



Prof. Dr. Antony Unwin, Dr. Ali Ünlü
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Stochastik I – Wahrscheinlichkeitstheorie (WS 2008/09)

Übungsblatt 5

Abgabe: Mittwoch 19. November 2008, bis spätestens 12.00 Uhr; Briefkasten: Wahrscheinlichkeitstheorie

1. 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:

$$P(X = k) = \frac{g_X^{(k)}(0)}{k!}.$$

2. An Studenten werden vor der Mensa 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) Die Lotterie beschließt, Lose an alle derzeit 15.000 Studenten zu verteilen.

- Welche Verteilung könnte man dann für den Durchschnitt des Gesamtgewinns einer zufälligen Stichprobe von 50 Personen anwenden?
- Welche Verteilung könnte man für den Gesamtgewinn pro Student benutzen, wenn die Anzahl m der Lose pro Student sehr groß wird?

3. Student A. der Vorlesung "W-theorie" hat den Virus `click_me.exe` als e-mail Attachment bekommen und doch tatsächlich seinen Computer infiziert. Der Virus sucht sich aus seinem Adressbuch neue Opfer aus, und schickt diesen eine infizierte e-mail. Die Anzahl der Kommilitonen, die das Attachment öffnen, und sich damit infizieren, sei Poisson-verteilt mit Parameter λ .

(a) Für welche Werte von λ werden nicht alle e-mail Benutzer infiziert?

(b) Mit welcher Wahrscheinlichkeit ist der Virus nach 7 Generationen verschwunden, wenn man annehmen kann, dass $\lambda = 0.7$ gilt?

- (c) Welche Voraussetzungen müssen gelten, damit vorherige Teilaufgaben (a) und (b) beantwortet werden können?
4. 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.

- (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 vom Umfang 40 zu ziehen. Wenden Sie vorherige Teilaufgaben (a) und (b) auf diesen neuen “Datensatz” an. Beschreiben Sie Ihre Ergebnisse.
5. Betrachtet wird die *Binomialapproximation der hypergeometrischen Verteilung*, d.h. die Annäherung der hypergeometrischen Verteilung an die Binomial-Verteilung. Plotten Sie in R die Wahrscheinlichkeitsfunktionen der Verteilungen

Hypergeometrisch mit Parametern $n = 10, N = 20, K = 5,$

Hypergeometrisch mit Parametern $n = 10, N = 40, K = 10,$

Hypergeometrisch mit Parametern $n = 10, N = 160, K = 40,$

Binomial mit Parametern $N = 10, p = 1/4$

in einem einzigen Plot, in Linienzügen verschiedener Farben oder Formen. Was sehen Sie anhand der Graphik? Welcher Parameter und welches Verhältnis zweier Parameter sind über alle drei hypergeometrischen Verteilungen hinweg konstant geblieben? Wie hängen letztere Beobachtungen mit der genannten Binomial-Verteilung zusammen? Begründen Sie intuitiv, warum man solch eine Annäherung erwarten würde.