

Qualitätskontrolle

Wir sehen uns ein weiteres Beispiel aus der Qualitätskontrolle an:

In einer Fabrik werden Vorprodukte von zwei verschiedenen Firmen (*A Productions* und *B-warez*) weiterverarbeitet. Dabei stammen 70% der Vorprodukte von Firma *A* und 30% von Firma *B*. Aus langjähriger Erfahrung wissen die Arbeiter der Fabrik:

- Die Ausschußquoten betragen 1% bei Firma *A* und 5% bei Firma *B*.

Die Arbeiter finden eine neutrale Kiste des Vorprodukts (mit sehr vielen Teilen) ohne weitere Information über den Hersteller. Sie kontrollieren $n = 100$ Stück und entdecken y Ausschußstücke. Ist aus diesem Ergebnis ein Rückschluss auf den Produzenten möglich?

Sei also:

- Ereignis A : Los kommt von Firma *A Productions*
- Zufallsvariable Y : "Anzahl der Ausschußstücke" bei $n = 100$

Für unsere Beobachtung können wir eine Verteilung angeben, die jedoch vom Eintreten von A abhängt. Es handelt sich um eine Binomial-Verteilung mit $n = 100$, wobei die Wahrscheinlichkeit des Eintretens p von A abhängt:

- Ist A eingetreten, also die Kiste kommt von Firma *A*, dann ist die Wahrscheinlichkeit für die Produktion eines Ausschußstückes $p_A = 0.01$
- Ist dagegen \bar{A} eingetreten, also die Kiste kommt aus Firma *B*, dann ist die Wahrscheinlichkeit $p_B = 0.05$

Bemerkung: Eigentlich wird hier Ziehen ohne Zurücklegen gemacht, sprich wir müssten die Hypergeometrische Verteilung benutzen. Da wir aber keine Angabe über die Anzahl der Teile in der Kiste haben (nur "sehr viele"), nehmen wir die Binomialverteilung als Annäherung.

Dementsprechend ergibt sich die *bedingte Verteilung* bzw. *Datenverteilung* von Y :

- $Y|A \sim B(n, p_A)$
- $Y|\bar{A} \sim B(n, p_B)$

mit der Wahrscheinlichkeitsdichte der Binomialverteilung:

$$P(Y = y|A) = \binom{n}{y} p^y (1 - p)^{n-y}$$

Satz von Bayes

Der Satz von Bayes lässt sich analog wie zuvor anwenden:

$$P(A|Y = y) = \frac{P(Y = y|A) \cdot P(A)}{P(Y = y)}$$

Posteriori-Wahrscheinlichkeiten

Es ergeben sich folgende Posteriori-Wahrscheinlichkeiten:

y	0	1	2	3	4	5	6
$P(A Y = y)$	0.993	0.965	0.842	0.505	0.164	0.036	0.007
$P(\bar{A} Y = y)$	0.007	0.035	0.158	0.495	0.836	0.964	0.993

Je nach Höhe von y können wir also mehr oder weniger gut angeben, aus welcher Firma das Los wahrscheinlich kommt.

Bei $y = 3$ können wir uns nicht wirklich zwischen den Firmen entscheiden, aber die Posteriori-Wahrscheinlichkeit $P(A|Y = y) = 0.505$ ist kleiner als unsere ursprüngliche Priori-Wahrscheinlichkeit $P(A) = 0.7$.

Weiter Lernen

In diesem Fall können wir die Posteriori-Wahrscheinlichkeit $P(A|Y = y)$ wiederum als Priori für eine neue Stichprobe verwenden.

Sei nun

- Zufallsvariable Z : "Anzahl der Ausschußstücke" bei weiteren $n = 100$

Offensichtlich gilt wieder $Z \sim B(n, p)$.

Nun lässt sich $P(A|Z = y, Y = y)$ wie folgt berechnen:

$$P(A|Z = z, Y = y) = \frac{P(Z = z|A) \cdot P(A|Y = y)}{P(Z = z)}$$

Die Wahrscheinlichkeitsverteilung $P(Z = z|A)$ der Daten hängt nicht von Y ab. Als Priori benutzen wir hier die vorherige Posteriori-Wahrscheinlichkeit.

Neue Posteriori-Wahrscheinlichkeiten

Sehen wir nach den $y = 3$ Stücken Ausschuß unter den ersten 100 untersuchten Stücken unter den nächsten 100 untersuchten Stücken erneut z Stücke Ausschuß, so erhalten wir folgende Posteriori-Wahrscheinlichkeiten:

z	0	1	2	3	4	5	6
$P(A \mid Z = z, Y = y)$	0.984	0.924	0.699	0.308	0.079	0.016	0.003
$P(\bar{A} \mid Z = z, Y = y)$	0.016	0.076	0.301	0.692	0.921	0.984	0.997

Zum Beispiel geht bei erneut $z = 3$ Ausschußstücken unsere Tendenz nun klarer in Richtung Firma B .