Von der Vierfeldertafel zum Baumdiagramm und zurück

Häufig betrachtet man Situationen, in denen 2 verschiedene „Merkmale“ untersucht werden und jedes dieser Merkmale zwei „Ausprägungen“ hat. Beispiel: Merkmal 1 = „Geschlecht“ mit den Ausprägungen „männlich“ oder „weiblich“ sowie Merkmal 2 = „Schulabschluss“ mit den Ausprägungen „erworben“ oder „nicht erworben“. Wegen der jeweils zwei Ausprägungen können dann zusammen $2 \cdot 2 = 4$ Fälle unterschieden werden: männlich mit Abschluss, männlich ohne Abschluss, weiblich mit Abschluss und weiblich ohne Abschluss.

Diese Daten können z.Bsp. durch eine Umfrage gewonnen werden und eine Möglichkeit sie übersichtlich darzustellen ist die „Vierfeldertafel“:

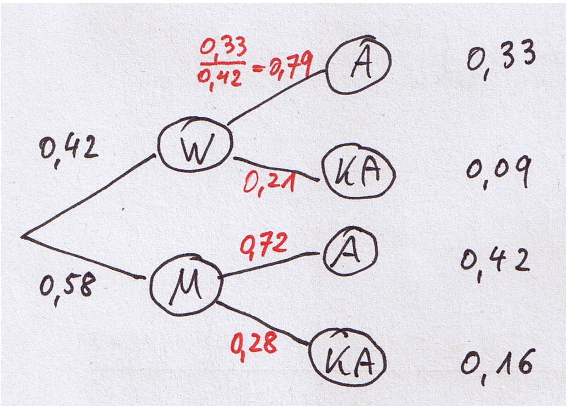
	männlich (M)	weiblich (W)	gesamt
Abschluss (A)	0,42	0,33	0,75
kein Abschluss (kA)	0,16	0,09	0,25
gesamt	0,58	0,42	1,00

Die Prozentangaben beziehen sich hier auf die *Gesamtzahl* der Schulabgänger (d.h. beiderlei Geschlecht und egal ob mit oder ohne Abschluss).

Das erkennt man auch daran, dass sich die vier Zahlen in der Mitte der Tafel zur Wahrscheinlichkeit 100% addieren (bzw. zu 1, wenn man die Dezimalpunktschreibweise wählt). Die Zeilen bzw. Spalten addieren sich zum Gesamtprozentsatz mit je *einer* Merkmalsausprägung.

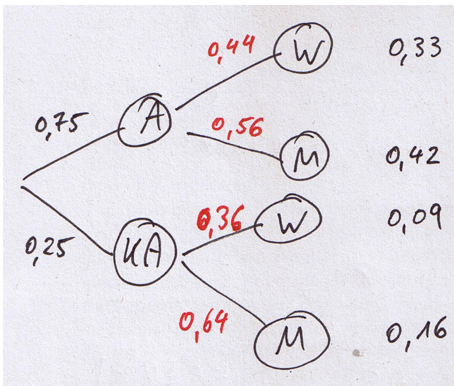
Was hat das Ganze mit Wahrscheinlichkeitsrechnung zu tun? Nun, man kann den (zugegeben) etwas künstlichen Fall betrachten, dass aus einer Gruppe von Schulabgängern *zufällig* Personen ausgewählt werden, bei denen man anschließend nach Geschlecht und Abschluss fragt. Dies ist dann aber gerade ein „Zufallsexperiment“, deren Ausgang in einem Baumdiagramm dargestellt werden kann. Die Werte für die Teilwahrscheinlichkeiten können aus der Vierfeldertafel berechnet werden:

Die Abbildung zeigt ein Baumdiagramm, das zu obiger Vierfeldertafel gehört. Die Wahrscheinlichkeiten auf der ersten Stufe können sofort aus der Vierfeldertafel abgelesen werden (und zwar aus der „gesamt Zeile/Spalte“ des ersten Merkmals). Die Pfadwahrscheinlichkeiten sind ebenfalls sofort abzulesen (etwa: 0,33 als Wahrscheinlichkeit für weiblich und Abschluss). Nur die Wahrscheinlichkeiten auf der *zweiten Stufe* müssen berechnet werden! Etwa mit Hilfe der Pfadregel: die gesuchte Teilwahrscheinlichkeit muss multipliziert mit der bekannten (ersten) Teilwahrscheinlichkeit die Pfadwahrscheinlichkeit ergeben. So kommt man an die „roten“ Ziffern (= Teilwahrscheinlichkeiten der zweiten Stufe).



Also zum Beispiel: mit einer Wahrscheinlichkeit von 42% wählt man eine Dame aus, die dann mit eine 21% ohne Schulabschluss ist.

Natürlich kann man auch das Merkmal „Abschluss“ als erste Stufe wählen. Dann ergeben sich aber andere Teilwahrscheinlichkeiten, wie im folgenden Diagramm dargestellt.
Dieses Baumdiagramm stellt man genauso wie das erste auf: zuerst die Teilwahrscheinlichkeit für die erste Stufe ablesen, dann die Pfadwahrscheinlichkeiten und nur die „zweiten“ Teilwahrscheinlichkeiten (rot) müssen berechnet werden. Man erkennt: Die Teilwahrscheinlichkeiten unterscheiden sich zwischen den beiden Baumdiagrammen. Wenn man *zuerst* ein Mädchen „wählt“ und *dann* nach dem Abschluss fragt, ist die Wahrscheinlichkeit für eine positive Antwort 79%. Fragt man *zuerst* nach dem Abschluss liegt die Wahrscheinlichkeit für eine positive Antwort bei 75%. Es kommt hier also auf die Reihenfolge an. Die Pfadwahrscheinlichkeit ganz rechts ändern sich natürlich (bis auf ihre Reihenfolge) nicht.

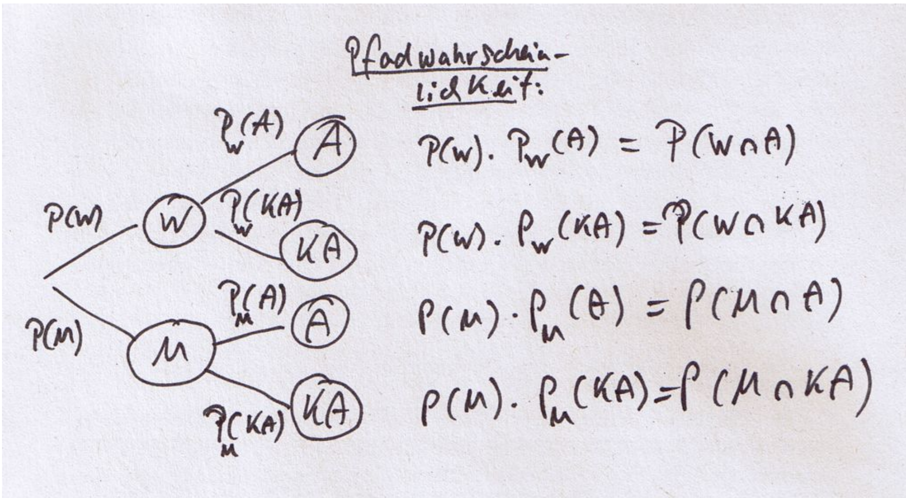


Symbolisch drückt man das wie folgt aus: $p_{\mathrm{w}}(\mathrm{A}) \neq p(\mathrm{A})$, sprich: Wahrscheinlichkeit für Ereignis „A“, wenn Ereignis „w“ vorliegt (unterer Index) ist ungleich der Wahrscheinlichkeit für „A“ *ohne* diese Bedingung.

Unabhängigkeit und Abhängigkeit von Ereignissen

Wenn die Wahrscheinlichkeit für „A“ davon abhängt, ob Ereignis „w“ vorliegt oder nicht, spricht man von **abhängigen Ereignissen**. Gilt dagegen $p_{\mathrm{w}}(\mathrm{A}) = p(\mathrm{A})$ nennt man die Ereignisse **unabhängig**. Ziel von Schulpolitik sollte sein, dass Geschlecht und erfolgreicher Schulbesuch unabhängig sind!

Wir geben noch eine andere Anwendung/Erläuterung dieses Begriffs: Die Pfadwahrscheinlichkeiten ergeben sich ja aus dem Produkt der Teilwahrscheinlichkeiten. Die folgende Abbildung drückt dies nur mit Formelbuchstaben aus:



So berechnet sich zum Beispiel der Pfad „weiblich“ und „Abschluss“ aus $p(w) \cdot p_w(A)$. Für dieses Produkt kann man auch $p(w \cap A)$ schreiben (sprich: Wahrscheinlichkeit für die Schnittmenge von „w“ und „A“, also die Ereignisse, bei denen *beide* Merkmalsausprägungen vorliegen). *Wären* die Ereignisse aber unabhängig, könnte man $p_w(A)$ einfach durch $p(A)$ ersetzen.

Regel: Bei **unabhängigen** Ereignissen A und B gilt: $p(A \cap B) = p(A) \cdot p(B)$.

Beispiele:

- Die Wahrscheinlichkeit für jede Augenzahl beim Würfeln ist $\frac{1}{6}$. Mit welcher Wahrscheinlichkeit würfelt man zwei Sechsen hintereinander?

  Antwort: Da beide Würfe unabhängig sind, berechnet sich die gesuchte Wahrscheinlichkeit als $p(6,6) = \frac{1}{6} \cdot \frac{1}{6}$.

- 27% der männlichen Bevölkerung Deutschlands sind über 1,80m. Gleichzeitig spielen in Deutschland 1,2% der männlichen Bevölkerung Basketball. Wie groß ist die Wahrscheinlichkeit, einen Basketballer anzutreffen, der über 1,80m ist?

  Offensichtlich wäre es blödsinn, die beiden Wahrscheinlichkeiten zu multiplizieren: $0,27 \cdot 0,012 = 0,00324$. Man erwartet, dass fast jeder Basketballer über 1,80m ist, und nicht nur die 27% des Bevölkerungsdurchschnitts. Körpergröße und Sportart sind eben **nicht** unabhängig.