Statistik für Informatiker **SS 2017**

Prof. Dr. Matthias Birkner
M.Sc. Frederik Klement

Blatt 3

Aufgabe 1: (2 + 3 Punkte)

a): Wenn Sie Ihren alten Rechner starten, dann fährt dieser bei jedem Versuch unabhängig von allen vorherigen Versuchen nur mit einer Wahrscheinlichkeit von 80% hoch. Sei X die Zahl ihrer glücklosen Versuche, bis es zum ersten Mal klappt. Mit welcher Wahrscheinlichkeit brauchen Sie genau 3 Versuche ($\{X = 2\}$ tritt ein)? Mit welcher Wahrscheinlichkeit brauchen Sie mindestens 3 Versuche ($\{X \geq 2\}$ tritt ein)?

b): Seien $p_1, ..., p_n \in (0, 1)$ mit $\sum_{j=1}^{n} p_j = 1$. Wie in Aufgabe 2 von Blatt 0 erzeugen wir einen zufälligen String $Str := (C_1, ..., C_k)$ aus der Menge $S := \{s_1, ..., s_n\}$. Die Character $C_1, ..., C_k$ sind unabhängig voneinander und für jeden Character C_i und jedes Symbol s_j gilt $P(C_i = s_j) = p_j$. Sei M_j die Häufigkeit des Symbols s_j im String Str für $j \in \{1, .., n\}$. Seien $m_1, ..., m_n$ mit $\sum_{i=1}^{n} m_i = k$. Geben Sie einen Ausdruck ohne Herleitung an für

$$P(M_1 = m_1, ..., M_n = m_n).$$

Seien nun $k = 5$, $S = \{s_1 = a, s_2 = b, s_3 = c\}$ und $p_1 = 0.5, p_2 = 0.2, p_3 = 0.3$. Bestimmen Sie den Wert von

$$P(M_1 = 2, M_2 = 1, M_3 = 2).$$

Wir interessieren uns nun nur für $M_3 + M_1$. Bestimmen Sie unter den gleichen Voraussetzungen wie zuvor den Wert von

$$P(M_3 + M_1 = 3).$$

Hinweis: Sie kennen die Verteilungen von X, $M = (M_1, ..., M_n)$ und $M_3 + M_1$ aus der Vorlesung.

Aufgabe 2 (2 + 1 + 2 Punkte)
Sei $(B_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen bernoulliverteilten Zufallsvariablen mit Erfolgsparameter $p \in (0, 1)$ ($B_i \sim \mathrm{Ber}_p$). Wir definieren für $k \in \mathbb{N}$

$$X_k := \min\left\{n > 0 : \sum_{i=1}^{n} B_i = k\right\} - k.$$

Wenn wir $(B_i)_{i\in\mathbb{N}}$ als eine Folge von unabhängigen Versuchen interpretieren, so ist X_k die Zahl der Misserfolge bis zum k-ten Erfolg. Gilt $X_k = n$, so bedeutet dies, dass der k-te Erfolg genau beim $n+k$-ten Versuch eingetroffen ist. Wir interessieren uns für die Verteilung von X_k für $k > 0$, wir wollen den Wert von $P(X_k = n)$ herleiten.

a): Man kann anhand der ersten $n + k$ Versuche, $(B_1, ..., B_{n+k})$, erkennen, ob das Ereignis $\{X_k = n\}$ eingetroffen ist oder nicht. Bestimmen Sie $A \subseteq \{0, 1\}^{n+k}$, so dass

$$\{X_k = n\} = \{(B_1, ..., B_{n+k}) \in A\}.$$

b): Sei $b = (b_1, ..., b_{n+k}) \in A \subset \{0, 1\}^{n+k}$. Bestimmen Sie $P(B_1 = b_1, ..., B_{n+k} = b_{n+k})$.

c): Überlegen Sie sich, wieviele Elemente A besitzt, und folgern Sie daraus zusammen mit **a)** und **b)** den Wert von $P(X_k = n)$.

Aufgabe 3: (2 + 2 + 2 Punkte)
Eine Internetfirma besitzt k Server, auf welche sich u User zufällig verteilen. Jeder User wählt einen Server, hierbei wählt er jeden Server mit gleicher Wahrscheinlichkeit. Daneben agiert jeder User unabhängig von den anderen Usern. Wenn X_i für $i \in \{1, ..., k\}$ die Zahl der User auf Server k entspricht, so ist die gemeinsame Verteilung von $X := (X_1, ..., X_k)$ durch die Multinomialverteilung mit Parametern u und $p_i = 1/k$ für $i \in \{1, ..., k\}$ gegeben. Ein Server gilt als überlastet, wenn sich mehr als 300 User auf diesem befinden.

a): Schreiben Sie eine Funktion `server(u,k)`, die zuerst den Vektor X mithilfe von `rmultinom` generiert und die Zahl Z der Server mit mehr als 300 Usern zurück gibt.

b): Führen Sie 10^4 Simulationen für $u = 60000$ und $k = 225$ der Zufallsvariable Z durch und erstellen Sie ein Histogramm-Plot Ihres Ergebnisses. Geben Sie den Plot eine passende Beschriftung und sorgen Sie mit `breaks=min(S):(max(S)+1)-0.5` für eine gute Balkeneinteilung, wobei der Vektor S die Ergebnisse ihrer Simulation enthalte.

c): Führen Sie jeweils 10^3 Simulationen für $u = 60000$ und $k = 220, 221, ..., 250$ der Zufallsvariable Z durch. Schätzen Sie basierend auf ihren Simulationen für jedes k die Wahrscheinlichkeit $P(Z > 0)$ und erstellen Sie einen Graphen mit `plot`, der die Abhängigkeit von $P(Z > 0)$ von k darstellt. Geben Sie dem Plot eine passende Beschriftung.

Achtung: Zusätzlich zur elektronischen Abgabe per Mail, drucken Sie bitte Ihren Code aus und reichen Sie diesen zusammen mit Ihrer restlichen Abgabe ein.

Abgabe: 10 Uhr 19.05.2017