Statistik & Datenanalyse

Diskrete Zufallsvariablen und ihre Verteilungen

Prof. Dr. Lukas F. Stoetzer

Diskrete Zufallsvariablen und ihre Verteilungen

Diese Woche

- **Statistikvorlesung**
  1. Diskrete Zufallsvariablen und ihre Verteilungen
  2. Verteilungsannahmen (Bernoulli, Binomial, Uniform)
- **Datenanalysevorlesung**
  1. Erzeugung zufälliger Rohdaten in R
  2. Diskrete Verteilungen (Bernoulli, Binomial, Uniform)
  3. Wahrscheinlichkeits-Masse-Funktion (PMF)

Warum Zufallsvariablen?

- Zufallsvariable ist eines der wichtigsten Konzepte in der Statistik.
- Eine Zufallsvariable ordnet den Ergebnissen eines Zufallsexperiments eindeutige numerische Werte zu. Dadurch entsteht ein Prozess, der unsichere Ergebnisse erzeugt.
- Zufallsvariablen sind mit stochastischen Prozessen verbunden, die eine wichtige Rolle bei der Verwaltung eines Unternehmens, bei Wirtschafts Prozessen und in der Politik spielen.

Beispiel für Anwendungen von Zufallsvariablen im Management:

Beispiel Managment, Startup Erfolg

Übersteht ein Startup die ersten drei Jahre? Das ist nicht immer eindeutig vorherzusehen. "Mehr als 80 Prozent aller Startups scheitern innerhalb von drei Jahren, einige Zahlen gehen auch von 90 Prozent und mehr aus."[Link Gruenderpilot].

Aufgabe: Beispiele

Überlegen sie sich jeweils ein Beispiel für Zufallsvariablen aus der Volkswirtschaftslehre und der Politikwissenschaft.

Definition Zufallsvariabel

Wie können wir unsere Beobachtungen formaler beschreiben?

Zufallsvariabel

Betrachten wir ein Zufallsexperiment mit einem Stichprobenraum Ω. Eine Funktion $X : \Omega \to \mathbb{R}$, die jedem Element $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet, wird Zufallsvariable genannt. Der beobachtete Wert von X in einem bestimmten Experiment wird mit dem entsprechenden Kleinbuchstaben x bezeichnet.

Mehrere Zufallsvariable

Mehrere Zufallsvariablen $X_1, X_2, \ldots, X_n$, die alle das Ergebnis verschiedener Experimente beschreiben, nennt man Zufallsstichprobe.

Diskrete und Kontinuierliche Zufallsvariabeln

Sei $Im(X) = \{y \in \mathbb{R} \mid \exists \omega \in \Omega : y = X(\omega)\}$ und definiert das Bild von X. Wenn $Im(X)$ abzählbar ist, heißt X diskrete Zufallsvariable. Andernfalls wird X als kontinuierliche Zufallsvariable bezeichnet.

Zufallsvariabel

Beispiel: Lebensdauer einer zufällig ausgewählten Batterie

- Stichprobenraum: $\Omega = [0, \infty)$
- Zufallsvariable $X(\omega) = \omega$ misst die Lebensdauer einer Batterie
- $Im(X) = [0, \infty) \implies X$ kontinuierliche Zufallsvariable
- Auch wichtig für Hersteller: Übersteigt die Lebensdauer die Grenze t?

  $\Rightarrow$ Definiere Zufallsvariable Y mit $Y(\omega) = \begin{cases} 0 & \text{if } \omega < t \\ 1 & \text{if } \omega \geq t \end{cases}$
- $Im(Y) = \{0, 1\} \Rightarrow Y$ diskrete Zufallsvariable

Wie kann man die Wahrscheinlichkeiten für Ergebnisse von Zufallsvariablen beschreiben?

Wahrscheinlichkeitsfunktion

Wahrscheinlichkeitsfunktion

Die Wahrscheinlichkeitsfunktion einer Zufallsvariablen X ist ein Wahrscheinlichkeitsmaß P, das Ereignissen, die X betreffen, Wahrscheinlichkeiten zuordnet.

Beispiel: Erzeugung von Batterien

Die Verteilung von Y (Lebensdauer über einer Grenze) erfüllt für $q \in [0,1]$:

$$
\begin{aligned}
P(Y=0) &= P(\omega < t) = q \\
P(Y=1) &= 1-q \\
P(Y \in \{0,1\}) &= 1
\end{aligned}
$$

Verteilungsfunktion

Können wir theoretische Entsprechungen zur empirischen Verteilungsfunktion und zu den Stichprobenquantilen definieren?

Kumulative Verteilungsfunktion

Für einen Zufallsvariable X mit der Verteilung P heißt eine Funktion $F_X : \mathbb{R} \to [0, 1]$, die durch $F_X(x) = P(X \leq x)$ definiert ist, kumulative Verteilungsfunktion von X.

Verteilungsfunktion

Quartile der Verteilungsfunktion

Für einen Zufallsvariable X mit der Verteilung P und einem beliebigen $p \in (0,1)$ ist das entsprechende p - Quantil $Q_p(X)$ definiert als der kleinste Wert x_p, der

$$p \leq F_X(x_p) = P(X \leq x_p)$$

$Q_{0.25}(X)$ nennt man das erste Quartil, $Q_{0.5}(X)$ das zweite Quartil oder den Median und $Q_{0.75}(X)$ das dritte Quartil der Verteilung von X.

Biespiel ff.: Lebensdauer

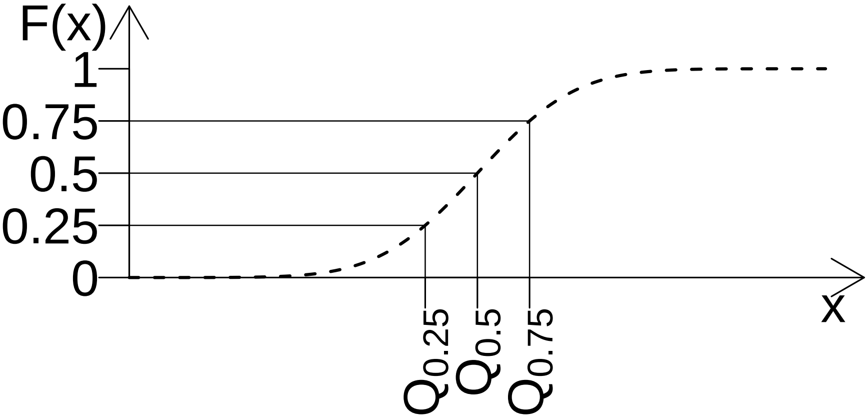
Definieren wir X wieder als die Lebensdauer einer Batterie.

- Wir können dann die Frage stellen wie wahrscheinlich ist es das die Batterie mindestens 60 Zeiteinheiten (t, z.B. Stunden) lang lebt? Dafür nutzen wir die kumulative Verteilungsfunktion. z.B.:

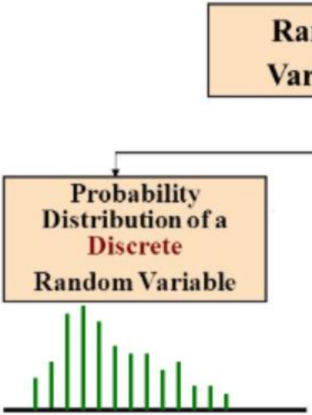
$$F_X(t) = P(X \geq 60) = 0.4$$

- Oder: die schwächsten 25% der Batterien halten nur bis zu? Dazu helfen die Quartile der Verteilungsfunktion.

$$Q_{0.25}(X) = 47.$$



F(x)
1
0.75
0.5
0.25
0
x
$Q_{0.25}$
$Q_{0.5}$
$Q_{0.75}$


Random Variables
Probability Distribution of a Discrete Random Variable
Probability Distribution of a Continuous Random Variable

Wahrscheinlichkeitsfunktion von kretekreten Zufallsvariablen

Für diskrete Zufallsvariablen bilden wir die Ausprägungen mit ihre Wahrscheinlichkeiten ab:

Wahrscheinlichkeitsfunktion einer diskrete Zufallsvariablen

Die Wahrscheinlichkeitsmassenfunktion (Dichtefunktion) einer diskreten Zufallsvariablen X ist eine Funktion $f_X : \mathbb{R} \to [0,1]$ mit $f_X(x) = P(X = x)$.

Wahrscheinlichkeitsfunktion von diskreten Zufallsvariablen

f_X erfüllt die folgenden Eigenschaften:

a) $f_X(x) > 0$ for $x \in Im(X)$

b) $P(X \in A) = \sum_{x \in A \cap Im(X)} f_X(x)$ for $A \subset \mathbb{R}$

c) $1 = P(X \in Im(X)) = \sum_{x \in Im(X)} f_X(x)$

d) $F_X(x) = P(X \leq x) = \sum_{y \in Im(X) \cap (-\infty, x]} f_X(y)$ for $x \in \mathbb{R}$

Beispiel: Zuällige gezogene Chips

Ziehen Sie einen von 6 Chips mit den Zahlen 1, 1, 1, 1, 2, 2 und 3.

- Beispielraum: $\Omega = \{1, 2, 3\}$
- Zufallsvariabel: $X(\omega) = \omega \Rightarrow Im(X) = \{1, 2, 3\}$
- Wahrscheinlichkeitsfunktion:

$$f_X(x) = P(X = x) = \begin{cases} \frac{3}{6} & \text{if } x = 1 \\ \frac{2}{6} & \text{if } x = 2 \\ \frac{1}{6} & \text{if } x = 3 \\ 0 & \text{else} \end{cases} = \frac{4-x}{6} \cdot \mathbb{I}_{\{1,2,3\}}(x)$$

- Wahrscheinlichkeit:

$$P(X \in [1.5, 4]) = \sum_{x \in [1.5,4] \cap \{1,2,3\}} f_X(x) = f_X(2) + f_X(3) = \frac{1}{2}$$

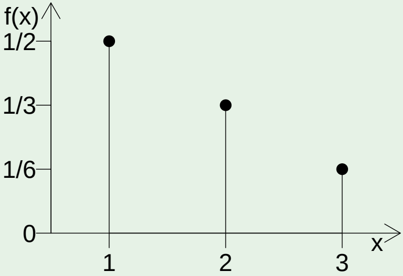
Beispiel: Zuällige gezogene Chips

- Verteilungsfunktion:
  $F_X(1) = P(X \leq 1) = f_X(1) = \frac{1}{2}$
  $F_X(2) = P(X \leq 2) = f_X(1) + f_X(2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$
  $F_X(3) = P(X \leq 3) = f_X(1) + f_X(2) + f_X(3) = 1$

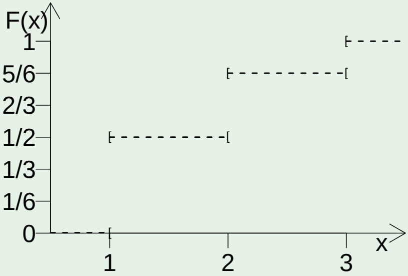
$$\Rightarrow F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{2} & \text{if } x \in [1, 2) \\ \frac{5}{6} & \text{if } x \in [2, 3) \\ 1 & \text{else} \end{cases}$$

Beispiel: Zuällige gezogene Chips

Wahrscheinlichkeitsfunktion:



Verteilungsfunktion:



Beispiel: Zuällige gezogene Chips

Wenn wir das Experiment oft wiederholen, werden sich die relativen Häufigkeiten h den Werten der Wahrscheinlichkeitsfunktion f_X annähern. Auch die empirischen Verteilungsfunktionen der Stichproben werden gegen die (theoretische) Verteilungsfunktion von X konvergieren.

	Wiederhole das Experiment n Mal			
n	10	100	1000	10000
$h(1)$	$\frac{4}{10} = 0.4$	$\frac{43}{100} = 0.43$	$\frac{462}{1000} = 0.462$	$\frac{5025}{10000} = 0.5025$
$h(2)$	$\frac{4}{10} = 0.4$	$\frac{35}{100} = 0.35$	$\frac{373}{1000} = 0.373$	$\frac{3336}{10000} = 0.3336$
$h(3)$	$\frac{2}{10} = 0.2$	$\frac{22}{100} = 0.22$	$\frac{165}{1000} = 0.165$	$\frac{1639}{10000} = 0.1639$

Erwartung & Varianz für diskrete Zufallsvariablen

Gibt es theoretische Entsprechungen zum Stichprobenmittelwert und zur Varianz?

Erwartungswert und Varianz

Sei X ein diskreter Zufallsvariable mit der Dichte f_X. Die Summe

$$E(X) = \mu_X = \sum_{x \in Im(X)} x f_X(x),$$

wird, sofern sie existiert, als Erwartungswert oder Erwartungswert von X bezeichnet. Er beschreibt die erwartete Lage der Werte von X.Die Summe

$$Var(X) = \sigma_X^2 = \sum_{x \in Im(X)} (x - E(X))^2 f_X(x),$$

wird, falls sie existiert, die Varianz von X genannt. Die Quadratwurzel $\sqrt{\sigma_X^2} = \sigma_X$ nennt man die Standardabweichung von X. Sie beschreiben die erwartete Variation der Werte von X.

Beispiel ff: Zuällige gezogene Chips

- Erwartungswert: $E(X) = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6} = \frac{5}{3}$
- Varianz und Standardabweichung:

$$
\begin{aligned}
Var(X) &= \left(1 - \frac{5}{3}\right)^2 \frac{3}{6} + \left(2 - \frac{5}{3}\right)^2 \frac{2}{6} + \left(3 - \frac{5}{3}\right)^2 \frac{1}{6} \\
&= \frac{4}{9} \cdot \frac{3}{6} + \frac{1}{9} \cdot \frac{2}{6} + \frac{16}{9} \cdot \frac{1}{6} \\
&= \frac{12 + 2 + 16}{54} = \frac{30}{54} = \frac{5}{9} \\
\sigma_X &= \frac{\sqrt{5}}{3}
\end{aligned}
$$

Exkurs

- Sei X eine diskrete Zufallsvariable mit Dichte f_X. Gegeben eine Funktion g erhalten wir eine neue Zufallsvariable $Y = g(X)$. Ihr Erwartungswert kann wie folgt berechnet werden

$$E(Y) = E(g(X)) = \sum_{x \in Im(X)} g(x) f_X(x),$$

  also ohne Kenntnis der Dichte f_Y.
- $Var(X) = E((X - E(X))^2)$, so dass die Varianz die erwartete quadratische Abweichung von X von seinem Erwartungswert ist.

$$\boxed{Var(X) = E\left(X^2\right) - E(X)^2}$$

Exkurs

Beweis:

$$
\begin{aligned}
Var(X) &= \sum_{x \in Im(X)} (x - E(X))^2 f(x) \\
&= \sum_{x \in Im(X)} \left(x^2 - 2xE(X) + E(X)^2\right) f(x) \\
&= \underbrace{\sum_{x \in Im(X)} x^2 f(x)}_{=E(X^2)} - 2E(X) \underbrace{\sum_{x \in Im(X)} x f(x)}_{=E(X)} \\
&\quad + E(X)^2 \underbrace{\sum_{x \in Im(X)} f(x)}_{=1} \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

Exkurs

Sei X eine Zufallsvariable und $a, b \in \mathbb{R}$:

$$\begin{aligned} E(aX + b) &= a\,E(X) + b \\ Var(aX + b) &= a^2\,Var(X) \end{aligned}$$

Diskrete Gleichverteilung

Diskrete Gleichverteilung

- Ein Verteilung P einer diskreten Zufallsvariable X nennen wir (diskrete) Gleichverteilung (mit m als Parameter), wenn X eine konstante Dichtefunktion besitzt:

$$f_X(x) = P(X = x) = \frac{1}{m}$$

  für all $x \in Im(X)$, wobei $m = \#\{\omega \in Im(X)\}$.
- Die diskrete Gleichverteilung wird abgekürzt mit: $X \sim \mathcal{DU}(m)$.
- Der Erwartungswert und die Varianz sind gegeben durch:
  - $E(X) = \frac{m+1}{2}$
  - $Var(X) = \frac{m^2-1}{12}$

Diskrete Gleichverteilung

Beispiel ff: 6-kantiger Würfel

- Definition: $\Omega = \{1, \ldots, 6\}, X(\omega) = \omega \Rightarrow Im(X) = \{1, \ldots, 6\}$
- Dichtefunktion: $f_X(x) = \frac{1}{6} \cdot \mathbb{I}_{\{1,\ldots,6\}}(x)$
- $E(X) = \sum_{x \in \{1,\ldots,6\}} x f_X(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = 3.5$
- $Var(X) = \sum_{x \in \{1,\ldots,6\}} (x - 3.5)^2 f_X(x) = \cdots = \frac{35}{12}$

Aufgabe: Diskrete Gleichverteilung

Zeichnen Sie die Dichtefunktion der Diskrete Gleichverteilung für:
$m = 10$
Was ist der Erwartungswert und Varianz?

Die Bernoulli Verteilung

Ein viel verwendetes Experiment mit Stichprobenraum $\Omega = \mathit{Erfolg}, \mathit{Misserfolg}$ wird als Bernoulli-Versuch bezeichnet. Dabei können wir die Zufallsvariable X als $X(\text{Erfolg}) = 1$ $X(\text{Misserfolg}) = 0$ definieren, und $Im(X) = \{0, 1\}$.

Die Bernoulli Verteilung

Die Bernoulli Verteilung

- Eine Verteilung P der diskreten Zufallsvariabel X mit $Im(X) = \{0, 1\}$ wird Bernoulli Verteilung genannt. Sie besitzt den Parameter p.
- Mit $p \in [0, 1]$ hat X die Dichte Funktion:

$$\begin{aligned} f_X(x) = P(X = x) &= \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} \\ &= p^x (1 - p)^{1-x} \end{aligned}$$

  für $x \in \{0, 1\}$.
- Wir schreiben auch $X \sim \mathcal{B}(p)$.
- Der Erwartungswert und die Varianz sind gegeben durch:

$$\begin{aligned} E(X) &= p \\ Var(X) &= p \cdot (1 - p) \end{aligned}$$

Exkurs

Für $X \sim \mathcal{B}(p)$:

$$\begin{aligned} E(X) &= 0 \cdot (1-p) + 1 \cdot p = p \\ Var(X) &= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 \\ &= p \cdot (1-p) \end{aligned}$$

Beispiel: Tombula mit 4 weißen und 6 schwarzen Bällen

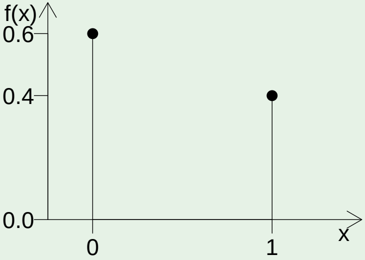
○ ○ ○ ○ ● ● ● ● ● ●

- Experiment: Zieh einen Ball
- Zufallsvariable: $X(\text{white chip}) = 1, X(\text{weißer Ball}) = 0 \Rightarrow X \sim \mathcal{B}(0.4)$
- Dichtefunktion:

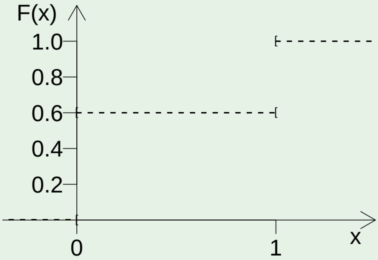
$$f_X(x) = P(X = x) \quad = \quad \begin{cases} \color{red}{0.4} & \text{if } x = 1 \\ \color{red}{0.6} & \text{if } x = 0 \\ 0 & \text{else} \end{cases}$$

Beispiel: Tombula mit 4 weißen und 6 schwarzen Bällen

Dichtefunktion:



Verteilungsfunktion:



$E(X) = 0.4, \; Var(X) = 0.4 \cdot 0.6 = 0.24$

Binomial Verteilung

Wir können die Anzahl der Erfolge aus n Bernoulli-Versuchen zählen: $X = \sum_{i=1}^{n} X_i$:

Binomial Verteilung

- Eine Verteilung P einer diskreten Zufallsvariable $X = \sum_{i=1}^{n} X_i$ aus unabhängigen Bernoulli-Versuchen $X_i \sim \mathcal{B}(p)$ nennen wir Binomial Verteilung, mit Parametern $n \in \mathbb{N}$ and $p \in [0, 1]$.
- X hat die Dichte[a]:

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- Abgekürzt schreiben wir: X $\sim \mathcal{B}(n, p)$.
- Erwartungswert un Varianz sind: $E(X) = np$, $Var(X) = np(1-p)$

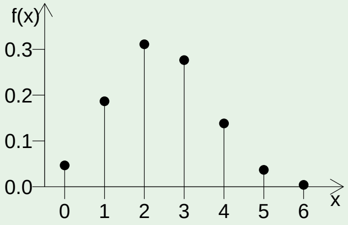
[a] Die Anzahl der möglichen Anordnung von x aus n Versuchen ist $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, wobei $y! = y \cdot (y-1) \cdots 2 \cdot 1$ für $y \in \mathbb{N}$.

Binomial Verteilung

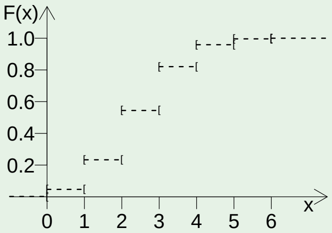
Beispiel ff.: Tombula mit 4 weißen und 6 schwarzen Bällen

- Experiment: Ziehen sechs Chips mit Zurücklegen
- Zufallsvariable: X=Anzahl der weißen Chips
- Parameter: $n = 6, p = 0.4$
- $E(X) = 6 \cdot 0.4 = 2.4$, $Var(X) = 6 \cdot 0.4 \cdot 0.6 = 1.44$

Dichtefunktion:

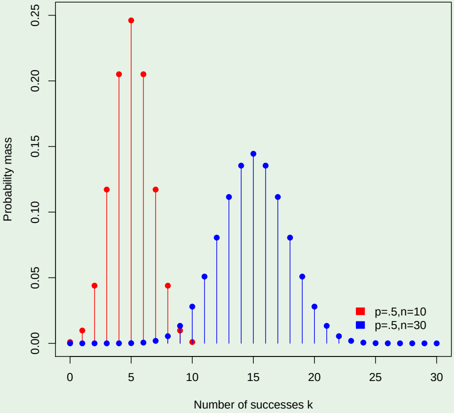


Verteilungsfunktion:



Binomial Verteilung

Beispiel: Varierende



Aufgabe: Binomial Verteilung

Zeichnen Sie die Dichtefunktion für die folgende Binomial Verteilung:
$p = 0.2, n = 8$
Was ist der Erwartungswert und Varianz?

Zusammenfassung

Zusammenfassung Woche 5 Diskrete Zufallsvariablen und ihre Verteilungen

- Definition Zufallsvariabel
- Wahrscheinlichkeitsmassenfunktion (Dichtefunktion) & Kumulative Verteilungsfunktion
- Erwartung & Varianz für diskrete Zufallsvariablen
- Diskrete Wahrscheinlichkeitsverteilungen
  - Diskrete Gleichverteilung
  - Bernoulli Verteilung
  - Binominal Verteilung