Lösungen zur Übung (2)

(1) Man beachte noch einmal die Definition der Verteilungsfunktion einer Zufallsvariablen $X$: $F_X(a) = P(X \leq a)$, der Funktionswert an der Stelle $a$ ist also die relative Häufigkeit oder allgemeiner Wahrscheinlichkeit für das Auftreten von $X-$ Werten, die $\leq a$ sind. Das ergibt für diskrete Verteilungen zwingend die Treppengestalt mit Sprüngen bei jedem Wert, der mit Wahrscheinlichkeit $> 0$ auftritt. (Das Anwenden dieser Definition bereitet gemeinhin recht große Probleme, die jedoch ziemlich verschwinden, wenn man einmal geruht, sich diese zentrale Definition zu merken.)

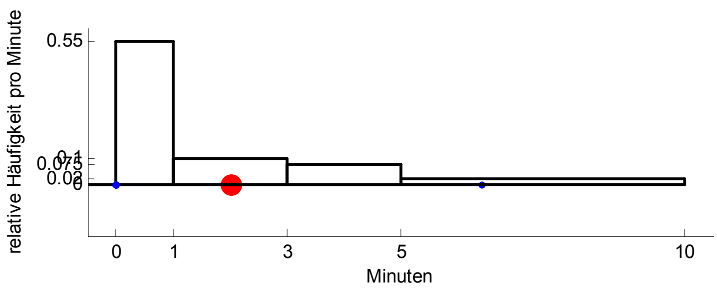
(2) Hier sind 'gruppierte Daten' zur Verteilung der Rededauer der Teilnehmer von Diskussionsgruppen:

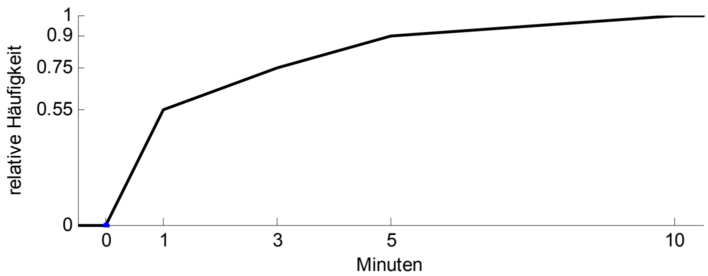
| Dauer (Minuten) | $0-1$ | $1-3$ | $3-5$ | $5-10$ |
|---|---|---|---|---|
| relative Häufigkeit | 0.55 | 0.2 | 0.15 | 0.1 |

(a) Zugehörige Tabelle der Dichtewerte - man beachte: Dichte ist jeweils die relative Häufigkeit geteilt durch Intervallbreite (!):

| Dauer (Minuten) | $0-1$ | $1-3$ | $3-5$ | $5-10$ |
|---|---|---|---|---|
| relative Häufigkeit | 0.55 | 0.1 | 0.075 | 0.02 |

Das Histogramm (und gleich parallel die Verteilungsfunktion sehen so aus:




Die Form sollte plausibel sein: In einer Gruppe sagen Viele wenig, und Wenige viel. (Es handelt sich (u.a.) um einen Rückkopplungseffekt.)

(b) Mittelwertberechnung (näherungsweise) - mit $X$ wird die Variable: 'Dauer der Diskussionsbeiträge in Minuten' bezeichnet:

$$\mu(X) \approx 0.5 \cdot 0.55 + 2 \cdot 0.2 + 4 \cdot 0.15 + 7.5 \cdot 0.1 = 2.025 \text{ [Minuten]}.$$

Näherungswerte für $\sigma^2(X)$ und $\sigma(X)$ :

$$\begin{aligned}
\sigma^2(X) &\approx (0.5 - 2.025)^2 \cdot 0.55 + (2 - 2.025)^2 \cdot 0.2 + (4 - 2.025)^2 \cdot 0.15 \\
&\quad + (7.5 - 2.025)^2 \cdot 0.1 \\
&= 4.861\,88, \text{ also} \\
\sigma(X) &\approx 2.205 \text{ (Auch 2.2 wäre genau genug!)}
\end{aligned}$$

(c) Der Bereich $\mu \pm 2\sigma$ ist im Beispiel $[-2.38; 6.435]$, und der enthaltene Populationsanteil berechnet sich (gemäß Histogramm, also Gleichverteilung auf den Gruppierungsintervallen voraussetzend!) zu

$$0.55 + 0.2 + 0.15 + \frac{1.435}{5} \cdot 0.1 \approx 0.93.$$

Das ist ganz wie von der Faustregel versprochen.

(d) Die Wertetabelle für die Verteilungsfunktion:

| Dauer (Minuten) $a$ | 0 | 1 | 3 | 5 | 10 |
|---|---|---|---|---|---|
| $F_X(a) = P(X \leq a)$ | 0 | 0.55 | 0.75 | 0.9 | 1 |

Graphisch ablesen und auch rechnen kann man den Median zu:

$$\frac{0.5}{0.55} \approx 0.91,$$

deutlich kleiner als der Erwartungswert, wie nach der Verteilungsform zu erwarten.

(e) Die Steigung der Verteilungsfunktion im Intervall $[1, 3]$ ist $\frac{0.2}{2} = 0.1$, und das ist genau der (gemäß Idealisierung konstante) Dichtewert auf diesem Intervall (relative Häufigkeit pro Minute).

(3) Symmetrische Glockenform: Zufalls-Messfehler bei physikalischen Messungen, allgemeiner Variablen wie Körperlänge oder Leistungsfähigkeiten, die durch Mittelung vieler unabhängiger Variablen entstehen. Asymmetrische Glockenformen mit Gipfel auf der linken Seite hat man sehr häufig bei Variablen, deren Werte etwa nur positiv sein können, nach oben hingegen 'ausreißen' können, während ein Gipfel etwas links vom Erwartungswert liegt. Typisch: Kinderzahl, tägliche Telefonminuten, Anzahl der Sechsen bei 100 Würfen mit einem Würfel. Typisch auch: Einkommensverteilung in einer einigermaßen entwickelten Wirtschaft. Zweigipflige Verteilungen treten typisch auf, wenn man zwei Populationsgruppen hat, innerhalb deren eine Variable $X$ Normalverteilungen besitzt, mit verschiedenen Mittelwerten für die beiden Gruppen. Beispiel: Körperlänge bei erwachsenen Menschen (die beiden Gruppen sind hier die Geschlechter). Für J-Form hatten wir oben ein Beispiel. Auch eine Einkommensverteilung in völlig unterentwickelter feudalistischer Wirtschaft sieht so aus. Eine J-Form in anderer Richtung hätte man dann bei einer Variablen, welche in einer solchen Population so etwas wie 'Armut' misst, derart, dass höhere Zahlenwerte größere Armut bedeuten. Man kann sich auch J-Form ansteigend vorstellen bei einer Variablen der Art: 'Benötigte Zeit für das Erbringen einer gewissen Leistung' bei einer Tätigkeit, welche große Übung erfordert, um Verbesserungen zu erzielen.

(4) Direkt nach dem qualitativen Begriff der Streuung hat man: Große Portionen der Population weit ab vom Mittelwert von $X$ bedeuten eine relativ große Streuung. Also hat man relativ kleine Streuung bei J-Form und bei Glockenform (in der Klasse der Glockenformen natürlich größere Streuung bei flacherer Glocke), man nähert sich der Gleichverteilung mit ihrer recht großen Streuung und landet bei der U-Form, deren (diskretes) Extrem (wie in früherer Aufgabe bereits gesehen) die maximal mögliche Streuung ergibt.

(5) **Hausaufgabe:**

(a) Die drei Zuordnungen lauten: $a \mapsto P(X = a)$, $a \mapsto$ Verteilungsdichte an der Stelle $a$, $a \mapsto P(X \leq a)$.

(b) (i) Abfallen einer Dichte auf einem Intervall $I$ bedeutet, dass die auf ein kleines Intervall um einen höheren Wert in $I$ eine geringere Wahrscheinlichkeit entfällt als auf ein kleines Intervall um einen kleineren Wert aus $I$. So ist es jenseits des Gipfels einer Normalverteilung z.B. Dagegen ist das Ereignis $X \leq a$ eine Teilmenge des Ereignisses $X \leq b$ für $a \leq b$, also $P(X \leq a) \leq P(X \leq b)$ für $a \leq b$. Eine Verteilungsfunktion kann also nirgends fallen.

(ii) $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$. Für eine Verteilung mit Dichte ist $P(a \leq X \leq b) = P(a < X \leq b)$, aber wenn $P(X = a) \neq 0$, so müsste man eine Zahl $c < a$ nehmen derart, dass $X$ keinen Wert im Bereich $[c, a[$ annimmt, und dann $F_X(a) - F_X(c) = P(a \leq X \leq b)$ bilden, oder natürlich auch $P(a \leq X \leq b) = F_X(c) - F_X(a) + P(X = a)$, was aber das Verlangte nicht rein in der Verteilungsfunktion ausdrückt.

(iii) Únter den gegebenen Voraussetzungen ist $F_X(a) = \sum_{k=1}^{i} p_k$. Ferner $F_X(a_n) = 1$.

(iv) Im angegebenen Falle ist $P(X = a) = 0$. Das weiß man dann für jeden Wert von $a$, und daraus erhält man keinerlei Information über die Verteilung. Zweite Frage: Die Streuung muss dann sehr klein sein. Dritte Frage: In einem Bereich sehr niedriger Dichtewerte ist die Verteilungsfunktion beinahe konstant.