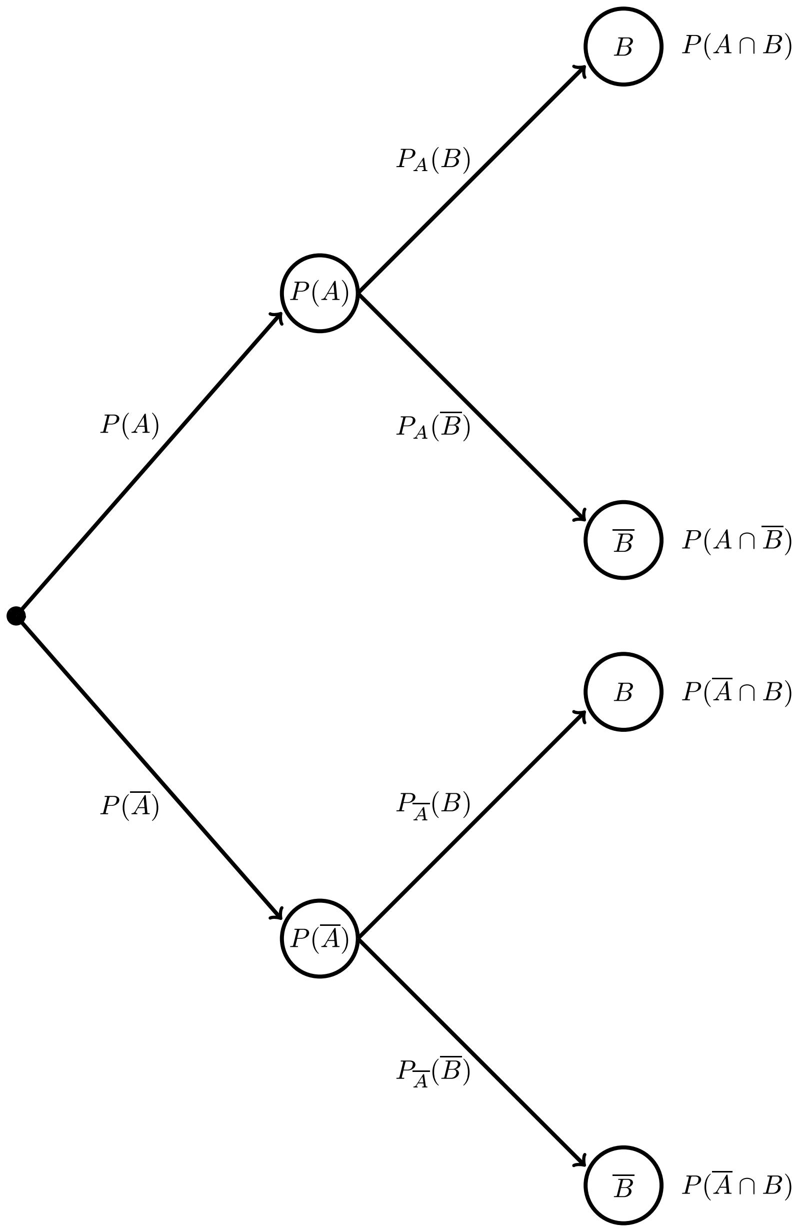
Stochastik

Marvin Baeumer **2023-12-06 10:22**

Baumdiagram

> Ein Baumdiagramm ist eine grafische Darstellung, die verwendet wird, um verschiedene mögliche Entscheidungswege oder Ergebnisse eines Ereignisses darzustellen. Es besteht aus einem Stamm, der das Ausgangsereignis darstellt, und Zweigen, die die verschiedenen möglichen Entscheidungspfade oder Ergebnisse repräsentieren.



Vierfeldertafel

> Innerhalb einer Vierfeldertafel stehen die $\cap$-Wahrscheinlichkeiten, unten und am Rand hingegen steht die Summe des jeweiligen Ereignisses, das sich mit der Summenregel berechnen lässt.

	$\mathbf{A}$	$\overline{\mathbf{A}}$	**Summe**
$\mathbf{B}$	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{\mathbf{B}}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
Summe	$P(A)$	$P(\overline{A})$	**100%**

Pfadregel

> Die Pfadregel besagt, dass man zusammenhängende Pfade eines Baumdiagramms miteinander multiplizieren kann, um somit die Wahrscheinlichkeit für ein gemeinsames Eintreten des Ereignisses zu erhalten. Dabei ist es wichtig, dass nur ein zusammenhängender Pfad betrachtet wird. $P(A) \cdot P_A(B) = P(A \cap B)$: Es wird die Summe mit der Bedingung multipliziert.

Summenregel

> Die Summenregel besagt, dass die Ergebnisse unabhängiger Pfade miteinander addiert werden dürfen, um die Summe bestimmter Ereignisse zu erhalten. Die Pfade müssen unabhängig voneinander sein, damit diese Regel angewendet werden kann. $P(A) = P(A \cap B) + P(A \cap \overline{B})$: Die Summe setzt sich zusammen aus der Addition der $\cap$-Wahrscheinlichkeiten.

Berechnung Wahrscheinlichkeiten Formeln

$$\begin{array}{ccccccc}
P(A \cap B) & = & P(A) & \cdot & P_A(B) & & \\
P_A(B) & = & \frac{P(A \cap B)}{P(A)} & & & & \\
P(A) & = & P(A \cap B) & + & P(A \cap \overline{B}) & & \\
P(A \cup B) & = & P(A) & + & P(B) & - & P(A \cap B)
\end{array}$$

Standardabweichung

> Die Standardabweichung ist ein Maß dafür, wie stark die Werte einer Datenmenge um den Durchschnitt streuen. Sie gibt an, wie weit die einzelnen Datenpunkte im Durchschnitt vom Mittelwert entfernt sind. Je größer die Standardabweichung, desto größer ist die Streuung der Daten um den Mittelwert.

Formel

$$\sigma(X) = \sqrt{n \cdot p \cdot (1 - p)}$$

```
StandardDeviation[BinomialDistribution[n, p]]
```

Erwartungswert

> Er repräsentiert den Durchschnitt oder die mittlere Wertung eines Zufallsprozesses oder einer Zufallsvariablen über eine große Anzahl von Experimenten oder Ereignissen.

$$\mu = n \cdot p$$

Fakultät

> Die Fakultät ist eine mathematische Funktion, die in der Kombinatorik oft verwendet wird. Sie wird oft verwendet, um die Anzahl der Möglichkeiten zu berechnen, wie eine bestimmte Anzahl von Objekten in einer bestimmten Reihenfolge angeordnet werden kann. Die Fakultät einer Zahl n, geschrieben als $n!$, ist das Produkt aller positiven ganzen Zahlen von 1 bis n. Mathematisch ausgedrückt:

$$n! = n \cdot (n - 1) \cdot (n - 2) \cdot \ldots \cdot 2 \cdot 1$$

Beispiel

> Angenommen, du hast 4 verschiedene Bücher (A, B, C und D), und du möchtest wissen, auf wie viele verschiedene Arten du diese Bücher auf ein Regal stellen kannst. Da die Reihenfolge wichtig ist (das Anordnen von A, B, C ist unterschiedlich von B, A, C), verwenden wir die Fakultät. Die Anzahl der Möglichkeiten, die Bücher zu arrangieren, ist $4!$.

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

> Das bedeutet, es gibt 24 verschiedene Arten, wie du die Bücher auf das Regal stellen kannst, da für das erste Buch 4 Möglichkeiten existieren, für das zweite Buch 3 Möglichkeiten, für das dritte Buch 2 Möglichkeiten und für das vierte Buch 1 Möglichkeit vorhanden ist.

Binomialkoeffizient

> Der Binomialkoeffizient ist eine wichtige mathematische Funktion, die in der Kombinatorik verwendet wird, um die Anzahl der Möglichkeiten zu berechnen, k Elemente aus einer Menge von n Elementen auszuwählen, ohne die Reihenfolge zu berücksichtigen. Er wird auch als "n über k" ausgesprochen und wird mathematisch als $\binom{n}{k}$ dargestellt. Die Formel für den Binomialkoeffizienten lautet:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

> Der Binomialkoeffizient wird oft verwendet, um die Anzahl der Kombinationen zu berechnen, die aus k Elementen aus einer Menge von n Elementen ohne Wiederholung ausgewählt werden können. Er wird auch in der Wahrscheinlichkeitstheorie verwendet, insbesondere in der Binomialverteilung, um die Anzahl der Erfolge in einer bestimmten Anzahl von unabhängigen Versuchen zu berechnen.

Beispiel

> Angenommen, du hast ein Fahrradschloss mit 4 Zahlenrädern, und jedes Rad hat 10 mögliche Zahlen (von 0 bis 9). Du möchtest die Kombination aus Zahlen wählen, um das Fahrradschloss zu öffnen. Wie viele verschiedene Kombinationen gibt es? Hier können wir den Binomialkoeffizienten verwenden, um die Anzahl der Möglichkeiten zu berechnen, die Kombination auszuwählen. Jedes Zahlenrad hat 10 mögliche Zahlen, und da die Reihenfolge der Zahlen wichtig ist, verwenden wir die Permutation. Die Anzahl der Möglichkeiten, k Zahlen aus n möglichen Optionen ohne Wiederholung auszuwählen, beträgt:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

> Für unser Beispiel ist $n = 10$ (Anzahl der möglichen Zahlen auf jedem Rad) und $k = 4$ (Anzahl der Räder). Also setzen wir die Werte in die Formel ein:

$$\binom{10}{4} = \frac{10!}{4! \cdot (10-4)!} = \frac{10!}{4! \cdot 6!} = 210$$

> Das bedeutet, es gibt 210 verschiedene Kombinationen von Zahlen, die du für das Fahrradschloss wählen kannst.

```
Binomial[n,k]
```

Binomialverteilung

> Die Binomialverteilung ist eine Wahrscheinlichkeitsverteilung, die in der Statistik verwendet wird, um die Anzahl der Erfolge in einer festen Anzahl von unabhängigen, identischen Bernoulli-Experimenten zu modellieren. Ein Bernoulli-Experiment ist ein zufälliges Experiment mit genau zwei möglichen Ergebnissen: Erfolg oder Misserfolg. Die Binomialverteilung hat zwei Parameter:
>
> 1. Die Anzahl der Versuche oder Stichproben n.
> 2. Die Wahrscheinlichkeit eines Erfolgs in jedem einzelnen Versuch p.
>
> Wenn wir X als die Anzahl der Erfolge in den n Versuchen definieren, dann ist die Wahrscheinlichkeitsfunktion der Binomialverteilung gegeben durch:

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

> Dabei steht $\binom{n}{k}$ für die Kombination n über k, auch als "n über k" bekannt, was die Anzahl der Möglichkeiten angibt, k Erfolge in n Versuchen zu haben. Die Formel $p^k \cdot (1-p)^{n-k}$ gibt die Wahrscheinlichkeit an, dass in einer bestimmten Reihenfolge k Erfolge und $n-k$ Misserfolge auftreten.
>
> Die Binomialverteilung ist wichtig, um die Wahrscheinlichkeit zu berechnen, eine bestimmte Anzahl von Erfolgen in einer festen Anzahl von Versuchen zu erhalten, wenn die Erfolgswahrscheinlichkeit p bekannt ist. Sie findet in vielen Bereichen Anwendung, wie z.B. in der Qualitätssicherung, bei Umfragen, in der Medizin, in der Finanzanalyse und vielen anderen Bereichen, in denen wiederholbare zufällige Experimente auftreten.

Berechnung in Mathematica wenn alle Werte bekannt sind

```
Probability[X >= Bedingung, X \[Distributed] BinomialDistribution[n, p]]
```

Berechnung in Mathematica wenn ein Wert fehlt

```
Solve[Probability[X >= Bedingung, X \[Distributed]
BinomialDistribution[n, p]]> P]
(* oder *)
Reduce[Probability[X >= Bedingung, X \[Distributed]
BinomialDistribution[n, p]]> P]
```

> Meistens fehlt n oder p. Wenn dies der Fall ist, berechnet man mit `Solve` die Wahrscheinlichkeit. Meistens hat man aber noch eine dritte Variable P, die das Gesamtereignis beschreibt.

Dokumentation

$$\begin{array}{rcc} X \sim B & & (n, p) \\ P(X = k) & = & p \end{array}$$

Signifikanzniveau

> Das Signifikanzniveau beschreibt die Wahrscheinlichkeit, die wir bereit sind einzugehen, um einen Fehler 1. Art zu machen, also die Nullhypothese zu verwerfen, obwohl sie stimmt. Das Signifikanzniveau wird mit α dargestellt und bei einem zweiseitigen Test auf beide Seiten aufgeteilt, das heißt, pro Seite gilt $\frac{\alpha}{2}$. Oft ist das Signifikanzniveau $0,05$ oder $0,01$.

Fehler 1. Art

> Ein Fehler 1. Art tritt auf, wenn die Nullhypothese fälschlicherweise abgelehnt wird, obwohl sie tatsächlich wahr ist. Die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, wird als Signifikanzniveau α festgelegt. Ein typischer Wert für das Signifikanzniveau ist $0,05$, was bedeutet, dass wir bereit sind, eine 5% Wahrscheinlichkeit zu akzeptieren, einen Fehler 1. Art zu machen.

Fehler 2. Art

> Ein Fehler 2. Art tritt auf, wenn die Nullhypothese nicht abgelehnt wird, obwohl sie tatsächlich falsch ist (d. h., die Alternativhypothese wahr ist).

Hypothesentest 2 Seitig

> Ein zweiseitiger Hypothesentest wird verwendet, um zu prüfen, ob ein Parameter signifikant von einem bestimmten Wert abweicht, ohne eine Richtung anzugeben. Mit anderen Worten, er untersucht, ob es einen Unterschied gibt, aber nicht, ob der Parameter größer oder kleiner ist als der festgelegte Wert.

Beispiel

> Ein Restaurant möchte herausfinden, ob sich die Vorlieben seiner Kunden hinsichtlich der gewählten Speisen geändert haben. In der Vergangenheit bestellte etwa ein Drittel der Kunden vegetarische Gerichte. Nach Einführung einer neuen Speisekarte mit dem Schwerpunkt auf veganen Gerichten wird eine Umfrage unter 150 Kunden durchgeführt. Das Restaurant möchte herausfinden, ob sich der Anteil der Kunden, die vegetarische Gerichte bevorzugen, signifikant verändert hat.

Vorgehen:

1. Null- und Alternativhypothese formulieren.
2. Die oberen und unteren Grenzen $[a, b]$ berechnen.
3. Entscheidung treffen durch Vergleich des Bereichs $[a, b]$.

Daten

$a \to 0,05$ (gegeben)

$n \to 150$

$p \to \frac{1}{3}$

Formulierung der Hypothesen:

> Nullhypothese (H_0): Die Wahrscheinlichkeit bleibt bei $\frac{1}{3}$, der Anteil der Kunden, die vegetarische Gerichte bevorzugen, hat sich nicht verändert.
>
> Alternativhypothese (H_1): Die Wahrscheinlichkeit ist ungleich $\frac{1}{3}$, der Anteil der Kunden, die vegetarische Gerichte bevorzugen, hat sich verändert.

Rechnung

$$P(X \leq a) \geq \frac{5}{2} = 38$$
$$P(X \geq b) \geq \frac{5}{2} = 62$$
$$a \to 38$$
$$b \to 62$$

> Die $\frac{5}{2}$ ergeben sich durch das Signifikanzniveau ÷ 2

```
(* fuer a *)
TableForm[
 Table[{a,
   N[Probability[X <= a,
     X \[Distributed] BinomialDistribution[150, 1/3]]]}, {a, 38, 38}]]

(* fuer b *)
TableForm[
 Table[{b,
   N[Probability[X >= b,
     X \[Distributed] BinomialDistribution[150, 1/3]]]}, {b, 62, 62}]]
```

> Bei diesem Vorgehen erstellt man eine Tabelle mit n und p Binomialverteilt. Dann sucht man den Bereich, in dem die Wahrscheinlichkeit $= 0,025$ beträgt oder am nächsten an $0,025$, aber darunter liegt, entsprechend dem Signifikanzniveau. In unserem Beispiel ergäben sich die Grenzen $a \to 38$ und $b \to 62$. Die Antwort auf unsere Nullhypothese wäre nun: Wenn die Anzahl der Kunden vorher zwischen 38 und 62 lag, hat sich nichts geändert an der Wahrscheinlichkeit $\frac{1}{3}$. Aber sagt uns der Restaurantbesitzer, es seien jetzt 100 Leute, können wir die Nullhypothese ablehnen und sagen, die Wahrscheinlichkeit hat sich verändert.

Hypothesentest 1 Seitig

> Ein einseitiger Hypothesentest lässt sich in einen linksseitigen und einen rechtsseitigen Test unterteilen. Bei einem rechtsseitigen Test möchte man prüfen, ob die Wahrscheinlichkeit der Nullhypothese größer geworden ist, also $H_0 : p < p_0$ oder $H_1 : p > p_0$. Bei einem linksseitigen Test möchte man prüfen, ob die Wahrscheinlichkeit der Nullhypothese kleiner geworden ist, also $H_0 : p > p_0$ oder $H_1 : p < p_0$. Um die Berechnung durchzuführen, für die obere und untere Grenze benötigen wir die Sigma-Regeln. Also gilt für $[a, b] = \mu \pm 2 \cdot \sigma$.
>
> **Wichtig** Das Signifikanzniveau bleibt auf einer Seite.

$2 \cdot \sigma$-Regel

> Die Zwei-Sigma-Regel umfasst einen Verteilungsbereich um den Erwartungswert herum. Bei einem Signifikanzniveau von 5% kann man die $2 \cdot \sigma$ -Regel anwenden, da diese einen Bereich von 95% umfasst. Sie wird berechnet, indem man die Standardabweichung mit 2 multipliziert.

Beispiel

> Angenommen, ein Schraubenhersteller behauptet, dass 90% seiner hergestellten Schrauben den Qualitätsstandard erfüllen. Um dies zu überprüfen, führen wir einen Hypothesentest durch. Wir untersuchen 100 Schrauben und bekommen 88 Heile schrauben geliefert.

Vorgehen

1. Null- und Alternativhypothese formulieren.
2. Erwartungswert berechnen
3. Standardabweichung berechnen
4. Berechnung der Ober- und Untergrenze
5. Entscheidung treffen ob $p > p_0$ oder $p < p_0$

Hypothesen

> Nullhypothese (H_0) : Die Garantie ist über $90\% \rightarrow p < p_0$
>
> Alternativhypothese (H_1) : Die Garantie ist unter $90\% \rightarrow p > p_0$

Rechnung

$$\begin{aligned}
\mu &= n \cdot p \\
\Leftrightarrow 100 \cdot 0.9 &= 90 \\
\mu &= 90 \\
\\
\sigma &= \sqrt{n \cdot p \cdot (1 - 0.9)} \\
\Leftrightarrow &= \sqrt{100 \cdot 0,9 \cdot (1 - 0.9)} \\
\sigma &= 3 \\
\\
& 2 \cdot \sigma \text{ Regel} \\
\\
a &= 90 - 6 \\
a &= 84 \\
b &= 90 + 6 \\
b &= 96
\end{aligned}$$

Entscheidung

> Wir lehnen die Nullhypothese ab, da wir 90 heile Schrauben erwarten, aber nur 88 haben. Somit befinden wir uns im Bereich, um die Nullhypothese ablehnen zu können. Außerdem wissen wir, dass die Garantie weniger als 90% beträgt.