} P(G|R) = 1$$

Kontinuierlich

$$\int_{s} p(s)\, ds = 1$$

$$\int_{s} p(s|K)\, ds = 1$$

Bayes'sches Gesetz

»Wk, dass die Straße glatt ist, wenn es schneit« $\neq$ »Wk, dass es schneit, wenn die Straße glatt ist«

$$P(G|R) \neq P(R|G)$$

Bayes

$$P(G|R)\,P(R) = P(R|G)\,P(G)$$

Verbundwahrscheinlichkeit $P(G,R)$

Statistische Unabhängigkeit (1)

»Zwei Ereignisse sind stat. unabh., wenn ihre Verbundwahrscheinlichkeit gleich dem Produkt der Einzelwahrscheinlichkeiten ist.«

$$P(A,B) = P(A)\,P(B) = P(A|B)\,P(B)$$

$$\Rightarrow P(A|B) = P(A)$$
$$\Rightarrow P(B|A) = P(B)$$

Statistische Unabhängigkeit (2)

Lautfolgen:

(Markov-)Annahme:
»Laute, zwischen denen mindestens ein Laut vorkommt, sind statistisch unabhängig.«

h OY t @

P(t | h,_) = P(t | _)

Statistisch unabhängig

Statistische Unabhängigkeit (3)

Welche Wk hat die Lautfolge h OY t @ *?*

$$P(\boxed{h, OY, t,}\ @) =$$

$$= P(\ @ \mid h, OY, t\ )\ P(\ \boxed{h, OY}, t\ ) =$$

$$= P(\ @ \mid h, OY, t\ )\ P(\ t \mid h, OY\ )\ P(\ h, OY\ ) =$$

$$= P(\ @ \mid \cancel{h}, \cancel{OY}, t\ )\ P(\ t \mid \cancel{h}, OY\ )\ P(\ OY \mid h\ )\ P(\ h\ ) \approx$$

$$\approx P(\ @ \mid t\ )\ P(\ t \mid OY\ )\ P(\ OY \mid h\ )\ P(\ h\ )$$

=> *einfaches Bigram-Modell*

Endlich: »Mit welcher Wahrscheinlichkeit ist die Strasse glatt?«

$$P(G) = \frac{0.92 \quad 0.05}{0.67} = 0.0687$$

ENDE

Statistischer Ansatz der Spracherkennung

$$W^* = \underset{W}{\mathit{argmax}} \, [\, P(W|s) \,]$$

$$= \underset{W}{\mathit{argmax}} \left[\frac{P(s|W)\, P(W)}{\cancel{p(s)}} \right]$$

$$= \underset{W}{\mathit{argmax}} \left[\, P(s|W)\, P(W) \, \right]$$