C1 - Stochastik

Aufgaben PLUS Tipps PLUS Lösungen PLUS

1.

1.1 ▶ **Zufallsvariable X definieren**

Du sollst die Zufallsvariable X definieren. Dies bedeutet, dass du angeben sollst, welche Größe von X beschrieben wird. Aus anderen Aufgaben, weißt du, dass X meistens die Anzahl der Gewinne beschreibt.
In unserem Fall beschreibt X die Anzahl der positiven Dopingtests.

▶ **Bedingungen für eine Binomialverteilung angeben**

Du kennst zwei Bedingungen, damit eine Zufallsvariable X als **binomialverteilt** angenommen werden kann:

1. Bei jedem Durchgang müssen die Wahrscheinlichkeiten für einen „Gewinn" gleich bleiben
2. Es darf nur zwei mögliche Ausgänge geben, beispielsweise „Gewinn" und „Niete"

Nun musst du diese Bedingungen nur noch auf den konkreten Sachzusammenhang übertragen.
Im Falle der Dopingtests bedeutet das konkret, X kann als binomialverteilt angenommen werden, wenn die folgenden Bedingungen erfüllt sind:

1. Bei jedem Spieler ist die Wahrscheinlichkeit eines positiven Dopingtests gleich
2. Es gibt nur zwei Möglichkeiten: „Der Test ist positiv" und „Der Test ist negativ"

1.2 ▶ **Bedeutung der Rechnung im Sachzusammenhang beschreiben**

Du sollst die Bedeutung der folgenden Rechnung im Sachzusammenhang beschreiben:

$$\sum_{i=10}^{19}\binom{5.000}{i}\cdot 0,003^i\cdot 0,997^{5.000-i}\approx 0,806$$

Du kennst bereits die Formel für die summierte Binomialverteilung mit den Parametern n und p:

$$\boxed{P(l\leq X\leq m)=\sum_{i=l}^{m}\binom{n}{i}\cdot p^i\cdot(1-p)^{n-i}}$$

Mit Hilfe dieser Formel lässt sich die Wahrscheinlichkeit dafür berechnen, dass die Zufallsvariable X einen Wert aus dem Intervall $[l,\dots,m]$ annimmt, also die Wahrscheinlichkeit für mindestens l und höchstens m Gewinne.
Nun sollte dir die Ähnlichkeit auffallen.
In unserem Fall kann X als binomialverteilt mit den Parametern $n=5.000$ und $p=0,003$ angenommen werden.
Die obige Rechnung gibt also an, dass die Wahrscheinlichkeit für mindestens 10 und höchstens 19 positive Dopingtests unter insgesamt 5.000 Tests ca. $0,806$ beträgt unter der Annahme, dass die Wahrscheinlichkeit, dass ein Test positiv ist $0,3\,\%$ beträgt.

2.

2.1 ▶ **Annahmebereich des französischen Labors berechnen**

Das französische Labor führt folgenden Test durch:

- Nullhypothese: $p_0=0,01$
- Gegenhypothese: $p_1<0,01$
- Signifikanzniveau: $\alpha=0,04$

Dabei ist der Stichprobenumfang nun $n=1.000$.
Der Annahmebereich gibt ein Intervall an, in welchem die Anzahl der positiven Tests liegen darf, wenn unter einem Signifikanzniveau von $0,04$ die Nullhypothese stimmen soll.
Liegt also die Anzahl der positiven Doping-Tests bei dem Test mit $n=1.000$ innerhalb des Annahmebereichs, so nehmen die französischen Forscher die Nullhypothese an, ansonsten lehnen sie diese ab.

Überlege dir nun zunächst welche Art von Hypothesentest durchgeführt wird.
Hierfür gibt es drei Möglichkeiten, welche du jeweils anhand der Formulierung der Gegenhypothese unterscheiden kannst:

- **linksseitiger Test**: wenn die Gegenhypothese lautet: $p_1 < p_0$
- **rechtsseitiger Test**: wenn die Gegenhypothese lautet: $p_1 > p_0$
- **zweiseitiger Test**: wenn die Gegenhypothese lautet: $p_1 \neq p_0$

Anschließend kannst du den Annahmebereich berechnen.

▶ 1. Schritt: Art des Tests auswählen

Du kannst sehen, dass die Nullhypothese genau dann abgelehnt wird, wenn die Anzahl der positiven Dopingtests signifikant zu klein ist.
Daher führt das französische Labor einen **linksseitigen Hypothesentest** durch.

▶ 2. Schritt: Annahmebereich bestimmen

Da das französische Labor einen linksseitigen Hypothesentest durchführt, hat der Annahmebereich die folgende Form:

$$\boxed{A = \{k, \ldots, 1.000\}}$$

Du suchst also das kleinste k, für das die Nullhypothese gerade noch angenommen wird. Dies ist das kleinste k, sodass die Wahrscheinlichkeit dafür, dass X kleiner als k ist gerade noch das Signifikanzniveau $0,04$ ist, also das k, welches die folgende Ungleichung erfüllt, wobei hier die Wahrscheinlichkeit $p = 0,01$ gelten soll.

$$P(X \leq k) \geq 0,04$$

Dieses k kannst du mit Hilfe der Tabelle im Anhang ermitteln und erhältst so:

$P(X \leq 4) = 0,0287 < 0,04$ und $P(X \leq 5) = 0,0661 > 0,04$

Damit ergibt sich der folgende Annahmebereich:

$$A = \{5, \ldots, 1.000\}$$

▶ Annahmebereich des Heidelberger Molekularbiologen berechnen

Im Gegensatz zum französischen Labor, führt der Molekularbiologe folgenden Test durch:

- Nullhypothese: $p_0 = 0,003$
- Gegenhypothese: $p_1 > 0,003$
- Signifikanzniveau: $\alpha = 0,04$

Er führt demnach einen **rechtsseitigen Hypothesentest** durch. Bei einem solchen, hat der Annahmebereich folgende Form:

$$\boxed{A = \{0, \ldots, k\}}$$

In diesem Fall suchst du demnach das größte k, sodass die Wahrscheinlichkeit, dass X größer als k ist gerade noch das Signifikanzniveau $0,04$ beträgt, also das k, für das die folgende Ungleichung erfüllt ist:

$$P(X \geq k) \geq 0,04 \Leftrightarrow 1 - P(X \leq k-1) \geq 0,04 \Leftrightarrow P(X \leq k-1) \leq 0,96$$

Diese kannst du wieder aus der Tabelle ablesen. Achte darauf, dass in diesem Fall die Wahrscheinlichkeit $p = 0,003$ angenommen wird. Du musst hier also in der anderen Spalte der Tabelle nachschauen. Dann erhältst du:

$P(X \geq 7) = 1 - 0,96672 = 0,03328 < 0,04$ und
$P(X \geq 6) = 1 - 0,91639 = 0,08361 > 0,04$

Damit ergibt sich der Annahmebereich mit:

$$A = \{0, \ldots, 6\}$$

▶ Neues Testverfahren beurteilen

Das neue Testverfahren hat unter 1.000 Teilnehmern 6 Sportlerinnen des Dopings überführt. Betrachtet man den Test des französischen Labors, so würden diese ihre Nullhypothese bei diesem Ergebnis annehmen. Sie würden also weiterhin von einer Wahrscheinlichkeit von $1\,\%$ ausgehen.
Nach dem Test des Molekularbiologen in Heidelberg, würde dessen Nullhypothese angenommen werden, da

6 im Annahmebereich liegt. Er würde sich also in seiner Vermutung ebenfalls bestätigt sehen, dass die Wahrscheinlichkeit immernoch bei $0,3\,\%$ liegt.
Insgesamt, gibt der Test bei dieser Stichprobe also beiden Recht und ist somit nicht aussagekräftig.

2.2 ► **Fehler 1. und 2. Art beschreiben**

Hier hast du zwei Möglichkeiten, entweder du wählst den Test des französischen Labors oder den des heidelberger Molekularbiologen, um den Fehler 1. und 2. Art zu beschreiben. Allgemein ergeben sich die Fehler 1. und 2. Art wie folgt:

- **Fehler 1. Art:** Die Nullhypothese wird fälschlicherweise abgelehnt.
- **Fehler 2. Art:** Die Nullhypothese wird fälschlicherweise angenommen.

►► Lösungsweg A: Test des französischen Labors

Im Falle des Französischen Labors bedeutet der **Fehler 1. Art** also, wenn weniger als 5 positive Dopingtests vorliegen, lehnen sie ihre Nullhypothese ab, obwohl diese womöglich tatsächlich zutrifft. Sie gingen also davon aus, dass ihre neue Testmethode gar nicht besser ist, als die alte, obwohl sie es in Wahrheit womöglich doch ist.
Der **Fehler 2. Art** kann passieren, wenn mehr als 4 Dopingtests positiv ausfallen. Dann würden die französischen Forscher ihre Hypothese, dass ihr neues Testverfahren besser ist, annehmen, obwohl vielleicht immer noch die alte Trefferquote gilt.

►► Lösungsweg B: Test des heidelberger Molekularbiologen

Im Falle des heidelberger Molekularbiologen kann der **Fehler 1. Art** auftreten, wenn mehr als 6 positive Dopingtests vorkommen. In diesem Fall würde der Molekularbiologe seine Nullhypothese, dass die französischen Forscher nicht Recht haben, ablehnen, obwohl diese vielleicht tatsächlich nicht richtig liegen.
Der **Fehler 2. Art** könnte auftreten, wenn weniger als 7 Dopingtests positiv sind. Dann bestünde der Fehler darin, dass der Molekularbiologe fälschlicherweise davon ausginge, dass die Trefferquote des neuen Testverfahrens nicht besser ist, obwohl sie vielleicht tatsächlich besser wäre.

► Fehler 2. Art berechnen

Wie du eben gesehen hast, besteht der Fehler 2. Art darin die Nullhypothese fälschlicherweise anzunehmen, obwohl eigentlich eine andere Wahrscheinlichkeit gilt. Der Fehler 2. Art ist also die Wahrscheinlichkeit dafür, dass das Ergebnis innerhalb des berechneten Annahmebereichs liegt, wenn die reale Wahrscheinlichkeit p_r gilt. Er ergibt sich also durch:

$$\boxed{P_{p_r}(X \in A)}$$

► Fehler 2. Art des französischen Labors berechnen

Du sollst hier annehmen, dass in der Realität eigentlich die Nullhypothese des Molekularbiologen gilt, also: $p_r = 0,003$.
Berechne nun $P_{0,003}(X \geq 5) = 1 - P_{0,003}(X \leq 4)$. Dies kannst du wieder aus der Tabelle im Anhang ablesen. Achte dabei wieder darauf, in der Spalte für $p = 0,003$ nachzuschauen.
Dann erhältst du das Ergebnis:

$$P_{0,003}(X \geq 5) = 1 - 0,81552 = 0,18448 = 18,448\,\%$$

Der Fehler 2. Art, also die Wahrscheinlichkeit fälschlicherweise davon auszugehen, dass die neue Methode bessere Ergebnisse liefert wenn eigentlich die Behauptung des Molekularbiologen aus Heidelberg stimmt, beträgt für das französische Labor ca. $18,448\,\%$.

► Fehler 2. Art des Molekularbiologen berechnen

Hierfür sollst du nun annehmen, dass tatsächlich die Nullhypothese des französischen Labors gilt, also $p_r = 0,01$.
Du suchst also:

$$P_{0,01}(X \leq 6)$$

Dies kannst du wieder aus der Tabelle im Anhang ablesen und erhältst so:

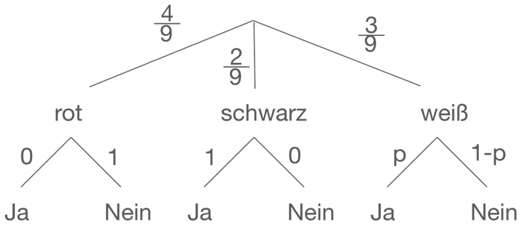
$$P_{0,01}(X \leq 6) = 0,1289 = 12,89\,\%$$

Die Wahrscheinlichkeit für einen Fehler 2. Art, also davon auszugehen, dass das neue Testverfahren nicht besser ist, obwohl eigentlich die neue und bessere Trefferquote des französischen Labors gilt, liegt für den heidelberger Molekularbiologen bei ca. $12,89\,\%$.

3.

3.1 ▶ **Baumdiagramm zeichnen**

Du sollst nun das Testverfahren der Medizinstudenten mit Hilfe eines Baumdiagramms darstellen. Überlege dir dazu zunächst, wie viele Ebenen der Baum haben muss und welche Ebene welchen Teil des Experiments darstellen soll.
Insgesamt liegt hier ein **zweistufiges Experiment** vor. Die erste Stufe besteht darin, zunächst aus der Urne eine Kugel zu ziehen. In der nächsten Stufe wird dann die Antwort gegeben.
Daraus ergibt sich also ein Baumdiagramm mit zwei Ebenen, p entspricht dabei der Wahrscheinlichkeit eines Sportlers, Dopingmittel zu nutzen.



▶ **Anteil der gedopten Athleten ermitteln**

Du weißt, dass insgesamt 1.093 von 3.010 getesteten Athleten mit „Ja“ geantwortet haben.
Hier kannst du mit der **Pfadadditionsregel** arbeiten. Du weißt, dass sich die Wahrscheinlichkeit für die Antwort „Ja“ aus der Summe der Wahrscheinlichkeiten aller Pfade ergibt, die zu der Antwort „Ja“ führen.
Die Wahrscheinlichkeiten der einzelnen Pfade kannst du mit Hilfe der **Pfadmultiplikationsregel** ermitteln.
So kannst du eine Gleichung in Abhängigkeit von p aufstellen und diese nach p lösen.
Es ergibt sich die Gleichung:

$$P(„Ja“) = \tfrac{2}{9}\cdot 1 + \tfrac{3}{9}\cdot p + \tfrac{4}{9}\cdot 0$$

Da $P(„Ja“)$ mit dem Anteil der „Ja“-Antworten der Stichprobe angenähert werden kann, ergibt sich:

$$\begin{aligned}
P(„Ja“) &= \frac{2}{9}\cdot 1 + \frac{3}{9}\cdot p \\
\frac{1.093}{3.010} &= \frac{2}{9} + \frac{1}{3}\cdot p && \mid -\tfrac{2}{9} \\
\frac{3.817}{27.090} &= \frac{1}{3}\cdot p && \mid \cdot 3 \\
0,4227 &\approx p
\end{aligned}$$

Der Anteil der Dopingsünder beträgt ca. $42,27\,\%$.

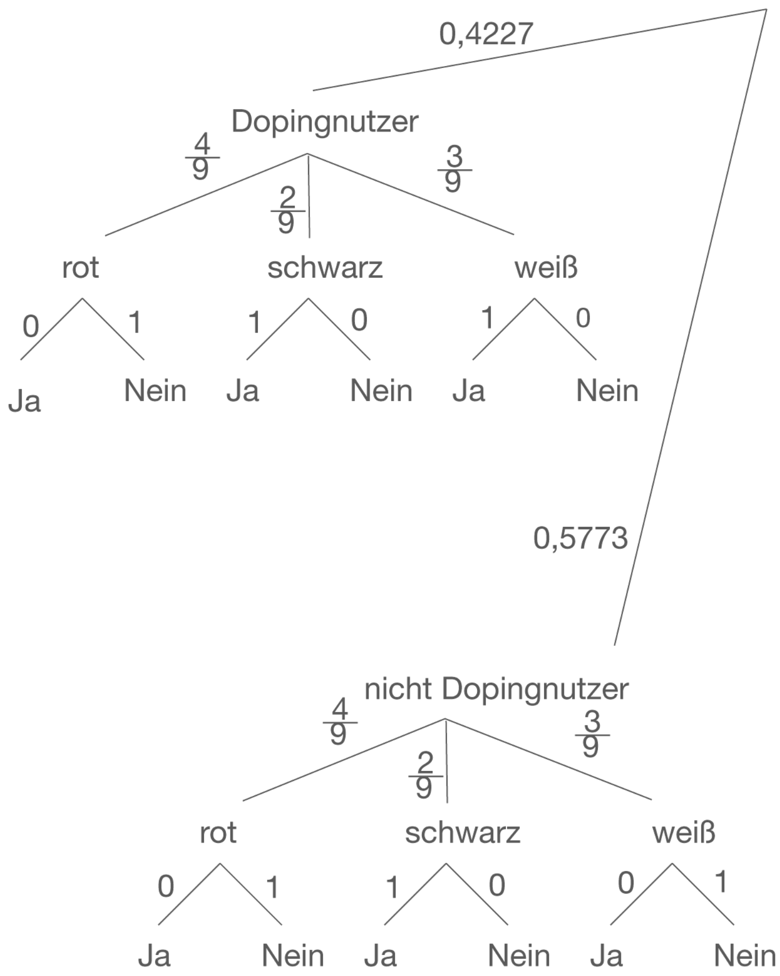
3.2 ▶ **Wahrscheinlichkeit für die richtige Antwort „Ja“ ermitteln**

Nun dreht sich das Experiment sozusagen um. Es geht nun darum, wie groß die Wahrscheinlichkeit dafür ist, dass ein Sportler, der mit „Ja“ geantwortet hat, auch tatsächlich gedopt hat. Du kennst inzwischen den Anteil derjenigen, die dopen: $p = 0,4227$.
Du kannst hier auf zwei verschiedene Arten vorgehen:

- **Lösungsweg A:** Berechne die Wahrscheinlichkeit mit Hilfe des Baumdiagramms
- **Lösungsweg B:** Nutze die Formel für die bedingte Wahrscheinlichkeit bzw. Die **Formel von Bayes**

▶▶ **Lösungsweg A: Baumdiagramm**

Zur Hilfe kannst du hier ein neues Baumdiagramm zeichnen, indem du eine Ebene oben einfügst, die beschreibt ob ein Spieler dopt oder nicht:



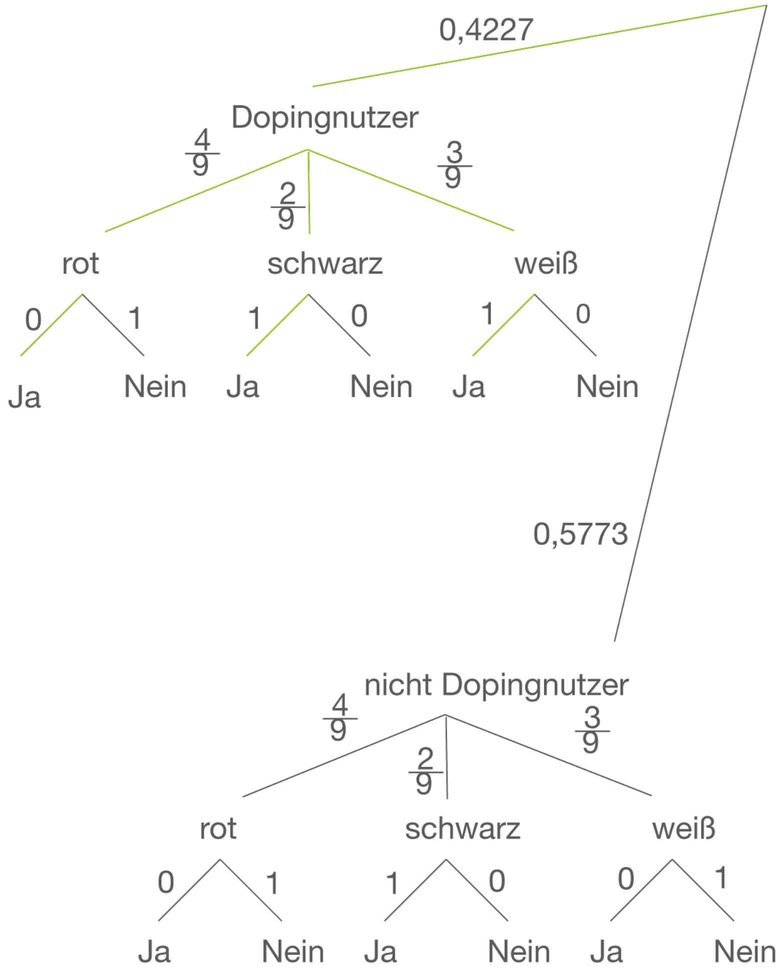
Diese Wahrscheinlichkeit ergibt sich dann, indem du alle Wahrscheinlichkeiten der Pfade des Baums addierst, bei denen ein Spieler tatsächlich dopt und auch mit „Ja“ antwortet. Diese teilst du durch die Summe aller Wahrscheinlichkeiten der Pfade bei denen ein Spieler mit „Ja“ antwortet:

$$P_{„Ja“}(„\text{gedopt}“) = \frac{P(„\text{gedopt}“ \cap „\text{Ja}“)}{P(„\text{Ja}“)}$$

Berechne die beiden benötigten Werte zunächst einzeln.

▶ 1. Schritt: $P(„\text{gedopt}“ \cap „\text{Ja}“)$ berechnen

$P(„\text{gedopt}“ \cap „\text{Ja}“)$ ergibt sich dabei durch die grünen Pfade im unten stehenden Baumdiagramm mit Hilfe der beiden **Pfadregeln**:

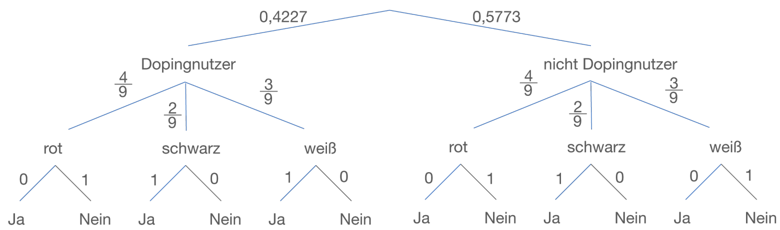


Dadurch erhältst du dann das Ergebnis:

$$\begin{aligned} P(\text{„ gedopt“} \cap \text{„ Ja“}) &= P(\text{„dopt“-„rot“-„Ja“}) + P(\text{„dopt“-„schwarz“-„Ja“}) + P(\text{„dopt“-„weiß“-„Ja“}) \\ &= 0,4227 \cdot \tfrac{4}{9} \cdot 0 + 0,4227 \cdot \tfrac{2}{9} \cdot 1 + 0,4227 \cdot \tfrac{3}{9} \cdot 1 \\ &\approx 0,235 \end{aligned}$$

▶ 2. Schritt: $P(\text{„ }Ja\text{“})$ berechnen

Berechne nun noch die gesamt Wahrscheinlichkeit für eine Antwort „Ja“, diese ergibt sich durch die blauen Pfade in der unten stehenden Abbildung:



Dadurch ergibt sich:

$$\begin{aligned} P(\text{„Ja“}) &= P_{\text{„}gedopt\text{“}}(\text{„Ja“}) + P_{\text{„}nichtgedopt\text{“}}(\text{„Ja“}) \\ &= 0,235 + 0,5773 \cdot \tfrac{4}{9} \cdot 0 + 0,5773 \cdot \tfrac{2}{9} \cdot 1 + 0,5773 \cdot \tfrac{3}{9} \cdot 0 \\ &\approx 0,363 \end{aligned}$$

▶ 3. Schritt: Gesamtergebnis ausrechnen

Nun musst du nur noch die berechneten Ergebnisse in die Formel einsetzen und erhältst so:

$$\begin{aligned} P_{\text{„Ja“}}(\text{„gedopt“}) &= \frac{P(\text{„ gedopt“} \cap \text{„ Ja“})}{P(\text{„Ja“})} \\ &= \frac{0,235}{0,363} \\ &\approx 0,647 \\ &= 64,7\,\% \end{aligned}$$

►► Lösungsweg B: Formel von Bayes

Die Formel von Bayes lautet wie folgt:

$$\boxed{P_B(A) = \frac{P_A(B) \cdot P(A)}{P(B)}}$$

Da du nun die Wahrscheinlichkeit $P_{\text{„Ja“}}(\text{„gedopt“})$ suchst, lautet die Formel übertragen auf den Sachverhalt wie folgt:

$$P_{\text{„Ja“}}(\text{„gedopt“}) = \frac{P_{\text{„gedopt“}}(\text{„Ja“}) \cdot P(\text{„gedopt“})}{P(\text{„Ja“})}$$

Du kannst nun beide Werte, die dir noch fehlen, getrennt und zum Schluss das Gesamtergebnis berechnen.

► 1. Schritt: $P_{\text{„gedopt“}}(\text{„Ja“})$

Überlege dir, wie groß die Wahrscheinlichkeit dafür ist, dass ein Spieler, der dopt, mit „Ja“ antwortet. Dafür gibt es nur zwei Möglichkeiten:

1. Der gedopte Spieler zieht eine schwarze Kugel aus der Urne
2. Der gedopte Spieler zieht eine weiße Kugel aus der Urne

In beiden Fällen antwortet er zu $100\,\%$ mit „Ja“. Zieht er eine rote Kugel, so antwortet er auf jeden Fall mit „Nein“. Dabei ist

$$P(A) = \frac{2}{9} \text{ und } P(B) = \frac{3}{9}.$$

Wegen der **Pfadadditionsregel** ergibt sich damit:

$$P_{\text{„gedopt“}}(\text{„Ja“}) = P(A) + P(B) = \tfrac{2}{9} + \tfrac{3}{9} = \tfrac{5}{9}$$

► 2. Schritt: $P(\text{„Ja“})$

Die Wahrscheinlichkeit, dafür, dass ein beliebiger Spieler mit „Ja“ antwortet, kannst du mit Hilfe des Baumdiagramms aus dem vorherigen Aufgabenteil und den **Pfadregeln** berechnen, indem du die Wahrscheinlichkeiten aller Pfade addierst, die zu „Ja“ führen.
Für p kannst du nun die berechnete Wahrscheinlichkeit $p = 0,4227$ einsetzen. Dann erhältst du das Ergebnis:

$$\begin{aligned} P(\text{„Ja “}) &= \tfrac{2}{9} \cdot 1 + \tfrac{3}{9} \cdot 0,4227 \\ &\approx 0,363 \end{aligned}$$

► 3. Schritt: Gesamtergebnis berechnen

Setze nun die berechneten Ergebnisse in die Formel von Bayes ein:

$$\begin{aligned} P_{\text{„Ja“}}(\text{„gedopt“}) &= \frac{P_{\text{„gedopt“}}(\text{„Ja“}) \cdot P(\text{„gedopt“})}{P(\text{„Ja“})} \\ &= \frac{\frac{5}{9} \cdot 0,4227}{0,363} \\ &\approx 0,647 \\ &= 64,7\,\% \end{aligned}$$

Etwa $64,7\,\%$ der Sportler, die mit „Ja“ antworten, nutzen tatsächlich Dopingmittel.