Übung 7

1. Die Tauben, die in der Stadt Zürich leben, seien erfahrungsgemäss häufig von der Krankheit A befallen. Um das Ausmass des Befalls abzuschätzen, werden zufällig 50 Tauben durch den Stadtjäger geschossen und auf die Krankheit A untersucht. Es wurde festgestellt, dass 12 Tauben an A erkrankt sind.

a) Gib eine Schätzung für die Wahrscheinlichkeit p an, dass eine Taube in der Stadt Zürich die Krankheit A hat. Mit welchem Modell lässt sich das Abschussexperiment beschreiben?

b) Es geht das Gerücht um, dass 50% der Tauben von der Krankheit A befallen sind. Kann das Gerücht stimmen? (Teste mit der Faustregel!)

c) Der P-Wert beträgt in b) 0.0003. Was folgt daraus?

d) Welche Parameterwerte p sind mit dem Ausgang des Experiments vereinbar? Bestimme einerseits mit dem Diagramm (s. Skript) und andererseits mit der 2σ-Regel ein 95%-Vertrauensintervall für p.

2. Bei einer Befragung von Leuten zwischen 20 und 50 hat man folgende Ergebnisse registriert:

- 60% sind Raucher (mehr als 5 Zigaretten durchschnittlich pro Tag)
- 70% sind Trinker (trinken regelmässig Alkohol)
- 35% sind Sportler (treiben regelmässig Sport)

Unter den Rauchern hat es 80% Trinker, aber nur 10% Sportler.

a) Wieviel Prozent sind Trinker und Raucher?

b) Wieviel Prozent treiben keinen Sport und rauchen?

c) Wieviel Prozent sind Raucher oder Trinker?

d) Gib eine möglichst gute obere Schranke an für den Prozentsatz der Leute, welche rauchen oder trinken oder keinen Sport betreiben.

3. Tests mit Lügendetektoren werden oft routinemässig bei Angestellten in den USA durchgeführt, um deren Vertrauenswürdigkeit zu überprüfen. Die möglichen Ereignisse bezeichnen wir wie folgt:

D	Detektor sagt: Lüge	L	Person lügt
D^c	Detektor sagt: keine Lüge	L^c	Person sagt die Wahrheit

Aus Studien über Lügendetektoren kennt man folgende Wahrscheinlichkeiten: $P[D|L] = 0.88$ und $P[D^c|L^c] = 0.86$.

a) Beschreibe die beiden bedingten Wahrscheinlichkeiten $P[D|L]$ und $P[D^c|L^c]$ in Worten.

b) Nehmen wir nun an, dass der Lügendetektor routinemässig eingesetzt wird. Bei einer gewissen Frage hat die Mehrheit der Befragten keinen Grund zum Lügen, also sei $P[L^c] = 0.99$, bzw. $P[L] = 0.01$.
Der Detektor zeigt bei einer Person „Lüge“ an. Wie gross ist nun die Wahrscheinlichkeit, dass die Person trotzdem die Wahrheit gesagt hat?

c) Kommentiere die Bedeutung dieses Resultates.

4. Die Poissonverteilung eignet sich gut, um eine Binomialverteilung $\text{Bin}(n, p)$ mit einem kleinen p und einem grossen n zu approximieren. Die Approximation durch eine Poissonverteilung mit Parameter $\lambda = np$ ist gut, falls $p < 0.05$ und $n > 10$ ist.
Man kann die Poissonverteilung deshalb auch als Verteilung von "seltenen Ereignissen" verstehen: Wenn wir sehr viele (n) mögliche Kandidaten haben, jeder aber nur mit kleiner Wahrscheinlichkeit (p) ein Erfolg ist, so ist das Resultat ungefähr Poisson-verteilt.
Das Ziel dieser Aufgabe ist, durch Simulationen ein Gespür für die Güte der Approximation zu bekommen.

a) Berechne die Wahrscheinlichkeiten $P(X = k)$ $k = 0 \dots 300$ für $X \sim \text{Bin}(300, 0.04)$ und für $X \sim \text{Poisson}(300 \cdot 0.04)$. Zeichne diese Wahrscheinlichkeiten in einen geeigneten Plot ein und berechne den maximalen Fehler, der durch diese Approximation gemacht wird.

b) Mache dasselbe wie in a) für die (kumulativen) Wahrscheinlichkeiten $P(X \leq k)$.

c) Berechne die Wahrscheinlichkeiten $P(X = k)$ $k = 0 \dots 20$ für $X \sim \text{Bin}(10, 0.5)$, $X \sim \text{Bin}(50, 0.1)$, $X \sim \text{Bin}(100, 0.05)$, $X \sim \text{Bin}(200, 0.025)$ und $X \sim \text{Poisson}(\lambda)$ (für ein geeignetes λ). Mache einen geeigneten Plot dieser Wahrscheinlichkeiten und kommentiere ihn.

d) Gibt es auch eine Möglichkeit, eine Binomialverteilung mit grossem p (p nahe bei 1) und grossem n durch die Poissonverteilung zu approximieren? Wie?

R-Anleitung:

a)
```
> binWkeiten <- dbinom(0:300, size=300, prob= 0.04)
                                    # W'keiten für die Binomialverteilung
> poisWkeiten <- dpois(0:300, lambda=300*0.04)
> plot(0:50,binWkeiten[1:51],type="l")
                                    # Liniendiagramm für die Binomialw'keiten
> lines(0:50,poisWkeiten[1:51],col=2,lty=2)
          # Linie für die Poissonw'keiten darüber (mit neuer Farbe 2 und neuem Linientyp 2)
> Fehler <- max(abs(binWkeiten-poisWkeiten))
```

b)
```
                                    # Verwende pbinom und ppois statt dbinom und dpois
```

Abgabe: Montag, 22. Mai in der Pause der Vorlesungsstunde