Prof. Antony Unwin, Alexander Pilhöfer
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
http://stats.math.uni-augsburg.de/

Wahrscheinlichkeitstheorie
Übungsblatt 4

Abgabe: *Donnerstag 17. November 2011, bis spätestens 12.00 Uhr*; Briefkasten: WTheorie oder per email

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

Achten Sie auf saubere mathematische Arbeitsweise!

1. (a) Sei $P(X = x_i) = \frac{c}{5^i}$ für (i=1,2,...). Für welches c liegt eine Wahrscheinlichkeitsverteilung vor? **(1P)**
   (b) Leiten Sie den Erwartungswert einer Poissonverteilten Zufallsvariable her! **(1P)**
   (c) Beweisen oder widerlegen Sie: Gilt $P(B|\overline{A}) = P(B|A)$, so sind A und B unabhängig! **(1P)**
   (d) Berechne mit Hilfe einer WEF die Varianz eines Würfelwurfs! **(1P)**

2. **WEF I (5P)**
   Beweisen Sie die nachfolgenden Aussagen über die wahrscheinlichkeitserzeugende Funktion.
   (a) Für X mit endlichem k. Moment ($k \in \mathbb{Z}^+$):
   $$E\left[\binom{X}{k}\right] = \frac{g_X^{(k)}(1)}{k!}.$$
   (b) Unter Benutzung vorheriger Teilaufgabe (a):
   $$\begin{aligned} E(X) &= g_X^{(1)}(1), \\ V(X) &= g_X^{(2)}(1) + g_X^{(1)}(1) - \left(g_X^{(1)}(1)\right)^2. \end{aligned}$$
   (c) Die Wahrscheinlichkeiten der Zufallsvariable werden in folgendem Sinne erzeugt:
   $$P(X = k) = \frac{g_X^{(k)}(0)}{k!}.$$

3. **Lose (5P)**
   An Studenten werden vor der Mensa je m Lose verteilt, die jeweils mit Wahrscheinlichkeit p ein Gewinn sind. Unter den Gewinnern werden Mensagutscheine im Wert von EUR 1, 2 und 3 verlost. Dabei ist die Wahrscheinlichkeit für einen Gutschein im Wert von EUR j gleich p_j, $j = 1, 2, 3$.
   (a) Bestimmen Sie die wahrscheinlichkeitserzeugende Funktion für die Gewinnsumme, die der Veranstalter pro Student bereitstellen muss. Leiten Sie daraus den Erwartungswert und die Varianz dieser Größe ab. Nehmen Sie dabei an, daß alle Personen ihren Gewinn einlösen.
   (b) Bestimmen Sie die WEF für den Gesamtgewinn pro Person, wenn die Anzahl der Lose pro Student Poissonverteilt ist mit Parameter λ!

(c) Die Lotterie beschließt, an jeden der derzeit 15.000 Studenten je m Lose zu verteilen.

   i. Welche Verteilung könnte man dann für den Durchschnitt des Gesamtgewinns einer zufälligen Stichprobe von 50 Personen anwenden? Bestimmen Sie die entsprechende WEF!

   ii. Welche Verteilung könnte man für den Gesamtgewinn pro Student benutzen, wenn die Anzahl m der Lose pro Student sehr groß wird? Bestimmen Sie die entsprechende WEF!

4. **Diskrete Verteilungen in R (5P)**

Die Umsetzung von Verteilungen in R erfolgt über Funktionen bestehend aus einem Schlüsselwort (z.B. *binom, nbinom, geom, pois, hyper*) und einem der Präfixe *d, p, q, r* (z.B. *qbinom*). Die Hilfe zu den Funktionen rufen Sie beispielsweise mittels `?qbinom` auf. Machen Sie sich zunächst mit der Umgangsweise vertraut.

(a) Erstellen Sie eine Grafik, die die Dichte einer Binomialverteilung mit $n = 20$ und $p = 0.2$ im Bereich 0 bis 20 zeigt.

```
barplot(dbinom(0:20,20,0.2),ylim=c(0,0.2))
```

(b) Vergleichen Sie obige Binomialverteilung mit der Dichte einer hypergeometrischen Verteilung mit gleichem Erwartungswert und $N = 100$. Verwenden Sie die gleiche y-Skala. Ist das Ergebnis erwartungsgemäß?

(c) Plotten Sie nun unter Zuhilfenahme des Befehls `par(mfrow = c(1,3))` die Differenzen der hypergeometrischen Dichten mit $N = 100, 200, 1000$ zur Binomialverteilungsdichte auf der y-Skala $[-0.02, 0.03]$! Interpretieren Sie Ihre Ergebnisse!

(d) In ähnlicher Weise wollen Sie nun die Binomialverteilung mit der Poissonverteilung vergleichen. Welche Parameter müssen Sie ändern, um wieder eine Annäherung zu erreichen? Führen Sie dies in R durch!

(e) Vergleichen Sie nun die Verteilungsfunktionen ($F(x) = P(X \leq x)$) von $NB(10, 0.25)$ und $Pois(\lambda = 30)$. Interpretieren Sie Ihr Ergebnis!

5. **Huftritte (5P)**

Eine in der Literatur häufig zitierte empirische Untersuchung behandelt die Anzahl der Soldaten eines preußischen Kavallerieregiments, die innerhalb eines Jahres an den Folgen eines Huftritts starben. Für 10 Regimenter wurden über einen Zeitraum von 20 Jahren die entsprechenden Zahlen ermittelt:

Anzahl Todesfälle	Beobachtete Häufigkeit
0	109
1	65
2	22
3	3
4	1

Folgende Teilaufgaben sollen in `R` behandelt werden.

(Zur "Erzeugung" der Daten in R verwenden Sie z.B. `x <- 0:4` und `y <- c(109, 65, 22,3,1)`)

(a) Was würden Sie als "mittlere Anzahl" von Todesfällen in einem Jahr ansehen? Berechnen Sie für eine Poisson-Verteilung mit dieser "mittleren Anzahl" die Wahrscheinlichkeiten dafür, die Werte 0, 1, 2, 3, und 4 anzunehmen.

(b) Plotten Sie eine geeignete Graphik, die diese berechneten Wahrscheinlichkeiten mit den beobachteten relativen Häufigkeiten vergleichend darstellt. Beurteilen Sie den Plot im Hinblick auf Übereinstimmung. Wie gut beschreibt die Poisson-Verteilung die Daten?

(c) Studieren Sie die `R`-Hilfeseite zur Funktion `sample()` und benutzen Sie diese Funktion, um eine zufällige Stichprobe (aus den Daten) vom Umfang 40 zu ziehen. Wenden Sie vorherige Teilaufgaben (a) und (b) auf diesen neuen "Datensatz" an. Beschreiben Sie Ihre Ergebnisse.