Wahrscheinlichkeitstheorie für Informatikstudien (Kapitel #6)

Siegfried Hörmann

TU Graz

WS 2020/21

Kursaufbau

1. Wahrscheinlichkeitsräume.
2. Laplace-Experimente.
3. Wichtige diskrete Verteilungen.
4. Zufallsvariablen.
5. Zufallsvektoren.
6. Momente.
7. Bedingte Wahrscheinlichkeiten.
8. Unabhängigkeit.
9. Konvergenz von Zufallsvariablen.
10. Der zentrale Grenzwertsatz.

Momente

6.1. Erwartungswert für diskrete ZVen.

6.2. Beispiele und Eigenschaften.

6.3. Erwartungswert für allgemeine ZVen.

6.4. Varianz.

6.5. Kovarianz und Korrelation.

Erwartungswert für für diskrete ZVen

Wir betrachten nun wieder skalare ZVen X.

Die ZV X ist durch ihre Verteilungsfunktion charakterisiert.

Oft will man die wichtigsten Eigenschaften einer Verteilung durch bestimmte Kennzahlen zusammenfassen.

Die zwei wichtigsten Kennzahlen die wir näher besprechen werden, sind der Erwartungswert für für diskrete ZVen und die Varianz einer ZV.

- Der Erwartungswert ist so etwas wie der "typische" Wert von X.
- Die Varianz ist ein Maß dafür, wie sehr X um den Erwartungswert streut.

Erwartungswert für für diskrete ZVen

Betrachte den W-Raum $(\Omega, \mathcal{A}, P)$ und eine diskrete ZV $X\colon \Omega \to \mathbb{R}$, also $\{X(\omega)\colon \omega \in \Omega\} = \{x_1, x_2, \ldots\}$.

Wir sagen, dass X integrierbar ist (kurz: $X \in L^1(\Omega, \mathcal{A}, P)$ oder $X \in L^1$) falls

$$\sum_{i \geq 1} |x_i| P(X = x_i) < \infty.$$

In diesem Fall setzen wir

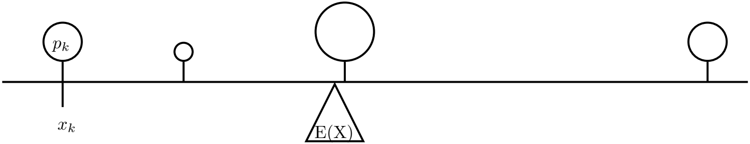
$$E(X) = \sum_{i \geq 1} x_i P(X = x_i).$$

Wir nennen $E(X)$ den *Erwartungswert von* X.

Erwartungswert für für diskrete ZVen

Bemerkung:

Angenommen wir legen Gewichte $p_k = P(X = x_k)$ auf die Position x_k der (gewichtslosen) x-Achse. Dann ist $E(X)$ *der Schwerpunkt* dieser Verteilung.



Erwartungswert für für diskrete ZVen

Beispiel (Roulette: setzen auf Farbe)

Wir spielen im Casino Roulette. Jemand setzt 10 Euro auf Rot (bei Rot bekommt man den Einsatz plus 10 Euro, also den doppelten Einsatz zurück). Wie groß ist der erwartete Gewinn?

Beispiel (Roulette: setzen auf Zahl)

Wir spielen im Casino Roulette. Jemand setzt 10 Euro auf eine bestimmte Zahl (für die richtige Zahl bekommt man den 36 fachen Einsatz zurück). Wie groß ist der erwartete Gewinn?

Beispiele und Eigenschaften

Beispiel (Indikatorfunktionen)

Sei $X(\omega) = I\{\omega \in A\}$. Dann nimmt X nur die Werte 0 und 1 an

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = P(A).$$

Beispiel (Binomialverteilung)

Sei $X \sim B_{n,p}$. Dann ist

$$E(X) = \sum_{k=1}^{n} \underbrace{k\binom{n}{k}p^k}_{np\binom{n-1}{k-1}p^{k-1}} (1-p)^{n-k}$$

$$= np\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = np.$$

Beispiele und Eigenschaften

Beispiel (Poissonverteilung)

Sei X die Anzahl der Tore von Sturm Graz in einem Meisterschaftsspiel. Wenn $X \sim P_2$, wie groß ist dann die erwartete Anzahl der Tore?

Beispiel (Konstanten)

Sei $X = c$ eine Konstante. Dann ist $E(X) = c$.

Beispiel (Geometrische Verteilung)

Sei X die Anzahl der benötigten Versuche bis wir mit einem 6-seitigen Würfel einen Sechser würfeln. Bestimme $E(X)$.

Beispiele und Eigenschaften

Proposition

Sei $Y = f(X)$. Falls $Y \in L^1$, dann gilt

$$E(Y) = \sum_{i \geq 1} f(x_i) P(X = x_i).$$

Beispiel

Sei $X \sim B_{2,p}$. Was ist $E(1 + X^2)^{-1}$?

Beispiel (Gleichverteilung)

Sei $X \sim \text{Unif}(\{0, \ldots, 3\})$. Bestimme $E(X^2)$.

Beispiel (Binomialverteilung)

Sei $X \sim B_{2,1/2}$. Bestimme $E(X^2)$.

Beispiel

Ein Konferenzhotel hat 20 Einzelzimmer. Der Preis pro Zimmer beträgt 100 Euro. Man weiß aus Erfahrung, dass im Schnitt jeder vierte Teilnehmer storniert (es gibt keine Stornogebühr) und überbucht deshalb oft. Für eine wissenschaftliche Tagung wurden 22 Zimmer verbucht. Falls es mehr als 20 Gäste gibt, muss man die entsprechende Anzahl an Zimmern im Nachbarhotel zur Verfügung stellen. Hier sind die Zimmer doppelt so teuer. Wie hoch ist der erwartete Einnahme? Wie hoch wäre diese wenn man nicht überbucht hätte?

Beispiele und Eigenschaften

Proposition

Seien X und Y in L^1 und $a \in \mathbb{R}$ konstant. Dann gilt

(a) *Monotonie: $X \leq Y$ impliziert $EX \leq EY$.*

(b) *Linearität: $aX \in L^1$ und $X + Y \in L^1$. Weiters,*

$$E(aX) = aE(X) \quad \textit{und} \quad E(X+Y) = E(X) + E(Y).$$

Beispiel

Sei $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ wobei alle $X_i \in L^1$. Dann ist $E(\overline{X}) = \frac{1}{n}\sum_{k=1}^{n} E(X_k)$. Haben alle X_i denselben Erwartungswert μ, dann ist $\overline{X}$ ein unverzerrter Schätzer von μ.

Beispiele und Eigenschaften

Aus Monotonie und Linearität lassen sich z.B. ganz leicht folgende Ungleichungen herleiten:

Korollar (Jensensche Ungleichung)
Für $X \in L^1$ gilt $|E(X)| \leq E(|X|)$.

Beweis. Da $X \leq |X|$ und $-X \leq |X|$, folgt mit der letzten Proposition die Aussage. $\square$

Beispiele und Eigenschaften

Korollar (Markovsche Ungleichung)

Angenommen $|X|^p \in L^1$ *für ein* $p > 0$. *Dann gilt* $\forall \varepsilon > 0$

$$P(|X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E(|X|^p).$$

Beweis. Wir haben

$$I\{|X| > \epsilon\} \leq \left(\frac{|X|}{\varepsilon} \right)^p.$$

Wegen der Monotonie folgt

$$P(|X| > \varepsilon) = E(I\{|X| > \varepsilon\}) \leq E\left(\frac{|X|^p}{\varepsilon^p} \right).$$

$\square$

Erwartungswert für allgemeine ZVen

Wir wollen dieses Konzept nun für Erwartungswerte allgemeiner ZVen erweitern.

Die Idee ist es, eine gegebene ZV X mit einer Folge diskreter ZVen $X_{(n)}$ zu approximieren und dann $E(X) = \lim_{n\to\infty} E(X_{(n)})$ zu setzen.

Genauer gesagt definieren wir $X_{(n)} = \lfloor nX \rfloor / n$. Hierbei ist $\lfloor x \rfloor$ die größte ganze Zahl $\leq x$.

Dann gilt $X_{(n)}(\omega) = \frac{k}{n}$ falls $\frac{k}{n} \leq X(\omega) < \frac{k+1}{n}$, $k \in \mathbb{Z}$.

Erwartungswert für allgemeine ZVen

Sei X eine ZV. Wir sagen, dass X einen Erwartungswert hat, falls $X_{(n)} \in L^1$. Wir schreiben $X \in L^1$ und definieren

$$E(X) = \lim_{n\to\infty} E(X_{(n)}).$$

Es ist nicht schwer zu zeigen, dass dieser Grenzwert existiert.

Die vorherige Proposition und deren Korollare gelten immer noch für allgemeine ZVen!

Erwartungswert für allgemeine ZVen

Betrachte nun den Spezialfall, in dem X eine Dichtefunktion hat. Dann folgt:

Proposition

Sei X eine ZV mit Dichte $f^X(x)$.

(a) *$X \in L^1$ dann und nur dann, wenn $\int_{-\infty}^{\infty} |x| f^X(x)dx < \infty$.*

(b) *In diesem Fall ist $E(X) = \int_{-\infty}^{\infty} x f^X(x)dx$.*

(c) *Wenn $g \colon \mathbb{R} \to \mathbb{R}$ und $g(X) \in L^1$, dann ist*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f^X(x)dx.$$

Erwartungswert für allgemeine ZVen

Beispiel

Eine Buslinie fährt in 15 minütigen Intervallen. Wir gehen zufällig zur Bushaltestelle. Sei X die Wartezeit bis der Bus kommt. Bestimme EX und EX^2.

Erwartungswert für allgemeine ZVen

Beispiel

Die Lebensdauer einer Glühbirne X sei verteilt gemäß $X \sim \mathrm{Exp}(\lambda)$. Zeige, dass

$$E(X) = \frac{1}{\lambda} \quad \text{und} \quad E(X^2) = \frac{2}{\lambda^2}.$$

Erwartungswert für allgemeine ZVen

Beispiel

Sei $X \sim N(0,1)$. Dann kann man zeigen, dass

$$E(X) = \int_{-\infty}^{\infty} x \left[\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right] dx = 0.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \left[\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right] dx = 1.$$

Beispiel

Sei $X \sim N(\mu, \sigma^2)$. Dann gilt

$$E(X) = \mu \quad \text{und} \quad E(X^2) = \mu^2 + \sigma^2.$$

Varianz

Wir haben gesehen, dass der Erwartungswert einer ZV uns sagt, welches das "erwartete" oder "durchschnittliche" Resultat ist.

Eine andere natürliche Frage wäre: Wie weit weicht die ZV im Durchschnitt von der Erwartung ab?

In anderen Worten: Ist $E(X) = \mu$, so interessiert uns $E(|X - \mu|)$.

Aus mathematischen Gründen ist es eleganter das Quadrat dieses Abstands zu berechnen: die Varianz.

Varianz

Sei X eine ZV, sodass $X^2 \in L^1$ (kurz $X \in L^2$). Dann nennen wir

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

die Varianz von X und $\sqrt{\mathrm{Var}(X)}$ die Standardabweichung.

Die Varianz ist das 2. zentrale Moment.

Allgemein:

- $E(X^k)$ k-tes Moment;
- $E((X-\mu)^k)$ k-tes zentrales Moment.

Wir setzen voraus $X \in L^k$, also $X^k \in L^1$.

Varianz

Lemma

Sei $X \in L^2$. Dann gilt

(a) $\mathrm{Var}(X) \geq 0$. $\mathrm{Var}(X) = 0$ *impliziert* $X = \mu$.

(b) $\mathrm{Var}(X) = E(X^2) - (E(X))^2$.

Daraus folgt mit (a), dass $E(X^2) \geq (E(X))^2$.

Beweis. Da $(X - \mu)^2 \geq 0$, folgt $\mathrm{Var}(X) \geq 0$ wegen der Monotonie. Mit der Markov-Ungleichung folgt für alle $\varepsilon > 0$

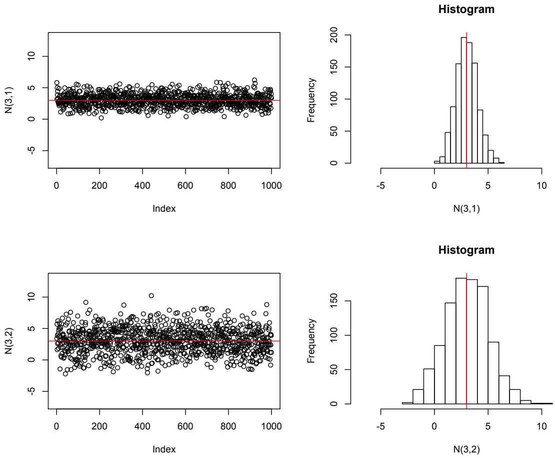
$$P(|X - \mu| > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Damit ist (a) gezeigt.

Dann, $E(X - \mu)^2 = E(X^2 + \mu^2 - 2\mu X)$. Der Beweis folgt dann mit der Linearität des Erwartungswerts. $\square$

Varianz

Aus den Herleitungen oben ergibt sich ganz leicht, dass eine $N(\mu, \sigma^2)$ Zufallsvariable Varianz σ^2 hat.



Varianz

Beispiel

Mit $\mathrm{Var}(X) = E(X^2) - (E(X))^2$ folgt für

- $X \sim \mathrm{Unif}(a, b)$:

$$\mathrm{Var}(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

- $X \sim \mathrm{Exp}(\lambda)$:

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

- $X \sim \mathrm{N}(0, 1)$:

$$\mathrm{Var}(X) = 1^2 - 0^2 = 1.$$

Varianz

Die Anzahl X der Lotto-Sechser bei einer Ziehung sei Poisson-verteilt mit $\lambda = 3$. Bestimme die Varianz von X.

Sei T die Wartezeit bis wir mit einem 6-seitigen Würfel einen Sechser würfeln. Bestimme die Varianz von T.

Sei $X \sim B_{1,p}$. Bestimme die Varianz von X.

Kovarianz und Korrelation

Sei (X, Y) ein Zufallsvektor mit Erwartungswert (μ, ν) und die Marginalen seien in L^2. Dann nennen wir

$$\mathrm{Cov}(X, Y) = E((X - \mu)(Y - \nu))$$

die Kovarianz zwischen X und Y. Ist $\mathrm{Cov}(X, Y) = 0$ so sagen wir, dass X und Y unkorreliert sind.

Wir schreiben dann: $X \perp Y$.

Kovarianz und Korrelation

Berechnung der Kovarianz:

- Es gilt im Allgemeinen
$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$
- Sind X und Y diskret, dann gilt
$$E(XY) = \sum_k \sum_\ell x_k y_\ell P(X = x_k, Y = y_\ell).$$
- Haben (X, Y) eine Dichte $f(x, y)$, dann gilt
$$E(XY) = \int \int xyf(x, y)dxdy.$$

Kovarianz und Korrelation

Lemma (Cauchy-Schwarz-Ungleichung)

Seien X und Y ZVen in L^2. Dann ist $XY \in L^1$ und

$$|E(XY)|^2 \leq E(X^2)E(Y^2).$$

Es folgt wegen der Cauchy-Schwarz-Ungleichung, dass

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)},$$

oder äquivalent dazu

$$-1 \leq \mathrm{Corr}(X, Y) := \frac{\mathrm{Cov}(\mathrm{X}, \mathrm{Y})}{\sqrt{\mathrm{Var}(\mathrm{X})}\sqrt{\mathrm{Var}(\mathrm{Y})}} \leq 1.$$

Wir nennen $\mathrm{Corr}(X, Y)$ die Korrelation zwischen X und Y.

Kovarianz und Korrelation

Die Korrelation wird oft als ein simples Maß der Abhängigkeit zwischen X und Y genutzt. Ist $\mathrm{Cov}(X, Y) = 0$, dann sagen wir dass X und Y unkorreliert sind.

Beispiel

Betrachte den Zufallsvektor (X, Y). Sowohl X als auch Y nimmt die Werte $0, 1, 2$ an. Hier ist die PMF:

$X \backslash Y$	0	1	2
0	0.1	0.2	0.1
1	0.2	0.1	0.1
2	0	0	0.2

Berechne $\mathrm{Corr}(X, Y)$.

Kovarianz und Korrelation

Betrachte einen Zufallsvektor mit Komponenten $X = (X_1, \ldots, X_n)$ und sei

$$\mathrm{Var}(X) = \Sigma = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \ldots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}.$$

Man nennt Σ die Kovarianzmatrix von X.

Es gelten:

- Σ ist nicht-negativ definit [$a'\Sigma a \geq 0$ für alle $a \in \mathbb{R}^n$];
- Σ ist symmetrisch.

Eine sehr nützliche Formel:

Lemma

Sei X ein Zufallsvektor in $\mathbb{R}^n$ und $A \in \mathbb{R}^{m \times n}$ und $b \in \mathbb{R}^m$. Sei $\Sigma = \mathrm{Var}(X)$. Dann gilt

$$\mathrm{Var}(AX + b) = A\Sigma A'.$$

Mit dieser Formel ist es sehr einfach die zwei Eigenschaften des nächsten Resultats zu zeigen, welches ein paar Rechenregeln für Varianzen und Kovarianzen zusammenfasst.

Kovarianz und Korrelation

Proposition

Seien $X, Y, X_i \in L^2$ und $a, b, c, d \in \mathbb{R}$. Dann gilt

(a) $aX + b$ *und* $cY + d$ *sind in* L^2 *und*

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y).$$

Im Besonderen gilt $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

(b) $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.
Sind die X_i *unkorreliert, dann*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Kovarianz und Korrelation

Beispiel

Seien $X_1, \dots, X_n$ unkorreliert und man nehme an: $X_i \sim B_{1,p}$.

Für den Mittelwert der Stichprobe $\bar{X} = \frac{S_n}{n}$ bekommen wir

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{p(1-p)}{n}.$$

Mit der Markov-Ungleichung bekommen wir

$$P(|\bar{X} - p| > \varepsilon) \leq \frac{\mathrm{Var}(\bar{X})}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Daher können wir für jedes $\varepsilon > 0$ die Wahrscheinlichkeit $P(|\bar{X} - p| > \varepsilon)$ beliebig klein machen!

Dies ist ein Spezialfall des Gesetzes der großen Zahlen.