Übung 4

1. In der nachfolgenden Tabelle wird die Wahrscheinlichkeit angegeben, dass eine männliche Person in der Altersklasse „10(i-1) Jahre < Alter ≤ 10i Jahre“ ($i = 1, ..., 11$) stirbt.

Altersklasse (Jahren)	Wahrscheinlichkeit	Kum. Verteilungsfunktion
00 < Alter ≤ 10	0.010	0.010
10 < Alter ≤ 20	$\ldots a)$	0.016
20 < Alter ≤ 30	0.016	0.032
30 < Alter ≤ 40	0.016	$\ldots b)$
40 < Alter ≤ 50	0.026	0.074
50 < Alter ≤ 60	$\ldots c)$	0.137
60 < Alter ≤ 70	0.153	0.290
70 < Alter ≤ 80	0.287	0.577
80 < Alter ≤ 90	0.322	$\ldots d)$
90 < Alter ≤ 100	0.098	0.997
100 < Alter ≤ 110	$\ldots e)$	1.000

Quelle : Sterbetafeln für die Schweiz 1988/1993, Bundesamt für Statistik

a) Ergänze die obige Tabelle. Zeichne die Wahrscheinlichkeitsfunktion und die kumulative Verteilungsfunktion $F(k)$ über die Altersklassen. Kommentiere die Diagramme. Welche Bedeutung hat die kumulative Verteilungsfunktion $F(k)$ in dieser Aufgabe? Welche Bedeutung hat $1 - F(k)$?

b) Wie gross ist die Wahrscheinlichkeit, dass jemand im Alter zwischen 30 und 60 Jahren ($30 <$ Alter ≤ 60) stirbt ? Berechne diese Wahrscheinlichkeit einmal mit Hilfe der Wahrscheinlichkeitsfunktion und einmal mit Hilfe der kumulativen Verteilungsfunktion.

c) Wie gross ist die Wahrscheinlichkeit, über den 80. Geburtstag hinaus zu leben?

2. Wir betrachten nochmals den Datensatz `milch` aus Übung 3, Aufgabe 4. Dieser Datensatz enthält $n = 400$ Beobachtungen mit arithmetischem Mittel $\overline{x} = 2.44$ und empirischer Varianz $s^2 = 4.6$. Wir haben bereits früher mit der Faustregel gesehen, dass hier die Poissonverteilung nicht passt. In dieser Aufgabe wollen wir diese Tatsache nochmals aus einem anderen Blickwinkel untersuchen.
Der Quotient zwischen der Varianz und dem Erwartungswert ist bei der Poissonverteilung eins (beide sind $= \lambda$). Bestimmt man die entsprechende empirische Grösse „empirische Varianz/arithmetisches Mittel“, so wird der Wert zufällig um 1 herum schwanken. Um mit diesen zufälligen Schwankungen vertraut zu werden, wollen wir sie mit Rsimulieren und dann in einem Histogramm darstellen. Dazu erzeugen wir uns $n = 400$ Pseudo-Zufallszahlen, die poissonverteilt mit Parameter $\lambda = 2.44$ sind, bestimmen daraus die beiden Kennzahlen, und bilden den Quotienten. Dann wiederholen wir dieses Vorgehen 300 mal.

a) Stelle die so erhaltenen 300 Quotienten in einem Histogramm dar.

b) Zwischen welchen Grenzwerten kann dieser Quotient bei einer Poissonverteilung noch ohne weiteres liegen?

R-Anleitung:

```
n ← 400; lambda ← 2.44                              ## Parameter für die Simulation
PZZpois ←  matrix(rpois(300*n, lambda), ncol=n)     ## Matrix mit 300 Zeilen und $n$ Spalten, die mit
                                                    ## Poisson−verteilten Pseudo−Zufallszahlen gefüllt ist
xquer ← apply(PZZpois, MARGIN=1, FUN=mean)          ## zeilenweise mitteln
sigma2 ← apply(PZZpois, MARGIN=1, FUN=var)          ## berechnet zeilenweise die Varianz
Qpois ← sigma2 / xquer                              ## berechnet den Quotienten
hist(Qpois, nclass=40, xlim=c(min(Qpois)-0.1,max(Qpois)+0.1))  ## Histogramm des simulierten
                                                               ## Quotienten für die Poisson−Verteilung
quantile(Qpois,probs=c(0.025,0.975))                ## die Grenzwerte
```

c) Berechne den Quotient $s^2/\overline{x}$ des Datensatzes `milch` (R-Anleitung: s. Aufgabe 4a)). Was kannst du mit Hilfe von b) schliessen?

3. Das „Standard-Zufalls-Experiment“ ist das Würfeln. Wenn wir mit einem „fairen“ Würfel würfeln, dann ist jede Augenzahl k ($k = 1, 2, \dots, 6$) gleichwahrscheinlich, also $P[X = k] = \frac{1}{6}$ für alle $k = 1, 2, \dots, 6$.
Allgemein können wir zufällige Ereignisse betrachten, welche die Werte $k = 1, 2, \dots, n$ alle mit gleicher Wahrscheinlichkeit $P[X = k] = \frac{1}{n}$ annehmen. Man nennt diese Verteilung auch *diskrete Gleichverteilung.* (Für $n = 2$ können wir uns z.B. das Werfen einer fairen Münze vorstellen, wobei “1” Kopf und “2” Zahl bedeutet; für $n = 6$ erhalten wir das Modell des Würfelns.)
Berechne für diese Wahrscheinlichkeitsverteilung (in Abhängigkeit von n) den Erwartungswert, die Varianz und die Standardabweichung.
Wie gross sind diese drei Kennzahlen numerisch für das Beispiel des Würfelns? Wie gross ist beim Würfeln die Wahrscheinlichkeit, dass sich eine Beobachtung ausserhalb des Erwartungswertes ± einer Standardabweichung befindet?

4. Falls die Poissonverteilung die Daten „schlecht“ beschreibt, ist die *Negative Binomialverteilung* eine naheliegende, einfache Alternative. Sie kommt in der Ökologie häufig vor. Diese Verteilung ist gegeben durch die Wahrscheinlichkeitsfunktion

$$P\langle X = x\rangle = \binom{x+\kappa-1}{x}\pi^x(1-\pi)^\kappa := \begin{cases} (1-\pi)^\kappa & : \text{ für } x = 0 \\ \frac{(x+\kappa-1)(x+\kappa-2)\dots\kappa}{x(x-1)\dots 1}\pi^x(1-\pi)^\kappa & : \text{ sonst} \end{cases}$$

für $x = 0, 1, 2, \dots$. Sie hat zwei Parameter, π mit $0 < \pi < 1$ und $\kappa > 0$. Der Erwartungswert ist $\kappa\pi/(1-\pi)$, die Varianz $\kappa\pi/(1-\pi)^2$.

Um die Parameter π und κ aus den Daten zu bestimmen, werden oft die theoretischen „Momente“ (Erwartungswert und Varianz) den empirischen (arithm. Mittel und empir. Varianz) gleichgesetzt. Daraus ergeben sich die Schätzungen

$$\widehat{\pi} = 1 - \bar{x}/s^2\,; \qquad \widehat{\kappa} = \bar{x}^2/(s^2 - \bar{x}).$$

Wir haben bereits in Übung 3, Aufgabe 3 und in der Aufgabe 2 in dieser Übung gesehen, dass der Datensatz `milch` nicht durch eine Poissonverteilung modelliert werden kann. Deshalb passen wir jetzt eine Negative Binomialverteilung an diese Daten an und überprüfen mit der Faustregel, wie gut sie passt.

a) Schätze die Parameter π und κ.
R-Anleitung:
```
> milch <- read.table.url("http://stat.ethz.ch/Teaching/Datasets/bacterial.clumps.dat",
header = T)
> milchdaten <- rep(milch[, "n"], milch[, "frequency"])          # sortierte "Rohdaten"
> mittel <- mean(milchdaten)                                      # x̄
> varianz <- var(milchdaten)
> PI <- ?                                                         # Parameter π̂
> kappa <- ?                                                      # κ̂
```

b) Berechne die Grössen $n\widehat{p}_k$
R-Anleitung: Im R wird für die Negative Binomial-Verteilung der Parameter $p = 1 - \pi$ verwendet!
```
> pk <- dnbinom(0:19,size=kappa,prob=1-PI)          # Wahrscheinlichkeit je Klasse
> npkErw <- ?                                       # Erwartete Anzahl je Klasse
```

c) Zeichne ein Stabdiagramm der Daten, trage die Punkte $n\widehat{p}_k$ ein und zeichne Linien ein, die den Intervallen $n\widehat{p}_k \pm 2 * \sqrt{n\widehat{p}_k}$ entsprechen. Beurteile, wie gut die Negative Binomial-Verteilung passt.
R-Anleitung:
```
> plot(milch[, "n"], milch[, "frequency"], type = "h")
> points(0:19 + 0.2, npkErw, pch = "*")
> for (i in 1:20)
> lines(rep(0:19 + 0.2, 2), npkErw[i] + 2 * sqrt(npkErw[i]) * c(-1,1), lty = 3)
```

Abgabe: Dienstag, 2. Mai, in der Übungsstunde