

3.5 Der Satz von Bayes

In der sogenannten Bayes-Statistik braucht man z.T. einen subjektiven Wahrscheinlichkeitsbegriff, zumindest in der radikalen Variante des Bayesianismus. Wir wollen in diesem Kapitel nur die Grundlagen für die Bayes-Statistik betrachten.

Die bedingte Wahrscheinlichkeit $\Pr(A \mid B)$ haben wir schon kennengelernt. Diese Wahrscheinlichkeit ist die *bedingte* Wahrscheinlichkeit des Ereignisses A gegeben Ereignis B .

Betrachten wir nun ein bekanntes Gebiet mit bedingten Wahrscheinlichkeiten, nämlich das Gebiet der *Diagnostik*.

In der Diagnostik reden wir z.B. von der Wahrscheinlichkeit einer *Hypothese* H (z.B. Gegenwart von Krankheit) gegeben *Daten* E oder gegeben *Test* T . Dazu schauen wir uns das *Problem des Umkehrschlusses* an. Unter diesem Problem versteht man das irrtümliche Gleichsetzen der beiden bedingten Wahrscheinlichkeiten $\Pr(H \mid E)$ und $\Pr(E \mid H)$. Diese Wahrscheinlichkeiten sind aber i.A. nicht identisch, können aber durch das *Theorem von Bayes* verbunden werden. Dieses folgt direkt aus (3.3.2) und (3.3.3) und lautet

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \Pr(H)}{\Pr(E)}. \quad (3.5.1)$$

Das Theorem ist benannt nach dem englischen Mathematiker und Pfarrer THOMAS BAYES, der diesen Satz erstmals 1763 in einer posthum veröffentlichten Arbeit beschrieb [3]. In der wichtigen Gleichung (3.5.1) stellt $\Pr(H \mid E)$ die *a posteriori*-Wahrscheinlichkeit von H dar, $\Pr(H)$ stellt die *a priori*-Wahrscheinlichkeit (z.T. subjektive Überzeugung) von H dar (die Wahrscheinlichkeit der Hypothese, *bevor* man im Besitz von Daten ist). $\Pr(H)$ ist im besten Fall eine gute Schätzung der *Prävalenz* einer Krankheit, wenn z.B. H für “krank” steht. H^C bedeutet das Komplement von H , “gesund”.

Aufgrund der totalen Wahrscheinlichkeit (3.3.6) können wir (3.5.1) schreiben als

$$\Pr(H \mid E) = \frac{\Pr(E \mid H) \Pr(H)}{\Pr(E \mid H) \Pr(H) + \Pr(E \mid H^C) \Pr(H^C)}. \quad (3.5.2)$$

Definition. Die *Likelihood* einer Hypothese, $L(H)$, ist eine Funktion der Hypothese und stellt die Wahrscheinlichkeit der beobachteten Daten unter der Hypothese dar,

$$L(H) = \Pr(E \mid H). \quad (3.5.3)$$

Die Likelihood ist also ein Funktion des zweiten Arguments in $\Pr(E \mid H)$. Die Daten wurden beobachtet und die Likelihood der Hypothese ist gleich der Wahrscheinlichkeit der beobachteten Daten unter dieser Hypothese.

Bei zwei komplementären Hypothesen H und H^C ist

- die Likelihood von H : $L(H) = \Pr(E | H)$
- die Likelihood von H^C : $L(H^C) = \Pr(E | H^C)$

Teilen wir die a posteriori-Wahrscheinlichkeit von H in (3.5.1) durch diejenige von H^C . Dann kann man (3.5.1) auch in der *Odds-Form* schreiben.

$$\underbrace{\frac{\Pr(H | E)}{\Pr(H^C | E)}}_{\text{Posterior Odds}} = \underbrace{\frac{\Pr(E | H)}{\Pr(E | H^C)}}_{\text{Likelihood Ratio}} \times \underbrace{\frac{\Pr(H)}{\Pr(H^C)}}_{\text{Prior Odds}}. \quad (3.5.4)$$

Wir können das kurz schreiben, mit LR als der *Likelihood Ratio*,

$$Odds(H | E) = LR \times Odds(H). \quad (3.5.5)$$

Die a posteriori-Chance ist also das Produkt aus den a priori-Odds und dem durch die Daten bestimmten LR . Oder: Die LR transformiert die a priori-Chance von H zur a posteriori-Chance von H .

Wenn wir noch auf beiden Seiten den Logarithmus nehmen, dann können wir (3.5.5) additiv schreiben,

$$\boxed{\log Odds(H | E) = \log LR + \log Odds(H)}. \quad (3.5.6)$$

Die **philosophische Konsequenz** ist:

Erkenntnis in der Wissenschaft: A Posteriori = A Priori + empirische Daten

Betrachten wir als Spezialfall für die empirischen Daten E einen Test T ($T = 1$ für positives und $T = 0$ für ein negatives Resultat) sowie einer Hypothese der Krankheit ($H = 1$ für Krankheit versus $H = 0$ für Gesundheit). Dann sind folgende aus der Diagnostik bekannten Größen wichtig:

- $\Pr(T = 1 | H = 1)$: Sn des Tests (Richtig-Positiv-Rate, TPR)
- $\Pr(T = 0 | H = 0)$: Sp des Tests (Richtig-Negativ-Rate, TNR)
- $\Pr(T = 0 | H = 1)$: $1-Sn$ (Falsch-Negativ-Rate, FNR)
- $\Pr(T = 1 | H = 0)$: $1-Sp$ (Falsch-Positiv-Rate, FPR)
- $\Pr(H = 1 | T = 1)$: PPV des Tests
- $\Pr(H = 0 | T = 0)$: NPV des Tests
- $LR+ = \frac{L(H=1)}{L(H=0)} = \frac{\Pr(T=1|H=1)}{\Pr(T=1|H=0)} = \frac{TPR}{FPR} = \frac{Sn}{1-Sp}$: LR bei positivem Test
- $LR- = \frac{L(H=1)}{L(H=0)} = \frac{\Pr(T=0|H=1)}{\Pr(T=0|H=0)} = \frac{FNR}{TNR} = \frac{1-Sn}{Sp}$: LR bei negativem Test

Abkürzungen: Sn : Sensitivität, Sp : Spezifität, PPV : Positiv prädiktiver Wert, NPV : Negativ prädiktiver Wert, LR : Likelihood ratio.

Das Problem des Umkehrschlusses ist vor allem bei *seltenen Krankheiten* relevant. Ist die Prävalenz $\Pr(H = 1)$ einer Krankheit klein, dann kann $\Pr(H = 1 \mid T = 1)$ z.T. (viel) kleiner werden als $\Pr(T = 1 \mid H = 1)$.

Beispiel 1. Die Tabelle 3.2 zeigt die Resultate einer Studie, die die Sensitivität Sn und die Spezifität Sp , also die Gütekriterien bestimmen wollte für einen neuen Test für die Diagnose einer Krankheit. Mit den Daten dieser Studie können die Gütekriterien, prädiktiven Werte und Likelihood Ratios geschätzt werden: $Sn = 0.9$, $Sp = 0.9$.

		Wahrheit		
		$H = 1$	$H = 0$	Total
Test	$T = 1$	90	100	190
	$T = 0$	10	900	910
	Total	100	1000	1100

Tabelle 3.1: Diagnostik bei einer seltenen Krankheit, $H = 1$: Krank, $H = 0$: Gesund, $T = 1$: Test positiv, $T = 0$: Test negativ. Die Berechnung der Gütekriterien ergibt $Sn = 0.9$ und $Sp = 0.9$.

Brauchen wir als (a priori) Prävalenz die beobachtete Prävalenz ($100/1100 = 0.091$), dann ist der $PPV = 0.474$ und der $NPV = 0.989$. Die Sensitivität Sn ist also viel grösser als der prädiktive Wert! Für den Patienten ist in diesem Fall ein positives Testresultat weniger dramatisch als man nach Betrachtung der hohen Sn und Sp meinen könnte.

Dieser Test eignet sich besser für den Ausschluss (weil NPV gross) und weniger für den Einschluss von Krankheit (PPV nicht gross). Für die LR 's ergibt sich $LR+ = 0.9/0.1 = 9$ und $LR- = 0.1/0.9 = 1/9$, die Wahrscheinlichkeit eines positiven Tests ist bei einem Kranken 9-mal grösser ist als bei einem Gesunden, und die Wahrscheinlichkeit eines negativen Tests ist bei einem Kranken 9-mal kleiner ist als bei einem Gesunden. (Allgemein ist natürlich $LR+$ nicht der Kehrwert von $LR-$).

Abbildung 3.10 zeigt für die Sn und Sp in diesem Beispiel, wie sich die Prävalenz auf den PPV und NPV auswirkt. Gestrichelt ist die beobachtete Prävalenz aus den erhobenen Daten.

```
p0obs <- 100/1100 #observed prevalence
p0 <- seq(0, by = 0.01, 1) #prevalences
Sn <- 0.9
Sp <- 0.9 #Sn and Sp
PPV <- (p0 * Sn)/(p0 * Sn + (1 - p0) * (1 - Sp)) #pos.pred. value
NPV <- (1 - p0) * Sp/((1 - p0) * Sp + p0 * (1 - Sn)) #neg. pred. value
plot(p0, PPV, xlab = "prevalence", ylab = "PPV", type = "l", ylim = c(0, 1))
abline(v = p0obs, h = (p0obs * Sn)/(p0obs * Sn + (1 - p0obs) * (1 - Sp)), lty = 2)
plot(p0, NPV, xlab = "prevalence", ylab = "NPV", type = "l", ylim = c(0, 1))
abline(v = p0obs, h = (1 - p0obs) * Sp/((1 - p0obs) * Sp + p0obs * (1 - Sn)), lty = 2)
```

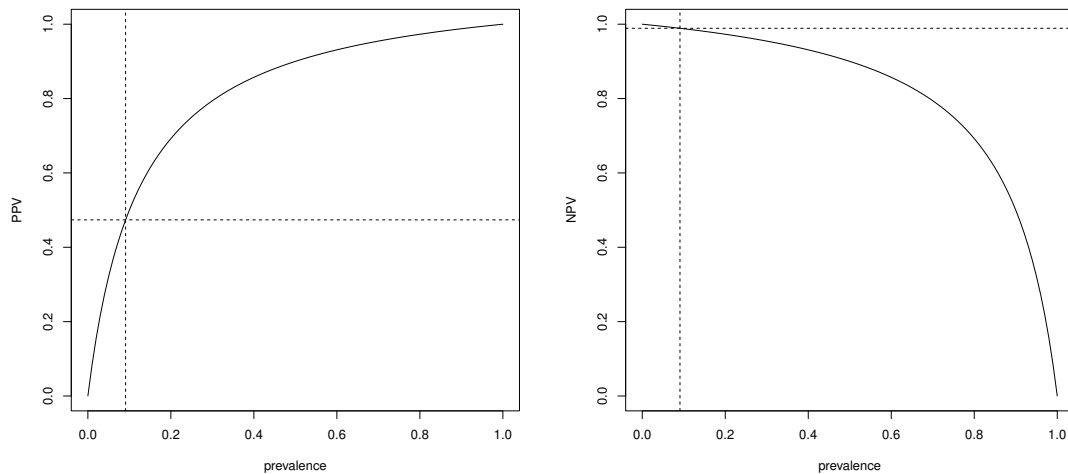


Abbildung 3.10: PPV und NPV als Funktion der Prävalenz, bei Test mit $Sn=0.9$, $Sp=0.9$. Gestrichelt: Beobachtete Prävalenz.

Bei niederprävalenten Merkmalen muss die Spezifität praktisch maximal sein (wenig Falsch-Positive), damit der positiv prädiktive Wert nicht absinkt. Denn schon wenige Falsch-Positive können den positiv prädiktiven Wert stark vermindern.

Will man also mit einem Test eine Krankheit (oder ein Merkmal wie “Blockade des Iliosakralgelenks”) *einschliessen*, muss man spezifisch sein, was gleichbedeutend ist mit wenig Falsch-Positiven und einem grossen positiven prädiktiven Wert. Ist dieses Merkmal zudem noch *selten*, muss die Spezifität fast maximal sein. Will man hingegen Krankheit *ausschliessen*, muss man sensitiv sein (wenig Falsch-Negative), ist zudem die Krankheit *häufig*, muss die Sensitivität fast maximal sein.

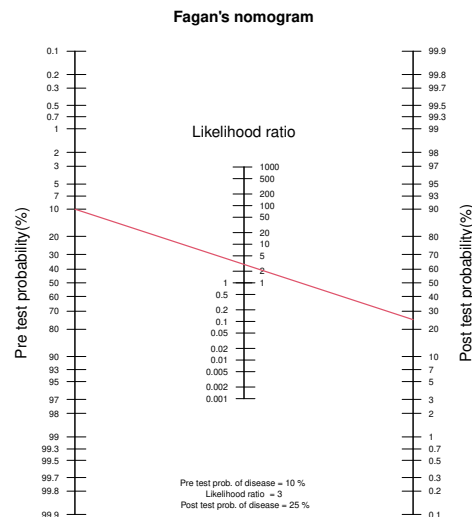
Beispiel 2. Im folgenden Beispiel veranschaulichen wir Bayes mit einem *Rechenschieber*, wie man ihn auch früher physisch zum Rechnen brauchte. Dieser basiert auf der additiven Bayes-Formel 3.5.6.

Wir nehmen eine Vortestwahrscheinlichkeit von Krankheit $\Pr(H = 1) = 0.1$ an, dies entspricht also einer prior Odds von 0.111. Zudem haben wir einen diagnostischen Test mit $Sn = \Pr(T = 1 | H = 1) = 0.3$ und $Sp = \Pr(T = 0 | H = 0) = 0.9$.

Die Likelihood ratio ($H = 1$ versus $H = 0$) **nach Beobachtung eines positiven Tests** (definiert als das Verhältnis der Wahrscheinlichkeit von einem positiven Test unter $H = 1$ zur Wahrscheinlichkeit eines positiven Tests unter $H = 0$) ist $LR+ = Sn/(1 - Sp) = 3$. Die resultierende Nachtestwahrscheinlichkeit ist $\Pr(H = 1 | T = 1) = 0.25$. Der Leser versuche, dies anhand (3.5.4) nachzurechnen.

3.5 Der Satz von Bayes

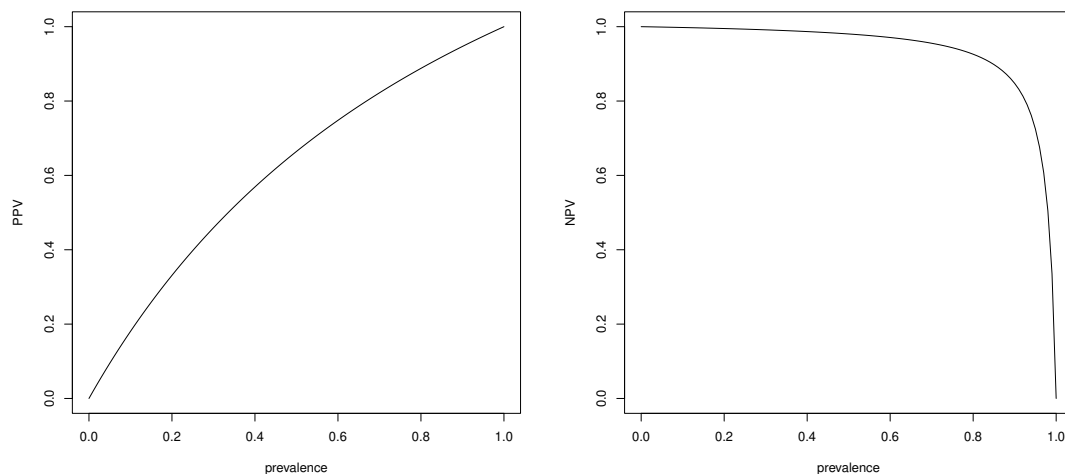
```
## install.packages('TeachingDemos') ## auskommentieren für Installation der package, dann Zeile
## löschen!
library(TeachingDemos) ## package laden
p0 <- 0.1
sn <- 0.3
sp <- 0.9
fagan.plot(probs.pre.test = p0, LR = sn/(1 - sp))
```



Mit `plotFagan2()` aus derselben package können die Vortestwahrscheinlichkeit, Sensitivität und Spezifität auch interaktiv eingegeben werden. Probiere es aus!

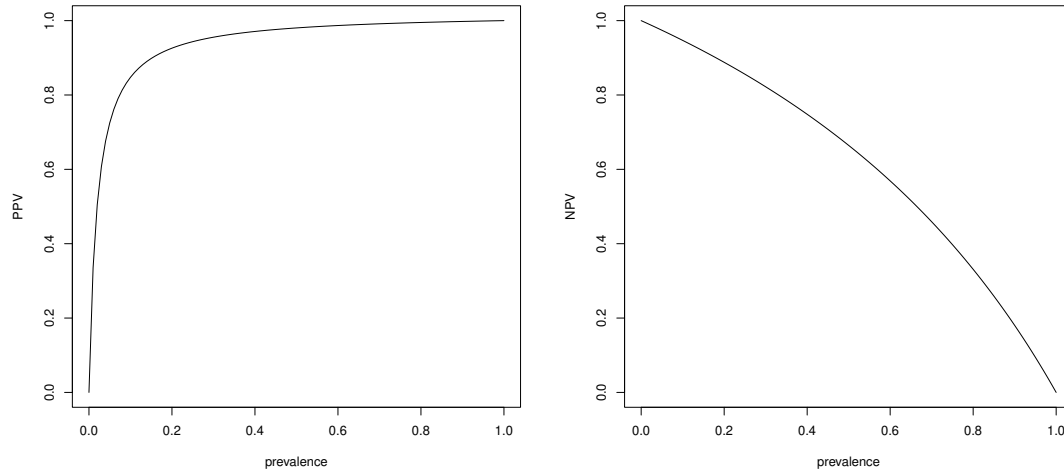
Visualisierung der Abhängigkeit von der Prävalenz bei unterschiedlichen Sn und Sp .

Beispiel 1: Prädiktive Werte bei hochsensitivem Test mit $Sn = 0.99$ und $Sp = 0.5$:



Test ist geeignet für das *Ausschliessen* von Krankheit, dazu brauchen wir eher einen sensitiven Test: “*SnNout*”, bei hoher *Sn*, wenn neg., dann ausschliessen

Beispiel 2: Prädiktive Werte bei hochspezifischem Test mit $Sn = 0.5$ und $Sp = 0.99$:



Test ist geeignet für das *Einschliessen* von Krankheit, dazu brauchen wir eher einen spezifischen Test: “*SpPin*”, bei hoher *Sp*, wenn pos., dann einschliessen.

Spezialfall: Gleichheit von a priori und a posteriori-Wahrscheinlichkeit. Wann sind diese beiden Wahrscheinlichkeiten gleich?

Aus (3.5.1) folgt: Damit $\Pr(H | T) = \Pr(H)$ gilt, muss $\Pr(T | H) = \Pr(T)$ gelten, d.h. die Wahrscheinlichkeit von T ist *unabhängig* von H . Das ist gleichbedeutend mit $LR = 1$ oder $\log LR = 0$:

- Positiver Test: $LR+ = \frac{\Pr(T=1|H=1)}{\Pr(T=1|H=0)} = \frac{TPR}{FPR} = 1 \Leftrightarrow TPR = FPR$.
- Negativer Test: $LR- = \frac{\Pr(T=0|H=1)}{\Pr(T=0|H=0)} = \frac{FNR}{TNR} = 1 \Leftrightarrow FNR = TNR$.

Spezialfall: Gleichheit von bedingten Wahrscheinlichkeiten. Wir haben gesehen, dass im Allgemeinen

$$\Pr(H | T) \neq \Pr(T | H).$$

Aus (3.5.1) folgt: Damit wir T und H in den bedingten Wahrscheinlichkeiten vertauschen dürfen, muss

$$\Pr(T) = \Pr(H)$$

gelten, wir haben dann *Symmetrie* bezüglich T und H . Das ist bei der Schätzung dann der Fall, wenn die Häufigkeiten symmetrisch sind, d.h. wenn wir gleich viele Falsch-Positive und Falsch-Negative Ereignisse haben. Das wäre der Fall bei der folgenden symmetrischen Vierfeldertafel (symmetrisch bezüglich der Hauptdiagonalen):

		Wahrheit		
		$H = 1$	$H = 0$	Total
Empirie	$T = 1$	a	b	a+b
	$T = 0$	b	d	b+d
	Total	a+b	b+d	a+2b+d

Tabelle 3.2: Symmetrie von T und H : $\Pr(H | T) = \Pr(T | H)$.

3.6 Bayesianische Statistik*.

Der Gebrauch vom Bayes-Theorem in der Diagnostik ist etabliert als *formalisierte klinische Argumentation*. Kontroverser ist die *Bayesianische Statistik*.

Wir werden später sehen, dass statistisches Schätzen und Testen eine Analogie hat zur Diagnostik.

Bei diagnostischen Tests ist die Hypothese wie oben z.B. “*Patient ist krank*”. In der schliessenden Statistik werden Hypothesen allgemeinere Aussagen sein, wie etwa Aussagen über einen unbekannten Parameter θ in einer Population, so wie “ $\theta \leq 177$ cm” für die unbekannte mittlere Körpergrösse in einer Population. Wir werden uns beim Schätzen (Kapitel 7) und Testen von Hypothesen (Kapitel 8) aber zunächst auf die klassische Statistik beschränken, auf die Quantifizierung der Wahrscheinlichkeit von empirischen Daten, gegeben eine Hypothese, also eigentlich auf obige Likelihood (3.5.3).

Die umgekehrte Wahrscheinlichkeit, die a posteriori-Wahrscheinlichkeit, ist nur über das Bayes-Theorem zu berechnen. Und damit wir die Wahrscheinlichkeit einer Hypothese berechnen können, brauchen wir grundsätzlich ein a priori. Die sogenannte Bayesianische Statistik braucht den Satz von Bayes in der statistischen Analyse, wo ein Parameter θ eine unbekannte Quantität ist (z.B. wie das obige θ). Die a priori Verteilung $p(\theta)$ muss dann spezifiziert werden. Dieser Schritt kann angesehen werden als eine natürliche Erweiterung auf die subjektive Interpretation von Wahrscheinlichkeit, siehe auch [27].

Nach diesem Einblick in das Bayes-Theorem wollen wir daher mit folgendem für jeden Praktiker heilsamen Satz schliessen:

Bayes: Ohne Subjektivität gibt es – strenggenommen – keine Erkenntnis.