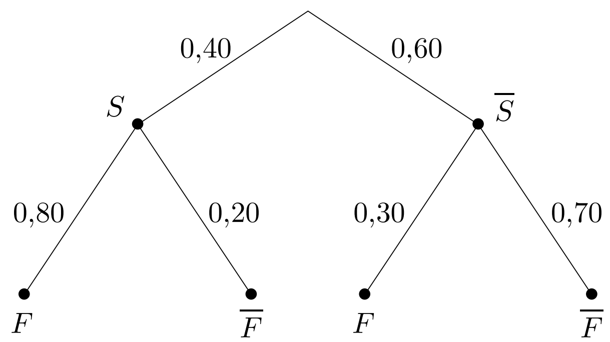
Bayes-Statistik

In einem Gerichtsprozess mutmaßt ein Geschworener, dass der Angeklagte zu 40 % schuldig ist.
Nun werden jedoch Fingerabdrücke des Angeklagten am Tatort gefunden. Der Chefermittler behauptet, die Wahrscheinlichkeit in vergleichbaren Fällen Fingerabdrücke am Tatort zu finden, liegt bei 0,80, wenn der Angeklagte schuldig ist, bei 0,30, wenn er unschuldig ist.
Wie bewertet der Geschworene die Schuld nun?



S Der Angeklagte ist schuldig.
F Ein Fingerabdruck wurde gefunden.

$$P(S \mid F) = \frac{P(S) \cdot P(F \mid S)}{P(S) \cdot P(F \mid S) + P(\overline{S}) \cdot P(F \mid \overline{S})} = 0{,}64 \qquad P(\overline{S} \mid F) = 0{,}36$$

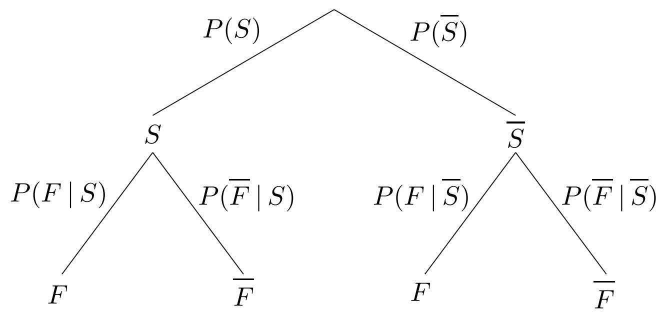
Mit Wahrscheinlichkeiten haben wir bisher relative Häufigkeiten modelliert, d. h. es besteht die Möglichkeit, die Wahrscheinlichkeitenshypothesen mit Versuchsreihen zu testen.

Nun gibt es Ereignisse, bei denen Versuchsreihen kaum oder gar nicht vorstellbar sind, denen man aber auch eine Wahrscheinlichkeit zuschreiben möchte, um deren Veränderung bei weiterer Information zu untersuchen. Wir erlauben daher, Wahrscheinlichkeit als Vermutungsgrad (subjektive Wahrscheinlichkeit) zu interpretieren. An den Rechenregeln ändert sich nichts.
Der Satz von Bayes bestimmt das Vorgehen.
Um in den aufwändigen Rechnungen bei komplizierten Verteilungen den Überblick zu behalten, wurden die folgenden Bezeichnungen eingeführt. Ich beziehe mich auf das Beispiel.

$P(S)$ und $P(\overline{S})$ nennen wir a-priori-Wahrscheinlichkeiten (lat. a priori vom Früheren her),
$P(S \mid F)$ und $P(\overline{S} \mid F)$ sind die a-posteriori-Wahrscheinlichkeiten (lat. a posteriori vom Späteren her).
Die bedingten Wahrscheinlichkeiten der 2. Stufe des Diagramms sind die Likelihoods (Mutmaßlichkeit) der Hypothesen S und $\overline{S}$. Die Likelihoods einer Hypothese geben Antwort auf die Frage: Wie wahrscheinlich ist das, was ich beobachtet habe, wenn ich das Zutreffen einer bestimmten Hypothese voraussetze?

Der Übergang von der a-priori-Verteilung (kurz Prior) zur a-posteriori-Verteilung (kurz Posterior) stellt dar, was wir aus der neuen Information (den beobachteten Daten) gelernt haben.
Die a-posteriori-Wahrscheinlichkeitsverteilung eignet sich als neuer Prior, wenn neue Informationen zur Verfügung stehen. a-priori-Verteilungen sind subjektiv. Je mehr Daten allerdings erhoben werden, desto mehr nähern sich die a-posteriori-Verteilungen einander auch dann an, wenn die ursprünglichen a-priori-Verteilungen verschieden waren. Bei fortlaufender Datensammlung verflüchtigt sich gleichsam die Subjektivität zugunsten zunehmender Objektivität.

Posterior $\propto$ Prior $\times$ Likelihood



$$P(S \mid F) \;=\; \frac{P(S) \cdot P(F \mid S)}{P(S) \cdot P(F \mid S) + P(\overline{S}) \cdot P(F \mid \overline{S})} \quad \propto P(S) \cdot P(F \mid S) \qquad\qquad \text{proportional } \propto$$

$$P(\overline{S} \mid F) \;=\; \frac{P(\overline{S}) \cdot P(F \mid \overline{S})}{P(S) \cdot P(F \mid S) + P(\overline{S}) \cdot P(F \mid \overline{S})} \quad \propto P(\overline{S}) \cdot P(F \mid \overline{S})$$

Die Wahrscheinlichkeiten der a-posteriori-Verteilung ergeben sich - bis auf einen Normierungsfaktor - als Produkt der Wahrscheinlichkeiten der a-priori-Verteilung und ihrer Likelihoods.
Wenn wir erkennen, dass die Produkte wieder eine bekannte Verteilung ergeben, kann der Normierungsfaktor unbeachtet bleiben, da die Summe aller Wahrscheinlichkeiten gleich 1 ist.

Sind die Stichprobendaten normalverteilt, ist es vorteilhaft, wenn auch die a-priori-Verteilung normal ist, denn dann ist es auch die a-posteriori-Verteilung. Sind die Daten binomialverteilt, ist es zweckmäßig, wenn die a-priori-Verteilung eine Betaverteilung ist, die a-posteriori-Verteilung ist es dann auch. In solchen Fällen spricht man von konjugierten Verteilungen.

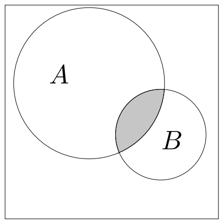
Wiederholung der bedingten Wahrscheinlichkeit:

Bedingte Wahrscheinlichkeit

Nach Laplace 1812, jedoch schon um 1700 von Jakob Bernoulli entdeckt, gilt:

$$P(A) = \frac{\textit{Anzahl der Möglichkeiten, die } A \textit{ festlegen}}{\textit{Anzahl aller Möglichkeiten}}$$

Bei geometrischen Wahrscheinlichkeiten tritt an Stelle von *Anzahl der Möglichkeiten* ein *Flächeninhalt*.



$A \cap B$

Die Wahrscheinlichkeit für das Ereignis A ändert sich (möglicherweise), falls wir voraussetzen können, dass das Ereignis B eingetreten ist:

$$P(A\,|\,B) = \frac{\textit{Anzahl der Möglichkeiten, die zu } A \textit{ und } B \textit{ gehören}}{\textit{Anzahl der Möglichkeiten, die } B \textit{ festlegen}}$$

Umgeformt:

$$P(A\,|\,B) = \frac{\dfrac{\textit{Anzahl der Möglichkeiten, die zu } A \textit{ und } B \textit{ gehören}}{\textit{Anzahl aller Möglichkeiten}}}{\dfrac{\textit{Anzahl der Möglichkeiten, die } B \textit{ festlegen}}{\textit{Anzahl aller Möglichkeiten}}} = \frac{P(A \cap B)}{P(B)}$$

Die Ereignisse A und B sind voneinander unabhängig, falls gilt:

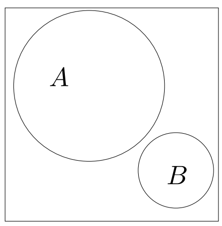
$$P(A\,|\,B) = P(A)$$

Dann folgt:

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

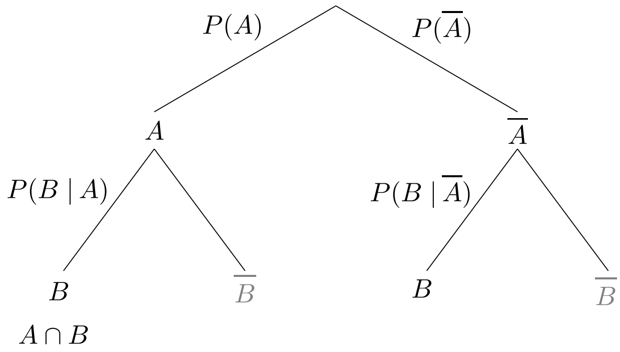
Nur im Falle disjunkter Ereignisse gilt:

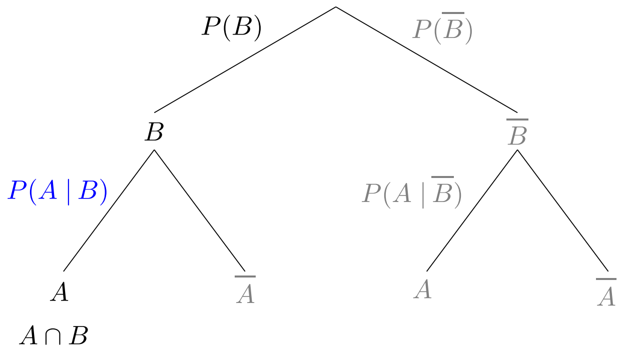
$$P(A \cup B) = P(A) + P(B)$$



$A \cap B = \emptyset$

Bedingte Wahrscheinlichkeit $P(A \mid B)$





Der zweistufige Prozess wird auch durch das umgekehrte Baumdiagramm dargestellt.

Es gilt:

$$P(A \cap B) = P(B) \cdot P(A \mid B) \qquad \text{2. Diagramm}$$

$$P(A \cap B) = P(A) \cdot P(B \mid A) \qquad \text{1. Diagramm}$$

$$P(B) = P(A) \cdot P(B \mid A) + P(\overline{A}) \cdot P(B \mid \overline{A}) \qquad \text{1. Diagramm}$$

Daraus folgt:

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

Die in dieser Formel benötigten Wahrscheinlichkeiten sind alle im 1. Diagramm enthalten.

Bedingte Wahrscheinlichkeiten sind stets die Wahrscheinlichkeiten der 2. Stufe des Baumdiagramms, bzw. dessen Umkehrung.

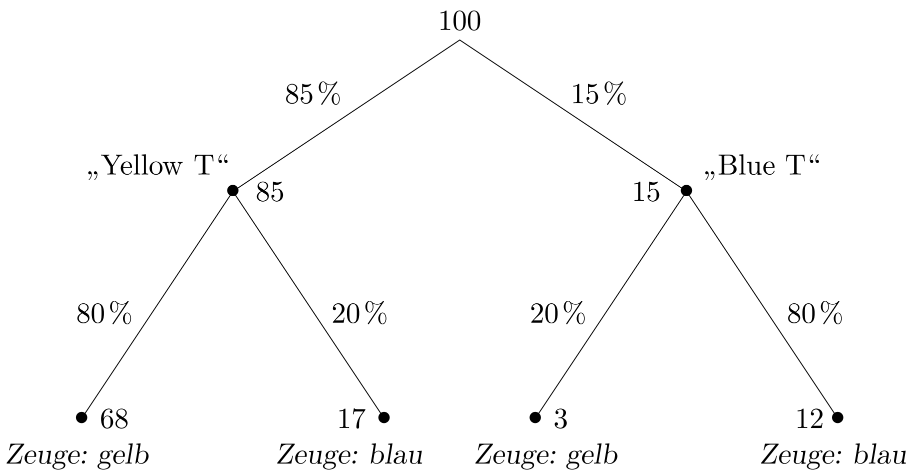
Gelbe und blaue Taxen

In einer Stadt existieren zwei konkurrierende Taxi-Unternehmen. Die Taxen des ersten Unternehmens „Yellow T“ sind gelb, die des zweiten Unternehmens „Blue T“ dagegen blau.

Ein Taxi verursacht einen Verkehrsunfall, der Fahrer begeht Fahrerflucht.
Ein Zeuge - der Einzige - gibt an, in der Dämmerung ein blaues Taxi erkannt zu haben.
Eine bewusste Falschaussage kann ausgeschlossen werden.
Wie groß ist die Wahrscheinlichkeit, dass der Fahrer dem Unternehmen „Blue T“ angehört?

Um diese zu ermitteln, führen die Kriminalbeamten mit dem Zeugen einen Test durch.
Es werden ihm Farbbeispiele vorgeführt, an die er sich später erinnern soll.
Dem Zeugen gelingt es, 80 % der Farbbeispiele richtig zuzuordnen.
Lässt sich daraus nun der Schluss ziehen, dass auch die Beschreibung des Unfallwagens mit 80 %iger Sicherheit korrekt ist?

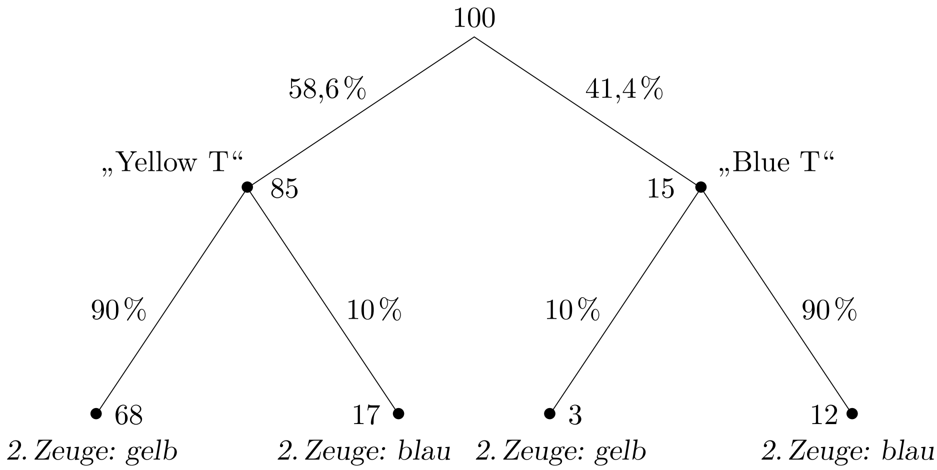
Der Marktanteil von „Yellow T“ liegt bei 85 %, den restlichen Marktanteil hält „Blue T“.
Ist diese Information für die Fragestellung relevant?



An den absoluten Zahlen ist zu erkennen, welche Bedeutung die Zeugenaussage hat.
17mal wäre die Zeugenaussage *blau* falsch, 12mal zutreffend.
Die Wahrscheinlichkeit dafür, dass das Unfalltaxi tatsächlich blau gewesen ist, beträgt
P(„Blue T“ | *Zeuge: blau*) = 41,4 %. Ohne die Zeugenaussage wären es 15 %.
Gelänge dem Zeugen eine 90 %ige korrekte Farbzuordnung, wäre die Wahrscheinlichkeit 61,4 %.

Den Berechnungen liegt die Annahme zugrunde, dass das Fahrverhalten aller Fahrer (*gelb*, *blau*) ähnlich ist, der Marktanteil also Auskunft über die Wahrscheinlichkeit gibt, an einem Unfall beteiligt zu sein.

Wie groß ist die Wahrscheinlichkeit, dass das Unfalltaxi tatsächlich blau gewesen ist,
wenn dies auch ein weiterer Zeuge mit 90 %iger korrekter Farbzuordnung aussagt?

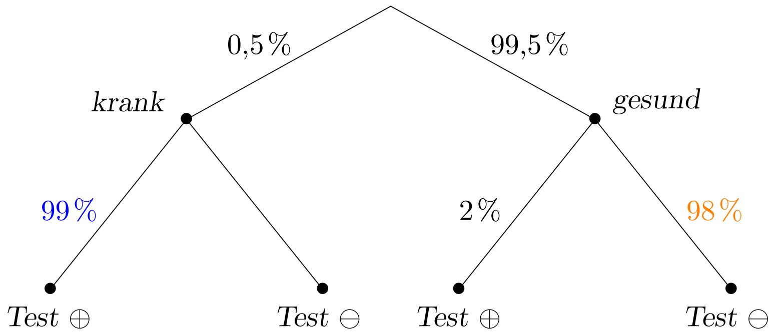


$P($„Blue T“$\,|\,$*auch 2. Zeuge: blau*$\,) = 86{,}4\,\%$.

Das Ergebnis hängt natürlich nicht von der Reihenfolge der Zeugenaussagen ab.

Weisen mehrere medizinische Tests auf eine Krankheit hin, wird man entsprechend vorgehen, um die Situation zu bewerten.

Eine Krankheit komme bei etwa 0,5 % der Bevölkerung vor. Ein Test zur Auffindung der Krankheit führt bei 99 % der Kranken zu einer Reaktion (Sensitivität des Tests), bei 98 % der Gesunden zu keiner Reaktion (Spezifität). Wie groß ist die Wahrscheinlichkeit, dass eine Person, bei der die Reaktion eintritt, die Krankheit tatsächlich hat?



Mit dem Pfaddiagramm ergibt sich: $P(krank \mid \oplus) = \dfrac{0{,}5\,\% \cdot 99\,\%}{0{,}5\,\% \cdot 99\,\% \;+\; 99{,}5\,\% \cdot 2\,\%} = 19{,}9\,\%$

Der Test wird wiederholt.
Wie groß ist die Wahrscheinlichkeit, dass die Person, bei der die Reaktion

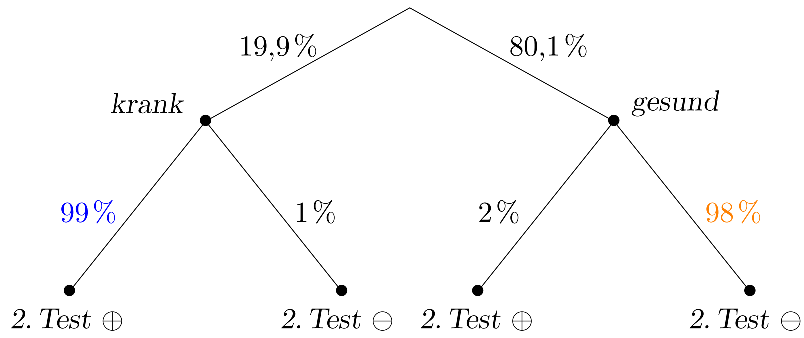
a) erneut
b) nur beim 1. Test

eintritt, in Wirklichkeit krank ist?

Der Test wird wiederholt.
Wie groß ist die Wahrscheinlichkeit, dass die Person, bei der die Reaktion

a) erneut
b) nur beim 1. Test

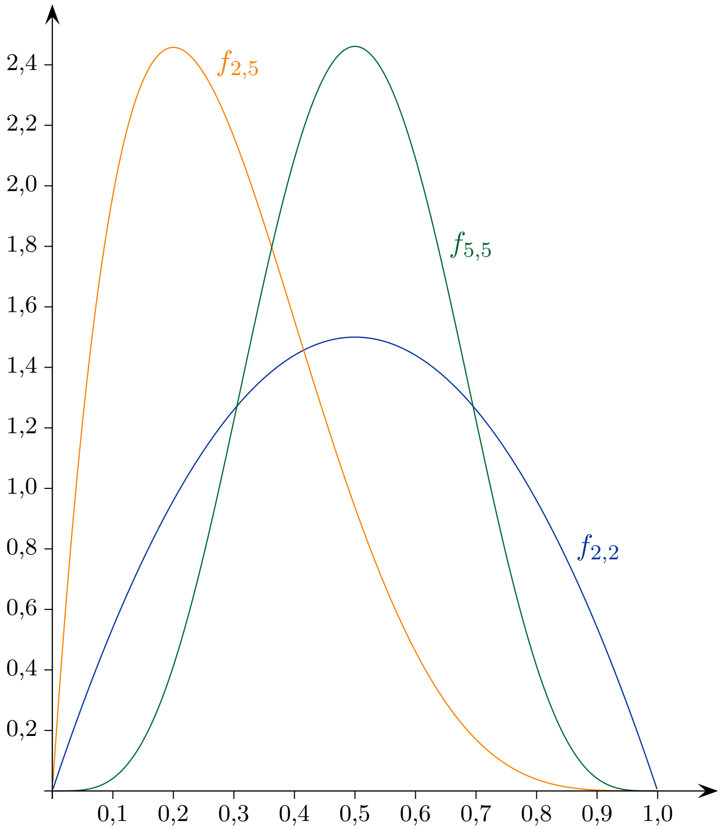
eintritt, in Wirklichkeit krank ist?



a) $P(\textit{krank} \mid \textit{auch 2. Test} \oplus) = \dfrac{19{,}9\,\% \cdot 99\,\%}{19{,}9\,\% \cdot 99\,\% \;+\; 80{,}1\,\% \cdot 2\,\%} = 92{,}5\,\%$

b) $P(\textit{krank} \mid \textit{nur im 1. Test} \oplus) = \dfrac{19{,}9\,\% \cdot 1\,\%}{19{,}9\,\% \cdot 1\,\% \;+\; 80{,}1\,\% \cdot 98\,\%} = 0{,}3\,\%$

Beta-Verteilung



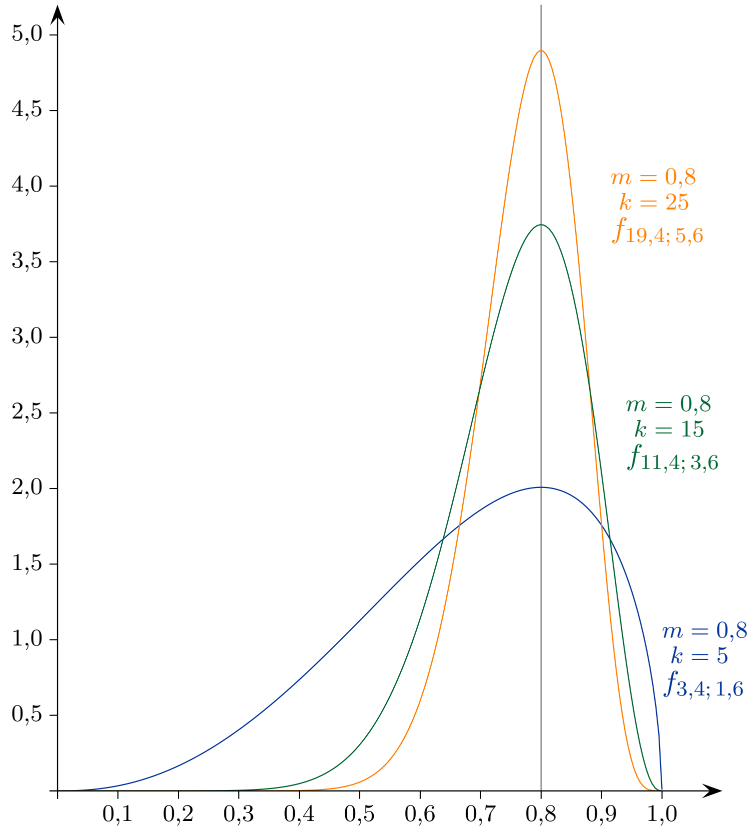
Die Betaverteilung ist mit der Binomialverteilung verwandt.
Anteile von Merkmalsträgern in einer Gesamtheit sind oft betaverteilt.
Die Verteilung ist durch zwei Parameter $a, b > 0$ festgelegt:

$$f_{a,b}(x) = \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}, \quad x \in\,]0,1[$$

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$$

Der Inhalt der Fläche unterhalb des Graphen von $f_{a,b}$ muss 1 sein.

Beta-Verteilung anpassen



Um mit der Betaverteilung eine Verteilung zu approximieren,
sind die Parameter $a, b > 0$ zu ermitteln.

$$f_{a,b}(x) = \frac{1}{B(a,b)}\,x^{a-1}(1-x)^{b-1}, \quad x \in \,]0,1[$$

Zunächst kann die Extremstelle (Modus) $m = \dfrac{a-1}{a+b-2}$ (Differenzialrechnung) festgelegt werden.
Mit $k = a + b$ kann die Abweichung von m in unmittelbarer Nähe beeinflusst werden, siehe Grafik.

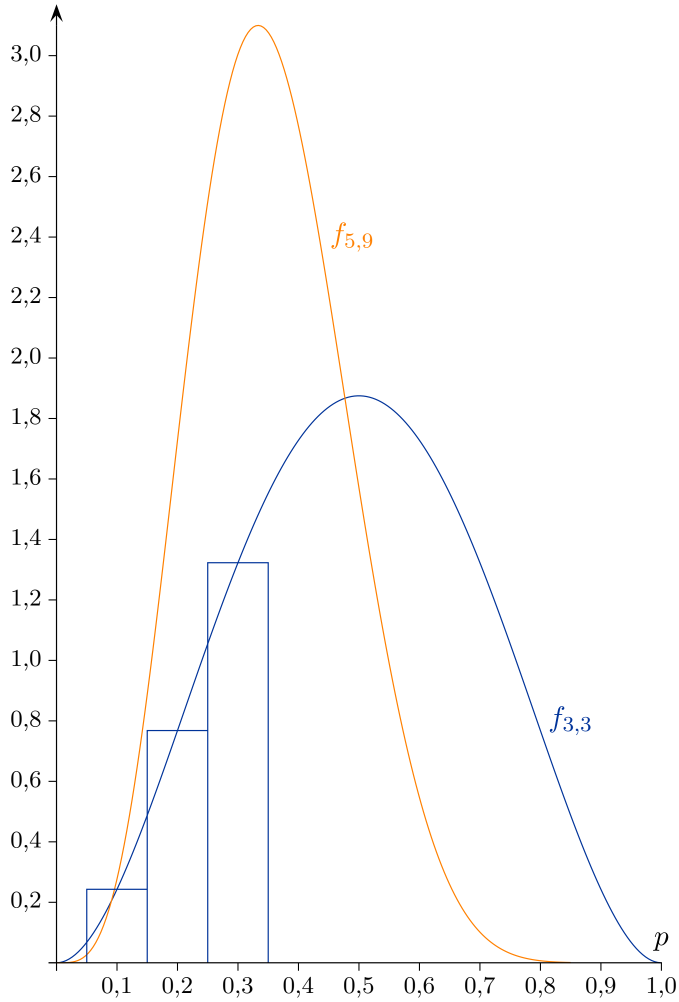
Leichte Umformungen ergeben:

$$\begin{aligned} a &= m(k-2)+1 \\ b &= (1-m)(k-2)+1 \end{aligned}$$

Alternativ können der Modus m $(0 < m < 1)$ und für einen Punkt $P(p^* \mid f(p^*))$ auf der Dichte das Verhältnis $\dfrac{f(p^*)}{f(m)} = \left(\dfrac{p^*}{m}\right)^{a-1}\left(\dfrac{1-p^*}{1-m}\right)^{b-1}$ festgelegt werden. Es ist dann:

$$a = \frac{\ln\dfrac{f(p^*)}{f(m)}}{\ln\dfrac{p^*}{m} + \dfrac{1-m}{m}\,\ln\dfrac{1-p^*}{1-m}} + 1, \qquad b = (a-1)\frac{1-m}{m} + 1$$

a-posteriori-Dichte



Gesucht ist der Anteil p von Merkmalsträgern in einer Gesamtheit.
Wir nehmen an, dass der Anteil betaverteilt mit der Dichte $f_{3,3}$ ist (a-priori-Verteilung).
In einer Stichprobe von $n = 8$ Elementen findet man $k = 2$ Merkmalsträger.
Welche a-posteriori-Verteilung ergibt das?

Die Anzahl X der Merkmalsträger in einer Stichprobe ist binomialverteilt:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}, \qquad q = 1 - p$$

Für die Likelihoods folgt

$$P(X = 2 \,|\, p\,) \propto p^k \cdot q^{n-k}, \qquad n = 8,\ \ k = 2,\ \ \binom{n}{k} = 28 \ \text{ ist von } p \text{ unabhängig.}$$

Für die bedingte a-posteriori-Dichte gilt:

$$f_{a,b}(p \,|\, X = k) \propto p^{a-1} \cdot q^{b-1} \cdot p^k \cdot q^{n-k} = p^{a+k-1} \cdot q^{b+n-k-1} = f_{a+k,\,b+n-k}(p)$$

$$f_{3,3}(p \,|\, X = 2) = f_{5,9}(p)$$

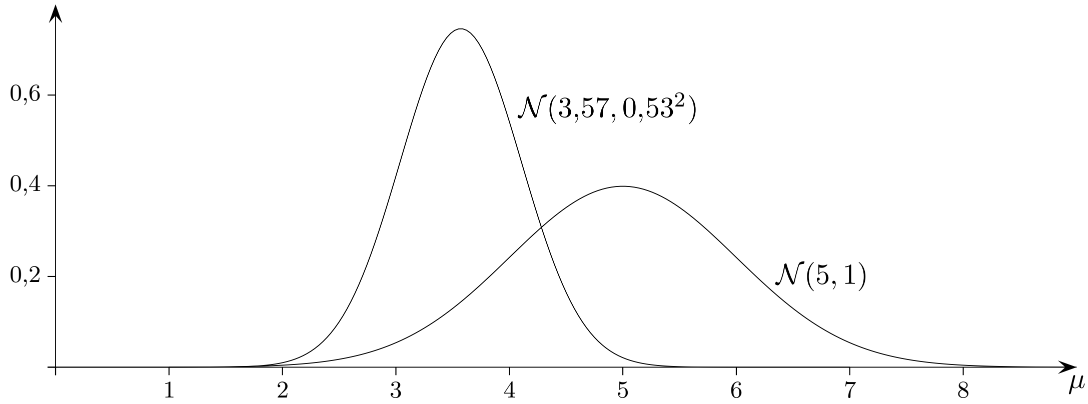
zum Nachrechnen: $f_{3,3}(p) = 30\,p^2 q^2$, $f_{5,9}(p) = \frac{28 \cdot 30 \cdot 429}{56}\, p^4 q^8$

Unsere bisherigen Überlegungen zu bedingten Wahrscheinlichkeiten bezogen sich auf diskrete Verteilungen, die Beta-Verteilung ist jedoch stetig. Eine stetige Verteilung kann beliebig genau durch eine diskrete approximiert werden, so dass sich die Überlegungen übertragen:

a-posteriori-Dichte $\propto$ a-priori-Dichte $\times$ Likelihood

Normalverteilung $\mathcal{N}(\mu, \sigma^2)$

Gesucht ist die a-posteriori-Verteilung des Erwartungswerts einer normalverteilten Zufallsgröße X, von der man den Mittelwert $\overline{x}$ einer Stichprobe vom Umfang n kennt, wenn man als a-priori-Verteilung eine Normalverteilung mit Mittelwert μ' und Varianz σ'^2 annimmt. Die Varianz σ^2 von X sei bekannt. Eine unbekannte Varianz ist aufwändiger zu handhaben.



Es gilt:

$$f(\mu) \propto e^{-\frac{1}{2}\left(\frac{\mu-\mu'}{\sigma'}\right)^2}$$ a-priori-Dichte

$$f(\overline{x} \,|\, \mu) \propto e^{-\frac{1}{2}\left(\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$ Likelihood, $\sqrt{n}$-Gesetz

$$f(\mu \,|\, \overline{x}) \propto e^{-\frac{1}{2}\left(\frac{\mu-\mu'}{\sigma'}\right)^2} e^{-\frac{1}{2}\left(\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$

a-posteriori-Dichte als Produkt der Dichten
Begründung: diskrete Approximation

$$= e^{-\frac{1}{2}\left(\frac{\mu-\mu''}{\sigma''}\right)^2}$$

Nach längerer Umformung ergeben sich die Parameter der a-posteriori-Dichte.

$$\mu'' = [\sigma'^2 \overline{x} + \frac{\sigma^2}{n}\mu']/D$$

$$\sigma''^2 = \sigma'^2 \frac{\sigma^2}{n} / D$$

$$D = \sigma'^2 + \frac{\sigma^2}{n}$$

Zahlenbeispiel:

$$\begin{aligned} n &= 10 \\ \overline{x} &= 3 \\ \sigma &= 2 \\ \sigma' &= 1 \\ \mu' &= 5 \end{aligned}$$

$$\begin{aligned} \mu'' &= 3{,}57 \\ \sigma'' &= 0{,}53 \end{aligned}$$

Siehe auch: Bedingte Wahrscheinlichkeit

Sek I

Sek II

Startseite