|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DISKRETE VERTEILUNGEN**  Bernoulliverteilung  X~{0,1}, E[X]=p, Var[X]=p(1-p), P(X=1)=p, P(X=0)=1-p  Binomialverteilung (Anz der Erfolge beim n-maligen Ziehen mit zurücklegen)  E[X]=np, Var[X]=Var[∑Xi]=∑Var[Xi]=nVar[Xi]=np(1-p)  dbinom(k,n,p)=P(X=k), pbinom(k,n,p)=F(k),  qbinom(q,n,p)=, rbinom(k,n,p) k bin-vertilte Zahl  Hypergeometrische Verteilung  n mal(ohneZurückleg), M-El-te mit Erfolg, N-El-te ohne  dhyper(k,M,N,n)=P(X=k), phyper(k,M,N,n)=F(k)  Poisson-Verteilung  P(X) in Zeit ∆t ist bekannt  E[X]=Var[X]=  dpois(k,)=P(X=k), ppois(k,)=F(k)  Gleichverteilung  Alle Werte sind gleichwahrscheinlich  sample(1:N,n) n-Anz ZV-len zwischen 1 und N  **STETIGE VERTEILUNGEN**  Normalverteilung  , E[x]=µ, Var[x]=,  max(f(x))=, Wendepunkte:  Fall: Standardnormalverteilung      dnorm(x,)=f(x), pnorm(x,)=F(x),  qnorm(  Exponentialverteilung  Modellierung von Lebensdauern, Wartezeiten.  Sei im Intervall [0,t] von t Zeiteinheiten, dann beschreibt die Exponentialverteilung die Wartezeit X bis Eintreten eines Ereignisses  : Durchschnit Eintreten eines Ereignisses pro Zeiteinheit **x**: Zeitabstand (höchstens) zw. zwei Ereignissen  dexp(x,)=f(x), pexp(x,)=F(x)  Gedächtnislos:  Gleichverteilung  dunif(x,a,b)=f(x), punif(x,a,b)=F(x),  runif(n) n Z-Zahlen zw (0;1)  Chi-Quadrat-Verteilung  dchisq(x,n)=f(x), pchisq(x,n)=F(x)  t-Verteilung  ist t-verteilt mit n Freiheitsgraden  dt(y,n)=f(y), pt(y,n)=F(y) | **BESCHREIBENDE STATISTIK**  Modalwert(e)  - am häufigsten auftretende Ausprägungen(insbes bei qualitativen Merkmahlen)  Mittelwert mean(x)    Schwerpunkt der Daten (empfindlich gegenüber Ausreißern)  Median median(x)  Nicht empfindlich (robust) gegenüber Ausreißern  Spannweite  Stichprobenvarianz var(x)  Stichprobenstandardabweichung s sd(x)  p-Quantil xp  quantile(x,p)  Boxplot    Korrelation – grafische Darstellung zw multivariaten Daten x und y durch ein Streudiagramm (Punkte)  Empirische Kovarianz sxy cov(x,y)  Für hat Punktewolke steigende , fallende Tendenz  Empirischer Korrelationskoeffizient r cor(x,y)  Näherungsweise lineare Zusammenhang zw x und y , falls |r|~1  Regressionsgerade    Ungleichung von Chebyschev  Sei eine Stichprobe mit s>0. Es sei ) mit N( der Anzahl der Elemente in . Dann gilt:  Z.B: k=2: Mehr als (1-1/4)\*100% der Daten liegen in der 2s Bereich usw.  **WAHRSCHEINLICHKEITSTEORIE**  De Morgan‘schen Regel  Axiome von Kolmogorov:  Laplace-Exmeriment  Zufallsexperiment mit n gleichwahrscheinlichen Elementarereignissen  Kombinatorik  **Allgemeines Zählprinzip**  Anz der Mögl-ten, für ein k-stufiges Zufallsexperiment mit Varianten im i-ten Schritt:  Anzahl der **Permutationen** einer n-elementigen Menge  n-maliges Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge  \* für n unterscheidbare Elemente: n!  \* für k Klassen mit je nicht unterscheidbaren Elementen(n=∑)  Anz der k-elementigen Teilmengen einer n-el-gen Menge k-maliges Ziehen aus einer n-elementigen Menge | | Bedingte Wahrscheinlichkeit  Die Wahrscheinlichkeit für das Eintreten von E, wenn F eingetreten ist.  Multiplikationsregel (E,F )  Satz von der totalen Wahrscheinlichkeit  Ereignisse bilden eine disjunkte Zerlegung bzw eine Partition von :    Formel von Bayes  Hilfreich, wenn man kennt, nicht aber  P  Unabhängigkeit  Die Ereignisse heißen **stochastisch** **unabhängig**, wenn die Info über das Eintreten des Ereignis die Wahr-keit für das Eintreten des anderen nicht ändert. Bedeutet **nicht unbedingt kausale** **Unabhängigkeit**!  Falls E, F unabhängig, gilt auch,F bzw. unabhäng  Disjunkheit zweier Ereignissen:  Aus Disjunkheit zweier Ereignissen folgt nicht dass sie unabhängig sind da, jedes kann sein  **ZUFALLSVARIABLEN**  Verteilungsfunktion F(x)  ,  Eigenschaften der Verteilungsfunktion:  rechtsseitig stetig, d.h:  monoton wachsend    ABER  Diskrete Zufallsvariablen  Eigenschaften der Wahrscheinlichkeitsverteilung:  Es gilt:  F(x) ist eine rechtsseitig stetige Treppenfunktion mit Sprüngen bei den Realisationen xi  Stetige Zufallsvariablen P(X=xi) = 0  Eigenschaften Wahrscheinlichkeitsdichte f(x):  Es gilt:  ist stetig und  Mehrdimensionale ZV  Erwartungswert  Erwartungswert E[X]=μ  Sei g(X) eine Funktion der ZV X, dann gilt: | **Eigenschaften des Erwartungswerts**:  Varianz und Kovarianz  *Standardabweichung*  **Eigenschaften der Varianz**  Kovarianz Cov[X,Y]  Wenn X, Y stochastisch unabhängig ⇒Cov[X,Y]=0  Verschiebungssatz:  **Eigenschaften der Kovarianz**  Quantile  Das p-Quantil ist der kleinste Wert ∈R für den gilt  p-Quantil einer stetigen ZV mit streng monoton wachsendem F(x):  Chebyshev Ungleichung  Es gilt für jedes beliebige :  Schwaches Gesetz der großen Zahlen  Mit Chebyshev-Ungleichung und  folgt:  ist eine Folge von unabhängigen, identisch verteilten ZV mit EW und Varianz .  Dann gilt für ein beliebig kleines :  d.h. der konvergiert stochastisch gegen den EW . **ZGWS** Wahrscheinlichkeitsaussagen über  n, und bekannt für n unabhängige identisch verteilte ZV  Verteilung nicht bekannt.  Für **hinreichend große n** gilt dann näherungsweise:  Faustregel für Größe von n:  n > 30: Verteilung ist schief aber ohne markante примечательный Ausreißer (Exponentialverteilung)  n > 15: Verteilung annähernd примерно symmetrisch (Binomialverteilung)  n <= 15: Verteilung annähernd normalverteilt  Stichprobenverteilungen für normalverteilte Grundgesamtheiten  Ja nachdem, was gegeben ist, wählen: |
| **PARAMATERSCHÄTZUNG, KONFIDENZINTERVAL**  Punktschätzer  für Erwartungswert: Stichprobenmittel  für Varianz: Stickprobenvarianz  **Eigenschaften**:   * Schätzwert für wahren Parameter, aber keine Aussage über Unsicherheit der Schätzung * Geringe Sicherheit, dass wahrer Parameterwert getroﬀen wird   Intervallschätzer  Mit Hilfe des ZGWS wird ein **Konﬁdenzintervall** konstruiert, das den wahren Parameter mit einer vorgegebenen Sicherheit, dem sog. **Konﬁdenzniveau** 1−α überdeckt  **Eigenschaften**:   * Intervall für wahren Parameter, aber mit vorgegebener * Sicherheit liegt dieser im Intervall Vorgabe einer großen Sicherheit (95% bzw. 99%) * α: Irrtungswahrscheinlichkeit (meist 1-5%)   Konfidenzintervall f. unbekannten EW μ bei bekannter Varianz  Länge Konfidenzintervall:  Stichprobenumfang n  Konfidenzniveau 1-α  Konfidenzintervall für unbekannten Erwartungswert µ bei unbekannter Varianz  **HYPOTESENTESTS**  Basierend auf n unabhängige identisch verteilte  Zufallsvariablen X1,...,Xn (Messungen) soll eine Entscheidung getroﬀen werden, ob eine Hypothese f¨ur einen unbekannten Parameter der Verteilung, z. B. der Erwartungswert µ, gültig ist oder nicht.  TG: Testgröße. z.B. Mittelwert  Nullhypothese : Angezweifelte Aussage, der nicht widersprochen werden kann, wenn die Stichprobe keinen Gegenbeweis liefert z.B:  Gegenhypothese : Gegenteil von  Kritischer Bereich C: Werte von TG, die für sprechen und mit **Wahrscheinlichkeit ≤α**, dem Signiﬁkanzniveau auftreten.   * α: Wahrsc-keit, dass verworfen wird, obwohl richtig => **Fehler 1. Art** * Die Wahrsc-keit, dass angenommen wird, obwohl sie falsch ist, ist der sog. **Fehler 2. Art**      * Durch das vorgegebene Signifikanzniveau wird versucht Fehler 1. Art klein zu halten * Beide Fehler können nicht gleichzeitig kontrolliert werden * Fehler 2. Art nicht kontrollierbar, wenn man μ\_0 nicht kennt   Klassische Parametertests  **Testprobleme**:   * Zweiseitiger Test:   H0 : µ = µ0 & H1 : µ 6= µ0   * Einseitige Tests:   H0 : µ ≥ µ0 & H1 : µ < µ0  bzw. H0 : µ ≤ µ0 & H1 : µ > µ0  **Testentscheidung** basierend auf Stichprobe {x1,...,xn}:   * H0 wird abgelehnt, falls tg = TG(x1,...,xn) ∈ C * H0 wird angenommen, falls tg = TG(x1,...,xn) ∈   *Wahrscheinlichkeit für Fehler 1. Art mit standartisierter Testgröße TG\*:*  Wird H0 verworfen => signifikante Schlußfolgerung  Wird H0 nicht verworfen => es lässt sich keine Schlußfolgerung über Fehler 2. Art treffen, eine schwache Schlußfolgerung.  Gauß-Test  Test für Erwartungswert einer Normalverteilung **bei bekannter Varianz** | | t-Test  Test für Erwartungswert einer Normalverteilung **bei unbekannter Varianz**      p-Wert „beobachtetes Signifikanzniveau.“  Wahrscheinlichkeit, bei Zutreffen von H0 den beobachteten Wert tg der Testgröße oder einen noch stärker von μ\_0 abweichenden Wert zu bekommen.   * Größter Wert von α, für den H0 nicht abgelehnt wird * Wenn H0 abgelehnt wird, ist das Ergebnis immer signifikant!   p-Wert <0,01: sehr hohe Signifikanz  0,01p-Wert <0,05: hohe Signifikanz  0,05p-Wert <0,1: schwache Signifikanz  p-Wert 0,1: keine Signifikanz  Zusammenhang Konfidenzintervall – Hypothesentest  Konfidenzintervall zum Konfidenzniveau (1-α) ist Annahmebereich für H0  Test:  H0 wird abgelehnt, falls ∉I  H0 wird angenommen, falls ∈I  NUMERISCHES RECHEN  Fehlerquellen:   * Rundungsfehler bei Darstellung als Maschinenzahl * Fehler aufgrund von Gleitpunktarithmetik * Diskretisierungsfehler   *Feinere Diskritisierung -> höhere Genauigkeit, erfordet mehr Rechnungen -> kann zu ungenauen Ergebnissen wieder führen*  Eigenschaften von Gleitpunktsystemen   * Sie sind endlich und diskret. ⇒ Rundungsfehler, da nicht alle Zahlen exakt * Sie sind nach oben und unten beschränkt. ⇒ Probleme mit Overﬂow bzw. Underﬂow * Maschinenzahlen haben nur zwischen zwei aufeinanderfolgenden Exponenten und gleichen Abstand, nämlich . n -Mantisselänge   Binärdarstellung von 0.3: von 10:  0.3 \* 2 = 0.6 10/2 = 5 R= 0  0.6 \* 2 = 1.2 5/2 = 2 R = 1  0.2 \* 2 = 0.4 2/2 = 1 R = 0  0.4 \* 2 = 0.8 1/2 = 0 R = 1  0.8 \* 2 = 1.6 1010  Absoluter Rundungsfehler:  Relativer Rundungsfehler:  Bei Addition von Gleitpunktzahlen aufsteigende Reihenfolge:  Auslöschung  Entsteht bei der Subtraktion von zwei fast gleich großen Zahlen  Erhöhung der Signifikanz weniger signifikanter Stellen  Kondition  **Berechnungen**:   * Kondition eines Problems hängt von x und f ab * Schlechte Kondition: * Falls ein Problem schlecht konditioniert -> es gibt keinen numerisch günstigen Alg das zu lösen | **INTERPOLATION**  **Interpolationsproblem**: Bestimmen Sie zu gegebenen Punkten (xi,yi), i = [0,n] mit für , eine Funktion G (diese ist nicht eindeutig!), so dass    **Approximation** durch eine Funktion, die den **Trend** der Daten beschreibt  **Interpolation** ist **nicht** geeignet **für verrauschte Daten**  Klassischer Ansatz / Vandermonde Ansatz    (n+1) Gleichungen für die unbekannten Polynom ntes Grades ist eindeutig bestimmt, falls  **Eigenschaften**:   * Die Vandermonde Matrix ist nicht singulär (falls alle xi verschieden). * Rechenaufwand für Lösung des LGS ist hoch: O() * Die Vandermonde Matrix kann schlecht konditioniert sein und ist deshalb als allgemeiner Ansatz nicht geeignet     Ansatz nach Lagrange  **Eigenschaften**:   * Anwendung z. B. für Numerische Integration * Rechenaufwand für Lagrangefunktionen hoch: O() * Bei Hinzunahme weiterer Stützpunkte müssen Lagrangefunktionen neu berechnet werden   *Vorteil von Lagrange: Systemgleichung muss man nicht lösen*    Ansatz nach Newton  **Interpolationsbedingungen**:  Rechenaufwand O()  Dividierte Diﬀerenzen | Eﬃzienz und numerische Eﬀekte der Polynominterpolation  Klassische Auswertung benötigt 2n-1 Multiplikationen:  Horner-Schema – n Multiplikationen  **Interpolationsfehler**  Falls f hinreichend glatt ist und das eindeutige Interpolationspolynom vom Grad n, dann gilt für den Interpolationsfehler:  Da θ nicht bekannt ist, kann der Fehler nur abgeschätzt werden  Je höher Grad -> besser Approxim -> größer lok Extreme  im Fall von **äquidistanten Stützstellen** Interpolationspolynom **nicht** **immer** gegen die zugrundeliegende stetige Funktion **konvergiert**, wenn die Anzahl der Stützstellen und damit der **Grad** des Polynoms **wächst**  Chebyshev-Punkte  haben genau diese Eigenschaft.  erhält man durch orthogonale Projektion von gleichverteilten Punkten auf dem Einheitskreis.    Auf einem Intervall [a,b]:  Spline-Interpolation   * 2n Interpolationsbedingungen * n−1 Bedingungen für die Stetigkeit der ersten Ableitungen * n−1 Bedingungen für die Stetigkeit der zweiten Ableitungen * 2 geeigneten Randbedingungen (RB), z. B. natürlichen RB (nichtanforderlich)   Rechenaufwand O(n) ﬂops |

|  |  |  |
| --- | --- | --- |
| **NUMERISCHE INTEGRATION**  Ansatz für numerische Integration (Quadratur) auf dem Intervall [0,1]  Ansatz für numerische Integration (Quadratur) auf dem Intervall [a,b]  Newton-Cotes Regeln  Das Integral des Interpolationspolynoms dient als Approximation für das Integral der Funktion:  Lagrange Ansatz:  ***Die Gewichte sich immer zu 1 aufsummieren*** *+ sie sind symmetrisch in i, d.h. und sie sind unabhängig von Intervall [a,b]*  Trapezregel    ***Integrationsregeln mit positiven Gewichten sind stabil***. Die Newton-Cotes-Regeln für k ≤ 7 und k = 9 haben positive  Simpson-Regel  Fehler der Quadratur  Eine Integrationsregel hat **Ordnung p**, wenn sie für Polynome vom  Grad ≤ p−1 exakte Werte liefert   * Ordnung der Trapezregel: exakte Lösung für const und lineare Funktionen, d.h. Ordnung 2 * Ordnung der Newton-Cotes Regeln: mind.(k+1). Entspricht der Anz der Knoten * Ordnung Simpson-Regel: exakte Lösung für Polynom 3tes Grades: Ordnung 4     Zusammengesetzte Regeln / Summenformeln  Zusammengesetzte Trapezregel  Zusammengesetzte Simpson-Regel  ,  Fehler der Summenformeln   * Der Fehler ist proportional zu , d. h. eine Halbierung der Intervalll¨ange reduziert den Fehler um den Faktor . * Ein Integral kann beliebig genau approximiert werden, wenn H entsprechend klein gewählt wird. Voraussetzung: H ist hinreichend glatt(d.h. Trapezregel zweimal stetig differenzierbar) !!!Nicht stetige Funktionen können nicht numerisch integriert werden!!!   Grenzen der Newton-Cotes-Regeln   * Bei Verwendung **vieler äquidistanter**(расположенный на одинаковом расстоянии от) **Knoten** treten die bekannten Probleme von Interpolationspolynomen höheren Grades auf **->** Gewichte werden negativ, also ***Verfahren instabil*** * Die sog. geschlossenen( Newton-Cotes-Regeln machen Funktionsauswertungen an den Grenzen des Intervalls erforderlich -> Problem mit Singularitäten(уникальность) * Die Newton-Cotes-Regeln erreichen aufgrund der äquidistanten Knoten nicht die größtmögliche Ordnung.   **Gauß-Quadratur zur Vermeidung dieser Probleme!** | Gauß-Quadratur  **Idee**: Wähle die Knoten und Gewichte so, dass man ein Verfahren möglichst großer Ordnung p erhält  (2k+2) Gleichungen für (2k+2) Unbekannten |  |