

Quantile von Zufallsvariablen I

Definition 9.11 (p -Quantil)

Seien X eine eindimensionale Zufallsvariable, $p \in (0, 1)$. Jeder Wert $x_p \in \mathbb{R}$ mit

$$P\{X \leq x_p\} \geq p \quad \text{und} \quad P\{X \geq x_p\} \geq 1 - p$$

heißt p -**Quantil** (auch p -**Perzentil**) von X . Man nennt Werte x_p mit dieser Eigenschaft **spezieller**

- **Median** von X für $p = 0.5$,
- **unteres Quartil** von X für $p = 0.25$ sowie
- **oberes Quartil** von X für $p = 0.75$.

Ist F_X die Verteilungsfunktion von X , so ist x_p also genau dann p -Quantil von X , wenn

$$F_X(x_p - 0) \leq p \leq F_X(x_p)$$

gilt, für stetige Zufallsvariablen X also genau dann, wenn $F_X(x_p) = p$ gilt.

Quantile von Zufallsvariablen III

- Auch ohne die Invertierbarkeit von F_X kann Eindeutigkeit von Quantilen zum Beispiel durch die Festsetzung

$$x_p := \min\{x \mid P\{X \leq x\} \geq p\} = \min\{x \mid F_X(x) \geq p\}$$

erreicht werden.

Man nennt die Abbildung

$$(0, 1) \rightarrow \mathbb{R}; p \mapsto x_p = \min\{x \mid F_X(x) \geq p\}$$

häufig auch *verallgemeinerte Inverse* von F_X und verwendet hierfür dann ebenfalls das Symbol F_X^{-1} sowie die Bezeichnung **Quantilsfunktion**.

- Diese Eindeutigkeitsfestlegung **unterscheidet** sich von der vergleichbaren Konvention aus der deskriptiven Statistik für empirische Quantile!

Quantile von Zufallsvariablen II

- p -Quantile sind nach Definition 9.11 eindeutig bestimmt, wenn die Verteilungsfunktion F_X der Zufallsvariablen X (dort, wo sie Werte in $(0, 1)$ annimmt) invertierbar ist, also insbesondere stetig und streng monoton wachsend.

Bezeichnet F_X^{-1} die Umkehrfunktion von F_X , so gilt dann

$$x_p = F_X^{-1}(p) \quad \text{für alle } p \in (0, 1).$$

F_X^{-1} wird in diesem Fall auch **Quantilsfunktion** genannt.

- Der Abstand $x_{0.75} - x_{0.25}$ zwischen unterem und oberem Quartil wird (wie auch bei empirischen Quartilen) auch **Interquartilsabstand (IQA)** genannt.

Spezielle diskrete Verteilungen

- Im Folgenden: Vorstellung spezieller (parametrischer) **Verteilungsfamilien**, die häufig Verwendung finden.
- Häufige Verwendung ist dadurch begründet, dass diese Verteilungen in vielen verschiedenen Anwendungen anzutreffen sind bzw. die Zufallsabhängigkeit interessanter Größen geeignet modellieren.
- Parametrische Verteilungsfamilien sind Mengen von (ähnlichen) Verteilungen Q_θ , deren Elemente sich nur durch die Ausprägung eines oder mehrerer **Verteilungsparameter** unterscheiden, d.h. die spezielle Verteilung hängt von einem Parameter oder einem Parametervektor θ ab, und zu jedem Parameter(vektor) gehört jeweils eine eigene Verteilung Q_θ .
- Die Menge aller möglichen Parameter(vektoren) θ , auch **Parameterraum** genannt, wird meist mit Θ bezeichnet. Die Verteilungsfamilie ist damit die Menge $\{Q_\theta \mid \theta \in \Theta\}$.
- Besitzt eine Zufallsvariable X die Verteilung Q_θ , so schreibt man auch kurz: $X \sim Q_\theta$.
- **Zunächst**: Vorstellung spezieller *diskreter* Verteilungen.

Bernoulli-/Alternativverteilung

- Verwendung:
 - ▶ Modellierung eines Zufallsexperiments (Ω, \mathcal{F}, P) , in dem nur das Eintreten bzw. Nichteintreten eines einzigen Ereignisses A von Interesse ist.
 - ▶ Eintreten des Ereignisses A wird oft als „Erfolg“ interpretiert, Nichteintreten (bzw. Eintreten von \bar{A}) als „Misserfolg“.
 - ▶ Zufallsvariable soll im Erfolgsfall Wert 1 annehmen, im Misserfallsfall Wert 0, es sei also

$$X(\omega) := \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \in \bar{A} \end{cases}$$

- ▶ Beispiel: Werfen eines fairen Würfels, Ereignis A : „6 gewürfelt“ mit $P(A) = \frac{1}{6}$.
- Verteilung von X hängt damit *nur* von „Erfolgswahrscheinlichkeit“ $p := P(A)$ ab; p ist also einziger Parameter der Verteilungsfamilie.
- Um triviale Fälle auszuschließen, betrachtet man nur Ereignisse mit $p \in (0, 1)$
- Der Träger der Verteilung ist dann $T(X) = \{0, 1\}$, die Punktwahrscheinlichkeiten sind $p_X(0) = 1 - p$ und $p_X(1) = p$.
- Symbolschreibweise für Bernoulli-Verteilung mit Parameter p : $B(1, p)$
- Ist X also Bernoulli-verteilt mit Parameter p , so schreibt man $X \sim B(1, p)$.

Binomialverteilung

- Verallgemeinerung der Bernoulli-Verteilung
- Verwendung:
 - ▶ Modellierung der **unabhängigen, wiederholten** Durchführung eines Zufallsexperiments, in dem nur die **Häufigkeit** des Eintretens bzw. Nichteintretens eines Ereignisses A interessiert („Bernoulli-Experiment“).
 - ▶ Eintreten des Ereignisses A wird auch hier oft als „Erfolg“ interpretiert, Nichteintreten (bzw. Eintreten von \bar{A}) als „Misserfolg“.
 - ▶ Zufallsvariable X soll die **Anzahl der Erfolge** bei einer vorgegebenen Anzahl von n Wiederholungen des Experiments zählen.
 - ▶ Nimmt X_i für $i \in \{1, \dots, n\}$ im Erfolgsfall (für Durchführung i) den Wert 1 an, im Misserfallsfall den Wert 0, dann gilt also $X = \sum_{i=1}^n X_i$.
 - ▶ Beispiel: 5-faches Werfen eines fairen Würfels, Anzahl der Zahlen kleiner 3.
 $\rightsquigarrow n = 5, p = 1/3$.
- Verteilung von X hängt damit *nur* von „Erfolgswahrscheinlichkeit“ $p := P(A)$ sowie der Anzahl der Durchführungen n des Experiments ab.
- Um triviale Fälle auszuschließen, betrachtet man nur die Fälle $n \in \mathbb{N}$ und $p \in (0, 1)$. Träger der Verteilung ist dann $T(X) = \{0, 1, \dots, n\}$.
- Symbolschreibweise für Binomialverteilung mit Parameter n und p : $B(n, p)$
- Übereinstimmung mit Bernoulli-Verteilung (mit Parameter p) für $n = 1$.

Bernoulli-/Alternativverteilung

$$B(1, p)$$

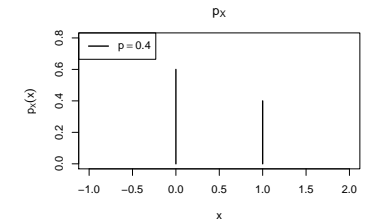
Parameter:

$$p \in (0, 1)$$

Träger: $T(X) = \{0, 1\}$

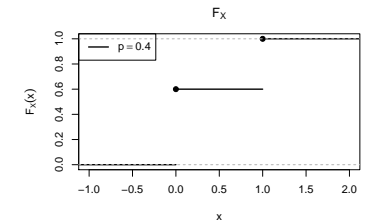
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} 1 - p & \text{für } x = 0 \\ p & \text{für } x = 1 \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}$$



Momente: $E(X) = p$

$$\gamma(X) = \frac{1-2p}{\sqrt{p(1-p)}}$$

$\text{Var}(X) = p \cdot (1 - p)$

$$\kappa(X) = \frac{1-3p(1-p)}{p(1-p)}$$

Binomialverteilung

$$B(n, p)$$

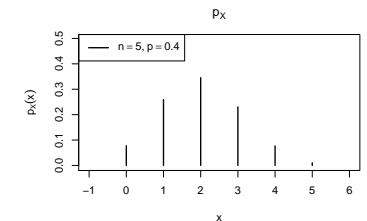
Parameter:

$$n \in \mathbb{N}, p \in (0, 1)$$

Träger: $T(X) = \{0, 1, \dots, n\}$

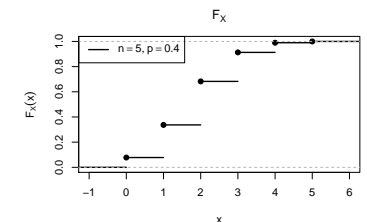
Wahrscheinlichkeitsfunktion: $p_X(x)$

$$= \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \sum_{\substack{x_i \in T(X) \\ x_i \leq x}} p_X(x_i)$$



Momente: $E(X) = n \cdot p$

$$\gamma(X) = \frac{1-2p}{\sqrt{np(1-p)}}$$

$\text{Var}(X) = n \cdot p \cdot (1 - p)$

$$\kappa(X) = \frac{1+(3n-6)p(1-p)}{np(1-p)}$$

Geometrische Verteilung

- Verwendung:
 - ▶ Modellierung der **unabhängigen, wiederholten** Durchführung eines Bernoulli-Experiments (nur das Eintreten bzw. Nichteintreten eines einzigen Ereignisses A ist von Interesse), bis das Ereignis A **zum ersten Mal** eintritt.
 - ▶ Zufallsvariable X zählt **Anzahl der Misserfolge**, ausschließlich des (letzten) „erfolgreichen“ Versuchs, bei dem Ereignis A zum ersten Mal eintritt.
 - ▶ X kann also nur Werte $x \in \mathbb{N}_0$ annehmen, man erhält die Realisation x von X , wenn nach genau x Misserfolgen (Nicht-Eintreten von A) in der $(x + 1)$ -ten Durchführung ein Erfolg (Eintreten von A) zu verzeichnen ist.
 - ▶ Ist $p := P(A)$ die „Erfolgswahrscheinlichkeit“ des Bernoulli-Experiments, so gilt offensichtlich $P\{X = x\} = (1 - p)^x \cdot p$ für alle $x \in \mathbb{N}_0$.
 - ▶ Beispiel (vgl. Folie 168): Anzahl des Auftretens von „Zahl“ beim Werfen einer Münze („Wappen“ oder „Zahl“), bis zum ersten Mal „Wappen“ erscheint
 $\rightsquigarrow p = 1/2$ (bei fairer Münze).
- Verteilung von X hängt damit *nur* von Erfolgswahrscheinlichkeit p ab.
- Um triviale Fälle auszuschließen, betrachtet man nur den Fall $p \in (0, 1)$. Träger der Verteilung ist dann $T(X) = \mathbb{N}_0 = \{0, 1, \dots\}$.
- Symbolschreibweise für geometrische Verteilung mit Parameter p : $\text{Geom}(p)$

Poisson-Verteilung

- „Grenzverteilung“ der Binomialverteilung
- Verwendung:
 - ▶ Approximation einer $B(n, p)$ -Verteilung, wenn n (sehr) groß und p (sehr) klein ist.
 - ▶ „Faustregeln“ zur Anwendung der Approximation:

$$n \geq 50, \quad p \leq 0.1, \quad n \cdot p \leq 10$$
 - ▶ Poisson-Verteilung hat einzigen Parameter $\lambda > 0$, der zur Approximation einer $B(n, p)$ -Verteilung auf $\lambda = n \cdot p$ gesetzt wird.
- Träger von Poisson-verteilten Zufallsvariablen X : $T(X) = \mathbb{N}_0 = \{0, 1, \dots\}$
- Wahrscheinlichkeitsfunktion für $x \in T(X)$: $p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, wobei $e = \exp(1)$ die Eulersche Zahl ist, also $e \approx 2.71828$.
- Gültigkeit der Approximation beruht auf Konvergenz der Punktwahrscheinlichkeiten. Es gilt nämlich für alle $x \in \mathbb{N}_0$:

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ n \cdot p \rightarrow \lambda}} \binom{n}{x} p^x (1 - p)^{n-x} = \frac{\lambda^x}{x!} e^{-\lambda}$$
- Symbolschreibweise für Poisson-Verteilung mit Parameter λ : $\text{Pois}(\lambda)$

Geometrische Verteilung

Parameter:

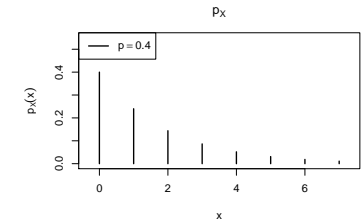
$\text{Geom}(p)$

$p \in (0, 1)$

Träger: $T(X) = \mathbb{N}_0 = \{0, 1, \dots\}$

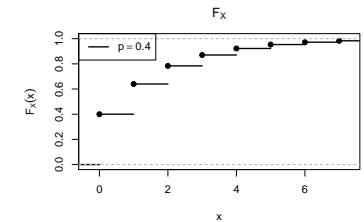
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} (1 - p)^x \cdot p & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - (1 - p)^{\lfloor x \rfloor + 1} & \text{für } x \geq 0 \end{cases}$$



Momente: $E(X) = \frac{1-p}{p}$ $\text{Var}(X) = \frac{1-p}{p^2}$
 $\gamma(X) = \frac{2-p}{\sqrt{1-p}}$ $\kappa(X) = \frac{p^2 - 9p + 9}{1-p}$

Poisson-Verteilung

Parameter:

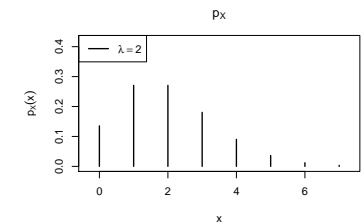
$\text{Pois}(\lambda)$

$\lambda > 0$

Träger: $T(X) = \mathbb{N}_0 = \{0, 1, \dots\}$

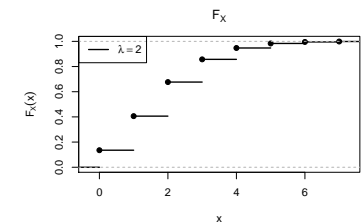
Wahrscheinlichkeitsfunktion:

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{für } x \in T(X) \\ 0 & \text{sonst} \end{cases}$$



Verteilungsfunktion:

$$F_X(x) = \sum_{\substack{x_i \in T(X) \\ x_i \leq x}} p_X(x_i)$$



Momente: $E(X) = \lambda$ $\text{Var}(X) = \lambda$
 $\gamma(X) = \frac{1}{\sqrt{\lambda}}$ $\kappa(X) = 3 + \frac{1}{\lambda}$