ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik
Musterlösung 8

1. Verteilung der Summe von Zufallsvariablen

Für zwei stetige Zufallsvariablen X und Y mit gemeinsamer Dichtefunktion $f_{X,Y}$ gilt (ohne Beweis), dass ihre Summe $S = X + Y$ die folgende Dichte hat

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(x, s-x)dx.$$

Im diskreten Fall, wenn $W_X, W_Y \subset \mathbb{Z}$, gilt äquivalent dazu

$$P(S = s) = \sum_{i=-\infty}^{\infty} P(X = i, Y = s - i).$$

Bemerke, dass für unabhängige X, Y gilt $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ für alle $(x, y) \in \mathbb{R}^2$, bzw. $P(X = x, Y = y) = P(X = x)P(Y = y)$ für alle $(x, y) \in \mathbb{Z}^2$.

a) Bestimme sowohl im stetigen als auch im diskreten Fall die kumulative Verteilungsfunktion von S.

b) Seien $X_i \sim \text{Poiss}(\lambda_i)$ unabhängig mit $\lambda_i > 0$ für $1 \le i \le 2$. Zeige, dass $S = X_1 + X_2$ einer $\text{Pois}(\lambda_1 + \lambda_2)$ Verteilung folgt.

Tipp: Verwende den binomischen Lehrsatz: $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

c) Sei $(X_1, X_2) \sim \mathrm{N}(\mu, \Sigma)$ ein zweidimensional normalverteilter Zufallsvektor, mit $\mu = (\mu_1, \mu_2)$ und $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho \\ \sigma_1\sigma_2\varrho & \sigma_2^2 \end{pmatrix}$, wobei $\varrho := \mathrm{Corr}(X_1, X_2)$. Zeige, dass gilt

$$S := X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\varrho).$$

Tipp 1: Betrachte zuerst den Fall $\mu_1 = \mu_2 = 0$. Verwende dann, dass für allgemeine $\mu_i \in \mathbb{R}$ gilt $(X_1, X_2) - (\mu_1, \mu_2) \sim \mathrm{N}(0, \Sigma)$.

Tipp 2: Sei $\eta := \sqrt{1 - \varrho^2}$ und $\mu_1 = \mu_2 = 0$, dann ist die Dichte von (X_1, X_2)

$$f_{X_1,X_2}(u, v) = \frac{1}{2\pi\sigma_1\sigma_2\eta} \exp\left(-\frac{1}{2\eta^2}\left[\frac{u^2}{\sigma_1^2} + \frac{v^2}{\sigma_2^2} - \frac{2\varrho uv}{\sigma_1\sigma_2}\right]\right).$$

Bitte wenden!

Tipp 3: Nach einsetzen von $u = x$ und $v = s - x$, bringe den Ausdruck innerhalb von exp auf die Form $\alpha(x^2+\beta s^2+\gamma xs)$. Danach ergänze quadratisch, so dass dieser Term als $\alpha(x - \delta s)^2 + \varepsilon s^2$ geschrieben werden kann.

Erinnerung: Das Integral über die Dichte einer eindimensionalen Normalverteilung ist

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1.$$

Lösung:

a) Stetiger Fall: sei $x \in \mathbb{R}$, dann gilt

$$F_S(x) = \int_{-\infty}^{x} f_S(s)ds = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(y, s-y)dyds.$$

Diskreter Fall: sei $x \in \mathbb{Z}$, dann gilt

$$F_S(x) = \sum_{s=-\infty}^{x} P(S=s) = \sum_{s=-\infty}^{x} \sum_{i=-\infty}^{\infty} P(X=i, Y=s-i).$$

b) Wir beweisen die Aussage durch nachrechnen mit der gegebenen Formel. Sei $S = X_1 + X_2$ und $s \in \mathbb{Z}$. Bemerke zuerst, dass für $s < 0$ gilt $P(S = s) = 0$, da $P(X_1 \geq 0) = 1$ und $P(X_2 \geq 0) = 1$. Für $s \geq 0$ haben wir

$$\begin{aligned}
P(S=s) &= \sum_{i=-\infty}^{\infty} P(X=i)P(Y=s-i) = \sum_{i=0}^{s} P(X=i)P(Y=s-i) \\
&= \sum_{i=0}^{s} e^{-\lambda_1} e^{-\lambda_2} \frac{\lambda_1^i \lambda_2^{s-i}}{i!(s-i)!} = e^{-(\lambda_1+\lambda_2)} \frac{1}{s!} \sum_{i=0}^{s} \lambda_1^i \lambda_2^{s-i} \frac{s!}{i!(s-i)!} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{1}{s!} \sum_{i=0}^{s} \lambda_1^i \lambda_2^{s-i} \binom{s}{i} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^s}{s!},
\end{aligned}$$

wobei wir den binomischen Lehrsatz im letzten Schritt verwendet haben. Dies ist genau die Wahrscheinlichkeitsfunktion der Poisson-Verteilung mit Parameter $\lambda_1 + \lambda_2$, somit haben wir gezeigt, dass $S \sim \text{Pois}(\lambda_1 + \lambda_2)$.

c) Wieder berechnen wir die Dichte von S mit der gegebenen Formel. Da wir bereits wissen, dass f_S eine Dichte ist (insbesondere also das Integral 1 sein muss), können wir alle konstanten Terme jeweils in der Variable c_j zusammenfassen und müssen nur die nicht-konstanten Terme beachten. Wenn der nicht-konstante Term am Ende demjenigen der gewünschten Normalverteilung entspricht, folgt sofort, dass auch der konstante Term gleich sein muss

Siehe nächstes Blatt!

(da es nur eine Konstante gibt so dass das Integral 1 ist). Wir nehmen zuerst an, dass $\mu_1 = \mu_2 = 0$. Dann gilt

$$
\begin{aligned}
f_S(s) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x, s-x)dx &= c_1 \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\eta^2}\left[\frac{x^2}{\sigma_1^2} + \frac{(s-x)^2}{\sigma_2^2} - \frac{2\varrho x(s-x)}{\sigma_1\sigma_2}\right]\right) dx \\
&= c_1 \int_{-\infty}^{\infty} \exp\left(-\underbrace{\frac{1}{2\eta^2\sigma_1^2\sigma_2^2}}_{=:d}\left[\underbrace{(\sigma_1^2+\sigma_2^2+2\varrho\sigma_1\sigma_2)}_{=:a} x^2 - 2\underbrace{(\sigma_1^2+\varrho\sigma_1\sigma_2)}_{=:b} sx + \sigma_1^2 s^2\right]\right) dx \\
&= c_1 \exp\left(-d\sigma_1^2 s^2 + ad(\tfrac{b}{a}s)^2\right) \int_{-\infty}^{\infty} \exp\left(-ad(x - \tfrac{b}{a}s)^2\right) dx \\
&= c_2 \exp\left(-s^2(d\sigma_1^2 - \frac{db^2}{a})\right) \\
&= c_2 \exp\left(-\frac{s^2}{2(\sigma_1^2+\sigma_2^2+2\varrho\sigma_1\sigma_2)}\right).
\end{aligned}
$$

Von Zeile 2 auf 3 haben wir quadratisch ergänzt und die Terme umsortiert. Von Zeile 3 auf 4 haben wir verwendet, dass das Integral eine Konstante unabhängig von s ergibt (vergleiche Erinnerung). Im letzten Schritt haben wir die die einzelnen Terme ausmultipliziert und vereinfacht. Mit der Argumentation von oben folgt, dass gilt

$$S \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2 + 2\varrho\sigma_1\sigma_2).$$

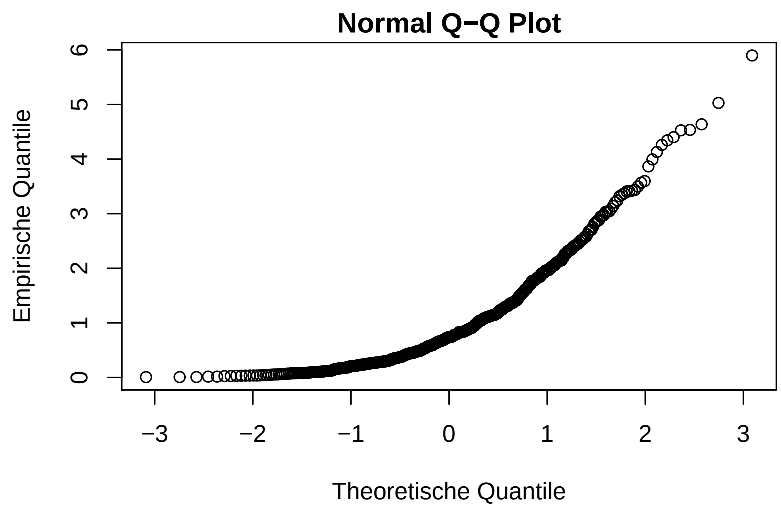
Seien nun $\mu_i, \in \mathbb{R}$. Dann gilt mit dem ersten Tipp $(X_1 - \mu_1, X_2 - \mu_2) \sim \mathrm{N}(0, \Sigma)$ und mit dem Resultat von oben also

$$X_1 + X_2 - \mu_1 - \mu_2 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2 + 2\varrho\sigma_1\sigma_2).$$

Nach Anwenden der linearen Transformation für die eindimensionale Normalverteilung folgt

$$S = X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\varrho\sigma_1\sigma_2).$$

2. Betrachten Sie folgenden Normalplot von einem Datensatz.



Welche der folgenden Aussagen sind richtig? Begründe deine Antwort.

a) Die Daten können gut mit einer Normalverteilung modelliert werden.

b) Als theoretische Verteilung wird im Normalplot die Standardnormalverteilung verwendet.

c) Die kleinste Beobachtung im Datensatz ist ungefähr 0.

d) Der empirische Median ist kleiner als 2.

Lösung:

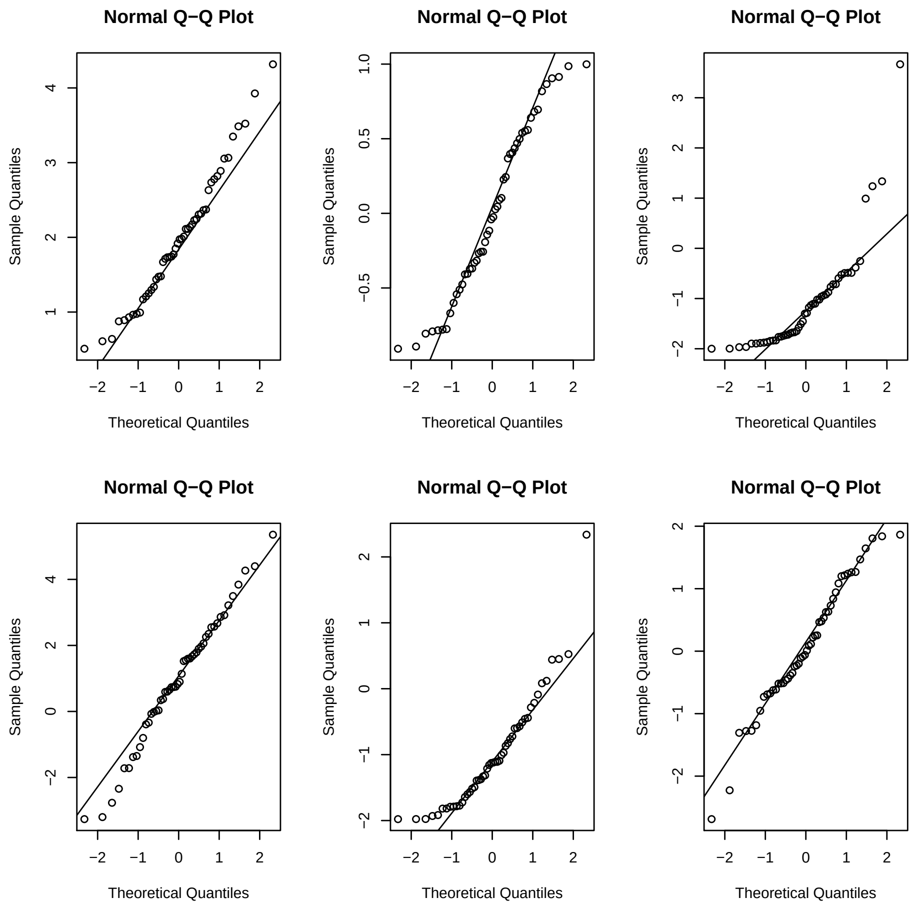
a) Falsch. Die Punkte weichen deutlich von einer Geraden ab.

b) Richtig. Siehe Definition vom Normalplot im Skript.

c) Richtig. Das kleinste empirische Quantil liegt ungefähr bei 0.

d) Richtig. Da die Standardnormalverteilung symmetrisch um 0 ist, liegt der theoretische Median bei 0. Der empirische Median liegt demzufolge ungefähr bei 1.

3. Die folgenden sechs Figuren stellen Quantil-Quantil-Plots mit Stichprobenumfang 50 verschiedener Verteilungen gegen die Normalverteilung dar. Es wurden also 50 Datenpunkte entsprechend einer gegebenen Verteilung simuliert (dies für sechs Verteilungen) und jeder dieser sechs Datensätze wird nun in einem Normalplot eingezeichnet.



Folgende Verteilungen wurden simuliert: Je einmal $\mathcal{N}(0,1), \mathcal{N}(2,1), \mathcal{N}(1,2^2)$, Uni $[-1,1]$ sowie zweimal die Verteilung von $Y := Z - 2$ wobei $Z \sim \operatorname{Exp}(1)$.

Bestimme bei jeder Figur welche Verteilung simuliert wurde und begründe deine Entscheidung.

Lösung: Folgende Zuordnung ist richtig:

1	2	3	=	$\mathcal{N}(2,1)$	Uni $[-1,1]$	$\operatorname{Exp}(1) - 2$
4	5	6		$\mathcal{N}(1,2^2)$	$\operatorname{Exp}(1) - 2$	$\mathcal{N}(0,1)$

Wie kommt man darauf?

Beim QQ-Plot werden die theoretischen Quantile (der Standardnormalverteilung $\mathcal{N}(0,1)$), auf der x-Achse aufgetragen, mit den empirischen Quantilen der zu testenden Verteilung, auf der y-Achse aufgetragen, verglichen.

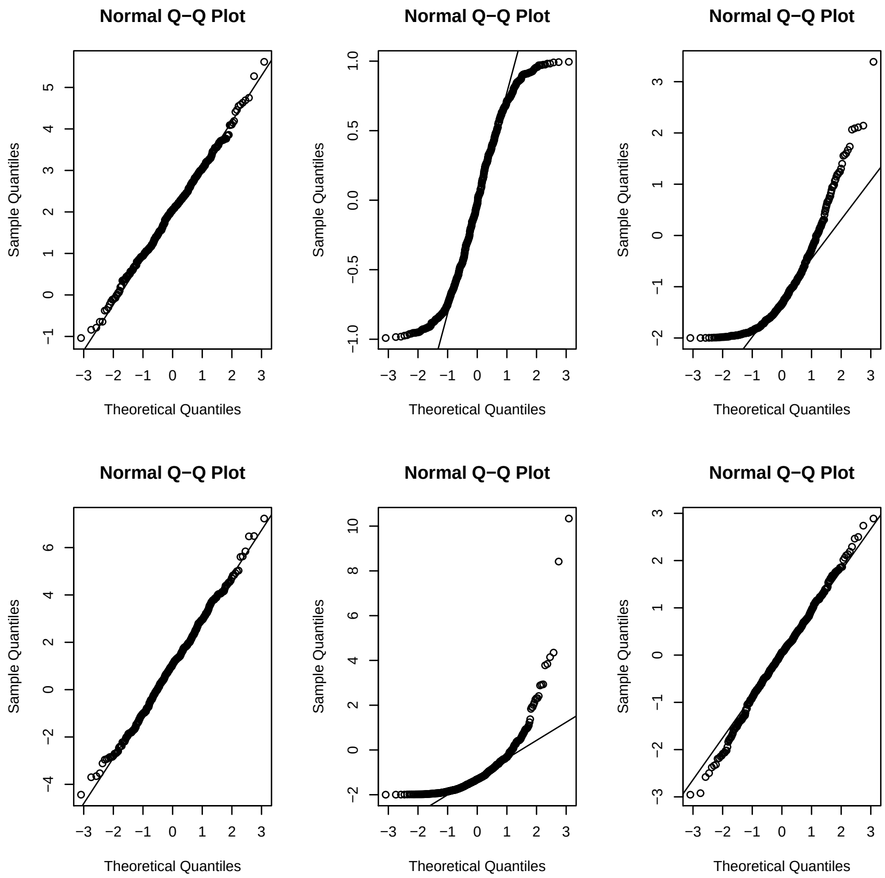
Falls es sich bei der zu testenden Verteilung um $\mathcal{N}(\mu,\sigma^2)$ handelt, dann berechnen sich die entsprechenden Quantile zu $q_{\mu,\sigma^2} = q_{0,1}\sigma + \mu$, die Gerade sollte also näherungsweise durch den Punkt $(0,\mu)$ gehen und Steigung σ haben.

Bitte wenden!

1. Die Punkte liegen einigermassen auf einer Geraden mit Steigung 1 die durch den Punkt $(0, 2)$ geht.
2. Man sieht schon auf der y-Achse, dass nur Werte zwischen -1 und 1 angenommen werden. Ausserdem wird die Steigung der Kurve immer geringer, je weiter es auf der x-Achse nach aussen geht. Das spricht für eine symmetrische Verteilung mit kürzeren Schwänzen als die Normalverteilung.
3. Die Punkte weichen stark von einer Geraden ab, und es werden Werte ausserhalb von $[-1, 1]$ angenommen, es handelt sich also weder um eine Normalverteilung, noch um Uni $[-1, 1]$. Ausserdem sieht man, dass die empirischen Quantile im linken und rechten Schwanz grösser sind als jene die man bei einer Normalverteilung erwarten würde. Die Verteilung ist also schief nach rechts. Man sieht auch, dass die empirischen Quantile bei -2 anfangen. Das alles spricht für die Verteilung $\text{Exp}(1) - 2$.
4. Die Punkte liegen einigermassen auf einer Geraden mit Steigung 2 die durch den Punkt $(0, 1)$ geht.
5. Siehe 3.
6. Die Punkte liegen einigermassen auf einer Geraden mit Steigung 1 die durch den Punkt $(0, 0)$ geht.

Bei einem Stichprobenumfang von 500 ist der Unterschied schon viel deutlicher wie man in der folgenden Grafik sehen kann:

Siehe nächstes Blatt!



4. Wir betrachten eine stetige Verteilung mit folgender Dichte:

$$f(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

wobei $\alpha > 0$ ein unbekannter Parameter ist. Wir wollen den Parameter α aus einer Stichprobe schätzen.

a) Bestimme die Likelihood- und die log-Likelihood-Funktion basierend auf n unabhängigen identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer Zufallsvariablen mit obiger Dichte.

b) Bestimme den zugehörigen Maximum-Likelihood-Schätzer für α. Schreibe zuerst die allgemeine Formel für n Beobachtungen hin und berechne den Schätzer dann für die folgende konkrete Stichprobe:

x_1	x_2	x_3	x_4	x_5
12.0	4.0	6.9	27.9	15.4

Bitte wenden!

c) Bestimme den Momentenschätzer für α, wieder zuerst allgemein basierend auf n unabhängigen Beobachtungen $x_1, \ldots, x_n$ und dann für obige Stichprobe.

Hinweis: Der Momentenschätzer setzt voraus, dass $\alpha > 1$ ist; warum?

d) Vergleiche den Maximum-Likelihood- und den Momentenschätzer für obige Stichprobe. Ist der Momentenschätzer hier sinnvoll?

Lösung:

a) Die Likelihood-Funktion ergibt sich aus dem Produkt der Dichten:

$$L(\alpha; x_1, \ldots x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\alpha}{x^{\alpha+1}} = \alpha^n \frac{1}{(\prod_{i=1}^{n} x_i)^{\alpha+1}} \ .$$

Beachte: wir können immer die Formel für $x_i \geq 1$ verwenden; der Fall $x_i < 1$ tritt nur mit Wahrscheinlichkeit 0 ein.

Die log-Likelihood-Funktion erhalten wir durch Logarithmieren obiger Formel:

$$\ell(\alpha; x_1, \ldots x_n) = \log L(\alpha; x_1, \ldots x_n) = n \log \alpha - (\alpha + 1) \sum_{i=1}^{n} \log x_i$$

b) Ableiten und Nullsetzen der log-Likelihood-Funktion gibt

$$\frac{\partial}{\partial \alpha} \ell(\alpha; x_1, \ldots x_n) = \frac{n}{\alpha} - \sum_{i=1}^{n} \log x_i = 0$$

$$\Rightarrow \widehat{\alpha}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n} \log x_i} = 0.4214$$

c) Der Erwartungswert einer Zufallsvariablen X mit obiger Dichte f ist

$$\mathbb{E}[X] = \int_{x=1}^{\infty} x f(x) dx = - \frac{\alpha}{\alpha - 1} \frac{1}{x^{\alpha-1}} \bigg|_{x=1}^{\infty} = \begin{cases} \frac{\alpha}{\alpha-1}, & \text{falls } \alpha > 1, \\ \infty, & \text{falls } \alpha \leq 1. \end{cases}$$

Da der Erwartungswert für $\alpha \leq 1$ gleich ∞ ist, ist der Momentenschätzer für diesen Fall nicht definiert.

Gleichsetzen von Erwartungswert und Stichprobenmittel, $\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^{n} x_i$, ergibt

$$\widehat{\alpha}_{\text{MoM}} = \frac{\overline{x}}{\overline{x} - 1} = 1.0817.$$

d) Der Momentenschätzer setzt $\alpha > 1$ voraus, was unplausibel erscheint, da der MLE deutlich kleiner als 1 ist.

Siehe nächstes Blatt!

5. Wir betrachten die geometrische Verteilung. Zur Erinnerung: dies ist eine diskrete Verteilung mit Wahrscheinlichkeitsfunktion

$$P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$$

wobei $0 < p < 1$ die Erfolgswahrscheinlichkeit ist. Eine geometrisch verteilte Zufallsvariable beschreibt die Anzahl Versuche bis zum ersten Erfolg bei unabhängigen Bernoulliversuchen mit Erfolgswahrscheinlichkeit p. Wir wollen den Parameter p aus einer Stichprobe schätzen.

a) Bestimme die Likelihood-Funktion basierend auf n unabhängigen identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer geometrisch verteilten Zufallsvariablen.

b) Bestimme den zugehörigen Maximum-Likelihood-Schätzer für p.

c) Bestimme den Momentenschätzer für p (wieder basierend auf n unabhängigen Beobachtungen $x_1, \ldots, x_n$). Vergleiche mit dem Maximum-Likelihood-Schätzer.

d) Angenommen du hast nur eine einzige Beobachtung x einer geometrisch verteilten Zufallsvariablen. Wir können das Experiment dann auch folgendermassen interpretieren: es wurden x unabhängige Bernoulliversuche durchgeführt und dabei wurde genau ein Erfolg beobachtet. Schreibe die Likelihood-Funktion für dieses Experiment auf: was ist der Unterschied zu der in a) gefundenen Likelihood-Funktion (für $n = 1$)? Warum erhälst du den gleichen Maximum-Likelihood-Schätzer?

Lösung:

a) Die Likelihood-Funktion ist

$$L(p; x_1, \ldots x_n) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{\sum_{i=1}^{n} x_i - n} p^n.$$

b) Die log-Likelihood-Funktion ist

$$\ell(p; x_1, \ldots, x_n) = \log L(p; x_1, \ldots, x_n) = \left(\sum x_i - n\right) \log(1-p) + n \log(p).$$

Bitte wenden!

Ableiten und Nullsetzen gibt

$$
\begin{aligned}
\frac{\partial}{\partial p}\ell(p;x_1,\ldots,x_n) = -\frac{\sum x_i - n}{1-p} + \frac{n}{p} = 0, \\
\Rightarrow \quad & \frac{-\sum_{i=1}^n x_i}{1-p} + \frac{n}{1-p} + \frac{n}{p} = 0 \\
\Rightarrow \quad & \frac{-p\cdot(\sum_{i=1}^n x_i) + n\cdot p + n\cdot(1-p)}{p\cdot(1-p)} = 0 \\
\Rightarrow \quad & -p\cdot\left(\sum_{i=1}^n x_i\right) + n\cdot p + n\cdot(1-p) = 0 \\
\Rightarrow \quad & -p\cdot\left(\sum_{i=1}^n x_i\right) + n = 0 \\
\Rightarrow \quad \widehat{p}_{\mathrm{MLE}} = \frac{n}{\sum_{i=1}^n x_i}
\end{aligned}
$$

c) Der Erwartungswert einer geometrisch verteilten Zufallsvariablen X ist $\mathbb{E}[X] = \frac{1}{p}$, siehe Skript.

Gleichsetzen der Momente, $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^n x_i$, ergibt

$$\widehat{p}_{\mathrm{MoM}} = \frac{n}{\sum_{i=1}^n x_i} \ ,$$

also $\widehat{p}_{\mathrm{MLE}} = \widehat{p}_{\mathrm{MoM}}$.

d) Die Likelihood-Funktion ist

$$L(p;x) = \binom{x}{1}(1-p)^{x-1}p = x(1-p)^{x-1}p.$$

Die Likelihood-Funktionen hier und in a) unterscheiden sich nur durch eine Proportionalitätskonstante, die nicht von p abhängt, und daher werden die Likelihoods als Funktionen in p für das gleiche Argument maximiert.