Dr. Quapp: Statistik für Mathematiker mit SPSS

Hilfe zur 7. Übung – Operationen mit Zufallsvariablen

1.] Es sei S die summierte Augenzahl beim Wurf mit 3 Würfeln.
a) Bestimmen Sie die Einzelwahrscheinlichkeiten $P(S=k)$ für $k=3,...,18$. Für welches k ist $P(S=k)$ maximal?
b) Erzeugen Sie dieses Experiment mit SPSS, indem Sie drei Variable X, Y und Z aus gleichmäßig verteilten Zufallszahlen erzeugen, und S als Summe dieser Zufallszahlen darstellen. Prüfen Sie das Resultat durch ein Balkendiagramm.
c) Mit B bezeichnen wir das Ereignis, daß bei einem Wurf alle Augenzahlen unterschiedlich sind. Bestimmen Sie $P(B)$ und bestimmen Sie $P(S=k|B)$ für $k=3,4,...,18$.

Da die Zufallsvariablen addiert werden, muß ihre Verteilungsdichte aus einer Faltung berechnet werden. Für zwei Zufallsvariable X und Y ergibt sich beim Würfel:

$$P_1(X=i)=P_1(Y=i)=\begin{cases}\frac{1}{6}, & \text{bei } i=1,...,6 \text{ und}\\ 0 & \text{sonst}\end{cases}$$

Bei $W=X+Y$ erhalten wir

$$P_2(W=l)=\begin{cases}\sum_{i=1}^{6} P_1(X=i)\,P_1(Y=l-i) & \text{bei } l=2,...,12, \text{und}\\ 0 & \text{sonst}\end{cases}$$

d.h. die beiden Gleichverteilungen der Würfe X und Y werden gegenläufig "gefaltet", daher der Name der Operation. $P_2(W)$ wird eine "Dreiecksverteilung". Wird noch eine dritte Zufallsvariable Z hinzugenommen, erhalten wir für $S=X+Y+Z=W+Z$ sukzessive

$$P_3(S=k)=\begin{cases}\sum_{l=2}^{12} P_2(W=l)\,P_1(Z=k-l) & \text{bei } k=3,...,18, \text{und}\\ 0 & \text{sonst}\end{cases}$$

oder

$$P_3(S=k)=\sum_{l=2}^{12}\sum_{i=1}^{6} P_1(X=i)\,P_1(Y=l-i)\,P_1(Z=k-l) \text{ bei } k=3,...,18, \text{und } 0 \text{ sonst.}$$

Es ist

k	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$P(k)\cdot 216$	1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1

Das Maximum von $P_3(S)$ liegt bei $k=10,11$.

zu b) Die Zufallsvariablen X, Y, Z sind aus einer Gleichverteilung in $[1.0, 6.9999]$ durch Abschneiden auf ganze Zahlen erzeugbar: TRUNC(RV.UNIFORM(1,6.9999)). Man verwende etwa 216 Werte, oder Vielfache davon: $n \times 216$.

zu c) Die Anzahl der günstigen Fälle ist eine Variation ohne Wiederholung mit n=6, l=3, also 120 Fälle, damit ist $P(B) = 120/216 = 5/9$. Die bedingte Wahrscheinlichkeit, wenn B eingetreten ist, kann leicht abgezählt werden: Bei $k = 3, 4, 5, 16, 17, 18$ ist sie Null (diese Fälle kommen nicht vor), bei $k = 6, 7, 14, 15$ ist sie je 3!, bei $k = 8, 13$ ist sie je $2 \times 3!$, und bei $k = 9, 10, 11, 12$ ist sie je $3 \times 3!$.

In SPSS kann eine neue Variable definiert werden unter Zuhilfenahme des Falls-Fensters: $SB = X+Y+Z$ $\rightarrow$ Falls $(X \sim= Y)\&(X \sim= Z)\&(Y \sim= Z)$ ist. Die Statistik von SB sollte obige Werte annähern. In einem Balkendiagramm kann man S oder SB ansehen, wobei bei letzterem noch unter $\rightarrow$ Optionen die Darstellung der "Fehlenden Werte" abgeschaltet werden sollte.

2.] a) Für welche Werte a ist $f(x) = a/(e^x + e^{-x})$ eine Dichte? Bestimmen Sie die Verteilungsfunktion F zu dieser Dichte!
b) Es sei $U \sim U(0, 1)$. Bestimmen Sie die Dichte und Verteilungsfunktion von $Y := F^{-1}(U)$, wobei F die Verteilungsfunktion aus a) ist (Inversionsmethode).
c) Bestimmen Sie mir SPSS Modalwert, Median, und Quartile von F. Benutzen Sie dazu N=200 Werte von $U \sim U(0, 1)$.

Es ist $a = 2/\pi$.

$$F(x) = \frac{2}{\pi} \int_0^{e^x} \frac{dz}{1 + z^2} = \frac{2}{\pi}\, arctan(e^x)$$

also wird

$$y = F^{-1}(u) = ln(tan(\frac{\pi}{2}\, u)).$$

Wenn $u \in U(0, 1)$ und $y = F^{-1}(u)$, dann ist

$$P(Y \leq y) = P(U \leq F(y)) = F(y) = \frac{2}{\pi} arctan(e^y)$$

Somit kann man aus der gleichverteilten Zufallsvariablen U mit obiger Umrechnung die "F(y)"-verteilte Zufallsvariable Y machen! Berechnung in SPSS:

$$Y = LN(SIN(\frac{\pi}{2}\, U)/COS(\frac{\pi}{2}\, U))$$

Sortiert man die Y-Werte, und ordnet ihnen den Wert \$casenum/N zu, so kann man auch wieder die empirische Verteilungsfunktion von Y darstellen.

3.] Berechnen Sie 200 Werte der Zufallsvariablen X mit Gleichverteilung $U(0,a)$ mit etwa a=10. Berechnen Sie die neue Zufallsvariable

$$Y = X^3 \ .$$

Analysieren Sie die Verteilung von Y. Bestimmen Sie theoretisch und in SPSS approximativ die Wahrscheinlichkeit $P(a < Y < 2a)$, z.B. für $a = 10$. Bestimmen Sie Erwartungswert und Varianz von Y !

Sei X eine Zufallsvariable und $Y = h(X)$, wobei h z.B. stückweise monoton und differenzierbar sei. Dann ist die Dichte von Y

$$f_Y(y) = f_X(h^{-1}(y)) \frac{1}{|h'(h^{-1}(y))|}$$

Da X gleichverteilt ist mit $U(0,a)$, wird

$$f_Y(y) = \frac{1}{3ay^{2/3}} \quad \text{bei} \ \ 0 < y \leq a^3 \ \text{und} \quad 0 \quad \text{sonst,}$$

also erhalten wir

$$P(a < Y < 2a) = \begin{cases} 0 & \text{bei } a < 1 \\ 1 - a^{-2/3} & \text{bei } 1 < a \leq \sqrt{2} \\ a^{-2/3}(2^{1/3} - 1) & \text{bei } \sqrt{2} < a \end{cases}$$

Werden die Ausprägungen von Y sortiert, so kann die Grenze bei Y=10 und bei Y=20 abgelesen werden, und es ergab sich im Test als Resultat die Wahrscheinlichkeit zu 0,2708-0,2156=0,0551. Theoretisch sollte herauskommen:

$$\frac{1}{\sqrt[3]{100}}(\sqrt[3]{2} - 1) = 0,0560.$$

Eine zweite Version ist die Belegung einer Zählvariable im Intervall $(10 < Y)\&(Y < 20)$ etwa mit 1, und ein Aufsummieren dieser Einsen, gegen die Gesamtzahl der Zeilen.