7 Wahrscheinlichkeitsverteilungen – Lösungshinweise

Aufgabe 7.1: Gegeben sei ein Würfel, der die Form eines Tetraeders hat und die Augenzahlen 1 bis 4 aufweist.

a) Bestimmen Sie die Wahrscheinlichkeitsverteilung X: *Augensumme* für den zweifachen Wurf des Tetraeders.

b) Bestimmen Sie Erwartungswert, Varianz und Standardabweichung von X.

Der Doppelwurf des Tetraeders wird nun 100 Mal durchgeführt. Es wird die Anzahl der Augensumme 6 betrachtet

c) Begründen Sie, dass X: *Anzahl der Augensumme 6* als binomialverteilte Zufallsgröße modelliert werden kann.

d) Bestimmen Sie Erwartungswert, Varianz und Standardabweichung für X_i: *Augensumme 6 im i-ten Doppelwurf.*

e) Bestimmen Sie Erwartungswert, Varianz und Standardabweichung für X: *Augensumme 6* für die 100 Doppelwürfe.

f) Zeichnen Sie die Wahrscheinlichkeitsverteilung von X.

g) Ab welcher Anzahl der Augensumme 6 würden Sie an dem Modell für p, das Sie oben aufgestellt haben, zweifeln?

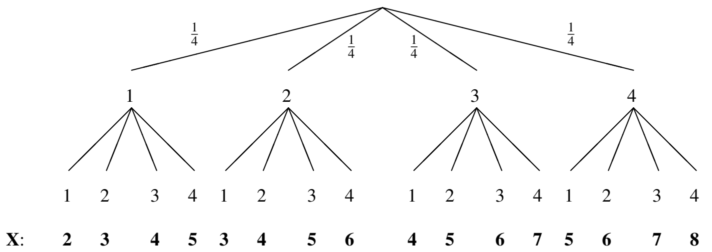
h) Bestimmen Sie die Wahrscheinlichkeit für das Intervall $[E(X)-\sigma; E(X)+\sigma]$.

i) Bestimmen Sie ein symmetrisch zum Erwartungswert von X konstruiertes Intervall $I=[a,b]$ so, dass die Wahrscheinlichkeit für $P(a \leq X \leq b) \approx 0,95$ ist.

Lösungsskizze Aufgabe 7.1 Für den Aufgabenteil a) verwenden wir einen Baum (nächste Seite). An den Pfaden des Baums ist zusätzlich der Wert der Zufallsgröße, also die Augensumme beider Würfe angetragen. Mit dem Baum gewinnt man folgende Wahrscheinlichkeitsverteilung der Zufallsgröße X:

x	2	3	4	5	6	7	8	Summe
$P(X=x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	
Produkt	$\frac{2}{16}$	$\frac{6}{16}$	$\frac{12}{16}$	$\frac{20}{16}$	$\frac{18}{16}$	$\frac{14}{16}$	$\frac{8}{16}$	$\frac{80}{16}=5$

Die Summe der Produkte der Werte der Zufallsgröße und der Wahrscheinlichkeiten dieser Werte ergibt den Erwartungswert der Zufallsgröße X: $E(X)=\sum_{i=1}^{n} x_i \cdot P(X=x_i)=5$.



In ähnlicher Weise wie oben den Erwartungswert bestimmen wir im Folgenden ausgehend von der Wahrscheinlichkeitsverteilung von X die Varianz von X:

x	2	3	4	5	6	7	8	Summe
$P(X=x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	
Quadrate	$(2-5)^2=9$	4	1	0	1	4	9	
Produkte	$\frac{9}{16}$	$\frac{8}{16}$	$\frac{3}{16}$	0	$\frac{3}{16}$	$\frac{8}{16}$	$\frac{9}{16}$	$V(X)=\frac{40}{16}=2{,}5$

Aus der Varianz von X folgt unmittelbar $\sigma(X)=\sqrt{V(X)}\approx 1{,}58$.

Da es nicht plausibel ist, den Würfeln ein Gedächtnis zuzusprechen, und auch keine andere Gründe für eine gegenseitige Beeinflussung der Doppelwürfe sprechen, ist das Modell der stochastischen Unabhängigkeit für die Hintereinanderausführung der Doppelwürfe geeignet. Da bei der Einschränkung der möglichen Ereignisse auf die beiden Ereignisse E: es fällt eine 6 und $\overline{E}$: es fällt keine 6, ist auch das Kriterium des Bernoulli-Experiments erfüllt. Nach Satz 13 (siehe Buch) ist damit X binomialverteilt.

Für jeden der Doppelwürfe, den ersten den zweiten und ebenso auch für den i-ten gilt die folgende Wahrscheinlichkeitsverteilung und die daraus resultierende Berechnung des Erwartungswerts und der Varianz von X_i: Anzahl der Sechsen im i-ten Doppelwurf:

	k	0	1	Summe
	$P(X_i=k)$	$1-p=\frac{13}{16}$	$p=\frac{3}{16}$	
Produkte I	$k\cdot P(X_i=k)$	0	$\frac{3}{16}$	$E(X_i)=\frac{3}{16}$
Quadrate	$(k-E(X_i))^2$	$\left(0-\frac{3}{16}\right)^2=\frac{9}{256}$	$\left(1-\frac{3}{16}\right)^2=\frac{169}{256}$	
Produkte II	$(k-E(X_i))^2\cdot P(X_i=k)$	$\frac{117}{4096}$	$\frac{507}{4096}$	$\frac{39}{256}\approx 0{,}152$

Aus $V(X_i)=\frac{39}{256}$ ergibt sich unmittelbar $\sigma(X_i)=\sqrt{V(X_i)}=\frac{\sqrt{39}}{16}\approx 0{,}390$.

Mit dem Wissen, dass X und damit auch X_i binomialverteilt sind, ließen sich die Parameter von X_i auch mit Hilfe der Formeln für die Binomialverteilung berechnen:

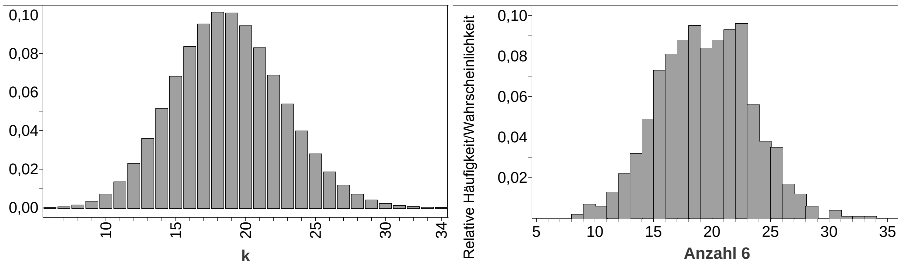
- $E(X_i)=n\cdot p=1\cdot\frac{3}{16}=\frac{3}{16}$

- $V(X_i) = n \cdot p \cdot (1-p) = 1 \cdot \frac{3}{16} \cdot \frac{13}{16} = \frac{39}{256}$

Für 100fachen Doppelwurf erhält man analog:

- $E(X) = n \cdot p = 100 \cdot \frac{3}{16} = \frac{300}{16} = 18{,}75$
- $V(X) = n \cdot p \cdot (1-p) = 100 \cdot \frac{3}{16} \cdot \frac{13}{16} = \frac{3900}{256} \approx 15{,}23$
- $\sigma(X) = \sqrt{V(X)} \approx 3{,}90$

Unten ist einerseits die Wahrscheinlichkeitsverteilung für X, andererseits eine Simulation mit 1000 Durchgängen von je 100 Doppelwürfen, bei den die Anzahl der Sechsen ermittelt wird, dargestellt.



Auf der Basis der Wahrscheinlichkeitsverteilung sowie auch der Simulation könnte man folgende Entscheidungsregel aufstellen: Ergeben die 100 Doppelwürfe eine Anzahl von Sechsen innerhalb des Intervalls $[11;27]$, so wird das Modell beibehalten. Bei einer Anzahl außerhalb des angegebenen Intervalls wird das Modell verworfen.

Da $P(11 \leq X \leq 27) \approx 0{,}972$ ist, würde man bei dieser Entscheidung in etwa 3 Prozent der Fälle einen Fehler begehen, wenn nämlich weniger als 11 oder mehr als 27 Sechsen erscheinen, obwohl die Wahrscheinlichkeit für die Augenzahl 6 tatsächlich $\frac{3}{16}$ beträgt.

Für das Intervall $[E(X)-\sigma; E(X)+\sigma]$ egibt sich:

$$P(E(X)-\sigma \leq X \leq E(X)+\sigma) \approx P(14{,}85 \leq X \leq 22{,}65) = P(15 \leq X \leq 22) \approx 0{,}696$$

Mit Hilfe des Rechners lässt sich ein annähernd symmetrisch um den Erwartungswert konstruiertes Intervall $I = [12;26]$ bilden, für das $P(12 \leq X \leq 26) \approx 0{,}95$ gilt. Würde man also eine Entscheidungsregel auf der Basis konventionalisierter Irrtumswahrscheinlichkeiten von 5 Prozent bilden, so wäre eine Anzahl von Sechsen außerhalb des Intervalls I Anlass, dass Modell für die Augenzahl 6 beim doppelten Tetraeder-Wurf zu verwerfen.

Aufgabe 7.2: Das Zahlenlotto 6 aus 49 soll analysiert werden. Bei diesem Zahlenlotto werden 6 Erfolgskugeln aus einer Urne mit 49 Kugeln mit den Zahlen von 1 bis 49 als Aufschrift zufällig und ohne Zurücklegen gezogen. Die Reihenfolge der Erfolgskugeln ist dabei unerheblich, da diese am Ende der Ziehung der Größe nach geordnet werden.

a) Ein Spieler tippt 6 Zahlen. Bestimmen Sie die Wahrscheinlichkeitsverteilung für X: *Anzahl der richtig getippten Zahlen*. Eine richtig getippte Zahl sei dann vorhanden, wenn sie mit der Aufschrift einer der Erfolgskugeln gleich ist.

b) Bestimmen Sie die Wahrscheinlichkeit, dass mindestens 3 richtige Zahlen bzw. höchstens 2 richtige Zahlen getippt werden.

Ein Spieler gibt jeden Samstag, also etwa 52 Mal im Jahr, jeweils einen Tipp von 6 Zahlen ab.

c) Begründen Sie, dass X: *Anzahl der Dreier (genau drei richtige Zahlen)* bei n Samstagen eine binomialverteilte Zufallsgröße ist.

d) Welche Anzahl von Dreiern kann der Spieler in einem, zwei, m Jahren erwarten?

Lösungsskizze Aufgabe 7.2 Wir ermitteln im Folgenden mit Hilfe der Hypergeometrischen Verteilung die Wahrscheinlichkeiten für die verschiedenen Anzahlen der Richtigen im Lotto (ohne Beachtung der Sonderregeln).

k	0	1	2	3	4	5	6
$P(X=k)$	0,436	0,413	0,132	0,018	0,001	0,00002	0,00000007

Aus der Wahrscheinlichkeitsverteilung ergibt sich unmittelbar

$$P(X \leq 2) \approx 0,981 \text{ und } P(X \geq 3) = 1 - 0,981 = 0,019$$

Beim Lotto kann davon ausgegangen werden, dass die Ziehung an einem Samstag unabhängig von jeder weiteren Ziehung ist. Weiterhin werden bei jeder Ziehung bezogen auf einen Tipp nur die beiden Ereignisse Dreier und kein Dreier beachtet. Damit ist die Zufallsgröße X: Anzahl der Dreier binomialverteilt.

Bei einem Tipp pro Ziehung und der Annahme von 52 Ziehungen pro Jahr, ergibt sich folgende Erwartung von Dreiern (statt mit dem gerundeten Wert 0,018 wurde mit dem exakten Wert 0,0176504 gerechnet):

Jahre	Erwartete Anzahl von Dreiern
1	$52 \cdot 0,018 = 0,918$
2	$52 \cdot 0,018 = 1,836$
10	$52 \cdot 0,018 = 9,178$
m	$m \cdot 0,018$

Aufgabe 7.3: Gegeben ist eine Urne mit 2 weißen und 8 schwarzen Kugeln. Aus dieser wird 4 Mal ohne Zurücklegen gezogen.

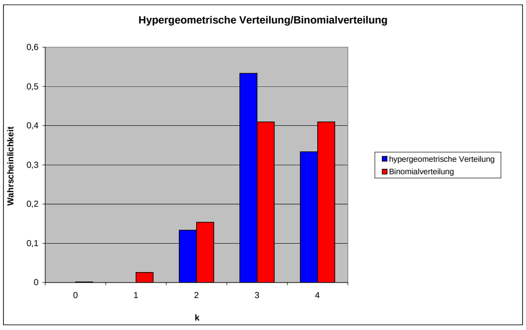
a) Bestimmen Sie die Wahrscheinlichkeitsverteilung für X: *Anzahl der schwarzen Kugeln*.

b) Simulieren Sie diesen Versuch 10, 100 und 1000 Mal und vergleichen Sie die simulierte Häufigkeit der schwarzen Kugeln mit der Wahrscheinlichkeitsverteilung.

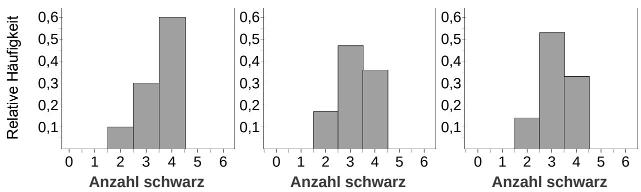
c) Bestimmen Sie die Wahrscheinlichkeitsverteilung von X unter der Voraussetzung, dass die Kugeln nach dem Zug wieder zurückgelegt wird. Vergleichen Sie die entstehende Verteilung mit der in a) ermittelten.

Lösungsskizze Aufgabe 7.3 Die Zufallsgröße X ist hypergeometrisch verteilt. Dadurch ergibt sich die erste Wahrscheinlichkeitsverteilung, die im Sinne von Aufgabenteil c) gleich mit dem Fall verglichen wird, bei dem die Kugeln nach dem Zug wieder zurückgelegt werden (binomialverteilte Zufallsgröße). Beide Verteilungen sind (geringe Größe der Grundgesamtheit, Stichproben ist halb so groß wie die Grundgesamtheit) stark unterschiedlich, wie sich numerisch und grafisch zeigt.

	k	0	1	2	3	4
X hypergeometrisch	$P(X=k)$	0	0	0,133	0,533	0,333
X binomial	$P(X=k)$	0,002	0,026	0,154	0,410	0,410



Simuliert man den Versuch mit unterschiedlichen Wiederholungszahlen, so ergibt sich etwa die Darstellung unten, bei der erkennbar ist, dass erst bei steigender Simulationsanzahl die Form der Wahrscheinlichkeitsverteilung sichtbar wird.



Aufgabe 7.4:

a) Teilen Sie beliebige Merkmale zu den Studierenden (siehe Zusatzmaterial) dichotom auf (mit zwei Merkmalsausprägungen) und bezeichnen diese als Erfolg/Misserfolg eines zufälligen Vorgangs. Setzen Sie anhand der relativen Häufigkeiten einer Merkmalsausprägung die Wahrscheinlichkeiten für Erfolg und Misserfolg fest.

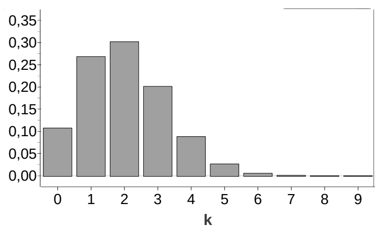
b) Berechnen Sie die Wahrscheinlichkeitsverteilungen jeweils für X: *Anzahl der Erfolge* bei Stichprobenumfängen von $n = 10$ und $n = 100$. Begründen Sie jeweils die Wahl des von Ihnen verwendeten Modells einer Wahrscheinlichkeitsverteilung.

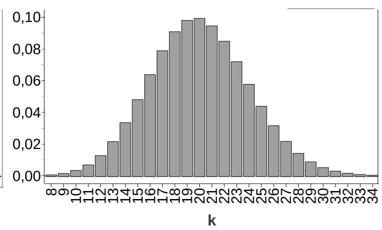
c) Bestimmen Sie die Wahrscheinlichkeiten für die Intervalle $I_1 = [E(X) - \sigma; E(X) + \sigma]$, $I_2 = [E(X) - 2\sigma; E(X) + 2\sigma]$, $I_3 = [E(X) - 3\sigma; E(X) + 3\sigma]$

Lösungsskizze Aufgabe 7.4 Wir skizzieren allein das Merkmal Wohnung mit den Ausprägungen Eltern und Nicht-Eltern. Durch Schätzung kann die Wahrscheinlichkeit für Studierende, die bei den Eltern wohnen bezogen auf die Freiburger Studierenden mit $P(E) = p = 0,20$ festgelegt werden.

Da die Stichprobengröße von 10 bzw. 100 (simulierten) Studierenden gegenüber der Grundgesamtheit von über 4000 Studierenden (der PH Freiburg) gering ist, kann sinnvoll das Modell einer binomialverteilten Zufallsgröße X: Anzahl der Studierenden, die bei den Eltern wohnen, verwendet werden.

In diesem Modell ergeben sich die unten aufgeführten Wahrscheinlichkeitsverteilungen (grafisch):





Im Aufgabenteil c) bestimmen wir die Wahrscheinlichkeiten der symmetrischen Intervalle allein für den Fall $n = 100$, bei dem die Wahrscheinlichkeitsverteilung selbst annähernd symmetrisch ist. Dort gilt mit $E(X) = n \cdot p = 100 \cdot 0,20 = 20$, $V(X) = n \cdot p \cdot (1-p) = 100 \cdot 0,20 \cdot 0,80 = 16$ und $\sigma(X) = 4$:

- $P(E(X) - \sigma \geq X \geq E(X) + \sigma) = P(16 \geq X \geq 24) \approx 0,740$
- $P(E(X) - 2\sigma \geq X \geq E(X) + 2\sigma) = P(12 \geq X \geq 28) \approx 0,967$
- $P(E(X) - 3\sigma \geq X \geq E(X) + 3\sigma) = P(8 \geq X \geq 32) \approx 0,998$

Diese drei σ-Umgebungen sind insofern von Bedeutung, als dass für annähernd symmetrisch Binomialverteilungen die Wahrscheinlichkeiten für die Intervalle (als Ereignis) nahezu invariant sind.

Aufgabe 7.5: „On August 18, 1913, at the casino in Monte Carlo, black came up a record tewnty-six times in succession [in roulette]. ... [There] was a near-panicky rush to bet on red, beginning about the time black had come up a phenomenal fifteen times. In application of the maturity [of the chances] doctrine, players doubled and tripled their stakes, this doctrine leading them to believe after black came up the twentieth time that there was not a chance in a million of another repeat. In the end the unusual run enriched the Casino by some millions of francs. “ (Huff & Geis, 1959, pp. 28–29)
Der Gewinn des Casinos an diesem Tag beruhte auf der auch unter Spielern weitverbreiteten Fehlvorstellung *gambler's fallacy* (vgl. Kap. 6.1, S. 118 Fußnote).

a) Bestimmen Sie die Wahrscheinlichkeit für die Farbe rot im 21. Spiel und die Wahrscheinlichkeit, dass die Farbe schwarz 21 Mal in Serie auftritt. (Zur Erinnerung: Das Rouletterad hat 37 Felder, die von 0 bis 36 durchnummeriert sind. 18 Felder sind rot, 18 Felder sind schwarz gefärbt. Die 0 („Zero“) ist grün.)

b) Wie wahrscheinlich tritt bei 100 Spielrunden eine Serie von 5 mal schwarz hintereinander auf? Simulieren Sie diesen Versuch 10, 100 und 1000 Mal und leiten Sie daraus Aussagen über die Wahrscheinlichkeit $P(\text{Anzahl von 5 mal schwarz hintereinander})$ bei einem Stichprobenumfang von $n = 100$ Spielen ab.

c) Entwickeln Sie eine rekursive Formel, mit der die Wahrscheinlichkeit errechnet werden kann, dass bei 100 Spielrunden k mal schwarz hintereinander auftritt. Setzen Sie $k = 5$ und überprüfen Sie Ihre Formel mit den Simulationsergebnissen aus der vorangehenden Teilaufgabe c).

So sehr es sich auch Spieler wie in diesem Casino wünschen, es gibt keinen plausiblen Grund anzunehmen, dass der Wurf einer Roulette-Kugel durch den vorhergehenden oder eine Serie von vorhergehenden Würfen beeinflusst wird. Einzig die Annahme der stochastischen Unabhängigkeit scheint eine plausible Annahme zu sein. Das bedeutet aber, das *unabhängig* von allen vorangegangenen Würfen die Wahrscheinlichkeit für die Farbe rot stets $\frac{18}{37}$ beträgt (Laplace-Modell: 37 gleichwahrscheinliche Zahlenfelder, 18 davon sind rot).

Im Gegensatz zu einer Prognose auf *einen* Wurf stellt sich die Situation bei einer Prognose einer *Serie* von Würfen anders dar. So beträgt die Wahrscheinlichkeit für eine Serie von 21 Mal schwarz $\frac{18}{37}^{21} \approx 0,00000003$, d.h. bei der enormen Anzahl von 100.000.000 Serien von 21 Würfen kann man lediglich drei solcher langer Serien mit schwarz erwarten.

Die Verwechslung der Vorhersage *eines* Wurfes *nach* einer Serie von Würfen und die Vorhersage der *Serie* von Würfen *vor* dieser Serie ist der Kern der „gamblers fallacy“. Aufgrund des plausiblen Modells der stochastischen Unabhängigkeit ist hier die Information, bestehend aus den Ergebnissen bereits geworfener Kugeln, nutzlos.

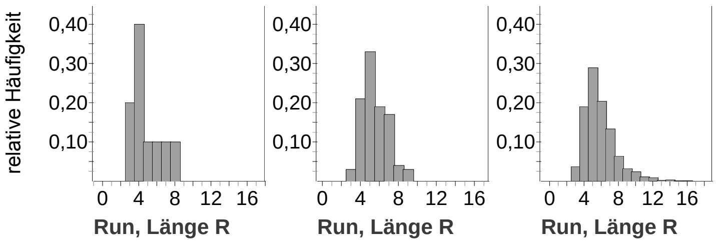
Interessant ist weiterhin die Analyse von sogenannten Runs einer bestimmten Länge, im Aufgabenteil der Länge 5, wobei wir hier als Run nur das k-fache erscheinen von S: schwarz in Reihenfolge bezeichnen. Wir simulieren zunächst 10, 100, 1000 Serien von je 100 Roulette-Spielen. Bei der Auswertung sind etwa folgende verschiedene Fragestellungen möglich:

- Wieviele Serien von Roulette-Spielen enthalten *mindestens* einen Run (von schwarz) der *genau* die Länge 5 hat?
- Wieviele Serien von Roulette-Spielen enthalten *genau* einen Run der *genau* die Länge 5

hat?

- Wieviele Serien von Roulette-Spielen enthalten *mindestens* einen Run der *mindestens* die Länge 5 hat?
- Wieviele Serien von Roulette-Spielen enthalten *genau* einen Run der *mindestens* die Länge 5 hat?
- Welche Verteilung hat die Anzahl der Runs der Länge 5 (bzw. der Länge von mindestens 5) in einer Serie von Roulette-Spielen?

Wir betrachten in der Simulation und im nächsten Aufgabenteil die einfachste Fragestellung, nämlich die Verteilung der Wahrscheinlichkeiten für mindestens einen Run der Länge R. Die Simulation von 10, 100 und 1000 Serien von Roulette-Spielen hat hierzu folgendes Bild ergeben:



Man erkennt, dass offenbar Roulette-Serien mit maximalen Serien von Schwarz mit $r < 4$ bzw. $r > 8$ selten sind. Wir erhalten numerisch:

Mindestens 1 Run mit der Länge	3	4	5	6	7	8	9	10	11	12	≥ 13
Häufigkeit (bei 1000 Simulationen	0,037	0,19	0,289	0,204	0,133	0,064	0,032	0,024	0,011	0,009	0,007

Wir nähern uns einer algorithmischen Lösung in mehreren Schritten und bestimmen im Folgenden jeweils die Wahrscheinlichkeit für Runs $R \geq k$. Für $k = 0$ ist das trivial, da jede Serie von Roulette-Spielen mindestens einen Run der Länge $R \geq 0$ enthält. Messen wir mit R die Länge des maximalen Runs in einer Serie von 100 Roulette-Spielen, so gilt also $P(R \geq 0) = 1$.

Auch für $R \geq 1$ ist das Problem schnell zu lösen, da

$$P(R \geq 1) = 1 - P(R = 0) = 1 - \left(\frac{19}{37}\right)^{100} \approx 1$$

gilt.

Für $R \geq 2$ gehen wir das Problem rekursiv an, wobei man sich einen Baum vorstelle, der sich stets mit zwei Ästen verzweigt (S: schwarz, $\overline{S}$: nicht schwarz) mit den Wahrscheinlichkeiten $p = \frac{18}{37}$ und $q := 1 - p = \frac{19}{37}$. Wir bezeichnen mit einem Index die Anzahl der Roulette-Spiele und erhalten unmittelbar $P(R_1 \geq 2) = 0$, da bei einem Roulette-Spiel kein Run mit zwei Mal schwarz geschehen kann.

Bezogen auf das zweifache Roulette-Spiel gilt $P(R_2 \geq 2) = p \cdot p = p^2$, die Wahrscheinlichkeit längs des Pfades mit zwei Mal S. Für $P(R_3 \geq 2)$ überlegen wir Folgendes:

- Die Wahrscheinlichkeit, in den ersten beiden Spielen einen 2er-Run (schwarz) zu erhalten war $P(R_2 \geq 2)$. Ob im dritten Spiel dieser Run zu einem 3er-Run wird oder abbricht, vernachlässigen wir.
- Zu dieser Wahrscheinlichkeit kommt der Pfad, an dem zuerst nicht-schwarz und dann zwei Mal schwarz erscheint $\overline{S}SS$ mit der Wahrscheinlichkeit $q \cdot p \cdot p = qp^2$. Insgesamt erhalten wir also $P(R_3 \geq 2) = P(R_2 \geq 2) + qp^2$.

Für $P(R_4 \geq 2)$ überlegen wir analog:

- Der 2er-Run kann im ersten Spiel begonnen haben ($P(R_2 \geq 2)$), im zweiten Spiel ($P(R_3 \geq 2)$) oder im dritten.
- Beginnt der 2er-Run im dritten Spiel, so hat das zweite Spiel nicht-schwarz ergeben, das erste schwarz oder nicht schwarz, das dritte und vierte jeweils schwarz. Da die Wahrscheinlichkeit für nicht-schwarz im 3ten (wie allgemein in jedem Versuch) q ist, ist also die Wahrscheinlichkeit für den 2er-Run mit Beginn im dritten Spiel wiederum qp^2.
- Insgesamt gilt also $P(R_4 \geq 2) = P(R_3 \geq 2) + qp^2$.

Bezogen auf einen imaginären Baum identifizieren wir also vom ersten Spiel beginnend Pfadbündel, die die Eigenschaft haben, dass *mindestens* ein Run der Länge von *mindestens* 2 auftritt. Ob dieses Pfadbündel später mehrere 2er-Runs oder längere Runs enthält, wird dagegen nicht beachtet. So hat etwa das am Beginn verarbeitete Pfadbündel, das mit SS beginnt, bei vier Roulette-Spielen die Pfade $SSSS$ (4er-Run), $SSS\overline{S}$ (3er-Run), $SS\overline{S}S$ (2er-Run), $SS\overline{SS}$ (2er-Run). Alle diese Pfade repräsentieren hier aber zusammenfassend ein Pfadbündel, in dem zumindest ein Run der Länge von mindestens 2 vorhanden ist.

Die rekursive Berechnung von $P(R_n \geq 2)$ lässt sich mit einer Erweiterung fortsetzen, die wir am Beispiel des 2er-Runs und $n = 5$ Spielen erläutern. Würde sich die Rekursion analog zu oben fortsetzen, so ergäbe sich $P(R_5 \geq 2) = P(R_4 \geq 2) + qp^2$, wobei der letzte Term alle die Pfade repräsentiert, in denen im dritten Spiel $\overline{S}$ auftritt und im vierten und fünften Spiel S. Hier muss man allerdings beachten, dass bei dieser Fortsetzung Pfade doppelt in die Berechnung eingehen, konkret der Pfad $SS\overline{S}SS$, der bereits am Anfang (im ersten und zweiten Spiel ereignet sich S) in die Berechnung eingeflossen ist. Ab diesem fünften Spiel, allgemeiner dem $2k+1$-ten Spiel müssen wir die doppelte Zählung also rückgängig machen und den entsprechenden Pfad mit der Wahrscheinlichkeit $p \cdot p \cdot q \cdot p \cdot p = P(R_2 \geq 2) \cdot qp^2$ in der Rekursionsgleichung abziehen. Es gilt also $P(R_5 \geq 2) = P(R_4 \geq 2) + qp^2 - P(R_2 \geq 2)qp^2$. Diese Rekursionsgleichung lässt sich nun beliebig fortsetzen:

- $P(R_6 \geq 2) = P(R_5 \geq 2) + qp^2 - P(R_3 \geq 2) \cdot qp^2$, da $\overline{S}SS\overline{S}SS$ bereits in der Betrachtung des Pfadbündels beginnend mit $\overline{S}SS$ in die Wahrscheinlichkeit $P(R_3 \geq 2)$ und damit indirekt in den ersten Term $P(R_5 \geq 2)$ eingegangen ist,
- $P(R_7 \geq 2) = P(R_6 \geq 2) + qp^2 - P(R_4 \geq 2) \cdot qp^2$,
- ...
- $P(R_{100} \geq 2) = P(R_{99} \geq 2) + qp^2 - P(R_{97} \geq 2) \cdot qp^2$.

In analoger Weise lassen sich die obigen Überlegungen auch für Run-Längen von mindestens 3,4,...,100 durchführen. So gilt insgesamt für die Existenz (mindestens) eines Runs der Länge k bei n Spielen

$$P(R_n \geq k) = P(R_{n-1} \geq k) + qp^2 - P(R_{n-k-1} \geq 2) \cdot qp^2$$

mit $P(R_i \geq k) = 0$ für $i < k$ und $P(R_i \geq k) = p^k$ für $i = k$ ($i \in \{0, \ldots, n\}$).

Mit dieser Rekursionformel erhalten wir auf drei Stellen gerundet:

k	≤ 2	3	4	5	6	7	8	9	10	11	12	≥ 13
$P(X_{100} \geq k)$	1	1	0,962	0,770	0,495	0,276	0,142	0,071	0,035	0,017	0,008	0,007

Aus dieser Tabelle erhalten wir auch die Wahrscheinlichkeit mindestens einen Run *genau* der Länge k durch $P(R_i = k) = P(R_i \geq k) - P(R_{i+1} \geq k)$ und damit:

k	≤ 2	3	4	5	6	7	8	9	10	11	12	≥ 13
$P(X_{100} \geq k)$	0,001	0,038	0,191	0,276	0,220	0,133	0,072	0,036	0,018	0,009	0,004	0,004

Die oben ermittelten Häufigkeiten entsprechen damit schon recht gut den hier ermittelten Wahrscheinlichkeiten.