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

Koordinator
Florian Krach

Stochastik

Interaktive Aufgabe für die Übung 8

1. Berechnung von π mittels Monte-Carlo-Simulation

In dieser Aufgabe wollen wir π durch eine *Monte-Carlo-Simulation* berechnen bzw. abschätzen. Die Monte-Carlo-Simulation ist ein Verfahren, bei dem von einer Zufallsvariable (viele) zufällige Realisierungen gezogen werden und das Gesetz der Grossen Zahlen (GGZ) verwendet wird um den Erwartungswert der Zufallsvariable abzuschätzen.

Wir betrachten im folgenden den $\mathbb{R}^2$, ein achsenparalleles Quadrat mit Seitenlänge 2 im Ursprung zentriert ist und den darin eingeschriebenen Kreis (mit Radius 1).

a) Berechne die Fläche vom Quadrat (Q) und dem Kreis (K).

b) Seien $U_i \sim \text{Unif}([-1,1])$, für $i = 1,2$ unabhängige Zufallsvariablen und definiere den Zufallsvektor $U = (U_1, U_2) : \Omega \to [-1,1]^2$. Berechne die Wahrscheinlichkeit, dass U im Kreis liegt, also $U \in K$.

c) Um das GGZ anwenden zu können, definieren wir die folgende Zufallsvariable

$$X = \begin{cases} 1, & U_1^2 + U_2^2 \leq 1 \\ 0, & otherwise \end{cases} .$$

Berechne den Erwartungswert von X.

d) Verwende das GGZ um π zu berechnen.

e) Angenommen wir verwenden N Realisierungen von X um π abzuschätzen. Verwende den Zentralen Grenzwertsatz (ZGWS) um ein (approximatives) 95%-Vertrauensintervall für den Schätzer von π zu berechnen (verwende σ für die Standardabweichung von X).

f) Ohne Realisierungen von X zu kennen, durch welchen Wert kann σ von oben abgeschätzt werden? Verwende diesen Wert, sowie $N = 10'000$ um die Grenzen des Vertrauensintervalls zu approximieren.

g) Überprüfe die Ergebnisse anhand von Simulationen mit dem beiliegenden R-Code.

Bitte wenden!

Lösung:

a) Sei $A : \mathcal{P}(\mathbb{R}^2) \to \mathbb{R}$ die Funktion die jeder Teilmenge von $\mathbb{R}^2$ ihre Fläche zuordnet. Dann gilt $A(Q) = 4$ und $A(K) = \pi$.

b) Da U uniform verteilt auf Q ist, mit $K \subset Q$, kann die gesuchte Wahrscheinlichkeit als Anteil des Flächeninhaltes berechnet werden (**Übung**: überprüfe das durch nachrechnen mit der Dichtefunktion von U). Es gilt also $P(U \in K) = \frac{A(K)}{A(Q)} = \pi/4$.

c) $\mathbb{E}[X] = 1P(X = 1) + 0P(X = 0) = P(U_1^2 + U_2^2 \leq 1) = P(U \in K) = \pi/4$.

d) Seien X_i, $1 \leq i \leq N$, i.i.d. Zufallsvariablen mit gleicher Verteilung wie X. Definiere $Y_i := 4X_i$. Dann gilt nach dem GGZ

$$\bar{Y}_N := \frac{1}{N}\sum_{i=1}^{N} Y_i \xrightarrow{N\to\infty} \mathbb{E}[4X] = \pi.$$

e) Für $Y := 4X$ gilt $\mu := \mathbb{E}[Y] = \pi$ und $\mathrm{Var}(Y) = 16\mathrm{Var}(X) = 16\sigma^2$. Nach dem ZGWS gilt somit $\bar{Y}_N \approx \mathrm{N}(\mu, \frac{16\sigma^2}{N})$. Wie auf Buch S. 132 beschrieben, ist das 95%-Vertrauensintervall für eine Normalverteilung gegeben durch $I = \bar{Y}_N \pm 1.96\frac{4\sigma}{\sqrt{N}}$.

f) Es gilt $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = P(U \in K) - P(U \in K)^2 \leq 1$. (Beachte dass $\mathrm{Var}(X) = \pi/4 - \pi^2/16$ gilt. Da unser Ziel jedoch ist π zu berechnen und hier eine Abschätzung für den Fehler herzuleiten, können wir den exakten Wert von π nicht verwenden.) Somit gilt also mit 95% Wahrscheinlichkeit, dass der wahre Wert von π nicht weiter von $\bar{Y}_N$ entfernt ist als $1.96\frac{4\sigma}{\sqrt{N}} \leq 1.96\frac{4}{\sqrt{N}}$. Einsetzen von $N = 10'000$ liefert den Wert 0.0784.

2. Ein Mathematikprofessor hat eine kleine Klasse, die aus 10 Studenten besteht. Er denkt, dass die Noten von typischen Studenten bei seiner Prüfung normalverteilt oder exponentialverteilt sind. Die Noten $x_{(k)}$ sind in der Tabelle ersichtlich.

k	1	2	3	4	5	6	7	8	9	10
$x_{(k)}$	1.00	1.75	2.25	2.75	3.00	3.75	4.25	5.50	5.75	6.00

Untersuchen Sie mit Hilfe zweier QQ-Plots für die angegebenen Noten, mit welcher Verteilung diese Noten besser beschrieben werden können.

Lösung:

Sei X die Zufallsvariable, welche eine typische die Note repräsentiert. Wir wollen wissen, ob X normalverteilt oder exponentialverteilt ist. Wir erstellen zwei QQ-Plots für die angegebenen Noten. Als theoretische Verteilung nehmen wir zuerst

Siehe nächstes Blatt!

die Exponentialverteilung und dann die Normalverteilung an. Die geordneten Beobachtungen $x_{(k)}$ sind gerade die empirischen α_k-Quantile $q^{emp}_{\alpha_k}$, wobei $\alpha_k = \frac{k-0.5}{n}$ mit $n = 10$.

k	1	2	3	4	5	6	7	8	9	10
$\alpha_k = \frac{k-0.5}{n}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$q^{emp}_{\alpha_k} = x_{(k)}$	1.00	1.75	2.25	2.75	3.00	3.75	4.25	5.50	5.75	6.00

Exponentialverteilung
Für die Exponentialverteilung haben wir die Dichte $f(x) = \lambda e^{-\lambda x}\ 1_{\{x\geq 0\}}(x)$. Wenn X exponentialverteilt ist, haben wir

$$\alpha = P(X \leq q_\alpha) = F(q_\alpha) = \int_{-\infty}^{q_\alpha} f(x)dx$$
$$= \int_{-\infty}^{q_\alpha} \lambda e^{-\lambda x}\ 1_{\{x\geq 0\}}\ dx = (1 - e^{-\lambda q_\alpha})1_{\{q_\alpha \geq 0\}}\,.$$

Das Exponential-α-Quantil ist also gegeben durch

$$q^{exp}_\alpha = F^{-1}(\alpha) = -\frac{1}{\lambda}\log(1-\alpha)\,.$$

k	1	2	3	4	5	6	7	8	9	10
$\alpha_k = \frac{k-0.5}{n}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$q^{exp}_{\alpha_k} = -\log(1-\alpha_k)$	0.05	0.16	0.29	0.43	0.60	0.80	1.05	1.39	1.90	3.00

Wir zeichnen daher $(x, y) = (q^{exp}_{\alpha_k}, q^{emp}_{\alpha_k})$ auf. Wir verwenden hier als Referenzverteilung die Exp(1) Verteilung. Wenn die Daten wirklich Exp(λ) verteilt sind, erwarten wir ungefähr eine Gerade mit Steigung $1/\lambda$.

Normalverteilung
Jeztz nehmen wir an, dass X normalverteilt mit $\mathcal{N}(\mu, \sigma)$ ist. Durch Standardisieren erhalten wir $Z \sim \mathcal{N}(0, 1)$ wobei $Z = \frac{X-\mu}{\sigma}$. Wir erhalten dann

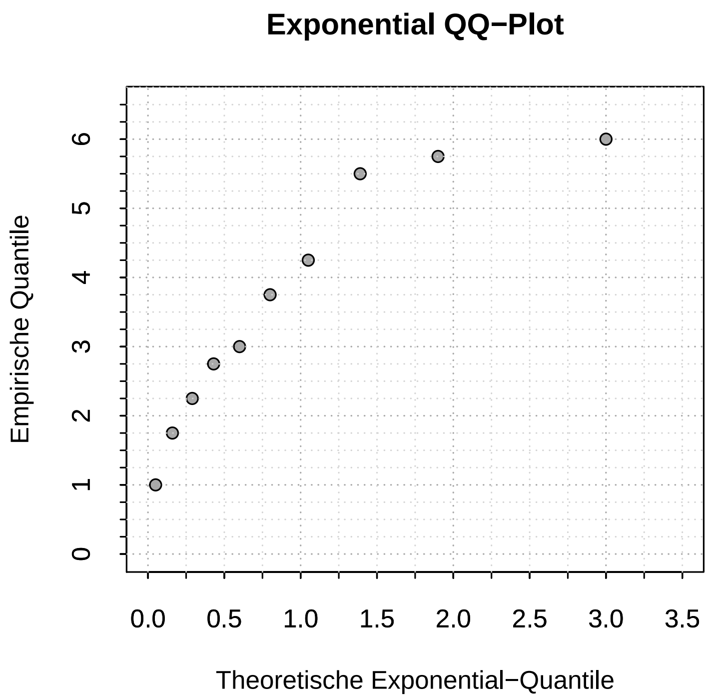
$$\alpha = P(X \leq q_\alpha) = P(Z \leq \frac{q_\alpha - \mu}{\sigma}) = \Phi\left(\frac{q_\alpha - \mu}{\sigma}\right)\,.$$

Das Normal-α-Quantil ist also gegeben durch

$$q^n_\alpha = \mu + \sigma\Phi^{-1}(\alpha)\,.$$

Die Quantile der Standardnormalverteilung $z_k = \Phi^{-1}(\alpha_k)$ findet man im Skript, siehe Tabelle auf Seite 126.

Bitte wenden!

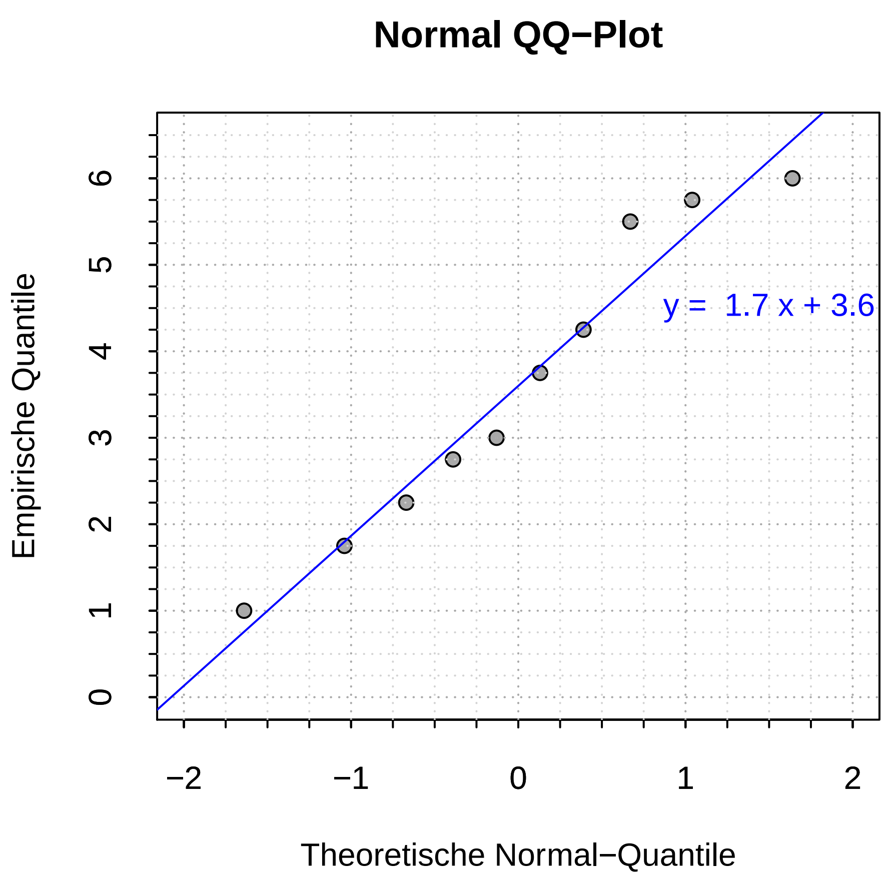


k	1	2	3	4	5	6	7	8	9	10
$\alpha_k = \frac{k-0.5}{n}$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
$z_k = \Phi^{-1}(\alpha_k)$	-1.64	-1.04	-0.67	-0.39	-0.13	0.13	0.39	0.67	1.04	1.64

Wir zeichnen daher $(q^n_{\alpha_k}, q^{emp}_{\alpha_k})$ auf. Wir verwenden hier als Referenzverteilung die $\mathcal{N}(0,1)$ Verteilung. Wenn die Daten wirklich $\mathcal{N}(\mu,\sigma)$-verteilt sind, erwarten wir ungefähr eine Gerade mit Steigung σ und Achsenabschnitt μ.

Mithilfe dieser zweie QQ-Plots können wir schliessen, dass die Noten eher $\mathcal{N}(\mu,\sigma)$-verteilt sind. Zusätzlich stellen wir fest, dass das arithmetische Mittel $\bar{x}$ in etwa μ entspricht und die empirische Standardabweichung $\bar{\sigma}$ in etwa σ entspricht.

$$\bar{x} = \frac{1}{10}\sum_{k=1}^{10} x_k = 3.6 \approx \mu \qquad \text{und} \qquad \bar{\sigma}^2 = \frac{1}{9}\sum_{k=1}^{10}(x_k - 3.6)^2 = 1.75 \approx \sigma.$$


Normal QQ−Plot
y = 1.7 x + 3.6
Empirische Quantile
0
1
2
3
4
5
6
−2
−1
0
1
2
Theoretische Normal−Quantile