# Start:Bayes!

Der Globus-Versuch

Sebastian Sauer





# Der Globus-Versuch

Wie übersetzen wir eine Forschungsfrage in ein (mathematisches) Modell, das uns mit der Bayes-Formel Antworten liefert?

# Lernziele

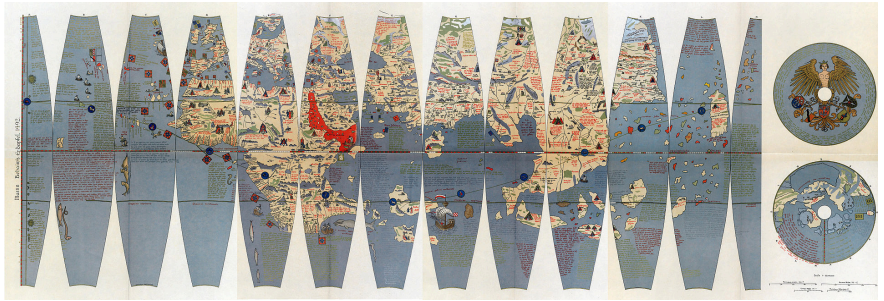
Nach diesem Kapitel können Sie …

- Unterschiede zwischen Modell und Realität erläutern
- die Binomialverteilung heranziehen, um einfache Modelle zu erstellen
- die weite Einsetzbarkeit anhand mehrerer Beispiele zeigen
- das Bayes-Modell anhand bekannter Formeln herleiten
- Post-Wahrscheinlichkeiten mit der Bayesbox berechnen

# Von Welten und Golems



# Kleine Welt, große Welt



- Kolumbus segelte 1492 los — auf seinem Globus existierte Amerika gar nicht
- Die *kleine Welt des Modells* entsprach nicht *der großen Welt*

**Merksatz**

Was in der kleinen Welt (dem Modell) funktioniert, muss nicht in der großen Welt (der Wirklichkeit) funktionieren.

# Der Golem als Sinnbild für Modelle



- Golem: mächtige, aber dumme Kreatur — führt Befehle wörtlich aus
- Wissenschaftliche Modelle sind ähnlich: mächtig, aber simpler als die Realität, Missbrauch leicht möglich

**Merksatz**

Wir bauen Golems: Modelle, die Befehle (Annahmen) wörtlich umsetzen — nicht mehr, nicht weniger.

# Kleine Welt vs. große Welt

| Kleine Welt | Große Welt |
|---|---|
| Die Welt, wie sie der Golem (das Modell) sieht | Die Welt, wie sie wirklich ist |
| ist das Modell, nicht zwangsläufig die Wirklichkeit | entspricht nicht zwangsläufig dem Modell |
| Verwenden wir beim Modellieren | Ist das, was wir modellieren |

# Ein erster Versuch: Wir werfen den Globus

# Forschungsfrage: Wasseranteil der Erde



- Wie hoch ist der Wasseranteil $\pi$ auf der Erdoberfläche?
- Versuch: Globusball werfen, auffangen, notieren: Wasser (W) oder Land (L)
- 9 Würfe, Ergebnis:

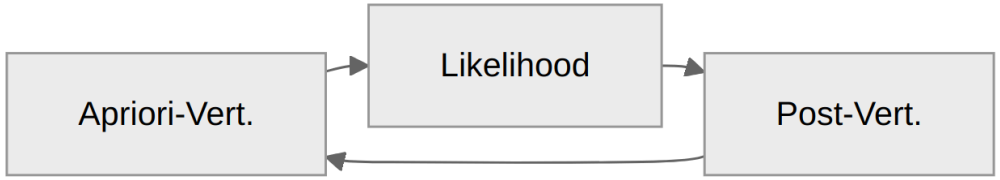
$$W \quad L \quad W \quad W \quad W \quad L \quad W \quad L \quad W$$

Also $W = 6$ Treffer, $L = 3$ Nieten, $n = 9$ Versuche.

# Das Prinzip: Bayes-Update

Der Bayes-Golem denkt vernünftig:

1. **Apriori**: Vorwissen zum Wasseranteil (hier: alle Werte gleich plausibel)
2. **Likelihood**: Wie wahrscheinlich sind die Daten unter einer Hypothese?
3. **Aposteriori**: abschließende, aktualisierte Meinung



# Definition: Apriori-Verteilung

**Definition: Apriori-Verteilung**

Für jede Hypothese haben wir ein Vorab-Wissen, das die jeweilige Plausibilität der Hypothese angibt: die *Apriori-Verteilung* (auch: Priori-Verteilung, Prior).

# Definition: Likelihood

**Definition: Likelihood**

Für jede Hypothese (jeden Parameterwert $\pi$) gibt die *Likelihood* an, wie wahrscheinlich die Daten sind, unter der Annahme, dass die Hypothese richtig ist. Kurz: wie gut die Daten zur Hypothese passen.

# Definition: Aposteriori-Verteilung

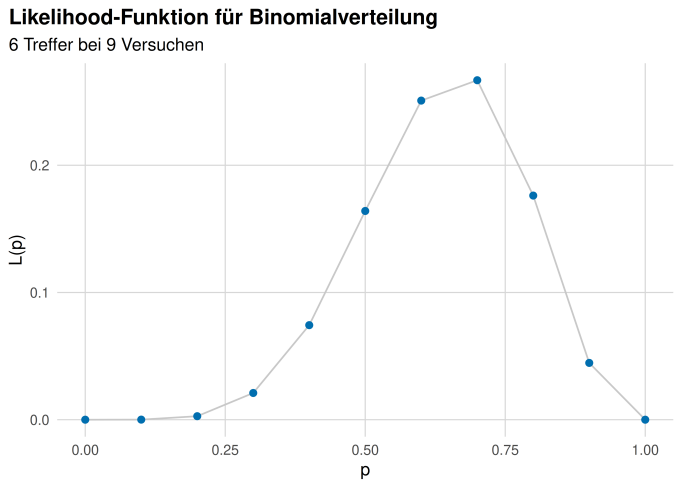
**Definition: Aposteriori-Verteilung**

Gewichtet man die Likelihood mit dem Vorabwissen, erhält man die *Aposteriori-Verteilung* (kurz “Post-Verteilung”), $Pr(H|D)$: “Wahrscheinlichkeit der Hypothese H gegeben der Daten D.”

# Likelihood berechnen

- Wie wahrscheinlich ist $W = 6$ Treffer bei $n = 9$ Würfen, gegeben einer Hypothese $\pi$?
- Annahme: Binomialverteilung (iid: unabhängige, identisch verteilte Würfe)

$$Pr(\pi = 1/2 \mid x = 6, n = 9) = \binom{9}{6} \cdot (1/2)^6 \cdot (1/2)^3$$
$$= 84 \cdot 1/512 = 0.16$$




Ein erster Versuch: Wir werfen den Globus

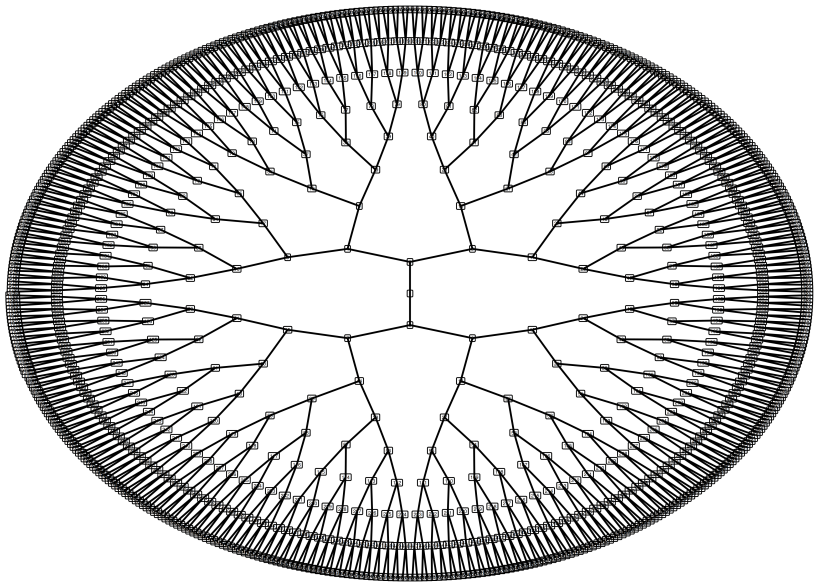
# Likelihood für verschiedene Hypothesen

- $\pi = .7$: $\binom{9}{6} \cdot (.7)^6 \cdot (.3)^3 = 0.267$ — Daten passen gut
- $\pi = 1/3$: deutlich kleiner — schlechte Hypothese
- $\pi = 0$: Likelihood $= 0$ — Daten sind unmöglich unter dieser Hypothese

**Merksatz**

Je höher die Likelihood, desto besser passt die Hypothese zu den beobachteten Daten.

# Warum nicht per Baumdiagramm?



Für $n = 9$ Würfe wird ein Baumdiagramm (512 Endknoten) schnell unübersichtlich — die Binomialverteilung übernimmt diese Arbeit kompakt.

# Unser Modell ist geboren

Ein Bayes-Modell besteht aus mindestens drei Komponenten:

1. **Likelihood**: Wahrscheinlichkeit der Daten unter der Hypothese
2. **Apriori-Verteilung(en)**: Wahrscheinlichkeit der Hypothese vor den Daten
3. **Aposteriori-Verteilung**: Wahrscheinlichkeit der Hypothese nach den Daten

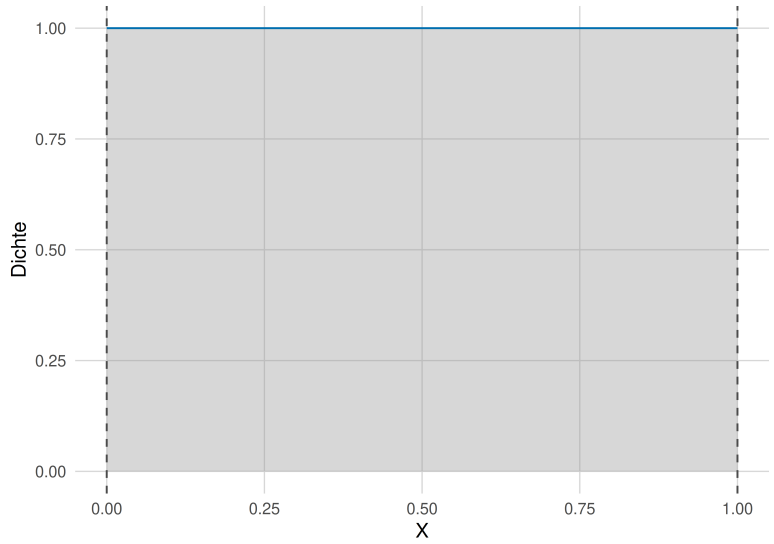




# Apriori-Verteilung im Beispiel

$$\pi \sim \text{Unif}(0, 1)$$

- Alle Werte von $\pi$ erscheinen uns gleich plausibel (“uniform”)
- Kein Vorwissen: apriori indifferent




Ein erster Versuch: Wir werfen den Globus

# Bayes' Theorem

# Die Formel von Bayes' Theorem

**Theorem: Bayes' Theorem**

$$Pr(H|D) = \frac{Pr(H) \cdot Pr(D|H)}{Pr(D)} = \frac{\text{Apriori} \cdot \text{Likelihood}}{\text{Evidenz}}$$

**Frage**: Was ist die Wahrscheinlichkeit der Hypothese H, jetzt wo wir die Daten kennen?

**Antwort**: Apriori-Wahrscheinlichkeit mal Likelihood, standardisiert durch die Evidenz.





# Definition: Evidenz

**Definition: Evidenz**

$Pr(D)$ heißt *Evidenz*: die Summe der Likelihoods über alle Parameterwerte (Hypothesen) $H_i$, gewichtet mit ihrer Apriori-Wahrscheinlichkeit.

$$Pr(D) = \sum_{i=1}^{n} Pr(D|H_i) \cdot Pr(H_i)$$

Die Evidenz sorgt nur dafür, dass $Pr(H|D)$ zwischen 0 und 1 liegt.





# Post = Priori × Likelihood

$$Pr_{\text{unPost}} = L \times \text{Priori}$$

$$\text{Posteriori} = \frac{\text{Likelihood} \times \text{Priori}}{\text{Evidenz}}$$

**Merksatz**

Sind alle Apriori-Wahrscheinlichkeiten gleich, ist die standardisierte Post-Verteilung identisch mit der (normierten) Likelihood.

# Golems können lernen

Nach jedem Wurf verändert sich die Post-Verteilung:

- Apriori (vor dem 1. Wurf): jeder Parameterwert gleich wahrscheinlich
- Mit jedem Datenpunkt (Wurf) wird die Verteilung schärfer/informativer
- Das Modell “lernt” — ob es *nützlich* ist, ist eine andere Frage





# Die Post berechnen mit der Bayesbox



# Die Bayesbox: Idee

Eine kleine Tabelle (“Bayesbox”, “Gitter-Methode”), um Post-Wahrscheinlichkeiten zu berechnen:

1. Wertebereich des Parameters in ein Gitter aufteilen (z. B. $0, 0.1, \ldots, 1$)
2. Apriori-Wahrscheinlichkeit je Parameterwert festlegen
3. Likelihood je Parameterwert berechnen
4. Unstandardisierte Post = Apriori × Likelihood
5. Standardisieren: durch die Summe aller unstandardisierten Werte teilen (Evidenz)





# Die Bayesbox für den Globusversuch

$k = 6$ Treffer, $n = 9$ Versuche, $Pr(H) = 1/11$

| id | p | Apriori | Likelihood | unstd. Post | Post |
|---|---|---|---|---|---|
| 1 | 0.0 | .091 | .000 | .000 | .000 |
| 4 | 0.3 | .091 | .021 | .002 | .021 |
| 6 | 0.5 | .091 | .164 | .015 | .164 |
| **8** | **0.7** | **.091** | **.267** | **.024** | **.267** |
| 9 | 0.8 | .091 | .176 | .016 | .176 |
| 11 | 1.0 | .091 | .000 | .000 | .000 |

Evidenz $= \sum \text{unstd. Post} = 0.09$





# Beispiel: Post für p = .7

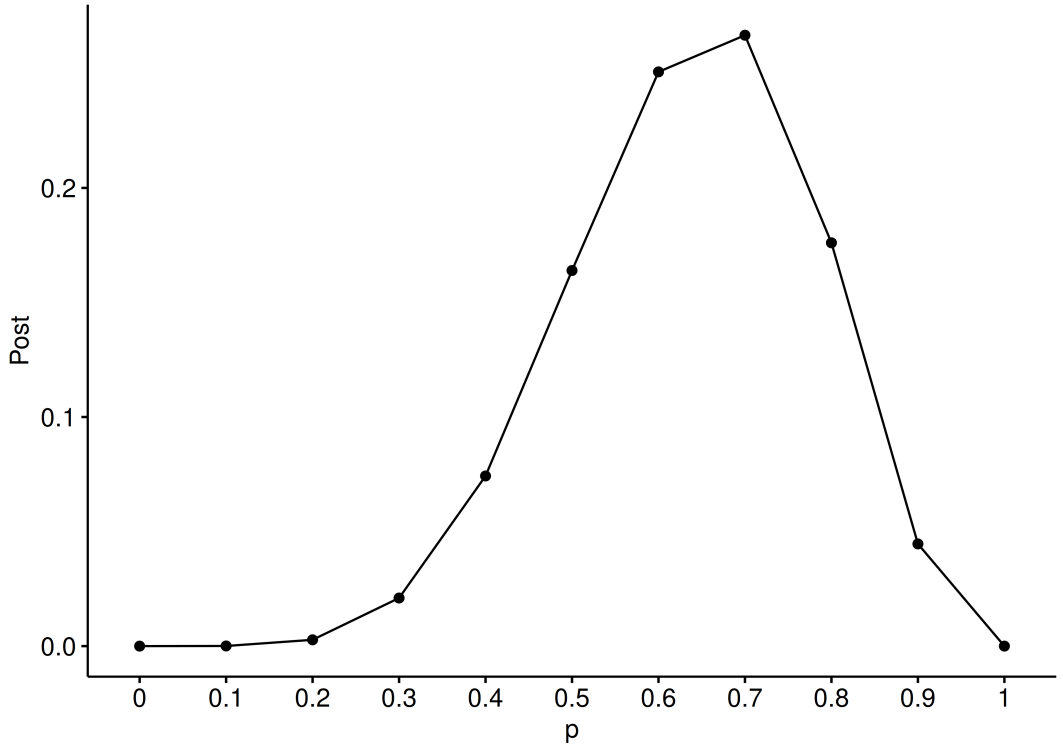
$$\text{Post}_{\text{unstand}} = Pr(H) \cdot Pr(D|H) = 1/11 \cdot 0.267 = 0.024$$

$$\text{Post} = \frac{0.024}{0.09} = 0.267$$

**Fazit**

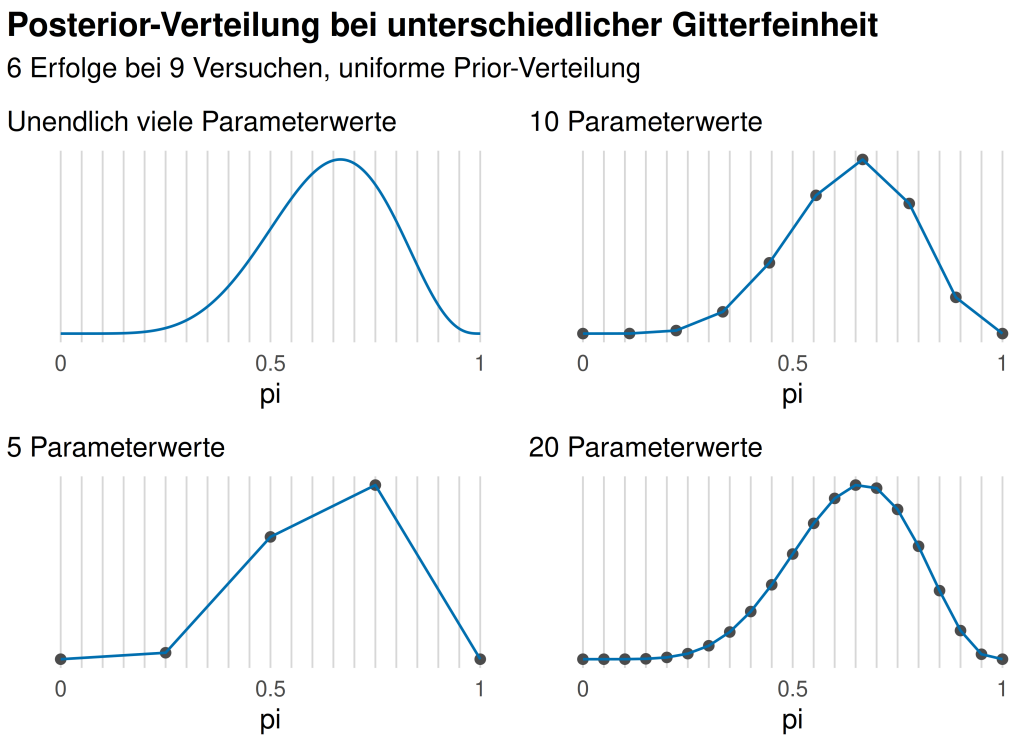
Nach dem Versuch hat sich unsere Meinung über $\pi = .7$ von 0.09 (Apriori) auf 0.27 (Post) aktualisiert — eine Verdreifachung.

# Post-Verteilung visualisiert




Der Bereich 0.6–0.8 ist ebenfalls recht wahrscheinlich.

# Wie viele Gitterpunkte?




Je mehr Parameterwerte im Gitter, desto “glatter” wird die Verteilung.

# Abschluss



# Der Globusversuch als allgemeines Modell

Prototypisch für viele zweiwertige Zufallsversuche:

- Produkt: von $n$ Exemplaren wurden $k$ verkauft — wie hoch ist die Akzeptanzrate $p$?
- Chat-Bot: von $n$ Fragen $k$ richtig beantwortet — wie hoch ist die Verständnisrate $p$?
- Krebstherapie: von $n$ Patient:innen wurden $k$ geheilt — wie hoch ist die Erfolgsrate $p$?

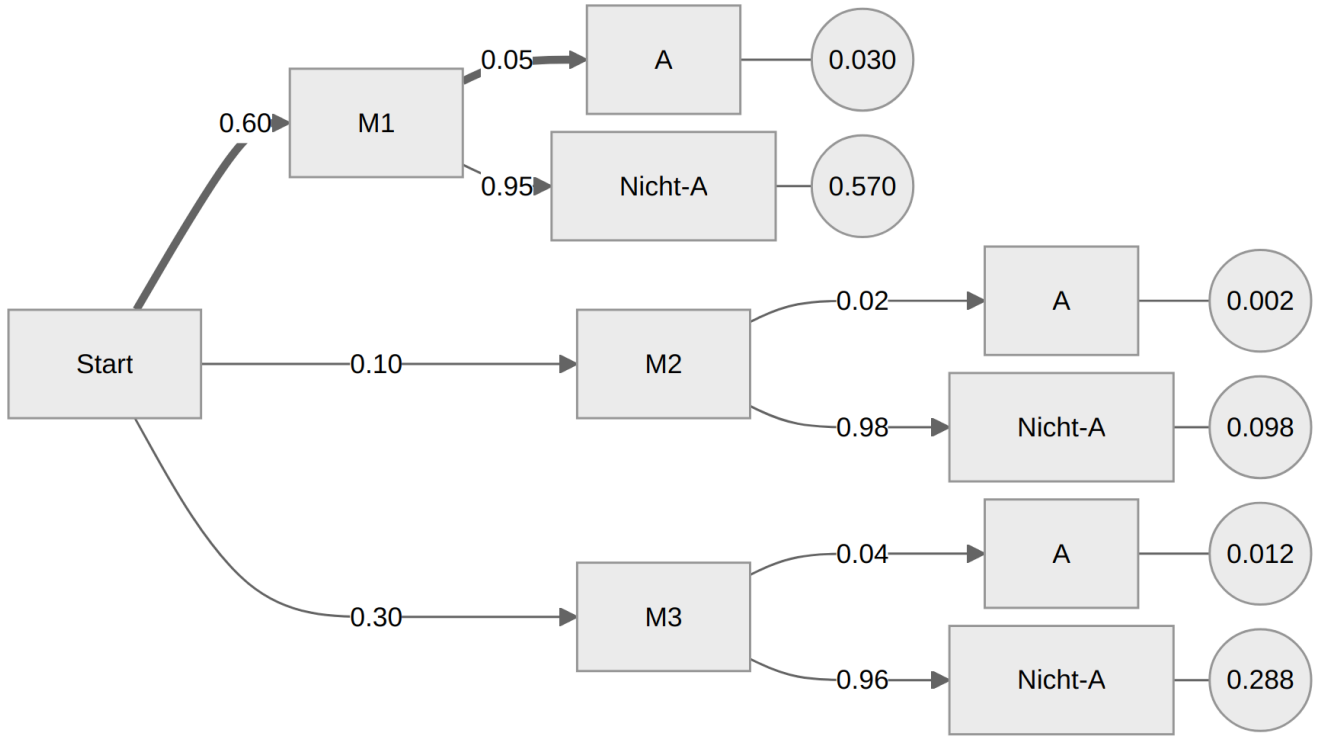
# Vertiefung

# Bayes als Baumdiagramm: Beispiel

Beispiel: Drei Maschinen $M_1, M_2, M_3$ produzieren Anteile 60 %, 10 %, 30 % — Ausschussquote 5 %, 2 %, 4 %.

Frage: Wie wahrscheinlich ist $Pr(M_1|A)$ — Teil von M1, gegeben Ausschuss?

# Bayes als Baumdiagramm



# Baumdiagramm-Lösung

$$Pr(M1|A) = \frac{Pr(M1 \cap A)}{Pr(A)} = \frac{0.6 \cdot 0.05}{0.03 + 0.002 + 0.012} = \frac{0.03}{0.044} \approx 0.68$$

- Zähler: der günstige (gesuchte) Pfad
- Nenner: die Summe aller relevanten Pfade — die Evidenz $Pr(A)$

# Bayes als bedingte Wahrscheinlichkeit

Bayes' Theorem ist nichts anderes als eine umgeformte bedingte Wahrscheinlichkeit:

$$Pr(H|D) = \frac{\overbrace{Pr(H \cap D)}^{\text{umformen}}}{Pr(D)} = \frac{\overbrace{Pr(H)}^{\text{Apriori}} \cdot \overbrace{Pr(D|H)}^{\text{Likelihood}}}{\underbrace{Pr(D)}_{\text{Evidenz}}}$$

Herleitung: $Pr(D \cap H) = Pr(H) \cdot Pr(D|H) = Pr(D) \cdot Pr(H|D)$ — nach $Pr(H|D)$ auflösen.

# Zusammengesetzte Hypothesen

Bayes' Theorem funktioniert auch für zusammengesetzte Hypothesen $H = H_1 \cap H_2$:

$$Pr(R \cap B \mid K) = \frac{Pr(R) \cdot Pr(B) \cdot Pr(K \mid R \cap B)}{Pr(D)}$$

Beispiel: Wahrscheinlichkeit von Regen ($R$) *und* Blitzeis ($B$), wenn es kalt ($K$) ist — gültig, wenn $R$ und $B$ unabhängig sind.





# Zusammenfassung

- Modell besteht aus Apriori, Likelihood und Post-Verteilung
- Der Golem updated sein Wissen: Apriori × Likelihood → Post (standardisiert durch die Evidenz)
- Die Bayesbox macht diese Rechnung für viele Hypothesen (Parameterwerte) explizit
- Unser Modell bildet die *kleine Welt* ab — ob es in der *großen Welt* nützlich ist, steht auf einem anderen Blatt

**Tipp**

Der Globusversuch ist ein Muster für alle zweiwertigen Zufallsversuche — in Wissenschaft, Business und Alltag.

# Vielen Dank!

Fragen?



