

Statistik & Wahrscheinlichkeitstheorie

Grundbegriffe

Wahrscheinlichkeitsbegriffe

- Statistischer Versuch, Wahrscheinlichkeiten, Ereignis, Merkmalraum
- klassische Wahrscheinlichkeitsdefinition
- geometrische Wahrscheinlichkeiten
- Grenzwert von Häufigkeiten (empirisches Gesetz der großen Zahlen)
- axiomatische Wahrscheinlichkeiten
- subjektive Wahrscheinlichkeiten

Wahrscheinlichkeitsräume

- diskrete Wahrscheinlichkeitsräume
- kontinuierliche Wahrscheinlichkeitsräume
- Additionstheorem
- bedingte Wahrscheinlichkeit
- Multiplikationstheorem
- Satz von der vollständigen Wahrscheinlichkeit & Bayes'sche Formel

stochastische Unabhängigkeit

stochastische Größen

X auf (M, E, W) , $X: M \rightarrow \mathbb{R}^k$

eindimensionale Verteilungen

Verteilungsfunktionen

Eigenschaften, Typen

diskrete Verteilungen

M_x höchstens abzählbar, ohne Häufungspunkt.

- Dirac Verteilung: nur ein bestimmter Wert
- diskrete Gleichverteilung: Zahlen $1..m$ gleiche Punktwahrscheinlichkeiten
- Alternativverteilung: 2 mögliche Werte
- Binomialverteilung: Alternativversuch n mal, Ziehung mit Zurücklegen
- Hypergeometrische Verteilung: Ziehung ohne Zurücklegen

- Poisson-Verteilung: N_0 , zB Qualitätskontrolle
- Geometrische Verteilung: Alternativversuch bis 1 eintritt, "gedächtnislos"

kontinuierliche (stetige) Verteilungen

Nehmen alle Werte eines Intervalls an.

- Uniforme Verteilung: in endlichem Intervall gleichmäßig variieren
- Exponentialverteilung: positive Werte, "gedächtnislos", zB Wartezeiten
- Standard-Normalverteilung: zentriert bei 0, Glockenkurve => zentraler Grenzwertungssatz
- Allgemeine Normalverteilung: nicht zentrierte Werte
- logarithmische Normalverteilung: positive Werte
- t-Verteilung t_n : Dichte symmetrisch zum Nullpunkt, verwendet Gammafunktion $\Gamma(x)$, für $n \rightarrow \infty$ konvergiert die Dichte gegen die Dichte der $N(0,1)$
- Chiquadrat-Verteilung χ^2_n : positive Werte, verwendet Gammafunktion
- F-Verteilung: tabelliert, verwendet Gammafunktion
- Beta-Verteilung: Werte im Intervall $[0,1]$, verwendet Gammafunktion
- Gamma-Verteilung $\text{Gam}(\alpha, \beta)$: positive Werte, verwendet Gammafunktion, Sonderfälle: Exponential- & Chiquadrat-Verteilung (= $\text{Gam}(n/2, 2)$), zB Lebensdauer für Kugellager

Erwartungswert

Funktionen von stochastischen Größen

- Satz vom unbewußten Statistiker
- Varianz
- Verschiebungssatz
- Standardisierung: Streuung, standardisierte stGr

Verteilung von Funktionen stGr

- $G(x) = F(\psi^{-1}(x))$
- $g(x) = f(\psi^{-1}(x)) \cdot |d/dx \psi^{-1}(x)|$
- Lineartransformation von $X \sim N(\mu, \sigma^2)$
- Computersimulation von Verteilungen mH der $U_{0,1}$: $F(\cdot)$ invertierbar, $Y \sim U_{0,1} \Rightarrow F^{-1}(Y) \sim F(\cdot)$

mehrdimensionale Verteilungen

Tschebyscheffsche Ungleichung

Abschätzung zum Beweis des Gesetzes der großen Zahlen, Fkt. weicht von ihrem Erwartungswert um mind. ε ab.

Stochastische Vektoren

- Randverteilungen

mehrdimensionale diskrete Verteilungen

- $p(x_1, \dots, x_m)$
- Randverteilungen

mehrdimensionale kontinuierliche Verteilungen

- $f(x_1, \dots, x_m)$
- 2-dimensionale Normalverteilung
- Randdichten

Erwartung von Funktionen von stV

- Satz vom unbewußten Statistiker: ψ meßbare Fkt, existiert Erwartungswert
- Linearität der Erwartungsbildung: Erwartungswert der Summe $a_i \cdot X_i = \text{Summe } a_i \cdot E(X_i)$

Kovarianz, Korrelation & Unabhängigkeit

Beschreibung von Zusammenhängen zwischen stGr

- Kovarianz: 2 Formeln, Varianz der Summe $c_i \cdot X_i$
- Korrelationskoeffizient: zwischen -1 & 1, Unkorreliertheit \Rightarrow Varianz Summe $X_i = \text{Summe } \text{Var}(X_i)$
- Stochastische Unabhängigkeit: Erwartungswert 2er Funktionen, unabhängige stGr sind unkorreliert

bedingte Verteilungen

Diskret, kontinuierlich, Regressionsfunktion (bedingte Erwartung)

Funktionen von stV

- Lineartransformation von NV: $\underline{Z} = A\underline{X} \sim N(A\underline{\mu}, A\underline{\Sigma}A^T)$
- unabhängige stGr, max, min
- Ordnungsstatistiken
- Faltungen von Wahrscheinlichkeitsverteilungen: unabhängige stGr, Additionstheorem für Poissonverteilung, NV & Gammaverteilung

Folgen stochastischer Größen

3 Arten von Konvergenz

1. Stochastische Konvergenz: (schwaches) Gesetz der großen Zahlen: UIV, $\text{Var}(X_i)$ existiert, Stichprobendurchschnitt mit $n \rightarrow \infty$ konvergiert in der Wahrscheinlichkeit gegen $E(X_i)$
2. Verteilungskonvergenz: zentraler Grenzwertungssatz: unendliche Folge, unabhängige stGr, $\text{Var}(X_i)$ existiert, Lindeberg-Bedingung (keine einzelne Verteilungsfunktion hat dominierenden Einfluß auf die Grenzverteilung), Z_n , Grenzwert $n \rightarrow \infty$ $W\{Z_n \leq x\} = \text{Standard-NV}(x)$
3. Fast sichere gleichmäßige Konvergenz: Fundamentalsatz: UIV, unbeschränkt, empirische Verteilungsfunktion für x_i , $W\{\text{Grenzwert } n \rightarrow \infty \sup |F_n^*(x) - F(x)| = 0\} = 1$

Stichproben & Statistiken

Begriffe

- Stichprobe, konkrete Stichprobe
- Statistik (wenn nicht von unbekanntem Parametern abhängig)
- Stichprobenmittel, Stichprobenvarianz, Stichprobenraum
- Wahrscheinlichkeitsverteilung $w(x_1, \dots, x_n) = \text{Produkt der Einzel-Wahrscheinlichkeiten}$
- Schätzfunktion (für charakteristische Größe der Wahrscheinlichkeitsverteilung)
- Parameter, Parameterraum, geraffter Parameter

klassische schließende Statistik

Glauben an den wahren Parameter, der Verteilung steuert, Parameter so gut wie möglich schätzen

Klassische Punktschätzungen

Qualitative Güteeigenschaften

- Unverzerrtheit (von Schätzfunktionen): erwartungstreu, Erwartungswert der Schätzfunktion trifft den zu schätzenden Parameter, zB Stichprobenmittel, Stichprobenvarianz
- Effizienz (von Schätzfunktionen): Varianz t^* minimal, zB Stichprobenmittel (wenn endliche Varianz)
- Konsistenz (von Folgen von Schätzfunktionen): UIV-Folge, Folge t_n von Schätzfunktionen, Grenzwert in der Wahrscheinlichkeit, zB Stichprobenmittel, $F_n^*(x)$
- Plausibilität (von Schätzwerten für Parameter): nach Beobachtung der konkreten Stichprobe, Plausibilitätsfunktion, plausibler Schätzwert, zB für NV Stichprobenmittel & $(n-1)/n \cdot \text{Stichprobenvarianz}$

Quantitative Güteeigenschaften

- Konfidenzbereiche: $1 - \alpha$, Konstruktion von Konfidenzbereichen mittels Pivotgrößen, zB für NV mit t-Verteilung für Erwartungswert, mit Chiquadrat-Verteilung für Varianz
- Pivotgröße: stGr, die eine Funktion einer Stichprobe & des Parameters ist, deren Verteilung nicht vom Parameter abhängt

Statistische Hypothesen & Tests

- Statistische Hypothese: Aussage über die Verteilung von X
- Gegenhypothese
- statistischer Test: Verfahren zur Entscheidung, ob eine statistische Hypothese angenommen oder verworfen wird
- Wahrscheinlichkeitspapiere: Transformation der $(x, F(x))$ -Ebene auf Gerade
- Fehlerarten & -wahrscheinlichkeiten: Fehler 1. Art, Fehler 2. Art
- Verwerfungsraum: Irrtumswahrscheinlichkeit, Teststatistik, kritischer Bereich

Tests für Normalverteilungen

- t-Test für Erwartungswert
- Test für die Varianz: Chiquadrat-Verteilung
- 2 Stichproben: t-Test für die Gleichheit der Erwartungswerte, F-Test für die Gleichheit der Varianzen

Chiquadrat-Anpassungstest

Für einfache Hypothesen, für zusammengesetzte Parameterhypothesen, für zusammengesetzte Hypothesen.

Zerlegung des Merkmalraumes, Berechnung von Wahrscheinlichkeiten dieser Zerlegungen.

ZB ob 100 zufällig aus $[0,1]$ gewählte Zahlen nach $U_{0,1}$ verteilt sind, Anzahl von Ausfällen eines Computertyps Poisson-verteilt

klassische Regressionsrechnung

Beschreibung kausaler, nicht deterministischer Zusammenhänge mit stochastischen Modellen. ZB Abhängigkeit des Bremsweges von der Geschwindigkeit

- lineare Regressionsfunktion
- Regressionsgeraden

Elemente der Bayes-Statistik

Alle unbekanntes Größen durch stGr beschrieben, auch die Parameter.

a-posteriori-Verteilung/Dichte ist proportional zur a-priori-Verteilung/Dichte * likelihood

- Suffizienz: Statistik $s(X_1, \dots, X_n)$ ist suffizient für einen Parameter, wenn die a-posteriori-Dichte von den Daten x_1, \dots, x_n nur über $s(x_1, \dots, x_n)$ abhängt. Statistik hängt nicht von unbekanntes Daten ab.
- konjugierte Verteilungsfamilien: Familie von Wahrscheinlichkeitsverteilungen für Parameter heißt zum Modell konjugierte a-priori-Familie, wenn a-posteriori-Verteilung selbst & für beliebige Daten Element der Familie sind.
- a-posteriori-Bayes-Schätzer: Erwartungswert von Theta-Schlange bedingt durch Daten