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

Koordinator
Florian Krach

Stochastik
Serie 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 $1 \leq i \leq 2$. 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)

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

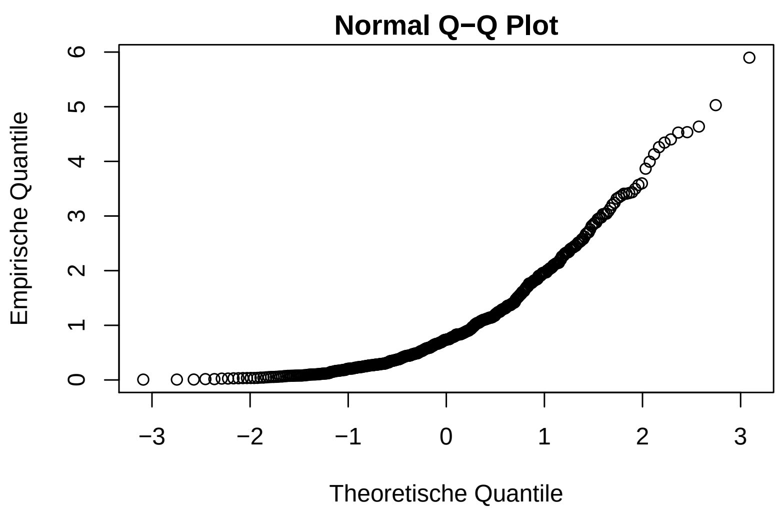
Bitte wenden!

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

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

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.

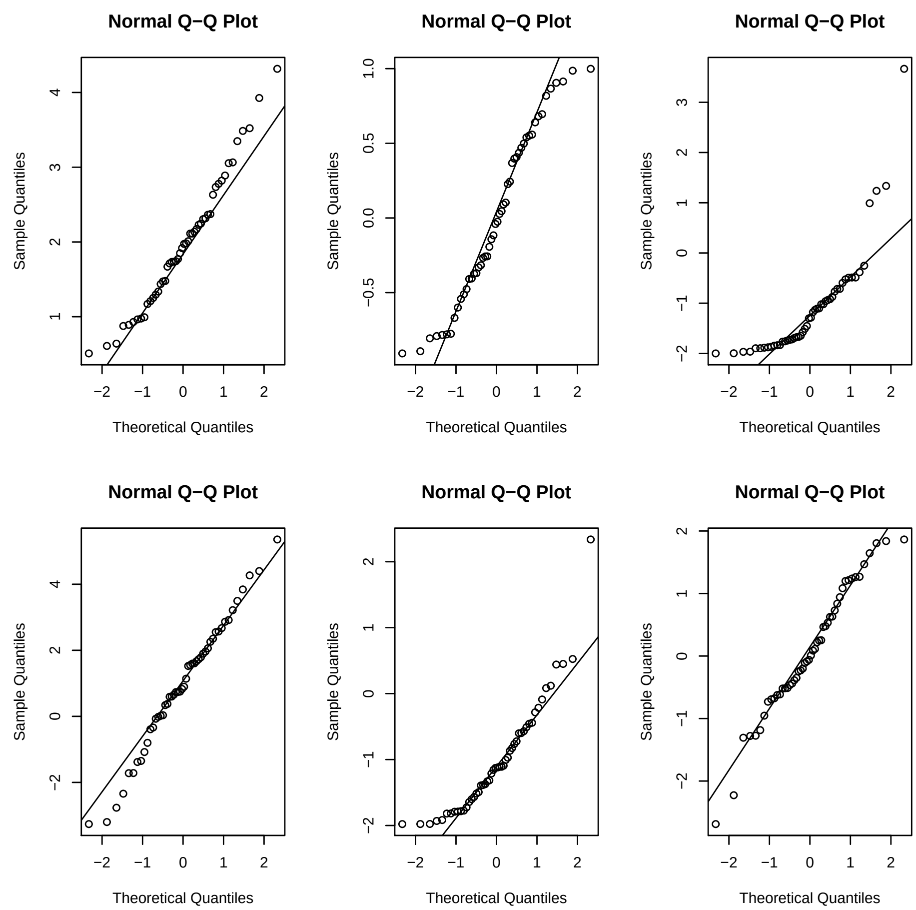
b) Als theoretische Verteilung wird im Normalplot die Standardnormalverteilung verwendet.

c) Die kleinste Beobachtung im Datensatz ist ungefähr 0.

d) Der empirische Median ist kleiner als 2.

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.

Siehe nächstes Blatt!



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 $Z \sim \text{Exp}(1)$.

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

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.

Bitte wenden!

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

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?

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?

Siehe nächstes Blatt!

Abgabe: 18. oder 19. November.
Präsenz: Montag und Donnerstag, 12:00-13:00 Uhr im HG G 32.6.
Homepage: `https://metaphor.ethz.ch/x/2019/hs/401-0603-00L/`