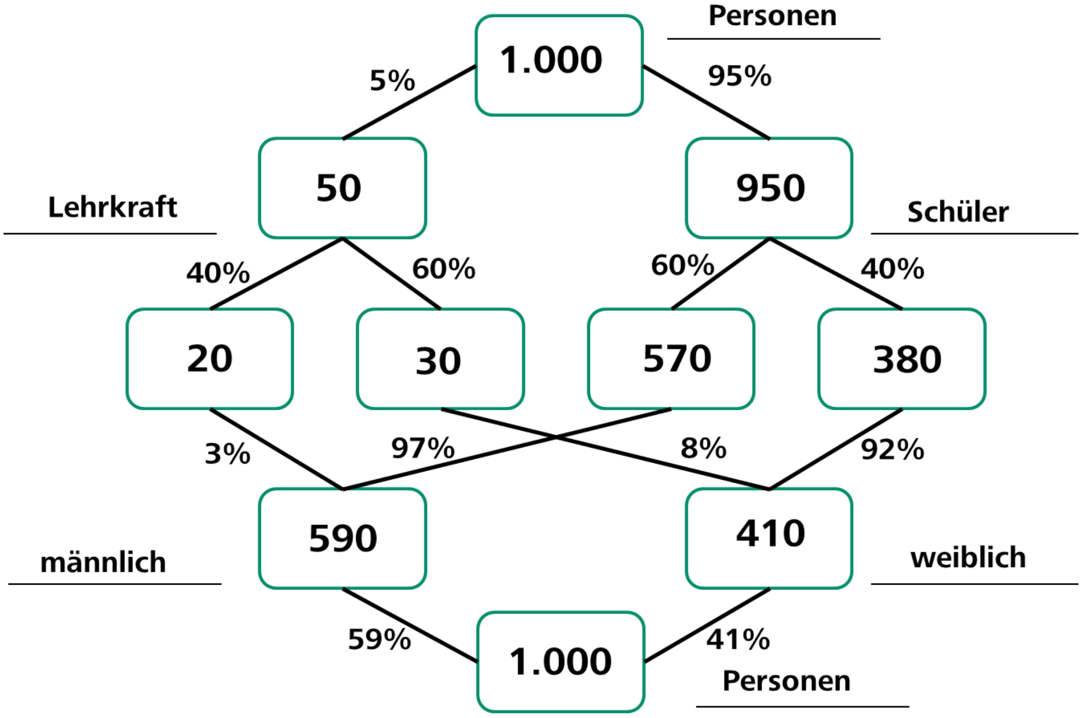
HÄUFIGKEITSDOPPELBAUM VS. VIERFELDERTAFEL

Häufigkeitsdoppelbaum (mit absoluten *und* relativen Häufigkeiten)



In *Häufigkeitsdoppelbäumen* ist die hierarchische Struktur der Daten besser abgebildet als in der Vierfeldertafel. Beide Möglichkeiten der sukzessiven Aufteilung der Stichprobe sind intuitiver sichtbar. Darüber hinaus lassen sich durch Lesen „von innen nach außen" alle (bedingten oder unbedingten) Wahrscheinlichkeiten ganz einfach „ablesen" und ggfs. sogar an den Pfaden bequem ergänzen.

Ist keine Gesamtstichprobe gegeben, so kann man sich eine „imaginäre Stichprobe" ausdenken (vgl. Folie 35). Dabei ist es immer möglich, diese imaginäre Zahl so zu wählen, dass sich letzlich im gesamten Häufigkeitsdoppelbaum nur ganze Zahlen ergeben (für eine Übungsaufgabe siehe Aufgabe 11 im Workshop)

Lösungen zu „Bayesianischen" Aufgaben (siehe Vortragsfolien) lassen sich dann „auf der anderen Seite" des Baumes ablesen. Wahrscheinlichkeiten von Schnittmengen (z.B. einer weiblichen Lehrerin) lassen sich durch „Überspringen einer Ebene" ablesen (z.B. 30 von 1.000). Für die Bestimmung von *beliebigen* Wahrscheinlichkeiten sind *weder* das Konzept der bedingten Wahrscheinlichkeit *noch* die Pfadregeln erforderlich.

Zudem zeigen empirische Studien, dass Häufigkeits(doppel)bäume die Lösungsfindung besser unterstützen als Vierfeldertafeln (Wassner, 2004).

Vierfeldertafel (mit absoluten Häufigkeiten)

	Lehrkraft	Schüler	
männlich	20	570	590
weiblich	30	380	410
	50	950	1.000

In einer Vierfeldertafel lassen sich dagegen *nicht* alle bedingten Wahrscheinlichkeiten bequem ergänzen. Wo beispielsweise sollte die Wahrscheinlichkeit P_{Lehrer}(männlich) = 20/50 eingetragen werden? Beim Doppelbaum gibt es für jede dieser bedingten Wahrscheinlichkeiten einen entsprechenden Ast!

Man beachte darüber hinaus, dass in einer Vierfeldertafel mit *relativen Häufigkeiten* im Inneren ausschließlich *Wahrscheinlichkeiten von Schnittmengen* eingetragen sind; ganz im Gegensatz zum Baumdiagramm mit Wahrscheinlichkeiten, an dessen unteren Äste gerade *bedingte Wahrscheinlichkeiten* stehen.

Dies hat zur Folge, dass die Vierfeldertafel *prinzipiell* ein weniger mächtiges didaktisches Werkzeug ist, als ein Baumdiagramm, da die Strategien zur Lösung von Aufgaben (z.B. „Wie viele Lehrkräfte sind männlich?“) identisch für beiden Arten von Vierfeldertafeln sind: Die zwei entsprechenden Zellen müssen in beiden Fällen auf die jeweils selbe Weise dividiert werden. Dies wird auch klar, wenn man die absoluten Häufigkeiten in obiger Vierfeldertafel durch relative Häufigkeiten ersetzt: Innen würden dann beispielsweise die Zahlen „0,02“; „0,57“; „0,03“ bzw. „0,38“ stehen (also abgesehen von einer Normierung auf 1 bzw. 100% dieselben Zahlen)! Somit eröffnet der Wechsel im Unterricht von einer Vierfeldertafel mit relativen Häufigkeiten zu einer mit absoluten Häufigkeiten keine zusätzliche Einsicht oder Lösungsstrategie, ganz im Gegensatz zum Baumdiagramm.

Außerdem ist der sukzessive Aufbau der Vierfeldertafel mithilfe der im Vortrag vorgeschlagenen sogenannten „imaginären Stichprobe“ wesentlich weniger intuitiv als mit dem Doppelbaum. Man stelle sich z. B. folgende „Bayesianische“ Aufgabe vor:

- Die Wahrscheinlichkeit, dass es sich bei einer Person um eine Lehrkraft handelt, beträgt 5 %.
- 2/5 der Lehrkräfte sind männlich.
- Die Wahrscheinlichkeit, dass ein Schüler männlich ist, beträgt 60 %.

Frage: Wie groß ist die Wahrscheinlichkeit, dass eine zufällig betrachtete weibliche Person eine Lehrkraft ist?

Ausgehend von einer imaginären Stichprobe von 1.000 Personen würde der Aufbau der kompletten Vierfeldertafel nun „Sprünge“ von rechts unten (die imaginären 1000) nach links unten (50) und dann nach links oben (20) erfordern. Im Doppelbaum dagegen ließen sich diese Zahlen sehr intuitiv „der Reihe nach“ entlang eines Pfades eintragen.