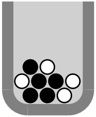
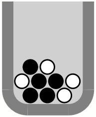
Bayes-Schätzer

Präludium

Es liegt eine Urne mit s schwarzen und w weißen Kugeln vor, die gut gemischt wurde. Aus dieser wird eine zufällige Kugel gezogen. Gesucht ist die Wahrscheinlichkeit, eine schwarze bzw. eine weiße Kugel zu erhalten.



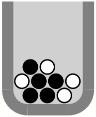
Es liegt eine Urne mit s schwarzen und w weißen Kugeln vor, die gut gemischt wurde. Aus dieser wird eine zufällige Kugel gezogen. Gesucht ist die Wahrscheinlichkeit, eine schwarze bzw. eine weiße Kugel zu erhalten.



Es sei $X(\omega) \in \{0, 1\}$ das Ergebnis der Ziehung, wobei 0 für *schwarz*, und 1 für *weiß* stehe. Die Gesamtzahl der Kugeln ist $s + w$. Für eine schwarze Kugel sind davon s günstig, für eine weiße sind es w. Man erhält somit

$$P(X = 0) = \frac{s}{s + w},$$
$$P(X = 1) = \frac{w}{s + w}.$$

Es liegt eine Urne mit s schwarzen und w weißen Kugeln vor, die gut gemischt wurde. Aus dieser wird eine zufällige Kugel gezogen. Gesucht ist die Wahrscheinlichkeit, eine schwarze bzw. eine weiße Kugel zu erhalten.



Es sei $X(\omega) \in \{0, 1\}$ das Ergebnis der Ziehung, wobei 0 für *schwarz*, und 1 für *weiß* stehe. Die Gesamtzahl der Kugeln ist $s + w$. Für eine schwarze Kugel sind davon s günstig, für eine weiße sind es w. Man erhält somit

$$P(X=0) = \frac{s}{s+w},$$
$$P(X=1) = \frac{w}{s+w}.$$

Wir fassen die beiden Fälle zusammen zu

$$P(X=x) = \frac{(1-x)s + xw}{s+w}.$$

Ein unbekannter Parameter

Die Zahl w der weißen Kugeln sei nun unbekannt, entstamme aber aus dem Bereich $\{0, ..., N-1\}$. Zumindest tun wir so, vergleichen später mit der wahren Zahl. Wir wollen w durch eine Ziehung aus der Urne schätzen.

Die Zahl w der weißen Kugeln sei nun unbekannt, entstamme aber aus dem Bereich $\{0,...,N-1\}$. Zumindest tun wir so, vergleichen später mit der wahren Zahl. Wir wollen w durch eine Ziehung aus der Urne schätzen.

Dazu sei W die Zufallsgröße für die Zahl der weißen Kugeln. Wir nehmen zu Anfang an, jede Zahl w wäre gleich wahrscheinlich, also $P(W=w)=\frac{1}{N}$ für jedes $w \in \{0,...,N-1\}$.

Die Zahl w der weißen Kugeln sei nun unbekannt, entstamme aber aus dem Bereich $\{0, ..., N-1\}$. Zumindest tun wir so, vergleichen später mit der wahren Zahl. Wir wollen w durch eine Ziehung aus der Urne schätzen.

Dazu sei W die Zufallsgröße für die Zahl der weißen Kugeln. Wir nehmen zu Anfang an, jede Zahl w wäre gleich wahrscheinlich, also $P(W = w) = \frac{1}{N}$ für jedes $w \in \{0, ..., N-1\}$.

Es wird eine Kugel aus der Urne gezogen, was die Evidenz $E := \{X = x\}$ zu festem $x = 0$ oder $x = 1$ schafft.

Dem Satz von Bayes nach gilt nun

$$P(W = w \mid E) = \frac{P(E \mid W = w)}{P(E)} P(W = w).$$

Dem Satz von Bayes nach gilt nun

$$P(W = w \mid E) = \frac{P(E \mid W = w)}{P(E)} P(W = w).$$

Die Terme besitzen hierbei die folgende Bedeutung:

$$\begin{aligned} P(W = w) &\text{ — } \textit{A-priori-Wahrscheinlichkeit} \\ P(W = w \mid E) &\text{ — } \textit{A-posteriori-Wahrscheinlichkeit} \\ \tfrac{P(E \mid W = w)}{P(E)} &\text{ — } \textit{Aktualisierungsfaktor} \\ P(E \mid W = w) &\text{ — } \textit{Likelihood} \end{aligned}$$

Dem Satz von Bayes nach gilt nun

$$P(W = w \mid E) = \frac{P(E \mid W = w)}{P(E)} P(W = w).$$

Die Terme besitzen hierbei die folgende Bedeutung:

$$\begin{aligned} P(W = w) & \text{ — } \textit{A-priori-Wahrscheinlichkeit} \\ P(W = w \mid E) & \text{ — } \textit{A-posteriori-Wahrscheinlichkeit} \\ \tfrac{P(E \mid W = w)}{P(E)} & \text{ — } \textit{Aktualisierungsfaktor} \\ P(E \mid W = w) & \text{ — } \textit{Likelihood} \end{aligned}$$

Zur Bestimmung der A-posteriori-Wahrscheinlichkeit benötigen wir den Aktualisierungsfaktor.

Die Likelihood ist bei Erhalt der Evidenz bekannt, da sie aufgrund der Bedingung $W = w$ die übliche Wahrscheinlichkeit der Ziehung aus der Urne bei Kenntnis des Parameters w ist. Das heißt,

$$P(X = x \mid W = w) = \frac{(1-x)s + xw}{s+w}.$$

Aber wie bestimmen wir $P(X = x)$?

Aber wie bestimmen wir $P(X = x)$?

Diese ergibt sich mit dem Gesetz der totalen Wahrscheinlichkeit zu

$$P(X = x) = \sum_{k=0}^{N-1} P(X = x \mid W = k) P(W = k),$$

lässt sich also auf die Likelihood und die A-priori-Wahrscheinlichkeit zurückführen, die beide bekannt sind.

Aber wie bestimmen wir $P(X = x)$?

Diese ergibt sich mit dem Gesetz der totalen Wahrscheinlichkeit zu

$$P(X = x) = \sum_{k=0}^{N-1} P(X = x \mid W = k) P(W = k),$$

lässt sich also auf die Likelihood und die A-priori-Wahrscheinlichkeit zurückführen, die beide bekannt sind.

Wir erhalten

$$P(X = x) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{(1 - x)s + xk}{s + k}.$$

Die A-posteriori-Wahrscheinlichkeit bestimmt sich also zu

$$P(W = w \mid X = x) = \frac{\frac{(1-x)s+xw}{s+w}}{\sum_{k=0}^{N-1} \frac{(1-x)s+xk}{s+k}}.$$

Mehrere Stichproben

Nun würden wir das Updaten der Wahrscheinlichkeit mit einer weiteren Stichprobe gerne abermals durchführen. Das Setting soll dabei natürlich unmodifiziert bleiben. Nach der jeweiligen Ziehung wird die Kugel also wieder zurück in die Urne gelegt, und diese daraufhin wieder gut gemischt.

Nun würden wir das Updaten der Wahrscheinlichkeit mit einer weiteren Stichprobe gerne abermals durchführen. Das Setting soll dabei natürlich unmodifiziert bleiben. Nach der jeweiligen Ziehung wird die Kugel also wieder zurück in die Urne gelegt, und diese daraufhin wieder gut gemischt.

Wir setzen dafür $X = (X_1, \dots, X_n)$, wobei ein $x = (x_1, \dots, x_n)$ aus den Stichproben $x_i \in \{0, 1\}$ besteht. Die X_i sind unter $W = w$ unabhängig und identisch verteilt. Was unser ursprüngliches x war, ist nun x_1.

Nun würden wir das Updaten der Wahrscheinlichkeit mit einer weiteren Stichprobe gerne abermals durchführen. Das Setting soll dabei natürlich unmodifiziert bleiben. Nach der jeweiligen Ziehung wird die Kugel also wieder zurück in die Urne gelegt, und diese daraufhin wieder gut gemischt.

Wir setzen dafür $X = (X_1, \ldots, X_n)$, wobei ein $x = (x_1, \ldots, x_n)$ aus den Stichproben $x_i \in \{0, 1\}$ besteht. Die X_i sind unter $W = w$ unabhängig und identisch verteilt. Was unser ursprüngliches x war, ist nun x_1.

Bestimmt wurde bislang also nur $P(W = w \mid X_1 = x_1)$. Als nächtes kommt $P(W = w \mid X_1 = x_1, X_2 = x_2)$. Oder besser gesagt, wir würden letztlich gern

$$P(W = w \mid X = x) = P(W = w \mid X_1 = x_1, \ldots, X_n = x_n)$$

bestimmen.

Dem Satz von Bayes nach rechnen wir

$$P(W = w \mid E_2) = \frac{P(E_2 \mid W = w)}{P(E_2)} P(W = w)$$

mit $E_2 := \{X_1 = x_1, X_2 = x_2\} = \{X_1 = x_1\} \cap \{X_2 = x_2\}$.

Dem Satz von Bayes nach rechnen wir

$$P(W = w \mid E_2) = \frac{P(E_2 \mid W = w)}{P(E_2)} P(W = w)$$

mit $E_2 := \{X_1 = x_1, X_2 = x_2\} = \{X_1 = x_1\} \cap \{X_2 = x_2\}$.

Aufgrund der stochastischen Unabhängigkeit der X_i bei $W = w$ gilt

$$P(E_2 \mid W = w) = P(X_2 = x_2 \mid W = w) P(X_1 = x_1 \mid W = w)$$

denn für unabhängige Ereignisse A, B gilt $P(A \cap B) = P(A)P(B)$.

Wir dürfen allerdings nicht einfach $P(E_2) = P(X_1 = x_1)P(X_2 = x_2)$ rechnen, da X_1, X_2 nicht unabhängig sein müssen, weil die Verteilung von W in ihre Verteilungen eingeht. Wir nutzen daher zuerst das Gesetz der totalen Wahrscheinlichkeit, zerlegen $P(E_2 \mid W = k)$ daraufhin ins Produkt. Das macht

$$\begin{aligned} P(E_2) &= \sum_{k=0}^{N-1} P(E_2 \mid W = k)P(W = k) \\ &= \sum_{k=0}^{N-1} P(X_2 = x_2 \mid W = k)P(X_1 = x_1 \mid W = k)P(W = k). \end{aligned}$$

Wir dürfen allerdings nicht einfach $P(E_2) = P(X_1 = x_1)P(X_2 = x_2)$ rechnen, da X_1, X_2 nicht unabhängig sein müssen, weil die Verteilung von W in ihre Verteilungen eingeht. Wir nutzen daher zuerst das Gesetz der totalen Wahrscheinlichkeit, zerlegen $P(E_2 \mid W = k)$ daraufhin ins Produkt. Das macht

$$\begin{aligned} P(E_2) &= \sum_{k=0}^{N-1} P(E_2 \mid W = k)P(W = k) \\ &= \sum_{k=0}^{N-1} P(X_2 = x_2 \mid W = k)P(X_1 = x_1 \mid W = k)P(W = k). \end{aligned}$$

Die hinteren Faktoren des Summanden formen wir jetzt um vermittels

$$P(X_1 = x_1 \mid W = k) = \frac{P(W = k \mid X_1 = x_1)}{P(W = k)} P(X_1 = x_1).$$

Der Faktor $P(X_1 = x_1)$ hängt nicht von k ab, darf also aus der Summe ausgeklammert werden.

Somit findet sich

$$P(W = w \mid E_2) = \frac{P(X_2{=}x_2|W{=}w)}{\sum_{k=0}^{N-1} P(X_2{=}x_2|W{=}k)P(W{=}k|X_1{=}x_1)} \underbrace{\frac{P(X_1{=}x_1|W{=}w)}{P(X_1{=}x_1)} P(W = w)}_{P(W=w|X_1=x_1)}.$$

Somit findet sich

$$P(W = w \mid E_2) = \frac{P(X_2{=}x_2|W{=}w)}{\sum_{k=0}^{N-1} P(X_2{=}x_2|W{=}k) P(W{=}k|X_1{=}x_1)} \underbrace{\frac{P(X_1{=}x_1|W{=}w)}{P(X_1{=}x_1)} P(W = w)}_{P(W{=}w|X_1{=}x_1)}.$$

Wir führen die Aktualisierungen also sequentiell aus, wobei die aktualisierte Wahrscheinlichkeit $P(W = w \mid X_1 = x_1)$ in der zweiten Aktualisierung die Rolle der A-priori-Wahrscheinlichkeit einnimmt.

Somit findet sich

$$P(W = w \mid E_2) = \frac{P(X_2{=}x_2|W{=}w)}{\sum_{k=0}^{N-1} P(X_2{=}x_2|W{=}k)P(W{=}k|X_1{=}x_1)} \underbrace{\frac{P(X_1{=}x_1|W{=}w)}{P(X_1{=}x_1)} P(W = w)}_{P(W=w|X_1=x_1)}.$$

Wir führen die Aktualisierungen also sequentiell aus, wobei die aktualisierte Wahrscheinlichkeit $P(W = w \mid X_1 = x_1)$ in der zweiten Aktualisierung die Rolle der A-priori-Wahrscheinlichkeit einnimmt.

Salopp gesprochen dürfen wir die ursprüngliche A-priori-Wahrscheinlichkeit nach der Aktualisierung verwerfen, und so tun, als wäre die aktualisierte die vorliegende A-priori-Wahrscheinlichkeit.

Somit findet sich

$$P(W = w \mid E_2) = \frac{P(X_2{=}x_2|W{=}w)}{\sum_{k=0}^{N-1} P(X_2{=}x_2|W{=}k)P(W{=}k|X_1{=}x_1)} \underbrace{\frac{P(X_1{=}x_1|W{=}w)}{P(X_1{=}x_1)} P(W = w)}_{P(W=w|X_1=x_1)}.$$

Wir führen die Aktualisierungen also sequentiell aus, wobei die aktualisierte Wahrscheinlichkeit $P(W = w \mid X_1 = x_1)$ in der zweiten Aktualisierung die Rolle der A-priori-Wahrscheinlichkeit einnimmt.

Salopp gesprochen dürfen wir die ursprüngliche A-priori-Wahrscheinlichkeit nach der Aktualisierung verwerfen, und so tun, als wäre die aktualisierte die vorliegende A-priori-Wahrscheinlichkeit.

Zu beachten ist hier, dass die Wahrscheinlichkeit zu *jedem* w bestimmt werden muss. Wir berechnen also nicht eine einzelne Wahrscheinlichkeit, sondern die gesamte Wahrscheinlichkeitsfunktion $w \mapsto P(W = w \mid X_1 = x_1)$, und daraufhin $w \mapsto P(W = w \mid E_2)$.

Bezüglich $E_0 := \Omega$ und $E_i := \{X_i = x_i\} \cap E_{i-1}$ ergibt sich also die Rekurrenz

$$P(W = w \mid E_i) = \frac{P(X_i = x_i \mid W = w)}{\sum_{k=0}^{N-1} P(X_i = x_i \mid W = k) P(W = k \mid E_{i-1})} P(W = w \mid E_{i-1}).$$

Nochmals mehrere Stichproben

Nun verhält es sich aber auch so, dass die Summe $Y = \sum_{i=1}^{n} X_i$ die Zufallsgröße des n-stufigen Bernoulli-Experiments darstellt. Alternativ lässt sich zur Abschätzung also auch der Ansatz

$$P(W = w \mid Y = y) = \frac{P(Y = y \mid W = w)}{P(Y = y)} P(W = w)$$

machen, wobei

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

mit $p = \frac{w}{s+w}$ die Wahrscheinlichkeitsfunktion der Binomialverteilung ist.

Nach Entfaltung der totalen Wahrscheinlichkeit im Nenner kürzt sich der Faktor $\binom{n}{y}$ heraus. Das macht

$$P(W = w \mid Y = y) = \frac{p^y(1-p)^{n-y}P(W = w)}{\sum_{k=0}^{N-1} p_k^y(1-p_k)^{n-y}P(W = k)}$$

mit $p = \frac{w}{s+w}$ und $p_k = \frac{k}{s+k}$.

Es tut sich die Frage auf, welcher der beiden Ansätze besser schätzt.

Es tut sich die Frage auf, welcher der beiden Ansätze besser schätzt.

Die Antwort darauf lautet *keiner*.

Es tut sich die Frage auf, welcher der beiden Ansätze besser schätzt.

Die Antwort darauf lautet *keiner*.

Beide Rechnungen führen – Spitzfindigkeiten zur numerischen Güte beiseitegelassen – exakt zum selben Resultat.

Ein paar Begrifflichkeiten

Wir haben einen *Parameter* $\theta \in \Theta$. Hier ist dies $\theta := w$. Als *Parameterraum* haben wir $\Theta := \{0, \dots, N-1\}$. Das Maß P_θ hängt von diesen Parameter ab. Dies ist hier

$$P_\theta(X = x) = \frac{(1-x)s + x\theta}{s+\theta}.$$

Wir haben einen *Parameter* $\theta \in \Theta$. Hier ist dies $\theta := w$. Als *Parameterraum* haben wir $\Theta := \{0, \ldots, N-1\}$. Das Maß P_θ hängt von diesen Parameter ab. Dies ist hier

$$P_\theta(X = x) = \frac{(1-x)s + x\theta}{s + \theta}.$$

Man nennt

$$L_x : \Theta \to [0,1], \quad L_x(\theta) := P_\theta(X = x)$$

die *Likelihood-Funktion*.

Wir haben einen *Parameter* $\theta \in \Theta$. Hier ist dies $\theta := w$. Als *Parameterraum* haben wir $\Theta := \{0, \ldots, N-1\}$. Das Maß P_θ hängt von diesen Parameter ab. Dies ist hier

$$P_\theta(X = x) = \frac{(1-x)s + x\theta}{s + \theta}.$$

Man nennt

$$L_x : \Theta \to [0,1], \quad L_x(\theta) := P_\theta(X = x)$$

die *Likelihood-Funktion*.

Da der wahre Parameter θ unbekannt ist, ermitteln wir unsere Schätzung für diesen durch eine Zufallsgröße $T(x)$, die man die *Schätzfunktion* für θ nennt. Sie ordnet jedem Stichprobenvektor x einen Schätzwert $t = T(x)$ zu, wobei wir uns $t \approx \theta$ wünschen.

Wir können $T(x)$ als Erwartungswert der A-posteriori-Verteilung ansetzen.

Wir können $T(x)$ als Erwartungswert der A-posteriori-Verteilung ansetzen.

Bezeichnet $p(\theta)$ die Wahrscheinlichkeitsfunktion der A-priori- und $p(\theta \mid x)$ die der A-posteriori-Verteilung, haben wir also

$$T(x) := \sum_{\theta \in \Theta} \theta p(\theta \mid x)$$

mit der Bayes-Rechnung

$$p(\theta \mid x) := \frac{L_x(\theta)}{\sum_{t \in \Theta} L_x(t) p(t)} p(\theta)$$

bezüglich $p(\theta) := \frac{1}{|\Theta|}$.

Frequentistische Deutung

Bei $P(W = w \mid X = x)$ blieb bislang vage, welche Bedeutung W zukommt. Wir gehen von einer unbekannten wahren Anzahl weißer Kugeln aus, die geschätzt werden soll. Es tut sich die Frage auf, wo hierbei der Zufall liegen soll, da es sich bei W um eine Zufallsgröße handelt, die definitionsgemäß ein zufälliges Ergebnis auf ein anderes abbildet.

Bei $P(W = w \mid X = x)$ blieb bislang vage, welche Bedeutung W zukommt. Wir gehen von einer unbekannten wahren Anzahl weißer Kugeln aus, die geschätzt werden soll. Es tut sich die Frage auf, wo hierbei der Zufall liegen soll, da es sich bei W um eine Zufallsgröße handelt, die definitionsgemäß ein zufälliges Ergebnis auf ein anderes abbildet.

Aus subjektivistischer Sicht lautet die Antwort darauf, dass die Verteilung von W die Unkenntnis der wahren Anzahl beschreibt. Die aktualisierte Verteilung beschreibt dahingehend die verringerte Unkenntnis unter der Beobachtung der Stichprobe.

Bei $P(W = w \mid X = x)$ blieb bislang vage, welche Bedeutung W zukommt. Wir gehen von einer unbekannten wahren Anzahl weißer Kugeln aus, die geschätzt werden soll. Es tut sich die Frage auf, wo hierbei der Zufall liegen soll, da es sich bei W um eine Zufallsgröße handelt, die definitionsgemäß ein zufälliges Ergebnis auf ein anderes abbildet.

Aus subjektivistischer Sicht lautet die Antwort darauf, dass die Verteilung von W die Unkenntnis der wahren Anzahl beschreibt. Die aktualisierte Verteilung beschreibt dahingehend die verringerte Unkenntnis unter der Beobachtung der Stichprobe.

Das eigentliche Wesen von W verbleibt dennoch im Halbschatten. Zur Klärung rücken wir den Sachverhalt in die frequentistische Sicht. Dieser nimmt dabei die Form eines übergeordneten, zweistufigen Zufallsexperiments an.

Im ersten Teilexperiment wird unter allen möglichen Urnen, also allen möglichen Konfigurationen von weißen Kugeln, zufällig eine ausgewählt. Hierbei sei jede Konfiguration gleich wahrscheinlich.

Im ersten Teilexperiment wird unter allen möglichen Urnen, also allen möglichen Konfigurationen von weißen Kugeln, zufällig eine ausgewählt. Hierbei sei jede Konfiguration gleich wahrscheinlich.

Im zweiten Teilexperiment zieht man n mal mit zurücklegen eine Kugel aus dieser Urne, um den Stichprobenvektor x zu erhalten.

Im ersten Teilexperiment wird unter allen möglichen Urnen, also allen möglichen Konfigurationen von weißen Kugeln, zufällig eine ausgewählt. Hierbei sei jede Konfiguration gleich wahrscheinlich.

Im zweiten Teilexperiment zieht man n mal mit zurücklegen eine Kugel aus dieser Urne, um den Stichprobenvektor x zu erhalten.

Als adäquate Ergebnismenge bietet sich also

$$\Omega = \{0, \ldots, N-1\} \times \{0, 1\}^n$$

an. Die Zufallsgrößen nehmen dahingehend die Rolle der Projektionen

$$\begin{aligned} W &: \Omega \to \{0, \ldots, N-1\}, \quad & W((w, x)) &:= w, \\ X &: \Omega \to \{0, 1\}^n, & X((w, x)) &:= x. \end{aligned}$$

ein.

Alle bisherigen Rechnungen bleiben natürlich unverändert gültig. Die Wahrscheinlichkeit $P(W = w)$ deutet sich nun als die zur Wahl der Urne mit w Kugeln. Die Likelihood $P(X = x \mid W = w)$ deutet sich als die Übergangswahrscheinlichkeit unter dieser Urne, so dass

$$P(X = x, W = w) = P(X = x \mid W = w)P(W = w).$$

Insofern die gesuchte Wahrscheinlichkeit $P(W = w \mid X = x)$ allerdings nicht direkt am Baumdiagramm ablesbar ist, muss sie über die erläuterte Bayes-Rechnung bestimmt werden.

Alle bisherigen Rechnungen bleiben natürlich unverändert gültig. Die Wahrscheinlichkeit $P(W=w)$ deutet sich nun als die zur Wahl der Urne mit w Kugeln. Die Likelihood $P(X=x\mid W=w)$ deutet sich als die Übergangswahrscheinlichkeit unter dieser Urne, so dass

$$P(X=x, W=w) = P(X=x\mid W=w)P(W=w).$$

Insofern die gesuchte Wahrscheinlichkeit $P(W=w\mid X=x)$ allerdings nicht direkt am Baumdiagramm ablesbar ist, muss sie über die erläuterte Bayes-Rechnung bestimmt werden.

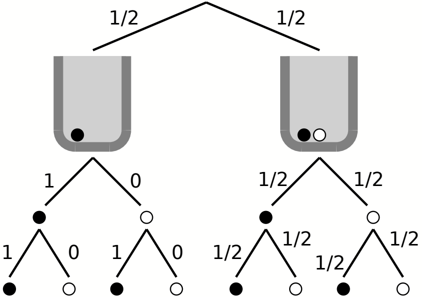
Darüber hinaus ist der Sachverhalt nun einer Simulation zugänglich. Wir führen das Experiment eine große Zahl, sagen wir eine Mio. mal durch. Aus den Ergebnissen $\omega\in\Omega$ werden diejenigen mit $X(\omega)=x$ zu einem fest gewählten x ausgesondert, deren Anzahl sei H_x. Zu einem festen w zählt man nun, wie häufig $X(\omega)=w$ darin vorkommt, dies sei H_{xw}. Man hat nun

$$P(W=w\mid X=x)\approx\frac{H_{xw}}{H_x},\ \text{sofern } H_x\neq 0.$$

Sind X_1, X_2 nun wirklich stochastisch abhängig? Betrachten wir dazu die Situation $s := 1$, $N := 2$, $n := 2$. Schon ein Gegenbeispiel (x_1, x_2) mit

$$P(X_1 = x_1, X_2 = x_2) \neq P(X_1 = x_1)P(X_2 = x_2)$$

genügt zur Widerlegung der Unabhängigkeit.



Für jedes i gilt die Rechnung

$$P(X_i = 1) = \sum_{k=0}^{1} P(X_i = 1 \mid W = k) W(W = k) = 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

womit $P(X_1 = 1)P(X_2 = 1) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$.

Für jedes i gilt die Rechnung

$$P(X_i = 1) = \sum_{k=0}^{1} P(X_i = 1 \mid W = k)W(W = k) = 0 \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4},$$

womit $P(X_1 = 1)P(X_2 = 1) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$.

Fürs gemeinsame Ereignis $\{X_1 = 1, X_2 = 1\}$ ergibt sich allerdings

$$\begin{aligned} P(X_1 = 1, X_2 = 1) &= \sum_{k=0}^{1} P(X_1 = 1, X_2 = 1 \mid W = k)P(W = k) \\ &= \sum_{k=0}^{1} P(X_1 = 1 \mid W = k)P(X_2 = 1 \mid W = k)P(W = k) \\ &= 0 \cdot 0 \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{8}. \end{aligned}$$

Für jedes i gilt die Rechnung

$$P(X_i = 1) = \sum_{k=0}^{1} P(X_i = 1 \mid W = k) W(W = k) = 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

womit $P(X_1 = 1)P(X_2 = 1) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$.

Fürs gemeinsame Ereignis $\{X_1 = 1, X_2 = 1\}$ ergibt sich allerdings

$$\begin{aligned}
P(X_1 = 1, X_2 = 1) &= \sum_{k=0}^{1} P(X_1 = 1, X_2 = 1 \mid W = k) P(W = k) \\
&= \sum_{k=0}^{1} P(X_1 = 1 \mid W = k) P(X_2 = 1 \mid W = k) P(W = k) \\
&= 0 \cdot 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.
\end{aligned}$$

Mit $P(X_1 = 1, X_2 = 1) = \frac{1}{8} \neq \frac{1}{16} = P(X_1 = 1)P(X_2 = 1)$ liegt also ein Gegenbeispiel zur stochastischen Unabhängigkeit von X_1, X_2 vor.

Ende.

Juli 2025