Einführung in die Bayes Statistik- Slides Begleitung zum Online Kurs

Einführung in die Wahrscheinlichkeit und Grundlagen

Martje Rave

Sommersemester 2025

ACHTUNG!

Diese Folien sind **NUR** Kurs begleitend und können noch viele Fehler beinhalten. Für die Klausurvorbereitung benutzt bitte den Online-Kurs und die Übungsblätter.
Bitte gebt mir Bescheid, wenn ihr Fehler findet.

Was ist Wahrscheinlichkeit?

Dieses Kapitel gibt eine kurze Einführung bzw. Wiederholung in die Theorie der Wahrscheinlichkeit.

Wir interessieren uns (vorerst) für *zufällige Ergebnisse* in einem Ergebnisraum Ω. Jedem Ereignis $A \subseteq \Omega$ kann eine Wahrscheinlichkeit $P(A)$ zugeordnet werden.

Axiome von Kolmogorov

Für Wahrscheinlichkeiten gelten die drei Axiome:

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. Für disjunkte Ereignisse A und B gilt:

$$P(A \cup B) = P(A) + P(B)$$

Beispiel: Niederschlag

Beispielereignis: *Niederschlag am morgigen Tag*.
Ergebnisraum:

$$\Omega = \{\text{"Es regnet"}, \text{"Es schneit"}, \text{"Es regnet und schneit nicht"}\}$$

Ereignis $A = \{\text{"Es regnet"}, \text{"Es schneit"}\}$

$$P(A) = P(\text{"Es regnet"}) + P(\text{"Es schneit"})$$

Folgerungen

Aus den Axiomen ergeben sich:

- $P(\emptyset) = 0$ (Wahrscheinlichkeit für das unmögliche Ereignis)
- $P(\bar{A}) = 1 - P(A)$ (Wahrscheinlichkeit des Gegenereignisses)

Beispiel: Würfeln

Wir werfen einen fairen, sechsseitigen Würfel.
Ereignis: "Es fällt eine 6"

$$P(A) = \frac{1}{6}$$

Allgemein für Laplace-Wahrscheinlichkeiten:

$$P(A) = \frac{|A|}{|\Omega|}$$

wobei $|A|$ die Anzahl der Elemente in A ist.

Odds (Chance)

Alternativ lassen sich Wahrscheinlichkeiten auch als *Odds* oder *Chance* darstellen:

Definition:

$$\text{Odds}(A) = \frac{P(A)}{P(\bar{A})}$$

wobei $\bar{A}$ das *Gegenereignis* zu A ist, also $P(\bar{A}) = 1 - P(A)$.

Ein Odds von 1 entspricht einer Wahrscheinlichkeit von 50 %.

Fortsetzung Beispiel: Würfel

Ereignis A: *"Der Würfel zeigt die Zahl 6"*

$$\text{Odds}(A) = \frac{1/6}{5/6} = 1 : 5$$

Man kennt Odds auch als **Wettquote**.
Bei $P(A) = \frac{1}{6}$ beträgt die Wettquote $5 : 1$.
Setzt man auf A, erhält man im Gewinnfall den Einsatz plus das Fünffache zurück:

$$\text{Auszahlung} = 6 \times \text{Einsatz}$$

Interpretation von Wahrscheinlichkeit

Der mathematische Begriff **Wahrscheinlichkeit** kann unterschiedlich interpretiert werden:

- **klassisch:** Gleichverteilung über Elementarereignisse (Laplace-Wahrscheinlichkeit)
- **frequentistisch:** als relative Häufigkeit bei Wiederholung (Frequentistischer Wahrscheinlichkeitsbegriff)
- **subjektiv:** Maß für den persönlichen Glauben (Bayesscher Wahrscheinlichkeitsbegriff)
- **propensitär:** Eigenschaft eines physikalischen Systems (z.B. nach Karl Popper)

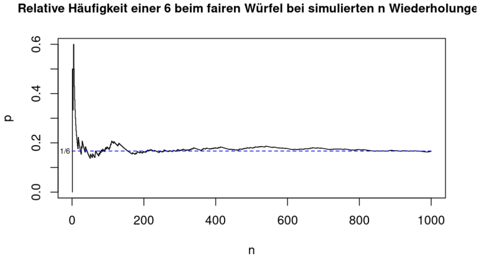
Frequentistischer Wahrscheinlichkeitsbegriff

Basierend auf dem **Gesetz der großen Zahlen**:

Wiederholt man ein Experiment n-mal, bei dem Ereignis A mit Wahrscheinlichkeit π eintritt, dann gilt für die relative Häufigkeit:

$$p = \frac{a_n}{n} \to \pi \quad \text{für } n \to \infty$$

wobei a_n die Anzahl der Experimente ist, in denen A eintrat.



Bayesscher Wahrscheinlichkeitsbegriff

Die Bayessche Wahrscheinlichkeit ist **subjektiv**:

- Maß für den *Grad persönlicher Überzeugung*, dass ein Ereignis eintritt
- Kann durch **Wetten** quantifiziert werden (z.B. Odds 5:1 für eine 6)
- Überzeugung basiert auf:
  - *Vorwissen* (z.B. Eigenschaften eines Würfels)
  - *Beobachtungen/Daten* (z.B. Häufigkeit in Wiederholungen)
- **Wahrscheinlichkeit = Wissen durch Information**

(Bayessche Wahrscheinlichkeit ist nicht exklusiv für die Bayessche Statistik)

Verteilung:

- Wenn wir für jedes mögliche Ereignis $A \subseteq \Omega$ die Wahrscheinlichkeit $P(A)$ definieren, sprechen wir insgesamt von einer Verteilung.
- Im Beispiel des fairen Würfels haben wir eine Gleich-Verteilung angenommen; jedes gleichgroße Ereignis A hat die selbe Wahrscheinlichkeit.

Zufallsvariablen:
Beispiel Würfel

- Ereignis A_1: "Die Seite mit einem Auge liegt oben"
- Ereignis A_2: "Die Seite mit zwei Augen liegt oben"
- Ereignis A_3: "Die Seite mit drei Augen liegt oben"
- Ereignis A_4: "Die Seite mit vier Augen liegt oben"
- Ereignis A_5: "Die Seite mit fünf Augen liegt oben"
- Ereignis A_6: "Die Seite mit sechs Augen liegt oben"

Sei die Zufallsvariable $\boldsymbol{X}$ "Anzahl der Augen", dann gilt für alle $i = 1, \ldots, 6$

- tritt A_i ein, dann ist $X = i$

Verteilung und Zufallsvariablen

Damit lassen sich auch für Zufallsvariablen Wahrscheinlichkeiten und damit Verteilungen definieren:

- $P(X = 1) = P(A_1) = \frac{1}{6}$
- $P(X \geq 5) = P(A_5) + P(A_6) = \frac{1}{3}$

Die Verteilung einer Zufallsvariablen wird dadurch spezifiziert, dass alle Wahrscheinlichkeiten für alle möglichen Realisation angegeben werden.

Verteilung und Zufallsvariablen

Binomialverteilung Sei $\boldsymbol{Y}$ die Zufallsvariable "Bei n-*maligen* Würfen, würfeln wir y-mal die 6. Dann ist $\boldsymbol{Y}$ Binomialverteilt, wir schreiben $Y \sim B(n, p)$, und es gilt für $0 \leq y \leq n$

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

wobei hier $p = \frac{1}{6}$.

- Wenn wir für jedes mögliche Ereignis $A \subseteq \Omega$ die Wahrscheinlichkeit $P(A)$ definieren, sprechen wir insgesamt von einer Verteilung.
- Im Beispiel des fairen Würfels haben wir eine Gleich-Verteilung angenommen; jedes gleichgroße Ereignis A hat die selbe Wahrscheinlichkeit.

Unabhängigkeit

Ein wichtiges Grundkonzept von zufälligen Ereignissen ist die **Unabhängigkeit**, dass also dass Eintreten eines Ereignisses A nicht das Eintreten eines anderen Ereignisses B beeinflusst. Interessanter ist in der Regel die **Abhängigkeit** von Ereignissen. In diesem Fall ändert sich die Wahrscheinlichkeit für das Eintreten von B durch das Eintreten von A. Bayesianisch betrachtet gewinnen wir durch das Eintreten von A Information über das Eintreten von B.

Stochastische Unabhängigkeit

Zwei Ereignisse heißen stochastisch unabhängig, wenn gilt

$$P(A \cap B) = P(A) \times P(B)$$

Dabei meint $A \cap B$, dass beide Ereignisse eintreten.
Zwei Zufallsvariablen heißen stochastisch unabhängig, wenn gilt

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

für alle möglichen Werte von x und y. Dabei meint "$X = x, Y = y$", dass beide Ereignisse eintreten, also das X den Wert x annimmt und Y den Wert y.

Stochastische Unabhängigkeit

Beispiel: Fairer Würfel
Sei A: “Es fällt eine 5 oder 6”, B: “Es fällt eine gerade Zahl”. Dann ist $A \cap B$: “Es fällt eine 6”. Es gilt

$$P(A) \times P(B) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6} = P(A \cap B)$$

Die beiden Ereignisse sind also voneinander unabhängig!
Wir nennen

$$P(A \cap B)$$

- gemeinsame Wahrscheinlichkeit $P(A)$
- bzw. P(B) Rand- oder marginale Wahrscheinlichkeit.

Bedingte Wahrscheinlichkeit

Ist dagegen A von B abhängig, ändert das Eintreten von B die Wahrscheinlichkeit für das Eintreten von A. Wir können dann die bedingte Wahrscheinlichkeit $P(A|B)$ für das Eintreten eines Ereignisses A unter der Bedingung, das ein Ereignis B eingetreten ist, definieren.

Definition

Wir nennen

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit von A gegeben B. $P(B)$ heißt Rand- oder marginale Wahrscheinlichkeit von B.

Beispiel: Fairer Würfel

Sei A: “Es fällt eine 4, 5 oder 6”, B: “Es fällt eine gerade Zahl”. Dann ist

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(4,6)}{P(2,4,6)} = \frac{2}{3}.$$

Unabhängigkeit

Sind A und B unabhängig, gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

Unter Unabhängigkeit verändert sich die Wahrscheinlichkeit von A nicht, egal ob B eintritt oder nicht.

Aus der Definition der bedingten Wahrscheinlichkeit $P(A|B) = P(A \cap B)P(B)$ ergeben sich folgende Zusammenhänge:

$$P(A \cap B) = P(A|B)P(B), \tag{1}$$

$$P(A \cap B) = P(B|A)P(A) \tag{2}$$

$$\Rightarrow P(B|A)P(A) = P(A|B)P(B) \tag{3}$$

Satz von Bayes

Daraus folgt direkt der für uns zentrale Satz von Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Satz von der totalen Wahrscheinlichkeit

$P(B)$ können wir dabei mit dem Satz von der totalen Wahrscheinlichkeit berechnen:

$$P(B) = P(B \cap A) + P(B \cap \overline{A}) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

Wir erhalten eine weitere Version des Satzes von Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

Beispiel: unfairer Würfel?

Als erstes Beispiel überlegen wir uns folgende Situation. Ein Bekannter von Ihnen besitzt zwei identische aussehende Würfel. Einer davon ist fair, einer würfelt immer die 6. Der Bekannte würfelt mit einem der beiden Würfel, es fällt die sechs. Ist das nun der unfaire Würfel?
Wir definieren uns folgende Ereignisse und Wahrscheinlichkeiten:

- Ereignis A: Der Würfel ist fair.
- Ereignis B: Eine 6 fällt.

Die Wahrscheinlichkeit von B hängt von A ab!

- A: Der Würfel ist fair: $\rightarrow P(B|A) = \frac{1}{6}$
- $\overline{A}$: Der Würfel ist unfair: $\rightarrow P(B|\overline{A}) = 1$

Nach Laplace gehen wir außerdem von $P(A) = P(\overline{A}) = 1/2$ aus; sprich: der Bekannte hat zufällig einen Würfel ausgewählt. (Das ist eine Vorannahme oder Vorwissen. Sie kennen den Bekannten besser, können diese Wahrscheinlichkeit vielleicht besser einschätzen!)

Anwendung des Satz von Bayes

Dann gilt mit dem Satz von Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})} \tag{4}$$

$$= \frac{\frac{1}{6} \times \frac{1}{2}}{\frac{1}{6} \times \frac{1}{2} + 1 \times \frac{1}{2}} = \frac{1}{7} \tag{5}$$

$$P(A|\overline{B}) = \frac{\frac{5}{6} \times \frac{1}{2}}{\frac{5}{6} \times \frac{1}{2} + 0 \times \frac{1}{2}} = 1 \tag{6}$$

Der Satz von Bayes

Der Satz von Bayes ermöglicht es uns, die bedingte Wahrscheinlichkeit "umzudrehen" (bis ins 20. Jahrhundert sprach man auch von inverser Wahrscheinlichkeit). Wir wissen die bedingte Wahrscheinlichkeit eines Ereignisses B gegeben das Ereignis A eingetreten ist. Daraus können wir schließen, wie die Wahrscheinlichkeit des Ereignisses A gegeben das Ereignis B eingetreten ist.
Der Satz von Bayes lautet in der einfachsten Form

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

oder auch:

$$\text{Posteriori} = \frac{\text{Bedingte Wahrscheinlichkeit d. Beobachtung} \times \text{Priori}}{\text{Marginale Wahrscheinlichkeit d. Beobachtung}}$$

Der Satz von Bayes

Kennen wir $P(B)$ nicht, so können wir die Wahrscheinlichkeit wie folgt über die bedingten Wahrscheinlichkeiten berechnen. Zusammengenommen lautet der **Satz von Bayes** dann:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}$$

Schaut bitte in den online Kurs für das Beispiel eines Medizinischen Test.

Interpretation

Nach Beobachtung des positiven Testergebnisses ist also die Wahrscheinlichkeit, dass die Person krank ist etwa 16.2%. Aus unserer Priori-Wahrscheinlichkeit wurde durch die Beobachtung die Posteriori-Wahrscheinlichkeit.

Die Posteriori-Wahrscheinlichkeit $P(A|B)$ ist hier relativ gering, weil schon die Priori-Wahrscheinlichkeit $P(A)$ sehr gering war.

Bayesianisches Lernen

Der Satz von Bayes hilft also, aus B für A zu lernen. Beim Beispiel "medizinischer Test" war

- a priori, also vor dem Test die Wahrscheinlichkeit, dass der Patient krank ist, bei 2%,
- a Posteriori, also nach positivem Testergebnis, lag die Wahrscheinlichkeit, dass der Patient krank ist, bei 16.2%.

Durch die Beobachtung (des Testergebnisses), haben wir Wissen (Information) hinzugewonnen. **Information wird durch eine Wahrscheinlichkeit, allgemeiner: eine Wahrscheinlichkeitsverteilung, ausgedrückt.**
Wir können allerdings nur dann Wissen hinzugewinnen, wenn A und B nicht (stochastisch) unabhängig sind. Sind A und B unabhängig, gilt nämlich

$$P(A \cap B) = P(A)P(B)$$

und nach Definition der bedingten Wahrscheinlichkeit

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Das heißt, das Eintreten (oder Nicht-Eintreten) von B ändert die Wahrscheinlichkeit von A nicht, liefert also keine Information über A.

Dichte

Wir brauchen daher den Begriff der **stetigen Dichte** für stetige Zufallsvariablen, also Zufallsvariablen mit überabzählbar vielen möglichen Ergebnissen.
Sei $\boldsymbol{Y}$ eine stetige Zufallsvariable. Dann nennen wir die Funktion

- $F(y) = P(Y \leq y)$ die **Verteilungsfunktion** von Y (gleich der Wahrscheinlichkeit, dass Y kleiner als y ist)
- $f(y) = \frac{d}{dy}F(y)$ die stetige Dichte von Y (gleich der Ableitung der Verteilungsfunktion)

Es gilt für ein Intervall $[y_1, y_2]$

$$P(Y \in [y_1, y_2]) = \int_{y_1}^{y_2} f(y)dy.$$

Diskrete und stetige Dichten

- Es existiert eine mathematische Theorie, die diskrete und stetige Verteilungen einheitlich definiert.
- Aus dieser Theorie heraus können wir diskrete (Wahrscheinlichkeits-)Dichten und stetige Dichten analog verwenden.
- Das gilt auch für bedingte Wahrscheinlichkeit, für die sich entsprechend bedingte Dichten definieren lassen:
$$f(x|\pi) = \frac{f(x,\pi)}{f(\pi)}$$
- Eigentlich sollte man hier $f(x|\Pi = \pi)$ schreiben, zur Übersichtlichkeit verwenden wir die Kurzschreibweise $f(x|\pi)$ für die bedingte Dichte von X gegeben $\Pi = \pi$.

Satz von Bayes mit Dichten

Der Satz von Bayes lässt sich entsprechend auch mit Dichten schreiben:

$$f(\pi|x) = \frac{f(x|\pi)f(\pi)}{f(x)}$$

- Für rein diskrete Probleme werden die Dichten durch Wahrscheinlichkeiten ersetzt und wir sind wieder bei der ursprünglichen Form.
- Oft ersetzt man die Schreibweise $f(\pi|x)$ durch $p(\pi|x)$, um klar zu machen, dass hier die Posteriori-Dichte berechnet wird.
- Analog wird auch für die Priori-Dichte $p(\pi)$ statt $f(\pi)$ benutzt:

$$p(\pi|x) = \frac{f(x|\pi)p(\pi)}{f(x)}$$

Normalisierungskonstante

Die Berechnung des Terms $f(x)$ (die sogenannte Normalisierungskonstante, den Grund dafür sehen wir später) hängt davon ab, ob Π diskret oder stetig ist:

- Ist Π diskret, dann gilt $f(x) = \sum_\pi f(x|\pi)p(\Pi = \pi)$ (wie zuvor Satz von der totalen Wahrscheinlichkeit)
- Ist Π stetig, dann gilt $f(x) = \int f(x|\pi)p(\pi)d\pi$ (wir "integrieren π raus")

Vorwissen über π

Kommen wir zurück zur Frage: Wie können wir vor dem Experiment Aussagen über die Zufallsvariable Π ("Punkt an der die weiße Kugel zu liegen kommt") machen.

- Wir nehmen an, dass die weiße Kugel keinen Ort bevorzugt (zumindest wissen wir nichts anderes).
- Für stetige Zufallsvariablen heißt das, die Dichte ist überall gleich!
- Statistisch gesehen nehmen wir die stetige Gleichverteilung für Π auf dem Intervall $[0, 1]$ an ($\Pi \sim U[0, 1]$). Die Priori-Dichte von Π ist dann also

$$p(\pi) = 1 \text{ für } 0 \leq \pi \leq 1$$

Für Werte von $\pi < 0$ und $\pi > 1$ ist die Dichte 0 - der Fall kann nicht eintreten!
Nebenbemerkung: Für jedes Intervall $[y_1, y_2]$ mit $0 \leq y_1 < y_2 \leq 1$ gilt dann:

$$P(\Pi \in [y_1, y_2]) = y_2 - y_1$$

Vorwissen oder: Die Priori

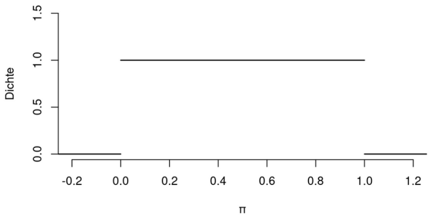


Figure: Priori Dichte

Vorwissen als Priori-Dichte: π liegt irgendwo zwischen 0 und 1

Diese Angabe ist die Vorinformation über Π vor (lateinisch a *priori*) Beobachtung der roten Kugeln, die sogenannte Priori(-Verteilung) von Π. Information liegt uns in Form einer Verteilung vor. Das ist ganz allgemein in der Bayes-Statistik der Fall: **Verteilung entspricht Information**.

Nebenbemerkung: Die Schreibweise $f(\pi)$ für eine Dichte erscheint Ihnen eventuell erstmal ungewohnt. In der Bayesianischen Statistik nehmen wir aber nicht nur Verteilungen für die Daten an, sondern auch für viele Parameter (hier π). Der Umgang mit den Dichten von Parameter unterscheidet sich nicht von dem mit Dichten von Daten, sowohl Daten als auch Parameter sind Bayesianische gesehen Zufallsgrößen!

Die Bayes-Formel

Nun können wir Datendichte und Priori(-dichte) kombinieren. Der Satz von Bayes lautet ja:

$$p(\pi|x) = \frac{f(x|\pi)p(\pi)}{f(x)} = \frac{f(x|\pi)p(\pi)}{\int f(x|\tilde{\pi})p(\tilde{\pi})d\tilde{\pi}}$$

- $p(\pi|x)$ ist die Dichte der Verteilung von Π gegeben $X = x$. Also genau das was wir suchen: Information über Π nach (lateinisch: a posteriori) unserer Beobachtung x. Information stellen wir – wie bei der Priori – als Verteilung dar. Wir nennen diese bedingte Verteilung $\Pi|X = x$ **die Posteriori-Verteilung**.
- Die Verteilung wird durch ihre Dichte festgelegt. Entsprechend bezeichnen wir $p(\pi|x)$ **als Posterior-Dichte**
- In der Praxis unterscheidet man sprachlich kaum zwischen Posteriori-Verteilung und Posteriori-Dichte (auch wenn es unterschiedliche Konzepte sind), sondern spricht nur von der Posteriori. Damit lässt sich die Bayes-Formel auch so darstellen:

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

- **Wichtig**: Im Beispiel ist $f(\pi)$ eine stetige Dichte; $f(x|\pi)$ eine diskrete oder Wahrscheinlichkeitsdichte. Für den Satz von Bayes ist es aber egal, ob wir diskrete oder stetige Dichten verwenden.
- Beachten Sie, dass wir in der Formel Π und X unterdrücken – vergessen Sie aber nicht, dass π und x Realisationen von Zufallszahlen sind. Formal richtig müssten wir schreiben:
- $f_{\Pi}(\pi)$ für die Dichte von Π
- $f_{X|\Pi=\pi}(x|\pi)$ für die Dichte von X gegeben $\Pi = \pi$

Normalisierungskonstante

- Die Bayes-Formel kombiniert also Datenmodell (Likelihood) und Priori zur Posteriori.
- Zusätzlich brauchen wir $f(x) = \int f(x|\tilde{\pi})f(\tilde{\pi})d\tilde{\pi}$. Die Funktion $f(x)$ hängt nicht von π ab, ist also bezüglich der Dichte von π eine Konstante. Wir nennen $f(x)$ Normalisierungskonstante, weil sie sicherstellt, dass $\int p(x|\pi)d\pi = 1$ gilt (für diskretes Π : das $\sum_\pi p(x|\pi) = 1$.
- Wir können im Satz von Bayes auch das Zeichen "$\propto$" (sprich: propotional zu) benutzen. Das Zeichen bedeutet, dass sich linke und rechte Seite bis auf eine Konstante entsprechen:

$$p(\pi|x) = \frac{f(x|\pi)p(\pi)}{f(x)} \propto f(x|\pi)p(\pi)$$

oder informell:

$$\text{Posteriori} \propto \text{Likelihood} \times \text{Priori}$$

- Wichtig dabei: "Konstant" in der Normalisierungskonstante bezieht sich auf π. Die Normalisierungskonstante ist eine Funktion in x, also der Beobachtung. Alle Aussagen gehen aber immer davon aus, dass wir die Beobachtung gemacht haben, x also bekannt und damit fest ist.

Konjugiertheit

Den Kern der Posterioridichte $\pi^x(1-\pi)^{n-x}$ ist wie gesagt eine Funktion in π bei gegebenen x. Würde man dagegen π festhalten und ließe x variieren, sieht der Kern nach einer Binomialverteilung aus, siehe Datenmodell.

- Das ist natürlich kein Zufall: die Posteriori entsteht ja genau aus Datendichte und Priori.
- In unserem Beispiel passen Datendichte (Binomialverteilung) und Prioridichte (Betaverteilung) genau zusammen.
- Die Posterioridichte hat daher die Form der Prioridichte, nur mit anderen Parametern. Bei beiden handelt es sich um Betaverteilungen.
- Wir nennen dies Konjugiertheit bzw. sprechen von der konjugierten Priori.

Nicht immer werden wir eine konjugierte Priori verwenden können oder wollen. Zur Wahl der Priori später mehr.

Konjugiertheit

Table: Conjugate Priors and Posterior Distributions for Common Exponential Family Distributions

Likelihood	Parameter	Prior	Posterior (after observing data)
Discrete Distributions			
Bernoulli(θ)	$\theta \in (0,1)$	Beta(α, β)	Beta($\alpha + x$, $\beta + n - x$)
Binomial(n, θ)	$\theta \in (0,1)$	Beta(α, β)	Beta($\alpha + x$, $\beta + n - x$)
Geometric(θ)	$\theta \in (0,1)$	Beta(α, β)	Beta($\alpha + r$, $\beta + \sum x_i$)
Poisson(λ)	$\lambda > 0$	Gamma(α, β)	Gamma($\alpha + \sum x_i$, $\beta + n$)
Neg. Binomial(r, θ)	$\theta \in (0,1)$	Beta(α, β)	Beta($\alpha + r$, $\beta + \sum x_i$)
Continuous Distributions			
Normal(μ, known σ^2)	$\mu \in \mathbb{R}$	Normal(μ_0, τ^2)	Normal$\left(\frac{\frac{\mu_0}{\tau^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\right)$
Normal(known μ, σ^2)	$\sigma^2 > 0$	Inv-Gamma(α, β)	Inv-Gamma($\alpha + n/2$, $\beta + \frac{1}{2}\sum(x_i - \mu)^2$)
Normal(μ, σ^2)	(μ, σ^2)	Normal-Inverse-Gamma	Normal-Inverse-Gamma with updated: $\kappa' = \kappa + n$ $\mu_0' = \frac{\kappa\mu_0 + n\bar{x}}{\kappa + n}$ $\alpha' = \alpha + n/2$ $\beta' = \beta + \frac{1}{2}\sum(x_i - \bar{x})^2 + \frac{\kappa n}{2(\kappa + n)}(\bar{x} - \mu_0)^2$
Exponential(λ)	$\lambda > 0$	Gamma(α, β)	Gamma($\alpha + n$, $\beta + \sum x_i$)
Gamma(β), known shape α	$\beta > 0$	Gamma(a, b)	Gamma($a + n\alpha$, $b + \sum x_i$)