

Bestimmen von Quantilen

Wie Rückwärtsdenken in der Stochastik hilft

Leitung: Tobias Wiernicki-Krips

Samstag, 10. Januar 2015

STIFTISCHES GYMNASIUM DÜREN



Lehr- und For-
schungsgebiet
Didaktik der
Mathematik

RWTHAACHEN
UNIVERSITY

Wie bestimmt man (empirische) p -Quantile \tilde{x}_p bzw. Q_p ?

- Invertieren/Rückgängigmachen am Beispiel der Quantilsbestimmung in der Stochastik
- Vorgehensweise:
 - realitätsbezogen
 - mathematisch formal
 - anspruchsvoll



Stochastik enthält

1 Statistik

(Analyse von Daten und Planung von empirischen Untersuchungen):

- **Beschreibende Statistik**
- **Schließende Statistik**

2 Wahrscheinlichkeitsrechnung

Hier:

- Quantile in der Beschreibenden Statistik
- Ausblick zu Quantilen in der Wahrscheinlichkeitsrechnung

- 1 Quantile in der Beschreibenden Statistik
 - Empirische Quantile, Boxplot
 - Empirische Verteilungsfunktion
 - Lorenzkurve

- 2 Quantile in der Wahrscheinlichkeitsrechnung

- 3 Fazit und Ausblick

1.1 Empirische Quantile, Boxplot

Empirische Quantile

Definition 1: Gegeben ist eine geordnete Urliste $x_1 \leq \dots \leq x_n$ eines quantitativen Merkmals. Dann nennt man für $0 < p < 1$

$$\tilde{x}_p := \begin{cases} x_{([np]+1)} & , \text{ falls } np \notin \mathbb{N} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & , \text{ falls } np \in \mathbb{N} \end{cases}$$

das *empirische p -Quantil* von x_1, \dots, x_n .

Bemerkung

Für ein empirisches p -Quantil gilt:

Mindestens $p \cdot 100\%$ der Listenwerte sind kleiner oder gleich \tilde{x}_p und mindestens $(1 - p) \cdot 100\%$ der Werte sind größer oder gleich \tilde{x}_p .

1.1 Empirische Quantile, Boxplot

Sonderfälle

- $\tilde{x}_{0,25}$ heißt *unteres Quartil*,
- $\tilde{x}_{0,5}$ *Median*,
- $\tilde{x}_{0,75}$ *oberes Quartil* und
- $\tilde{x}_{\frac{k}{10}}$ *k-tes Dezantil* für $k \in \{1, \dots, 9\}$.

Für den Median ist folgende Definition äquivalent zu Definition 1:

Definition 2: Gegeben ist eine geordnete Urliste $x_1 \leq \dots \leq x_n$ eines quantitativen Merkmals. Dann nennt man

$$\tilde{x}_{0,5} := \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{falls } n \text{ gerade} \end{cases}$$

den *Median* von x_1, \dots, x_n .

1.1 Empirische Quantile, Boxplot

Boxplot erstellen:

a) Pulsmessung in Ruhe

- 1 Geordnete Urliste $x_1 \leq \dots \leq x_n$
- 2 Median, unteres und oberes Quartil
- 3 Minimum, Maximum
- 4 Boxplot zeichnen

b) Pulsmessung nach 15 Kniebeugen und zweites Boxplot über derselben Skala zeichnen



1.1 Empirische Quantile, Boxplot

Urlisten und Berechnung von Median und unterem Quartil:

(ungeordnete) Urliste: 61, 72, 69, 73, 91, 65, 74, 67, 51, 63
[in Ruhe] ↑ nicht in Ruhe

(geordnete) Urliste: 51, 61, 63, 65, 67, 69, 73, 74, 77, 91
3/10 3/10
3/10 1/10

$n=10$, $p=0,5$ (Median $\tilde{x}_{0,5}$)
 $np=10 \cdot 0,5=5 \in \mathbb{N} \Rightarrow \tilde{x}_{0,5} = \frac{1}{2} (x_{(5)} + x_{(6)}) = 68$

$n=10$, $p=0,25$ ($\tilde{x}_{0,25}$)
 $np=10 \cdot 0,25=2,5 \notin \mathbb{N} \Rightarrow \tilde{x}_{0,25} = x_{(2,5)} = x_{(3)} = 63$

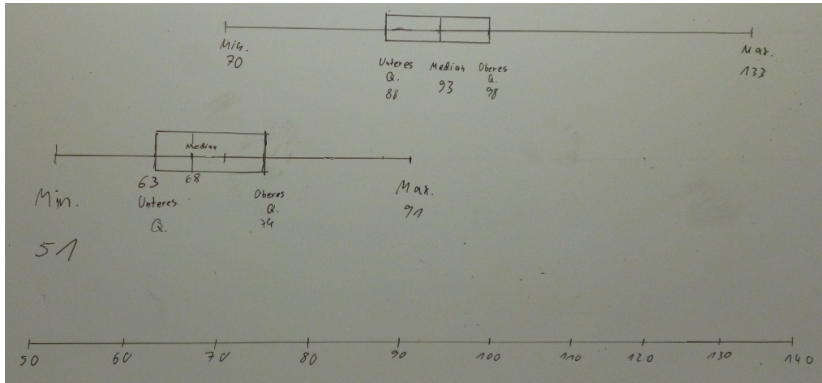
[X] = $\max \{9 \in \mathbb{Z} \mid 2 \leq x\}$

(ungeordnete) Urliste: 88, 98, 92, 102, 133, 79, 91, 94, 96, 70
[nach 15 Minuten]

(geordnete) Urliste: 70, 79, 88, 91, 92, 94, 96, 98, 102, 133

1.1 Empirische Quantile, Boxplot

Boxplots zur Pulsmessung aller 10 Anwesenden:



⇒ Auf einen Blick Vergleich möglich

1.1 Empirische Quantile, Boxplot

Zusammenfassung zur

Quantilsbestimmung mit Hilfe einer geordneten Urliste:

- Berechnen von kumulierten relativen Häufigkeiten als **bekannte Richtung:**
 $x_k \rightarrow$ Anteil $\sum_{i=1}^k h_i$ der ersten k Werte am Stichprobenumfang
- Quantilsbestimmung als **Umkehraufgabe:**
 $x_k \leftarrow$ Anteil \tilde{x}_p der ersten k Werte am Stichprobenumfang

1.2 Empirische Verteilungsfunktion

Empirische Verteilungsfunktion (Kumulierte Häufigkeitsfunktion)

Definition 3: Zu einer geordneten Urliste $x_1 \leq \dots \leq x_n$ seien $u_1 < \dots < u_m$ die m verschiedenen Merkmalsausprägungen eines quantitativen Merkmals. Dann nennt man die Funktion

$F_n : \mathbb{R} \rightarrow [0, 1]$ mit

$$F_n(x) = \begin{cases} 0 & , \text{ falls } x < u_1 \\ \sum_{i: u_i \leq x} h_i & , \text{ falls } u_1 \leq x \leq u_m \\ 1 & , \text{ falls } x > u_m \end{cases}$$

empirische Verteilungsfunktion zu $x_1 \leq \dots \leq x_n$.

1.2 Empirische Verteilungsfunktion

Empirische Verteilungsfunktion

anhand des Bsp. Pulsmessung (in Ruhe):

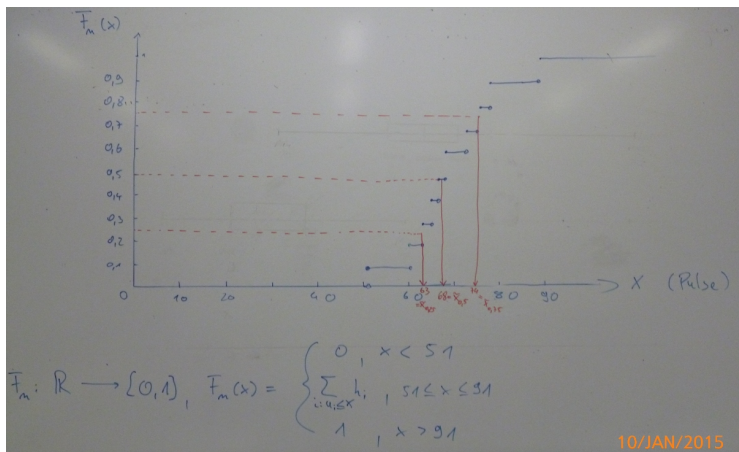
- relative Häufigkeiten berechnen
- relative Häufigkeiten aufsummieren
- Graphen zeichnen

Wie kann man empirische Quantile am Graphen ablesen?

Oft nicht: siehe nächste Folie

1.2 Empirische Verteilungsfunktion

Rückwärtslesen (rot), $n = 10$:



⇒ Problem der Nicht-Eindeutigkeit hier bei $\tilde{x}_{0.5}$

⇒ Bestimmung empirischer Quantile nicht sinnvoll mit F_n möglich

1.2 Empirische Verteilungsfunktion

Eigenschaften der empirischen Verteilungsfunktion F_n :

- monoton steigend
- nicht umkehrbar bzw. nicht injektiv
- Treppenfunktion
- Sprunghöhe bei $u_i = h_i$
- rechtsseitig stetig
- $F_n \xrightarrow{x \rightarrow \infty} 1$ und $F_n \xrightarrow{x \rightarrow -\infty} 0$

OECD

Kluft zwischen Arm und Reich wächst

Berlin. In Deutschland und zahlreichen anderen Ländern wächst der Abstand zwischen Arm und Reich, teilte die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) mit. In Deutschland hätten Mitte der 80er Jahre die reichsten zehn Prozent der Bevölkerung fünf Mal so viel verdient wie die ärmsten zehn Prozent. Heute liegt das Verhältnis nach OECD-Angaben bei sieben zu eins. (afp)

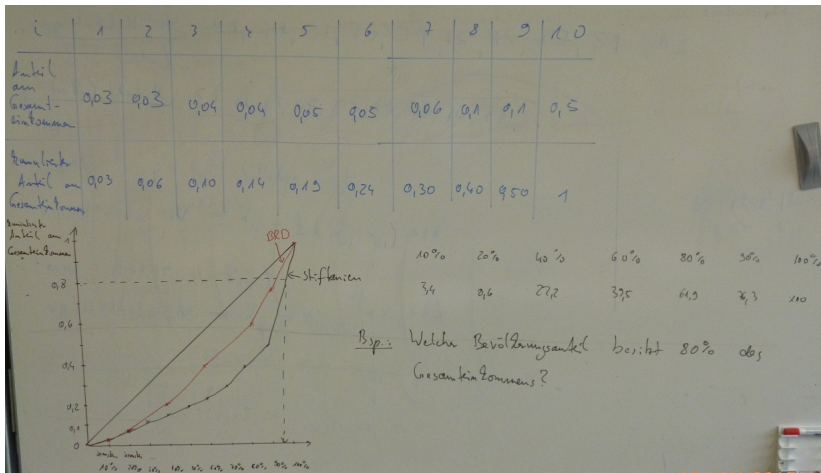
(Kölner Stadt-Anzeiger 9.12.2014)

Wie kann man die Einkommenskonzentration einer Bevölkerung darstellen?

- Beispiel Stiftnien:
Verteilung von 100 Chips als Gesamteinkommen an die 10 Anwesenden, die jeweils für 10% der Bevölkerung von Stiftnien stehen
- Lorenzkurve:
Graph einer kumulierten und linear interpolierten Klassenverteilung (stückweise linear)
- Vergleich der Lorenzkurven von Stiftnien und Deutschland:
Je größer die Fläche zwischen Diagonale und Lorenzkurve, desto größer die Ungleichheit

1.3 Lorenzkurve

Lorenzkurven in Stiftanien und Deutschland:



blaue Tabelle: Stiftanien; schwarze Tabelle: Deutschland (Cingano 2014, S. 36)

Lorenzkurven:

- Quantile \tilde{x}_p für $0 < p < 1$ (scheinbar) eindeutig ablesbar
- Aber: für Bereiche zwischen u_i nur Schätzung möglich (linear interpoliert)
- Je mehr verschiedene Merkmalsausprägungen u_i vorliegen, desto genauer kann man \tilde{x}_p ablesen.

2. Quantile in der Wahrscheinlichkeitsrechnung

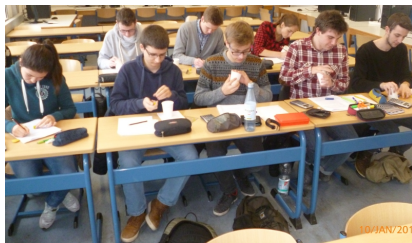
Einteilung der Wahrscheinlichkeitsrechnung:

- **Diskrete Zufallsgrößen** (Merkmale): z.B. beim Würfelwurf interessiert die Augensumme, das sind nur 11 Ausprägungen
- **Stetige Zufallsgrößen** (Merkmale): z.B. Körpergröße, hier sind es viel mehr mögliche Ergebnisse

2. Quantile in der Wahrscheinlichkeitsrechnung

Würfeldoppelwurf: Merkmal Augensumme

- Subjektive Wahrscheinlichkeitsverteilung? (EA, 30s)
- absolute und relative Häufigkeiten **aller** Schülerinnen und Schüler
- Neue Schätzung auf Grundlage der relativen Häufigkeiten
- Berechnung der Wahrscheinlichkeiten



2. Quantile in der Wahrscheinlichkeitsrechnung

Würfeldoppelwurf zweigeteilter Prozess:

1 Würfeln (*hier* Zufall):

- 36 Würfelkombinationen: $\{(1, 1), \dots, (1, 6), \dots, (6, 6)\}$
- Wahrscheinlichkeit für Kombination: $\frac{1}{36}$ (Laplace-Experiment)
- $P(A) = \frac{\text{Anzahl der günstigen Ergebnisse}}{\text{Anzahl der möglichen Ergebnisse}} = \frac{|A|}{36}$

2 Augensumme bestimmen („ X beschreibt die Augensumme“)

- Augensummen: 2, ..., 12
- Welche Kombinationen werden den Augensummen zugeordnet? $2 \mapsto \{(1, 1)\}$, $3 \mapsto \{(2, 1), (1, 2)\}$, ..., $12 \mapsto \{6, 6\}$
- Wahrscheinlichkeit für Augensummen: $P(X = 2) = \frac{1}{36}$,
 $P(X = 3) = \frac{2}{36}$, ..., $P(X = 12) = \frac{1}{36}$

2. Quantile in der Wahrscheinlichkeitsrechnung

Verteilungsfunktion (Kumulierte Wahrscheinlichkeitsverteilung)

Definition 4: Zu einer Zufallsgröße X mit Wahrscheinlichkeitsverteilung P^X heißt

$F : \mathbb{R} \rightarrow \mathbb{R}$ mit $F(x) = P(X \leq x)$ die *Verteilungsfunktion* von P^X .

Quantile

Definition 5: Zu einer Verteilungsfunktion F heißt die Funktion $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ mit $F^{-1}(y) = \min\{x \in \mathbb{R} \mid F(x) \geq y\}$

Quantilfunktion von F .

$Q_p := F^{-1}(p)$ nennt man *p-Quantil* von F .

2. Quantile in der Wahrscheinlichkeitsrechnung

Verteilungsfunktion für die Augensumme (Würfeldoppelwurf)
und Rückwärtslesen für $Q_{0,1}$

Augensumme x	2	3	4	5	6	7	8	9	10	11	12
Relative Häufigkeit	$\frac{1}{36}$ $\approx 0,04$	$\frac{3}{36}$ $\approx 0,07$	$\frac{5}{36}$ $\approx 0,08$	$\frac{7}{36}$ $\approx 0,11$	$\frac{9}{36}$ $\approx 0,16$	$\frac{11}{36}$ $\approx 0,15$	$\frac{13}{36}$ $\approx 0,12$	$\frac{15}{36}$ $\approx 0,12$	$\frac{17}{36}$ $\approx 0,06$	$\frac{19}{36}$ $\approx 0,05$	$\frac{16}{36}$ $\approx 0,04$
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$P(X \leq x)$	$\frac{1}{36}$ $\approx 0,03$	$\frac{3}{36}$ $\approx 0,08$	$\frac{6}{36}$ $\approx 0,17$	$\frac{10}{36}$ $\approx 0,30$	$\frac{15}{36}$ $\approx 0,42$	$\frac{21}{36}$ $\approx 0,58$	$\frac{26}{36}$ $\approx 0,70$	$\frac{30}{36}$ $\approx 0,83$	$\frac{33}{36}$ $\approx 0,92$	$\frac{35}{36}$ $\approx 0,95$	$\frac{36}{36}$ 1

$F(x)$

$F: \mathbb{R} \rightarrow [0,1], F(x) = P(X \leq x)$

2. Quantile in der Wahrscheinlichkeitsrechnung

Eigenschaften der Verteilungsfunktion F

(analog zur empirischen Verteilungsfunktion):

- monoton steigend
- im Allgemeinen nicht umkehrbar bzw. nicht injektiv
- Treppenfunktion bei diskreten Zufallsgrößen
- Sprunghöhe bei $x_i = P(X = x_i)$
- rechtsseitig stetig
- $F \xrightarrow{x \rightarrow \infty} 1$ und $F \xrightarrow{x \rightarrow -\infty} 0$

Ablezen der Quantile durch Rückwärtslesen

(vgl. Cramer/Kamps, S. 198):

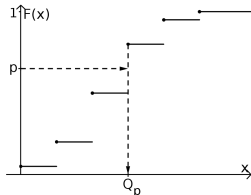
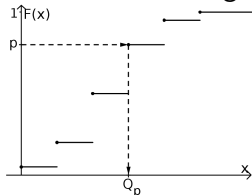
- vgl. letzte Folie: $Q_{0,1} = 4$
- bei Mehrdeutigkeiten Einigung auf Minimum:
z.B. bei $Q_{1/12} = 3$

2. Quantile in der Wahrscheinlichkeitsrechnung

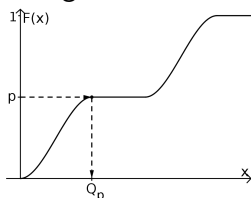
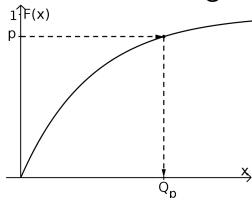
Ablezen der Quantile durch Rückwärtslesen:

(Abbildungen mit GeoGebra erstellt nach Cramer/Kamps, S. 198):

■ Diskrete Verteilung:



■ Ausblick zu stetigen Verteilungen:



3. Fazit und Ausblick

In Beschreibender Statistik:

- Empirische Quantile insbesondere zur Erstellung eines Boxplots wichtig
- Empirische Quantile i.A. bei empirischen Verteilungsfunktionen nicht eindeutig ablesbar, geordnete Urliste notwendig

Beispiel Einkommenskonzentration:

- Einkommenskonzentration durch Lorenzkurve beschreibbar
- Ablesen von Quantilen aus Lorenzkurve nur scheinbar genau und eindeutig ablesbar: Lorenzkurve entsteht durch lineare Interpolation von Klassenwerten

3. Fazit und Ausblick

Bei diskreten Wahrscheinlichkeitsverteilungen:

- Ablesen von Quantilen bei Verteilungsfunktionen
Eindeutigkeitsproblematik wie bei empirischen Verteilungsfunktionen
- Einigung auf das Minimum in Frage kommender Werte

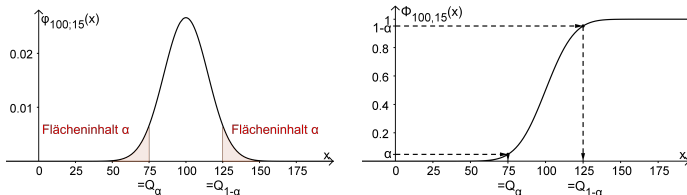
Anknüpfungspunkt bei diskreten und stetigen Wahrscheinlichkeitsverteilungen in Q2:

- Hypothesentests in Schließender Statistik
- i.d.R. Binomialverteilungen und Normalverteilungen als Näherungen

3. Fazit und Ausblick

Ausblick auf stetige Verteilungen:

- Quantile von streng monoton steigenden Verteilungsfunktionen eindeutig ablesbar
- Beispiel Normalverteilung des Intelligenzquotienten:
 - Youtube-Video von DorFuchs (Gesehen am 5.2.2015):
<https://www.youtube.com/watch?v=m-nEbS4QpYw>
 - DGhK (Gesehen am 5.2.2015):
<https://dghk-hh.de/index.php?id=5>



(nach Cramer/Kamps, S. 199)

- Auer, B., Rottmann, H.** (2015). *Statistik und Ökonometrie für Wirtschaftswissenschaftler*. 3. Auflage. Wiesbaden: Springer Gabler.
- Cingano, F.** (2014). *Trends in Income Inequality and its Impact on Economic Growth*. OECD Social, Employment and Migration Working Papers Nr. 163. Im Internet erhältlich unter: <http://www.oecd.org/els/soc/> [02.02.2015]
- Cramer, E., Kamps, U.** (2014). *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. Ein Skript für Studierende der Ingenieur- und Wirtschaftswissenschaften*. 3. Auflage. Berlin, Heidelberg: Springer.
- Henze, N.** (2012). *Stochastik für Einsteiger*. 9. Auflage. Wiesbaden: Vieweg+Teubner.

Fotos: T. Wiernicki-Krips