Aufgaben und Lösungsvorschlag

1. Multiple choice **[7 Punkte]**

(a) **[1.5 Punkte]** Gilt $\mathbb{E}\left[\frac{X}{Y}\right] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$?

i. Ja, $\mathbb{E}\left[\frac{X}{Y}\right] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt für alle Zufallsvariablen X, Y.

ii. $\mathbb{E}\left[\frac{X}{Y}\right] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt für alle unabhängigen Zufallsvariablen X, Y. Es gibt nicht unabhängige Zufallsvariablen für die $\mathbb{E}\left[\frac{X}{Y}\right] \neq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt.

iii. Nein, es gibt unabhängige sowie nicht unabhängige Zufallsvariablen für die $\mathbb{E}\left[\frac{X}{Y}\right] \neq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt. Es gibt Spezialfälle in denen $\mathbb{E}\left[\frac{X}{Y}\right] = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt.

iv. Nein, für alle Zufallsvariablen X, Y gilt $\mathbb{E}\left[\frac{X}{Y}\right] \neq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$.

Lösung:

1.(a)iii. Sei zB $X = 1$ und $Y \sim \mathcal{U}(1,3)$, dann ist $\mathbb{E}\left[\frac{X}{Y}\right] = \mathbb{E}\left[\frac{1}{Y}\right] = \int_1^3 \frac{1}{2y} dy = \frac{\log(3)}{2} \approx 0.5493 \neq \frac{1}{2} = \frac{1}{\mathbb{E}[Y]} = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$, obwohl die beiden Zufallsvariablen unabhängig sind. Somit sind **1.**(a)i und **1.**(a)ii falsch. Das vorigen Beispiel ist schon ein Beispiel für eine unabhängige Zufallsvariablen für die $\mathbb{E}\left[\frac{X}{Y}\right] \neq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt. Ein Beispiel für abhängige Zufallsvariablen für die $\mathbb{E}\left[\frac{X}{Y}\right] \neq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$ gilt wäre zB $\Omega, \mathcal{F}, \mathbb{P}$ ein Laplace-Modell auf $\{-1, 1\}$ mit $X : \Omega \to \mathbb{R}, \omega \mapsto X(\omega) = \omega$ und $Y : \Omega \to \mathbb{R}, \omega \mapsto Y(\omega) = \begin{cases} 1, & \text{falls } \omega = 1 \\ 0, & \text{falls } \omega = 0 \end{cases}$, dann gilt nämlich $\mathbb{E}\left[\frac{X}{Y}\right] = \frac{1}{2} \neq 0 = \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$. **1.**(a)iv ist falsch, weil es auch Spezialfälle geben kann wo die Gleichheit gilt, zB wenn $X = 0$ die deterministische Zufallsvariable, die immer 0 ausgibt ist.

(b) **[2 Punkte]** Welche dieser Funktionen ist eine Wahrscheinlichkeitsdichte f_X einer reellwertigen Zufallsvariable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$?

i. $x \mapsto \begin{cases} 1 - x, & \text{falls } x \in [0,1] \\ 10, & \text{falls } x \in \left(1, \frac{21}{20}\right] \\ 0, & \text{sonst} \end{cases}$

ii. $x \mapsto \begin{cases} \sin(x), & \text{falls } x \in [0, 3\pi] \\ 0, & \text{sonst} \end{cases}$

iii. $x \mapsto \begin{cases} \frac{1}{32}, & \text{falls } x \in [-1, 0) \\ \frac{1}{3} + x, & \text{falls } x \in \left[0, \frac{1}{4}\right) \\ 1, & \text{falls } x \geq \frac{1}{4} \\ 0, & \text{sonst} \end{cases}$

iv. $x \mapsto \frac{1}{100} e^{-x^2}$ für $x \in \mathbb{R}$

v. $x \mapsto \begin{cases} 2x, & \text{falls } x \in \left[0, \frac{1}{2}\right) \\ \frac{1}{2} + \frac{x}{2}, & \text{falls } x \in \left[\frac{1}{2}, 1\right) \\ 1, & \text{falls } x \geq 1 \\ 0, & \text{sonst} \end{cases}$

vi. $x \mapsto \begin{cases} x, & \text{falls } x \in \left[0, \frac{1}{2}\right] \\ 1, & \text{falls } x > \frac{1}{2} \\ 0, & \text{sonst} \end{cases}$

vii. $x \mapsto x$ für $x \in \mathbb{R}$

viii. $x \mapsto 1$ für $x \in \mathbb{R}$

Lösung:

1.(b)i. f_X aus **1.**(b)i ist eine Wahrscheinlichkeitsdichte, weil sowohl $f_X(x) \geq 0$ für alle $x \in \mathbb{R}$ gilt als auch $\inf_{-\infty}^{\infty} f_X(x)dx = 1$ gilt.

1.(b)ii und **1.**(b)vii sind keine Dichten, weil auch negative Werte angenommen werden.

1.(b)iii, **1.**(b)v, **1.**(b)vi und **1.**(b)viii sind keine Wahrscheinlichkeitsdichten, weil das Integral über diese Funktionen unendlich anstatt 1 ist.

1.(b)iv ist keine Wahrscheinlichkeitsdichte, weil das Integral über die Funktion viel kleiner als 1 ist ($\frac{\sqrt{\pi}}{100}$), was man direkt daran erkennt, dass der Vorfaktor stark von dem der Normalverteilungsdichte abweicht.

(c) [**1.5 Punkte**] Welche der Funktionen aus Frage **1.**b ist eine Verteilungsfunktion[1] F_X einer reellwertigen Zufallsvariable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$?

Lösung:

Die Funktion aus **1.**(b)iii. Die Funktion F_X aus **1.**(b)iii ist eine Verteilungsfunktion, weil F_X alle Bedinungen aus Theorem 1.18 aus dem Skript erfüllt.

1.(b)i, **1.**(b)ii, **1.**(b)iv und **1.**(b)v sind nicht monoton steigend.

1.(b)vi ist nicht rechts-stetig.

1.(b)vii und **1.**(b)viii erfüllen nicht $\lim_{x \to -\infty} F_X(x) = 0$

(d) [**2 Punkte**] Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Wahrscheinlichkeits-Modell auf $\Omega = \{1, 2, 3, 4\}$. Sind die Ereignisse $A = \{1, 2\}$ und $B = \{4\}$ unabhängig?

i. A und B sind unabhängig für jedes $\mathbb{P}$.

ii. A und B sind abhängig für jedes $\mathbb{P}$.

iii. A und B sind unabhängig, falls $\mathbb{P}[\{\omega\}] \neq 0 \ \forall \omega \in \Omega$ erfüllt, aber **1.**(d)i gilt nicht.

iv. A und B sind abhängig, falls $\mathbb{P}[\{\omega\}] \neq 0 \ \forall \omega \in \Omega$ erfüllt, aber **1.**(d)ii gilt nicht.

Lösung:

[1] Die Verteilungsfunktion F_X wird in der Literatur häufig auch als Wahrscheinlichkeitsverteilungsfunktion, kumulative (Wahrscheinlichkeits-)Verteilungsfunktion, **c**umulative (probability) **d**istribution **f**unction, cdf oder CDF bezeichnet.

1.(d)iv. Falls $\mathbb{P}[\{\omega\}] \neq 0\ \forall\omega \in \Omega$ erfüllt, gilt auch $\mathbb{P}[A] \neq 0$ und $\mathbb{P}[B] \neq 0$, wegen der Montonie eines Masses. Dann gilt auch $\mathbb{P}[A]\,\mathbb{P}[B] \neq 0 = \mathbb{P}[\varnothing] = \mathbb{P}[A \cap B]$. In diesem Fall sind die beiden Ereignisse alsi abhängig. Im Fall dass beispielsweise $\mathbb{P}[B] = 0$ ist (zB wenn $\mathbb{P} : \mathcal{F} \to [0,1], E \mapsto \mathbb{P}[E] = \frac{|A\cap E|}{2}$), würde $\mathbb{P}[A]\,\mathbb{P}[B] = 0 = \mathbb{P}[A \cap B]$ gelten – in so einem Spezialfall wären die beiden Ereignisse dann unabhängig.

2. S**,** T**,** U **und** V **[6.5 Punkte]**

(a) Sei $S \sim \mathcal{N}(-5, 4^2)$ und $T \sim \mathcal{N}(10, 3^2)$ unabhängig. [**4 Punkte**]

i. Berechne $\mathbb{P}[S < T]$.

ii. Wäre die Berechnung von $\mathbb{P}[S < T]$ (**2.**(a)i) auch ohne der Unabhängigkeits-Annahme korrekt? (Die Antwort muss nicht begründet werden.)

iii. Berechne die Varianz σ_R^2 von $R := S - 2T$.

iv. Wäre die Berechnung von σ_R^2 (**2.**(a)iii) auch ohne der Unabhängigkeits-Annahme korrekt? (Die Antwort muss nicht begründet werden.)

(b) Sei $U \sim \mathcal{U}(1, 3)$ und $V \sim \mathcal{U}(0, 4)$ unabhängig. [**2.5 Punkte**]

i. Berechne $\mathbb{E}[2U + V^3]$.

ii. Wäre die Berechnung von $\mathbb{E}[2U + V^3]$ (**2.**(b)i) auch ohne der Unabhängigkeits-Annahme korrekt? (Die Antwort muss nicht begründet werden.)

Lösung:

(a) i. $\mathbb{P}[S < T] = \mathbb{P}[S - T < 0]$. Wegen der Unabhängigkeit gilt $(S-T) \sim \mathcal{N}(-5-10, 4^2 + 3^3) = \mathcal{N}(-15, 25)$ (siehe „Properties of normal random variables“ auf Seite 50 des Skripts). Wir betrachten die normierte Zufallsvariable $Z := \frac{S-T+15}{5} \sim \mathcal{N}(0, 1)$.

$$\mathbb{P}[S < T] = \mathbb{P}[S - T < 0] = \mathbb{P}\left[\frac{S - T + 15}{5} < \frac{15}{5}\right] = \mathbb{P}[Z < 3] = \Phi(3) \approx 0.9987$$

ii. Nein. (In „Properties of normal random variables“ auf Seite 50 des Skripts wird die entsprechende Rechenregel nur für *unabhängige* normal-verteilte Zufallsvariablen gezeigt. Im allgemeinen gilt die Rechenregel nicht: Wäre zum Beispiel $T = \frac{3}{4}(S+5)+10$, dann würde immer noch $T \sim \mathcal{N}(10, 3^2)$ gelten, aber $S - T = S - \frac{3}{4}(S + 5) - 10 = \frac{1}{4}S - \frac{3}{4}5 - 10 = \frac{1}{4}S - \frac{55}{4}$ wäre dann $\mathcal{N}(-\frac{5}{4} - \frac{55}{4}, \frac{4^2}{4^2}) = \mathcal{N}(-15, 1)$, wegen der Abhängigkeit der beiden Zufallsvariablen. Ja nach Abhägnigkeit kann die Summe/Differenz von normalverteilten Zufallsvariablen auch ganz anderen Verteilungen folgen, die sich nicht als Normalverteilung darstellen lassen.)

iii. $\sigma_R^2 = \sigma_{S-2T}^2 \overset{\text{unab.}}{=} \sigma_S^2 + 2^2\sigma_T^2 = 4^2 + 2^2 3^2 = 16 + 6^2 = 52$

iv. Nein. (In Proposition 2.42 im Skript wird die entsprechende Rechenregel nur für unabhängige Zufallsvariablen gezeigt.[a] Im allgemeinen gilt die Rechenregel nicht: Wäre zum Beispiel $T = \frac{3}{4}(S + 5) + 10$, dann würde immer noch $T \sim \mathcal{N}(10, 3^2)$ gelten, aber $R = S - 2T = S - \frac{3}{2}(S + 5) - 20 = -\frac{1}{2}S - \frac{3}{2}5 - 20 = -\frac{1}{2}S - \frac{55}{2}$ hätte dann nur noch eine Varainz von $\sigma_R^2 = \frac{1}{2^2}\sigma_S^2 + 0 = \frac{4^2}{2^4} = 4$, wegen der Korrelation der beiden Zufallsvariablen.)

(b) i.

$$\mathbb{E}[U] = \frac{1+3}{2} = 2$$

Die Dichte von V ist $f_V(v) = \frac{1}{4}\mathbb{1}_{[0,4]}$

$$\mathbb{E}\left[V^3\right] = \int_{-\infty}^{\infty} f_V(v)v^3 dv = \int_0^4 \frac{1}{4}v^3 dv = \frac{1}{4\cdot 4}4^4 - \frac{1}{4\cdot 4}0^4 = 4^2 = 16.$$

$$\mathbb{E}\left[2U + V^3\right] = 2\mathbb{E}\left[U\right] + \mathbb{E}\left[V^3\right] = 4 + 4^2 = 20$$

ii. Ja. (Linearität des Erwartungswerts, siehe Theorem 2.34 aus dem Skript)

[a] Diese Rechenregel würde noch etwas allgemeiner auch für unkorrelierte Zufallsvariablen gelten, das wurde aber in der Vorlesung nicht behandelt.

3. Rechteck **[6.5 Punkte]**

Gegeben sei ein Rechteck mit den zufälligen Seitenlängen X und Y. Die gemeinsame Dichtefunktion von X und Y ist gegeben durch

$$f_{X,Y}(x,y) := \begin{cases} C(x^2+y^2) & 0 \le x, y \le 1 \\ 0 & \text{sonst.} \end{cases}$$

(a) Bestimmen Sie den Parameter C. [**1.5 Punkte**]

(b) Berechnen Sie die Randdichten von X und Y. [**2 Punkte**]

(c) Sind X und Y unabhängig? Begründen Sie Ihre Antwort. [**1 Punkt**]

(d) Berechnen Sie die Wahrscheinlichkeit, dass die Seite X mehr als doppelt so lang wie die Seite Y ist. [**2 Punkte**]

Lösung:

(a)

$$1 \stackrel{!}{=} \int_0^1 \int_0^1 C(x^2+y^2)dxdy = C \cdot \int_0^1 \left(\frac{1}{3} + y^2\right) dy = C\left(\frac{1}{3}+\frac{1}{3}\right) = \frac{2}{3}C\ .$$

Damit findet man $C = \frac{3}{2}$.

(b) Für die Randdichte von X hat man für $0 \le x \le 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{3}{2}\int_0^1 (x^2+y^2)dy = \frac{3}{2}x^2 + \frac{1}{2}.$$

Also hat man

$$f_X(x) = \begin{cases} \frac{3}{2}x^2 + \frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{sonst.} \end{cases}$$

Für die Randdichte von Y hat man für $0 \le y \le 1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{3}{2}\int_0^1 (x^2+y^2)dx = \frac{3}{2}y^2 + \frac{1}{2}.$$

Also hat man

$$f_Y(y) = \begin{cases} \frac{3}{2}y^2 + \frac{1}{2} & 0 \le y \le 1 \\ 0 & \text{sonst.} \end{cases}$$

(c) X und Y sind unabhängig voneinander genau dann wenn $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ gilt, was nicht der Fall ist. Folglich sind X und Y NICHT unabhängig.

(d)

$$\begin{aligned} P[X > 2Y] &= \frac{3}{2}\int_0^1 \int_0^{\frac{x}{2}} (x^2+y^2)dydx = \frac{3}{2}\int_0^1 \left(\frac{x^3}{2} + \frac{x^3}{2^3 \cdot 3}\right)dx \\ &= \frac{3}{2}\left(\frac{1^4}{2^3} + \frac{1^4}{2^5 \cdot 3}\right) = \frac{3}{2^4}\left(1 + \frac{1}{2^2 \cdot 3}\right) = \frac{3}{2^4}\left(\frac{13}{2^2 \cdot 3}\right) = \frac{13}{2^6} = \frac{13}{2^6} = 0.203125. \end{aligned}$$

4. Urnen-Problem **[3 Punkte]**

Wir haben zwei Urnen A und B. Urne A enthält eine schwarze und eine weisse Kugel und Urne B enthält zwei schwarze und eine weisse Kugel.

Eine Urne wird zufällig gewählt und daraus eine Kugel gezogen. Was ist die Wahrscheinlichkeit, dass Urne A gewählt wurde, gegeben, dass die gezogene Kugel weiss ist?

Lösung:

Wir definieren die Ereignisse

$$\begin{aligned} A &= \text{„Urne } A \text{ wird ausgewählt“}, \\ B &= \text{„Urne } B \text{ wird ausgewählt“}, \\ W &= \text{„gezogene Kugel ist weiss“}. \end{aligned}$$

Gemäss Aufgabenstellung haben wir

$$\mathbb{P}[A] = \mathbb{P}[B] = \frac{1}{2}$$

sowie

$$\mathbb{P}[W \mid A] = \frac{1}{2} \qquad \text{und} \qquad \mathbb{P}[W \mid B] = \frac{1}{3}.$$

Gesucht ist nun $\mathbb{P}[A \mid W]$. Der Satz von Bayes liefert uns

$$\mathbb{P}[A \mid W] = \frac{\mathbb{P}[W \cap A]}{\mathbb{P}[W]} = \frac{\mathbb{P}[W \mid A]\ \mathbb{P}[A]}{\mathbb{P}[W \mid A]\ \mathbb{P}[A] + \mathbb{P}[W \mid B]\ \mathbb{P}[B]} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2}} = \frac{3}{5}.$$

5. Medikament **[7 Punkte]**

Ein Pharmainstitut behauptet, ein bestimmtes Medikament wirke mit einer Wahrscheinlichkeit von mindestens 90%. Daraufhin wird das Medikament an 100 Personen verabreicht. Verwende für die folgenden Aufgaben den Zentralen Grenzwertsatz um schwer berechenbare Werte zu approximieren und begründe deine Antworten!

(a) In dieser ersten Testreihe zeigt das Mittel bei 97 der 100 Personen Wirkung. Ist damit die Behauptung des Pharmainstituts auf dem Signifikanzniveau von 5% statistisch bewiesen? [**5 Punkte**]

(b) Einige Zeit, nachdem das neue Medikament zugelassen ist, bekommt der angesehene Medizinprofessor Zweifel den Verdacht, das Pharmainstitut habe die Studie gefälscht und das Mittel wirke doch nicht so gut wie behauptet. Er lässt daraufhin erneut einen Test an 100 Personen durchführen. Formuliere für Professor Zweifels Test Nullhypothese und Gegenhypothese. Wie muss seine Entscheidungsregel lauten, wenn er seine Vermutung auf dem Signifikanzniveau von 5% statistisch belegen will? [**2 Punkte**]

Lösung:

(a) Die Nullhypothese wird so gewählt, dass das, was man selbst beweisen will, in der Gegenhypothese steht. Das Pharmainstitut versucht natürlich zu beweisen, dass das Medikament gut ist.

Die Nullhypothese ist dann:

$$H_0 : p < 0.9,$$

was bedeutet, dass das Medikament mit einer geringeren Wahrscheinlichkeit als 0.9 wirkt.

Die Gegenhypothese ist

$$H_1 : p \geq 0.9,$$

also dass das Medikament mit einer Wahrscheinlichkeit von mindestens 90% wirkt. Im folgenden berechnen wir die Wahrscheinlichkeiten unter der Nullhypothese[a]

$$\mathbb{P}_{p=0.9}[X \geq k] \leq 0.05 \Leftrightarrow 1 - \mathbb{P}_{p=0.9}[X \leq k-1] \leq 0.05 \Leftrightarrow \mathbb{P}_{p=0.9}[X \leq k-1] \geq 0.95.$$

X ist Binomialverteilt mit Parametern $p = 0.9$ und $n = 100$. Die Binomialverteilung ist für grosses n numerisch schwierig zu berechnen, kann aber gut mit einer Normalverteilung approximiert werden, wie man einfach mit dem Zentralen Grenzwertsatz beweisen kann:

$$X = X_1 + \cdots + X_n,$$

wobei X_i i.i.d. Bernoulli-Variablen sind mit Parameter 0.9. Somit gilt $\mathbb{E}[X_i] = 0.9$ und $\sigma^2_{X_i} = 0.9 \cdot (1 - 0.9) = 0.09$. Daher gilt laut dem Zentralen Grenzwertsatz (CLT)

$$\mathbb{P}_{p=0.9}[X \leq k-1] = \mathbb{P}_{p=0.9}\left[\frac{X_1 + \cdots + X_n - 0.9n}{\sqrt[2]{0.09n}} \leq \frac{k-1-0.9n}{\sqrt[2]{0.09n}}\right] \overset{\text{CLT}}{\approx} \Phi\left(\frac{k-1-0.9n}{\sqrt[2]{0.09n}}\right)$$

Als nächstes bringen wir k in der folgenden Ungleichung auf die linke Seite:

$$\begin{aligned}\Phi\left(\frac{k-1-0.9n}{\sqrt[2]{0.09n}}\right) \geq 0.95 \Leftrightarrow \frac{k-1-0.9n}{\sqrt[2]{0.09n}} &\geq \Phi^{-1}(0.95) \\ \Leftrightarrow k &\geq \Phi^{-1}(0.95)\sqrt[2]{0.09n}+1+0.9n \\ \Leftrightarrow k &\geq 1.65\sqrt[2]{0.09\cdot 100}+1+0.9\cdot 100 \\ \Leftrightarrow k &\geq 1.65\cdot 3+1+90 \\ \Leftrightarrow k &\geq 4.95+91 \\ \Leftrightarrow k &\geq 95.95\end{aligned}$$

Die Nullhypothese wird also bei 95 oder wenigern Treffern im Test, also Personen bei denen das Medikament wirkt, angenommen und bei 96 oder mehr abgelehnt.

Im Test wirkte das Medikament bei 97 Personen, das liegt für das Pharmainstitut im günstigen Ablehnungsbereich.

Das Ergebnis der Testreihe ist damit statistisch signifikant bewiesen.

Alternative 1: Man kann auch nachrechnen ob $\mathbb{P}_{p=0.9}[X \geq 97] \leq 0.05$ gilt:

$$\begin{aligned}\mathbb{P}_{p=0.9}[X \geq 97] &= 1-\mathbb{P}_{p=0.9}[X \leq 96] \\ &= 1-\mathbb{P}_{p=0.9}\left[\frac{X_1+\cdots+X_n-0.9n}{\sqrt[2]{0.09n}} \leq \frac{96-0.9n}{\sqrt[2]{0.09n}}\right] \overset{\text{CLT}}{\approx} 1-\Phi\left(\frac{96-0.9n}{\sqrt[2]{0.09n}}\right).\end{aligned}$$

Einsetzen ergibt:

$$\begin{aligned}\Phi\left(\frac{96-0.9n}{\sqrt[2]{0.09n}}\right) &= \Phi\left(\frac{96-90}{\sqrt[2]{9}}\right) \\ &= \Phi\left(\frac{6}{3}\right) \\ &= \Phi(2) \approx 0.9772.\end{aligned}$$

Somit erhält man $\mathbb{P}_{p=0.9}[X \geq 97] \approx 1-0.9772 = 0.0228 < 0.05$ und kann daher die Nullyhpothese auf einem Signifikanz-Niveau von 5% verwerfen.

Alternative 2: Man kann auch mit einer bisschen anderen Approximation nachrechnen ob $\mathbb{P}_{p=0.9}[X \geq 97] \leq 0.05$ gilt:

$$\begin{aligned}\mathbb{P}_{p=0.9}[X \geq 97] &= 1-\mathbb{P}_{p=0.9}[X < 97] \\ &= 1-\mathbb{P}_{p=0.9}\left[\frac{X_1+\cdots+X_n-0.9n}{\sqrt[2]{0.09n}} < \frac{97-0.9n}{\sqrt[2]{0.09n}}\right] \overset{\text{CLT}}{\approx} 1-\Phi\left(\frac{97-0.9n}{\sqrt[2]{0.09n}}\right).\end{aligned}$$

Einsetzen ergibt:

$$\Phi\left(\frac{97-0.9n}{\sqrt[2]{0.09n}}\right) = \Phi\left(\frac{97-90}{\sqrt[2]{9}}\right)$$
$$= \Phi\left(\frac{7}{3}\right) \approx 0.9901.$$

Somit erhält man $\mathbb{P}_{p=0.9}[X \geq 97] \approx 1 - 0.9901 = 0.0099 < 0.05$ und kann daher die Nullyhpothese auf einem Signifikanz-Niveau von 5% verwerfen.

Alternative 3:

$$\mathbb{P}_{p=0.9}[X \geq k] \leq 0.05 \Leftrightarrow 1 - \mathbb{P}_{p=0.9}[X < k] \leq 0.05 \Leftrightarrow \mathbb{P}_{p=0.9}[X < k] \geq 0.95.$$

X ist Binomialverteilt mit Parametern $p = 0.9$ und $n = 100$. Die Binomialverteilung ist für grosses n numerisch schwierig zu berechnen, kann aber gut mit einer Normalverteilung approximiert werden, wie man einfach mit dem Zentralen Grenzwertsatz beweisen kann:

$$X = X_1 + \cdots + X_n,$$

wobei X_i i.i.d. Bernoulli-Variablen sind mit Parameter 0.9. Somit gilt $\mathbb{E}[X_i] = 0.9$ und $\sigma^2_{X_i} = 0.9 \cdot (1 - 0.9) = 0.09$. Daher gilt laut dem Zentralen Grenzwertsatz (CLT)

$$\mathbb{P}_{p=0.9}[X < k] = \mathbb{P}_{p=0.9}\left[\frac{X_1 + \cdots + X_n - 0.9n}{\sqrt[2]{0.09n}} < \frac{k - 0.9n}{\sqrt[2]{0.09n}}\right] \overset{\text{CLT}}{\approx} \Phi\left(\frac{k-0.9n}{\sqrt[2]{0.09n}}\right)$$

Als nächstes bringen wir k in der folgenden Ungleichung auf die linke Seite:

$$\begin{aligned}
\Phi\left(\frac{k-0.9n}{\sqrt[2]{0.09n}}\right) \geq 0.95 \Leftrightarrow \frac{k-0.9n}{\sqrt[2]{0.09n}} &\geq \Phi^{-1}(0.95) \\
&\Leftrightarrow k \geq \Phi^{-1}(0.95)\sqrt[2]{0.09n} + 0.9n \\
&\Leftrightarrow k \geq 1.65\sqrt[2]{0.09 \cdot 100} + 0.9 \cdot 100 \\
&\Leftrightarrow k \geq 1.65 \cdot 3 + 90 \\
&\Leftrightarrow k \geq 4.95 + 90 \\
&\Leftrightarrow k \geq 94.95
\end{aligned}$$

Die Nullhypothese wird also bei 94 oder wenigern Treffern im Test, also Personen bei denen das Medikament wirkt, angenommen und bei 95 oder mehr abgelehnt.

Im Test wirkte das Medikament bei 97 Personen, das liegt für das Pharmainstitut im günstigen Ablehnungsbereich.

Das Ergebnis der Testreihe ist damit statistisch signifikant bewiesen.

Diskussion der unterschiedlichen Lösungen: Der Zentrale Grenzwertsatz ist für endliches n nur eine Approximation und nicht exakt. Mit einer exakteren numerisch aufwendigeren Approximation der Binomialverteilung würde man folgende Werte erhalten:

$$\mathbb{P}_{p=0.9}[X \geq 96] \approx 0.023711$$
$$\mathbb{P}_{p=0.9}[X \geq 95] \approx 0.057577$$

Mit exakten Berechnung wird also für $k \geq 96$ die Nullhypothese verworfen und für $k \leq 95$ kann die Nullhypothese nicht verworfen werden. Somit ist in diesem Fall die erste Lösung die bessere Approximation als Alternative 3. Mit einer besseren Approximation der Binomialverteilung erhält man $\mathbb{P}[X \geq 97] \approx 0.007836 < 0.05$. Hier wäre also Alternative 2 die präzisere Approximation als Alternative 1. Im allgemeinen ist es schwer zu sagen, was die bessere Approximation ist. Daher sollte man bei knappen Ergebnissen eine bessere Approximation verwenden oder den Fehler der Approximation abschätzen.

(b) Die Nullhypothese wird so gewählt, dass das, was man selbst beweisen will, in der Gegenhypothese steht. Hier wird versucht zu überprüfen, ob das Medikament nicht doch schlechter wirkt, als behauptet.
Die Nullhypothese ist somit

$$H_0 : p \geq 0.9,$$

also dass das Medikament mit 90% Wahrscheinlichkeit oder mehr wirkt.
Die Gegenhypothese lautet

$$H_1 : p < 0.9.$$

Wir wollen also k finden, sodass

$$\mathbb{P}_{p=0.9}[X \leq k] \leq 0.05$$

gilt. Analog zu **5.**a können wir den zentralen Grenzwersatz anwednen

$$\mathbb{P}_{p=0.9}[X \leq k] = \mathbb{P}_{p=0.9}\left[\frac{X_1 + \cdots + X_n - 0.9n}{\sqrt[2]{0.09n}} \leq \frac{k - 0.9n}{\sqrt[2]{0.09n}}\right] \overset{\text{CLT}}{\approx} \Phi\left(\frac{k - 0.9n}{\sqrt[2]{0.09n}}\right)$$

und dann analog weiter rechnen:

$$\begin{aligned}
\Phi\left(\frac{k - 0.9n}{\sqrt[2]{0.09n}}\right) \leq 0.05 &\Leftrightarrow \frac{k - 0.9n}{\sqrt[2]{0.09n}} \leq \Phi^{-1}(0.05) \\
&\Leftrightarrow k \leq \Phi^{-1}(0.05)\sqrt[2]{0.09n} + 0.9n \\
&\Leftrightarrow k \leq -1.65\sqrt[2]{0.09 \cdot 100} + 0.9 \cdot 100 \\
&\Leftrightarrow k \leq -1.65 \cdot 3 + 90 \\
&\Leftrightarrow k \leq -4.95 + 90 \\
&\Leftrightarrow k \leq 84.05
\end{aligned}$$

Die Nullhypothese wird bei 84 oder weniger Patienten, bei denen das Medikament wirkt, abgelehnt und demnach bei 85 oder mehr Patienten angenommen.[b]

[a]Die Nullhypothese ist genau genommen $p < 0.9$ und nicht $p = 0.9$. Es reicht allerdings aus den Fall $p = 0.9$ zu betrachten, weil wenn ein grosser Wert k von X unter der Annahme $p = 0.9$ unwahrscheinlich isst, dass ist dieser Wert k für jedes $p < 0.9$ aus der Nullhypothese mindestens gleich unwahrschienlich – i.e. für alle $\tilde{p} < 0.9$ gilt: $\mathbb{P}_{p=0.9}[X \geq k] \leq 0.05 \Rightarrow \mathbb{P}_{p=\tilde{p}}[X \geq k] \leq 0.05$.

[b]Das gleiche Ergebnis würde man auch für exakte Werte erhalten, da eine präzisere Approximation der Binomialverteilung ergibt: $\mathbb{P}[X \leq 85] \approx 0.07257 > 0.05$ und $\mathbb{P}[X \leq 84] \approx 0.03989 < 0.05$. Im allgemeinen kann es aber oft vorkommen, dass man kleine Approximationsfehler erhält. Bei der Approximation einer ganzzahligen Zufallsvaraible durch eine kontinuierliche Zufallsvariable kann man kaum vermeiden, das manchmal fälschlicherweise eine benachbarte ganze Zahl ausgewählt wird, weil $\mathbb{P}[X < k] = \mathbb{P}[X \leq k-1]$ für ganzzahlige Zufallsvariablen gilt, aber typischerweise nicht für die kontinuierliche Approximation von X.

Tabelle der Standardnormalverteilung

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Zum Beispiel ist $P[Z \leq 1.96] = 0.975$.