

# Zusammenfassung: Statistik

## Attribute und ihre Werte

- *qualitativ*: Familienstand, Geschlecht, Konfession, ...
- *Rangmerkmal*: Schulnote, Dienstgrad, ...
- *quantitativ-diskret*: Anzahl der Fachsemester, ...
- *quantitativ-stetig*: Größe, Entfernung, ...

## Skalierungen

- *normalskaliert*: keine Rangfolge, z. B. Familienstand, Geschlecht, Konfession
- *ordinalskaliert*: Anordnung/Rangfolge, z. B. Schulnote, Dienstgrad
- *kardinalskaliert*: Werte sind reelle Zahlen; Aussagen wie stark sich zwei Werte
  - *intervallskaliert*: Nullpunkt willkürlich, Verhältnisse nicht sinnvoll, z. B. Temperatur
  - *verhältnisskaliert*: natürlicher Nullpunkt, Verhältnisse sinnvoll, z. B. Körpergröße, Alter

## Kennwerte einer Stichprobe und Streuungsmaße

arithmetisches Mittel:  $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$

geometrisches Mittel:  $\bar{x}_{geom} = \sqrt[n]{\prod_{i=1}^n x_i}$

Median:  $\tilde{x} = \begin{cases} x_{(n+1)/2} & n \text{ ungerade} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & n \text{ gerade} \end{cases}$

(50 % der Werte sind kleiner)

$\rho$ -Quantil:  $\tilde{x}_\rho = \begin{cases} x_{[np]+1} & np \notin \mathbb{N} \\ \frac{1}{2}(x_{np} + x_{np+1}) & np \in \mathbb{N} \end{cases}$

( $\rho$ -Anteil der Werte sind kleiner)

Modalwert/Modus: häufigster Wert

Spannweite:  $R = x_{\max} - x_{\min}$

emp. Varianz:  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

emp. Standardabweichung:  $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

## Lineare Korrelation

gegeben ist zweidimensionale Stichprobe mit Wertepaaren  $(x_1, y_1) \dots (x_n, y_n)$

emp. Kovarianz:  $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  emp. Korrelationskoeffizient:  $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$

(emp. Korrelationskoeffizient bestimmt Grad der linearen Abhängigkeit)

## Regression

Regressionsgerade:  $f(x) = kx + d$  mit  $k = r_{xy} \cdot \frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2}$  und  $d = \bar{y} - k \cdot \bar{x}$

## Methode der kleinsten Quadrate

allgemeiner Ansatz:  $f(x) = \text{allgemein} \rightarrow$  Minimum von  $\sum_{i=1}^n (f(x_i) - y_i)^2$  finden

$\rightarrow$  nach jedem Parameter einmal partiell Ableiten und Ableitungen 0 setzen  $\rightarrow$  LGS lösen

für Parabel  $y = a \cdot x^2 + b$ :  $\Delta(a, b) = \sum_{i=1}^n (x_i^2 \cdot a + b - y_i)$

**Beispiel: Berechnung linearer Ausgleichskurve mit Methode der kleinsten Quadrate**

geg.:  $x_1=1960, \dots, x_4=1990$ ;  $y_1=1.5, \dots, y_4=5$  ges.:  $y=a \cdot x+b$

$$Q:=(1.5-(a \cdot 1960+b))^2+\dots+(5-(a \cdot 1990+b))^2$$

Berechne  $a$  und  $b$  mit  $\frac{dQ}{da}=\dots=0$  und  $\frac{dQ}{db}=\dots=0$

evtl. zur Bewertung am Ende  $\Delta$  berechnen: alle Werte in  $Q$  einsetzen

**Wahrscheinlichkeitsrechnung**

**Begriffe**

*Ergebnismenge  $\Omega$*  : Menge möglicher Ergebnisse, z. B. bei 2 Würfeln  $\Omega=\{2,3,\dots\}$

*Elementarereignisse*: Elemente  $x \in \Omega$

*Ereignis  $A$* : Teilmenge  $A \subseteq \Omega$ , z. B. ungerader Wurf mit 2 Würfeln  $A=\{3,5,7,9,11\}$

*Ereignisraum  $(\Omega, \mathcal{Y})$*  : Skript S.36 - TODO: anschauliche Erklärung

es gilt:  $\Omega \in \mathcal{Y}$

ist  $\Omega$  endlich und  $\{x\} \in \mathcal{Y} \forall x \in \Omega$  dann gilt  $\mathcal{Y}=P(\Omega)$

$\cup_i A_i$  ist Ereignis "  $A_1$  oder  $A_2$  oder ..." "

*Wahrscheinlichkeit*: jedem Ereignis  $A \in \mathcal{Y}$  wird Wahrscheinlichkeit  $P(A) \in \mathbb{R}_+$  zugeordnet, sodass Gesamtwahrscheinlichkeit  $P(\Omega)$  eins ist

*Wahrscheinlichkeitsmaß*:  $\mathcal{Y} \rightarrow [0,1] \subset \mathbb{R}_+$

Laplace: alle Elementarereignisse haben gleiche Wahrscheinlichkeit

$$\rightarrow \text{prob}(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

Wahrscheinlichkeitsraum:  $(\Omega, \mathcal{Y}, P)$

$$P(\bar{A})=1-P(A) \text{ und } P(\emptyset)=0$$

es gilt: und  $P(A \cup B)=P(A)+P(B)-P(A \cap B)$

und  $P(A) \leq P(B)$  für  $A \subseteq B$

**Bedingte Wahrscheinlichkeit**

Wie wahrscheinlich ist es, dass  $B$  eintritt, wenn  $A$  schon eintrat?  $P(B|A) = \frac{P(A \cap B)}{P(A)}$

Multiplikationsgesetz:  $P(A \cap B)=P(A) \cdot P(B|A)=P(B) \cdot P(A|B)$

*Unabhängige Ereignisse*: es gelten folgende Eigenschaften

$$P(B|A)=P(B), P(A|B)=P(A) \text{ und } P(A \cap B)=P(A) \cdot P(B)$$

*Formel von Bayes*:

$$P(E|A)=P(A|E) \cdot \frac{P(E)}{\sum_{k=1}^n P(E_k) \cdot P(A|E_k)} \leftarrow \text{totale Wahrscheinlichkeit}$$

$$P(E|A) = \frac{P(E) \cdot P(A|E)}{P(E) \cdot P(A|E) + P(\bar{E}) \cdot P(A|\bar{E})}$$

nützliche Umformung:  $P(A|B)=1-P(\bar{A}|B)$

Bsp.:  $P(\text{ok}), P(\text{pos}|\text{ok}), P(\text{neg}|\overline{\text{ok}}) \rightarrow P(\overline{\text{ok}}|\text{neg}) = \frac{P(\overline{\text{ok}}) \cdot P(\text{neg}|\overline{\text{ok}})}{P(\overline{\text{ok}}) \cdot P(\text{neg}|\overline{\text{ok}}) + P(\text{ok}) \cdot P(\text{neg}|\text{ok})}$   
=  $1 - P(\text{pos}|\text{ok})$

## Totale Wahrscheinlichkeit

$$P(A) = \sum_{k=1}^n P(A \cap E_k) = \sum_{k=1}^n P(E_k) \cdot P(A|E_k)$$

$$P(A) = P(E) \cdot P(A|E) + P(\bar{E}) \cdot P(A|\bar{E})$$

## Zuverlässigkeit von Systemen

Serielle Systeme:  $R_s = R_1 \cdot R_2 \cdot \dots \cdot R_n$

Ausfallwahrscheinlichkeit:  $1 - R_s$

Parallele Systeme:  $R_s = 1 - (1 - R_1) \cdot (1 - R_2) \cdot \dots \cdot (1 - R_n)$

## Zufallsvariablen

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n}{k} \cdot \frac{n-1}{k-1} \cdot \frac{n-2}{k-2} \cdot \frac{n-3}{k-3} \dots$$

k Faktoren

## Auswählen

ohne Zurücklegen:  $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

mit Zurücklegen:  $\binom{n+k-1}{k}$

## Grundlegende Definitionen

Verteilungsfunktion  $F(x)$ : Werte von 0 bis 1, monoton steigend, Fläche unter Graph 1

Dichtefunktion  $f(x)$ : Ableitung der Verteilungsfunktion (Wahrscheinlichkeit bei  $x$ )

Unabhängigkeit:  $\text{prob}(X \leq x, Y \leq y) = \text{prob}(X \leq x) \cdot \text{prob}(Y \leq y)$

## Erwartungswert

$$\mu = E(X) = \begin{cases} \sum_i \text{prob}(X = x_i) \cdot x_i & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{falls } X \text{ stetig mit Dichte } f \end{cases}$$

- $E(X+Y) = E(X) + E(Y)$  und  $E(aX) = a \cdot E(X)$  (Linearität)
- $E(aX+b) = a \cdot E(X) + b$
- $E(X \cdot Y) = E(X) \cdot E(Y)$  (nur für unabhängige Zufallsvariablen)
- wenn stetische Zufallsvariable symmetrisch um Punkt  $c$ , also  $f(c-x) = f(c+x)$  für alle  $x$ , dann gilt  $E(X) = c$

## Varianz

$$\sigma^2 = \text{Var}(X) = E((X - \mu)^2) = \begin{cases} \sum_i (x_i - \mu)^2 \cdot p_i & \text{falls } X \text{ diskret ist} \\ \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx & \text{falls } X \text{ stetig ist} \end{cases}$$

$$\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2$$

Standardabweichung:  $\sigma = \sqrt{\sigma^2}$

- wenn  $Y = a \cdot X + b$ :  $\text{Var}(Y) = a^2 \cdot \text{Var}(X)$  bzw.  $\sigma_Y = |a| \cdot \sigma_X$   
(z. B. bei °C/°F-Umrechnung nützlich)  
 $\text{Var}(X+Y) = \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y)$   
wobei:  $\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(X \cdot Y) - \mu_X \mu_Y$
- Kovarianz:  
 $\text{Cov}(X, Y)^2 \leq \text{Var}(X) \cdot \text{Var}(Y)$   
wenn  $X$  oder  $Y$  konst. oder  $Y = a \cdot X + b$
- für unabhängige Zufallsvariablen:  
 $\text{Cov}(X, Y) = 0$  und  $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

## Beispiel zu $E(X)$ und $\text{Var}(X)$ - diskret: $n$ -mal Würfeln

Sechseckiger Würfel wird 80 mal geworfen:

$X_i$  ist Augenzahl des  $i$ -ten Wurfs, somit ist  $S := \sum_{i=1}^{80} X_i$  Augensumme und  $\bar{X} := S/80$

$$\rightarrow E(X_i) = \frac{1}{6} \sum_{k=1}^6 k = 3.5, \quad E(S) = \sum_{i=1}^{80} E(X_i) = 80 \cdot 3.5 = 280, \quad E(\bar{X}) = \frac{E(S)}{80} = 3.5$$

$$\rightarrow \text{Var}(X_i) = \frac{1}{6} \sum_{k=1}^6 k^2 - \left(\frac{7}{2}\right)^2 = \frac{35}{12}, \quad \text{Var}(S) = \sum_{i=1}^{80} \text{Var}(X_i) = 80 \cdot \frac{35}{12} = \frac{700}{3}, \quad \text{Var}(\bar{X}) = \frac{\text{Var}(S)}{80^2} = \frac{7}{192}$$

### Weiteres Beispiel zu E(X) und Var(X) - stetig: Dichte gegeben

$$f(x) = \frac{2}{3}x \text{ für } 1 < x < 2 \rightarrow E(X) = \int_1^2 x \cdot \frac{2}{3}x \, dx = \dots = \frac{14}{9} \text{ und } \text{Var}(X) = \int_1^2 x^2 \cdot \frac{2}{3}x \, dx - \left(\frac{14}{9}\right)^2 = \dots$$

### Weiteres Beispiel: Wahrscheinlichkeiten aus Erwartungswert und Varianz

geg.: X kann Werte 1, 2 und 3 annehmen;  $\mu = 2$  und  $\sigma^2 = 0.5$

$$\rightarrow \text{LGS lösen: } p_1 + p_2 + p_3 = 1 \text{ (I), } 1p_1 + 2p_2 + 3p_3 = 2 \text{ (II), } (1^2p_1 + 2^2p_2 + 3^2p_3) - 2^2 = 0.5 \text{ (III)}$$

### Momente

$$k\text{-te Moment: } m_k(X) = E(X^k) \quad k\text{-te zentrierte Moment: } m_k(X - \mu) = E((X - \mu)^k)$$

### Standardisierung einer Zufallsvariablen

$Z = X - \mu$ ,  $E(Z) = 0$ ,  $\text{Var}(Z) = 1$  (für Beispiel siehe Standardnormalverteilung)

für Verteilungsfunktion gilt:  $F_X(\sigma \cdot z + \mu) = F_Z(z)$  für p-Quantile gilt:  $x_p = \sigma \cdot z_p + \mu$

### Tschebyscheff und das Gesetz der großen Zahlen™

stochastische Konvergenz:  $\lim_{t \rightarrow \infty} \text{prob}(|\bar{X} - \mu| \geq c) = 1$  Ungleichung:  $P(|\bar{X} - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$

Theorem von Bernoulli:  $\lim_{t \rightarrow \infty} \text{prob}(|f_t - p| \leq c) = 1$  emp. Verteilf.:  $\bar{F}(x) = \frac{|\{i \mid x_i \leq x\}|}{n}$

### Gleichverteilung (Rechteckverteilung) in Intervall [a; b]

$$F(X) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{b-a} & \text{für } a < x < b \\ 1 & \text{für } x \geq b \end{cases} \text{ bzw. } f(X) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{sonst} \end{cases} \quad \mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$$

### Beispiel: $T_1, \dots, T_{81}$ unabhängige Zufallsvariablen auf Intervall [5; 15]

$$\text{außerdem geg.: } \bar{T} := \frac{1}{81} \cdot (T_1 + \dots + T_{81}) \text{ und } S := T_1 + \dots + T_{81}$$

$$\text{Mittelwert von } \bar{T} \text{ und } S: E(\bar{T}) = \frac{1}{81} \cdot 81 \cdot \frac{5+15}{2} = 10, \quad E(S) = 81 \cdot \frac{5+15}{2} = 810$$

Standardabweichung von  $\bar{T}$  und S:

$$\text{Var}(\bar{T}) = \frac{1}{81^2} \cdot 81 \cdot \frac{(15-5)^2}{12} = \frac{25}{243} \rightarrow \sigma_{\bar{T}} \approx 0.032, \quad \text{Var}(S) = 81 \cdot \frac{(15-5)^2}{12} = 675 \rightarrow \sigma_S \approx 25.98$$

Berechnung von Wahrscheinlichkeit mit zentralem Grenzwertsatz Analog zu Bsp. für Exponentialverteilung.

### Exponentialverteilung (beschreibt z. B. Lebensdauer von Bauteilen)

$$F(X) = \begin{cases} 1 - e^{-kx} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \text{ bzw. } f(X) = \begin{cases} k \cdot e^{-kx} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases} \quad \mu = \frac{1}{k}, \quad \sigma^2 = \frac{1}{k^2}$$

### Beispiel: $T_1, \dots, T_{102}$ unabhängige Zufallsvariablen mit Rate $k = 3$

$$\text{außerdem geg.: } \bar{T} := \frac{1}{102} \cdot (T_1 + \dots + T_{102}) \text{ und } S := T_1 + \dots + T_{102}$$

Mittelwert von  $\bar{T}$  und  $S$ :  $E(\bar{T}) = \frac{1}{102} \cdot 102 \cdot \frac{1}{3} = \frac{1}{3}$ ,  $E(S) = 102 \cdot \frac{1}{3} = 34$

Standardabweichung von  $\bar{T}$  und  $S$ :

$$\text{Var}(\bar{T}) = \frac{1}{102^2} \cdot 102 \cdot \frac{1}{3^2} = \frac{1}{918} \rightarrow \sigma_{\bar{T}} \approx 0.033, \quad \text{Var}(S) = 102 \cdot \frac{1}{3^2} = \frac{34}{3} \rightarrow \sigma_S \approx 3.3665$$

Berechnung von Wahrscheinlichkeit mit zentralem Grenzwertsatz, z. B.

$$P(S > 40) = 1 - F(40) = 1 - \Phi\left(\frac{40 - 34}{3.3665}\right) \approx 0.0375$$

## Diskrete Verteilungen

### Hypergeometrische Verteilung: Ziehen einer Stichprobe *ohne* Zurücklegen

$N$ : Gesamtzahl,  $M$ : Elemente mit bestimmter Eigenschaft,  $n$ : Größe der Stichprobe

$$\text{prob}(X=x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad \mu = E(X) = n \cdot \frac{M}{N}, \quad \sigma^2 = \text{Var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

Bsp.: Batterieladegeräte kann gleichzeitig 5 Batterien prüfen. Unter 25 Batterien sind 2 fehlerhaft. Wie groß ist die Wahrscheinlichkeit, dass diese gleich beim ersten Test entdeckt werden?  $\rightarrow x = 2, N = 25, n = 5, M = 2$

### Binomialverteilung: Ziehen einer Stichprobe *mit* Zurücklegen $X \sim \text{Bi}(n; p)$

*Bernoulli-Experiment*: Zufallsereignis mit genau 2 möglichen Ereignissen: A tritt ein oder nicht.

*Bernoulli-Kette* der Länge  $n$ :  $n$ -maliges Wiederholen eines Bernoulli-Experiments unter selben Bedingungen

$n$  := Anzahl der Ziehungen,  $x$  := Anzahl der Ereignisse,  $p$  := Wahrscheinlichkeit für Ereignis

$$\text{prob}(X=x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \quad \mu = E(X) = np \quad \sigma^2 = \text{Var}(X) = np \cdot (1-p)$$

Bsp.: Anteil fehlerhafter Einheiten ist gleichbleibend 4 %. Wie groß ist Wahrscheinlichkeit, dass in zufälliger Sichtprobe von 50 Einheiten genau 2 fehlerhafte sind?  $\rightarrow x = 2, n = 50, p = 4 \%$

Ist Stichprobengröße  $n$  klein gegenüber  $N$  ( $20 \cdot n \leq N$ ), so kann man die hypergeometrische Verteilung gut durch die Binomialverteilung annähern.

### Poisson -Verteilung: Ereignisse pro Einheit $X \sim \text{Po}(\lambda)$

$x$  := Anzahl der tatsächlichen Ereignisse pro Einheit

$\lambda$  := Anzahl der erwarteten Ereignisse pro Einheit

$$\text{prob}(X=x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad \mu = E(X) = \lambda \quad \sigma^2 = \text{Var}(X) = \lambda$$

Bsp.: durchschnittlich 0.05 Staubteilchen pro  $\text{cm}^2$ , Wie groß ist Wahrscheinlichkeit, auf  $100 \text{ cm}^2$  genau 3 Staubteilchen zu finden?  $\rightarrow \lambda = 5$  (Teilchen pro  $100 \text{ cm}^2$ ),  $x = 3$

Für großes  $n$  und kleines  $p$  kann  $\text{Bi}(n; p)$  durch  $\text{Po}(\lambda)$  mit  $\lambda = np$  genähert werden.

## Spezielle stetige Verteilungen

### Normalverteilung $X \sim N(\mu; \sigma^2)$

Dichtefunktion:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$

Verteilungsfunktion:

$$F(x) = P(X \leq x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} dt$$

Bei jeder Normalverteilung ist die Wahrscheinlichkeit, dass:

- $X \in [\mu - 1\sigma; \mu + 1\sigma] \rightarrow P = 68.3\%$
- $X \in [\mu - 2\sigma; \mu + 2\sigma] \rightarrow P = 95.5\%$
- $X \in [\mu - 3\sigma; \mu + 3\sigma] \rightarrow P = 99.7\%$

### Standardnormalverteilung $X \sim N(0; 1)$

Festlegung:  $\mu = 0, \sigma^2 = 1 \rightarrow f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, F(x) = \Phi$  es gilt:  $\Phi(-x) = 1 - \Phi(x)$

Standardisierung:  $F(x) = \text{prob}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right), f(x) = \frac{1}{\sigma} \cdot \Phi\left(\frac{x-\mu}{\sigma}\right)$

Quantile: siehe Tabelle, es gilt:  $z_{1-p} = -z_p$

### Normalverteilung als Näherung

- wenn  $X \sim (\mu_x; \sigma_x^2)$  und  $Y \sim (\mu_y; \sigma_y^2)$  unabhängige normalverteilte Zufallsvariablen  $\rightarrow X+Y \sim N(\mu_x + \mu_y; \sigma_x^2 + \sigma_y^2)$
- *Zentraler Grenzwertsatz*: Sind  $X_1, X_2, \dots, X_n$  unabhängige, identisch verteilte Zufallsvariablen, dann hat die Summe  $S = \sum_{i=1}^n X_i$  den Erwartungswert  $n \cdot \mu$  und die Varianz  $n \cdot \sigma^2$  für große  $n$  auch  $\bar{X} = 1/n \cdot S$  ist näherungsweise normalverteilt mit  $\mu^* = \sigma^* = \sigma^2/n$ .

für zugehörige standardisierte Zufallsvariable  $Z := \frac{S - \mu}{\sqrt{n\sigma}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  gilt:

$$\lim_{n \rightarrow \infty} \text{prob}(Z \leq z) = \Phi(z) \quad (\text{für Beispiel siehe unten bei Tests})$$

- Näherung einer Binomialverteilung

$$\mu = n \cdot p, \sigma^2 = n \cdot p(1-p), F_B(x) \approx F_N(x+0.5) = \Phi\left(\frac{x+0.5-np}{\sqrt{np(1-p)}}\right) \quad \text{Faustregel: } \sigma^2 \gtrsim 9$$

- Näherung einer Poisson-Verteilung

$$\mu = \lambda, \sigma^2 = \lambda, F_p(x) \approx F_N(x+0.5) = \Phi\left(\frac{x+0.5-\lambda}{\sqrt{\lambda}}\right) \quad \text{Faustregel: } \lambda \gtrsim 9$$

### Chi-Quadrat-Verteilung mit $m$ Freiheitsgraden $Z \sim \chi^2(m)$

$Z := \sum_{i=1}^m X_i^2$  ( $X_1, \dots, X_m$  sind standardnormalverteilte, unabhängige Zufallsvariablen)

- Erwartungswert und Varianz:  $E(\chi^2(m)) = m, \text{Var}(\chi^2(m)) = 2m$
- Seien  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  das arithmetische Mittel und  $S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  die empirische Varianz einer zufälligen Stichprobe  $X_1, \dots, X_n$  aus einer normalverteilten Grundgesamtheit, dann ist  $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2$  Chi-Quadrat-verteilt.
- für große  $m$  Näherung durch Standardnormalverteilung:  $\chi^2(m) \approx N(m; 2m)$

- Quantile: siehe Tabelle; für  $m > 39$ :  $\chi_{m;p}^2 \approx m \left( 1 - \frac{2}{9m} + z_p \sqrt{\frac{2}{9m}} \right)^3$

### T-Verteilung $T \sim t(m)$

Wobei  $Z$  und  $X$  unabhängige Zufallsvariablen sind

und  $Z \sim N(0; 1)$  standardnormalverteilt ist und  $X \sim \chi^2(m)$

$$T := \frac{Z}{\sqrt{\frac{X}{m}}}$$

- Erwartungswert und Varianz:  $E(T) = 0$  für  $m > 1$ ,  $\text{Var}(T) = \frac{m}{m-2}$  für  $m > 2$
- normalverteilte Grundgesamtheit ist t-verteilt mit  $n-1$  Freiheitsgraden:  $T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$
- für große  $m$  Näherung durch Standardnormalverteilung:  $t(m) \approx N(0; m/(m-2))$
- Quantile: siehe Tabelle; für  $m > 39$ :  $t_{m;p} \approx z_p \left( 1 + \frac{1+z_p^2}{4m} \right)$   
außerdem gilt:  $t_{m;p} = -t_{m;1-p}$
- wenn Stichprobe groß, Standardnormalverteilung anstatt nutzbar (ZGS)

### F-Verteilung $X \sim F(m_1; m_2)$

Wobei  $X_1$  und  $X_2$  unabhängige Zufallsvariablen sind

und  $X_1 \sim \chi^2(m_1)$  und  $X_2 \sim \chi^2(m_2)$

$$X := \frac{\sqrt{X_1/m_1}}{\sqrt{X_2/m_2}}$$

$$E(X) = \frac{m_2}{m_2-2} \text{ für } m_2 > 2 \text{ und } \text{Var}(T) = \frac{m_2^2 \cdot (m_1 + m_2 - 2)}{m_1 \cdot (m_2 - 4) \cdot (m_2 - 2)^2} \text{ für } m_2 > 4$$

## Zufallsstichproben: Folge von unabhängigen Zufallsvariablen

### Punktschätzungen

- Merkmal  $X$  einer Verteilung genügt (z. B. Normalverteilung), deren Parameter ? man nicht kennt (z. B. Erwartungswert, Varianz)
- aus Merkmalswerten der Stichproben wird mit *Schätzfunktion* geschätzter Wert bestimmt

### Eigenschaften von Schätzfunktionen

1. *erwartungstreu/unverzerrt* falls  $E(T) = ?$  *systematischer Fehler*:  $E(T) - ?$
2. *asymptotisch erwartungstreu*, falls  $\lim_{n \rightarrow \infty} E(T_n) = 0$
3. *konsistent*, falls sie stochastisch gegen ? Konvergiert:  $\lim_{n \rightarrow \infty} \text{prob}(|T_n - ?| < \epsilon) = 0$
4. *konsistent im quadratischen Mittel*, falls  $\lim_{n \rightarrow \infty} E((T_n - ?)^2) = 0$   
(wenn Punkte 1. bis 3. gelten)

### Im quadratischen Mittel konsistente und erwartungstreue Schätzer sind

1.  $\bar{X}$  (arithmetisches Mittel) für Erwartungswert  $\mu$  einer reellen Zufallsvariablen  $X$
2.  $\bar{P}$  (empirische Anteil) für den Anteil  $p$  an einer Grundgesamtheit mit einer bestimmten Eigenschaft
3.  $\bar{F}$  (empirische Verteilungsfunktion) für die Verteilungsfunktion  $F$

### Erwartungstreuer und im quadratischen Mittel konsistenter Schätzer für Varianz $\text{Var}(X)$

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2 \right)$$

## Konfidenzintervall für Erwartungswert $\mu$ eines normalverteilten Merkmals bei *bekanntem* $\sigma$

1. wähle Konfidenzniveau:  $1-\alpha$ , z. B. 0.99 ( $\rightarrow \alpha=0.01$ )
2. ziehe eine Stichprobe vom Umfang  $n$  und berechne  $\bar{x}$
3. bestimme Quantil  $z_{1-\alpha/2}$  der Standardnormalverteilung (im Bsp.:  $z_{0.995}=2.5758$ )
4. Intervalle des gesuchten Erwartungswerts  $\mu$  mit der Wahrscheinlichkeit  $1-\alpha$ :
  - zweiseitiges Konfidenzintervall:  $\left[ \bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$
  - einseitige Konfidenzintervalle:  $\left( -\infty, \bar{x} + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} \right]$  und  $\left[ \bar{x} - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}, +\infty \right)$
5. Umkehrung: Wahrscheinlichkeit  $1-\alpha$  für gegebenes, einseitiges Konfidenzintervall:  $z_{1-\alpha} = (\text{Grenze} - \bar{X}) \cdot \frac{\sqrt{n}}{\sigma} \rightarrow p = \Phi(z_{1-\alpha})$   
(bei oberer Grenze noch Vorzeichen vertauschen)

**Länge des Konfidenzintervalls**  $L = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Notwendiger Stichprobenumfang für vorgegebene Maximallänge des

Konfidenzintervalls:  $n \geq \left( \frac{2 z_{1-\alpha/2} \cdot \sigma}{L_{\max}} \right)^2$

## Konfidenzintervall für Erwartungswert $\mu$ eines normalverteilten Merkmals bei *unbekanntem* $\sigma$

Wie bei bekanntem  $\sigma$ , bis auf:

- $s$  schätzen, wenn nicht bekannt (mit Schätzer für Varianz)
- anderes Quantil (bei 3.) zu bestimmen:  $t_{n-1; 1-\alpha/2}$  (wenn  $n > 30$  kann auch Standardnormalverteilung benutzt werden)
- Konfidenzintervall:  $\left[ \bar{x} - t_{n-1; 1-\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1; 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \right]$

## Konfidenzintervall für $\mu$ eines *beliebig verteilten* Merkmals bei *großen* Stichprobenumfang ( $n \gtrsim 30$ )

Wie bei bekanntem  $\sigma$  (wenn  $\sigma$  nicht bekannt schätzen)

## Konfidenzintervall für Vergleich der Erwartungswerte von zwei *Normalverteilungen mit bekannten Standardabweichungen*

Wie bei bekanntem  $\sigma$  nur mit Variablen für beide Normalverteilungen

Konfidenzintervall:  $\left[ \bar{x}_1 - \bar{x}_2 - z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$

## Konfidenzintervall für den Vergleich der Erwartungswerte von zwei *Normalverteilungen mit unbekanntem, aber gleichen Standardabweichungen*

$[g_-, g_+]$  mit  $g_{\pm} = \bar{x}_1 - \bar{x}_2 \pm t_{n_1+n_2-2; 1-\alpha/2} \sqrt{\left( \frac{n_1+n_2}{n_1 \cdot n_2} \right) \left( \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \right)}$



## Konfidenzintervall für $\sigma^2/\sigma$ eines *normalverteilten* Merkmals

$$\text{für } \sigma^2: \left[ \frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}} \right] \quad \text{für } \sigma: \left[ \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}} \right]$$

## Approximatives Konfidenzintervall für eine Wahrscheinlichkeit bzw. einen Anteilswert $p$ bei *großem Stichprobenumfang* ( $n \geq 20$ )

$$[g_-, g_+], g_{\pm} = \frac{n}{n+z_{1-\alpha/2}^2} \cdot \left( \bar{p} + \frac{z_{1-\alpha/2}^2}{2n} \pm z_{1-\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}} \right)$$

## Maximum-Likelihood-Methode

Man wählt Parameter  $\theta$  für den beobachteter Wert von  $X=(X_1, \dots, X_n)$  die größte Wahrscheinlichkeit hat

Likelihood-Funktion wenn  $X_i$  diskret:  $L=(x_1, \dots, x_n, \theta) := \prod_{i=1}^n \text{prob}(X_i=x_i | \theta)$

Likelihood-Funktion wenn  $X_i$  stetig:  $L=(x_1, \dots, x_n, \theta) := \prod_{i=1}^n f(x_i, \theta)$

Beispiel (für stetig):  $F(x) = \begin{cases} 1 - e^{-(\lambda x)^2} & \text{für } X \geq 0 \\ 0 & \text{sonst} \end{cases} \rightarrow f(x) = \begin{cases} 2\lambda^2 x e^{-(\lambda x)^2} & \text{für } X \geq 0 \\ 0 & \text{sonst} \end{cases}$

konkrete Stichprobe: 0.38, 0.41, 0.51  $\rightarrow L(\lambda) = f(0.38) \cdot f(0.41) \cdot f(0.51) = \dots$

Allg.:  $L(\lambda) = \prod_{i=1}^n (2\lambda^2 x_i e^{-(\lambda x_i)^2}) = 2^n \lambda^{2n} \cdot \prod_{i=1}^n (x_i) \cdot e^{-\lambda^2 \sum_{i=1}^n x_i^2}$  in Formel einsetzen

$\ln(L(\lambda)) = n \ln(2) + 2n \ln(\lambda) + \ln\left(\prod_{i=1}^n x_i\right) - \lambda^2 \sum_{i=1}^n x_i^2$  ln anwenden

$0 = \frac{d}{d\lambda} \ln(L(\lambda)) = 2 \frac{n}{\lambda} - 2\lambda \sum_{i=1}^n x_i^2 \Leftrightarrow \lambda = \sqrt{n / \sum_{i=1}^n x_i^2}$  Ableitung nach  $\lambda$  0 setzen  
(Am Ende ist nur Betrag relevant.)

## Tests

### Fehlerarten und Güte

	entschieden für $H_0$	entschieden für $H_1$
$H_0$ wahr	ok	Fehler 1. Art ( $\alpha$ -Fehler)
$H_1$ wahr	Fehler 2. Art ( $\beta$ -Fehler)	ok

Die Wahrscheinlichkeit für einen  $\alpha$ -Fehler wird durch die Wahl des Signifikanzniveaus begrenzt bzw. festgelegt.

Die Wahrscheinlichkeit für einen  $\beta$ -Fehler kann in der Regel nicht vorgegeben werden.

### $p$ -Wert

Wahrscheinlichkeit, bei Gültigkeit von  $H_0$ , den beobachteten Prüfwert oder einen in Richtung von  $H_1$  extremen Wert zu erhalten. ( $H_0$  wird verworfen, falls  $p$ -Wert kleiner als Signifikanzniveau.)

### Test für $\mu$ eines *normalverteilten* Merkmals bei *bekanntem* $\sigma$ (Gauß-Test)

- formuliere Hypothesen
  - $H_0: \mu = \mu_0$  gegen  $H_1: \mu \neq \mu_0$  (zweiseitig)
  - $H_0: \mu \geq \mu_0$  gegen  $H_1: \mu < \mu_0$  (einseitig)
  - $H_0: \mu \leq \mu_0$  gegen  $H_1: \mu > \mu_0$  (einseitig)

- z. B.  $H_0: \mu \leq 100$  (also Fall b))
- wähle Signifikanzniveau  $\alpha$  (z. B. 0.10, 0.05 oder 0.01)
  - ziehe Stichprobe vom Umfang  $n$ , berechne daraus  $\bar{x}$  und  $s$  sowie Prüfwert
- $$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \quad (\text{Die zugehörige Prüfgröße } Z \text{ ist standardnormalverteilt})$$

- $H_0$  verwerfen, falls
  - $|z| > z_{1-\alpha/2}$
  - $z < -z_{1-\alpha}$
  - $z > z_{1-\alpha}$

### t-Test für $\mu$ eines *normalverteilten* Merkmals bei unbekanntem $\sigma$ mit Beispiel

- Signifikanzniveau und Hypothesen wie bei Gauß-Test, z. B.  $H_0: \mu \leq 100$  (also Fall b)
- ziehe Stichprobe vom Umfang  $n$ , berechne daraus  $\bar{x}$  und  $s$  sowie Prüfwert

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \quad (\text{Die zugehörige Prüfgröße } T \text{ ist dann } t\text{-verteilt mit } n-1 \text{ Freiheitsgraden})$$

- z. B. -1.833
- $H_0$  analog zu Gauß-Test verwerfen  
 nur andere Quantile: a)  $t_{n-1; 1-\alpha/2}$  bzw. b), c)  $t_{n-1; 1-\alpha}$  z. B. -1.593  
 im Beispiel wird also  $H_0$  beibehalten, da  $-1.593 < -1.833$

Bemerkung:  $H_0$  wird genau dann abgelehnt, wenn  $\mu_0$  nicht im einseitigen Konfidenzintervall zum Niveau  $1-\alpha$  ist.

### $\chi^2$ -Test für $\sigma^2$ eines *normalverteilten* Merkmals

- wähle Signifikanzniveau, Hypothesen wie bei Gauß-Test nur mit  $\sigma^2$
- ziehe Stichprobe vom Umfang  $n$ , berechne  $s^2$  sowie Prüfwert  $y = \frac{n-1}{\sigma_0^2} \cdot s^2$
- $H_0$  verwerfen, falls
  - $y < \chi_{n-1; \alpha/2}^2$  oder  $y > \chi_{n-1; \alpha/2}^2$
  - $y < \chi_{n-1; \alpha}^2$
  - $y < \chi_{n-1; 1-\alpha}^2$

### Beispieltest mit Normalapproximation der Binomialverteilung

geg.: 4000 Teile überprüft; genau 1049 nicht von erster Wahl

Aufgabe: Teste, dass 75 % der Teile der Produktion nicht von erster Wahl sind mit Signifikanzniveau von 5 %

$$H_0: P(\text{Bauteil von erster Wahl}) = p = 0.75$$

$$X := \text{Anzahl Teile 1. Wahl } k \text{ bei 4000 Teilen} \rightarrow X \sim b(4000; p), P(X=k) = \binom{4000}{k} p^k (1-p)^{4000-k}$$

$$E(X) = n \cdot p = 4000 \cdot 0.75 = 3000, \text{ Var}(X) = n \cdot p(1-p) = \dots = 750 \quad \text{für } k = 4000 - 1049 = 2951$$

$$\Rightarrow P(X \leq (4000 - 1049)) \approx \Phi\left(\frac{k + 0.5 - 3000}{\sqrt{750}}\right) \approx \Phi(-48.5 / \sqrt{750}) \approx 1 - \Phi(1.771) \approx 0.038$$

$\Rightarrow H_0$  wird abgelehnt, da  $0.038 < 5\%$

### $\chi^2$ -Anpassungstest mit Beispiel

Testet für disjunkte Mengen, ob Zufallsvariable  $X$  bestimmte Verteilung besitzt.

- bilde Zerlegung  $A_i$  mit  $1 \leq i \leq N$  in disjunkte Mengen, die alle möglichen Werte von  $X$  umfasst (z. B. „Werkzeuge“  $\rightarrow N = 5$ )
- berechne Wahrscheinlichkeiten  $p_i$  für Ereignisse  $X \in A_i, 1 \leq i \leq N$  unter angenommener Verteilung (im Bsp.  $p_i = 0.2$  für gleichmäßige Verteilung)

$$H_0: P(X \in A_i) = p_i, \quad 1 \leq i \leq N \quad \text{gegen}$$

3. formuliere Hypothese:  $H_1: P(x \in A_i) \neq p_i$  für mindestens ein  $i$

4. wähle Signifikanzniveau  $\alpha$  (z. B. 0.05)

5. ziehe Stichprobe vom Umfang  $n$  stelle Anzahl  $h_i$  der Stichprobenwerte in  $A_i$  fest

und berechne daraus Prüfwert: 
$$T = \sum_{i=1}^N \frac{(n_i - np_i)^2}{np_i} = \frac{1}{n} \left( \sum_{i=1}^N \frac{n_i^2}{p_i} \right) - n$$

(z. B. 37, 29, 26, 27, 31  $\rightarrow n = 150, T = 1/150 \cdot 5 \cdot (37^2 + 29^2 + \dots + 31^2) - 150 = 2.5\bar{3}$ )

6.  $H_0$  verwerfen, falls  $T > \chi_{N-1; 1-\alpha}^2$  (im Bsp.  $\chi_{4; 0.95}^2 = 9.488 \rightarrow$  also Hypothese beibehalten)

Für das Quantil die Anzahl der Mengen minus 1 verwenden und *nicht* den Stichprobenumfang!

### Weiteres Beispiel: Exponentialverteilung

Stichprobe vom Umfang 200: 78 Werte in  $[0; 1]$ , 45 in  $]1; 2]$ , 54 in  $]2; 4]$ , 23 in  $]4; \infty[$

Nullhypothese:  $X$  ist exponentialverteilt mit Parameter 0.5; Signifikanzniveau  $\alpha = 0.05$

1. Wahrscheinlichkeiten für die einzelnen Intervalle mit Verteilfunktion  $F$  für Exponentialverteilung berechnen

2. Prüfwert berechnen: 
$$T = 1/200 \left( \frac{78^2}{0.393} + \dots + \frac{23^2}{0.134} \right) - 200$$

3. Quantil berechnen:  $\chi_{3; 1-0.95}^2 = 7.815 \rightarrow$  Nullhypothese wird beibehalten

### Logarithmenregeln

$$\log_a(u \cdot v) = \log_a u + \log_a v$$

$$\log_a \frac{u}{v} = \log_a u - \log_a v$$

$$\log_a u^v = v \cdot \log_a u$$

$$\log_b r = \frac{\log_a r}{\log_a b}$$