Wahrscheinlichkeitstheorie für Informatikstudien (Kapitel #5)

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.

Zufallsvektoren

5.1. Zusammenfassung.

5.2. Zufallsvektoren.

5.3. Marginale und gemeinsame Verteilung.

Zusammenfassung

- Wir haben Zufallsvariablen eingeführt:
  Funktionen von Ω nach $\mathbb{R}$.
- Wir nennen $P(X \leq x) = P(\{\omega\colon X(\omega) \leq x\})$ die Verteilung von X.
- Verteilung ist charakterisiert über die Verteilungsfunktion: $F^X(x) = P(X \leq x)$.
- Im Modellierungskontext reicht es oft, nur $F^X(x)$ zu betrachten.
- Ein einfacher Weg um Verteilungsfunktionen zu bekommen sind Dichtefunktionen: $F(x) := \int_{-\infty}^{x} f(t)dt$.
- Wir haben ein paar wichtige Verteilungen betrachtet.

Zufallsvektoren und Momente

Bis jetzt haben wir nur einzelne (univariate) Zufallsvariablen betrachtet.

Oft haben wir einen Zufallsvektor

$$\omega \mapsto \boldsymbol{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega))'$$

oder einen Zufallsprozess

$$\omega \mapsto (X_k(\omega) \colon k \geq 1)$$

den wir analysieren wollen.

Wichtig: wir betrachten nur Mengen von Zufallsvariablen, die alle auf demselben W-Raum definiert sind.

Beispiel

Wir haben N_1 **schwarze** Kugeln, N_2 rote Kugeln und N_3 blaue Kugeln in einer Urne. Wir ziehen 10 Kugeln mit Zurücklegen. Sei

- X_1 die Anzahl der **schwarzen** Kugeln,
- X_2 die Anzahl der roten Kugeln und
- X_3 die Anzahl der blauen Kugeln.

Dann beschreibt

$$\boldsymbol{X} = (X_1, X_2, X_3)'$$

wie viele Kugeln jeder Farbe gezogen wurden.

Zufallsvektoren

Beispiel

Wir werfen eine Münze 10 mal. Sei $\Omega = \{0, 1\}^{10}$. Sei X_i das Resultat des i-ten Wurfs und $X_i(\omega) = \omega_i$. Der Vektor (X_1, X_5) gibt das Resultat des ersten und fünften Wurfs an.

Beispiel

Wir werfen eine Münze ∞ oft. Sei $\Omega = \{0, 1\}^{\infty}$. Sei wieder X_i das Resultat des i-ten Wurfs und $X_i(\omega) = \omega_i$. Das ist ein Zufallsprozess.

Beispiel

Wir werfen eine Münze und wetten 1 € auf Kopf oder Zahl. Sei X_i der Gewinn/Verlust des i-ten Wurfs, dann ist $X_i = \pm 1$ und

$$S_n := X_1 + \ldots + X_n, \quad S_0 = 0,$$

ist der Gewinn/Verlust nach n Würfen. Der Prozess $(S_n\colon\ n \geq 1)$ ist ein Random Walk.

Marginale und gemeinsame Verteilung

Für einen gegebenen Zufallsvektor $(X_1, \ldots, X_n)'$ oder Zufallsprozess $(X_k : k \geq 1)$ definieren wir die marginale PMF oder CDF:

$$p_{kj} = P(X_k = x_j) \quad \text{oder} \quad F_k(x) = P(X_k \leq x).$$

Beachte aber, dass dies i.A. nicht ausreicht um zu wissen, wie die ZV miteinander agieren (im stochastischen Sinne).

Seien $X_h = 1$, wenn es heute um h Uhr in Graz regnet. Sonst ist $X_h = 0$. Die Prognose sagt z.B. $P(X_{15} = 1) = 0.9$.

Kann man aus der Prognose links die Wahrscheinlichkeit errechnen, dass es zwischen 15 und 18 Uhr durchregnet?

Marginale und gemeinsame Verteilung

Versuche folgende zwei Szenarien zu modellieren:

1. Eine breite Kaltfront zieht durch. Mit 10% Wahrscheinlichkeit zieht sie an Graz vorbei. Allerdings, wenn Sie nicht vorbeizieht, dann regnet es in den nächsten 10 Stunden durchgehend.

2. Das Wetter ist sehr wechselhaft und schaueranfällig. Ob es zu einer bestimmten Uhrzeit regnet oder nicht hat keinen Einfluss auf das nachfolgende Wetter.

Marginale und gemeinsame Verteilung

Wir brauchen also die gemeinsame PMF oder CDF

$$p(x_1,\dots,x_n) = P(X_1 = x_1,\dots,X_n = x_n),$$

$$F(x_1,\dots,x_n) = P(X_1 \le x_1,\dots,X_n \le x_n).$$

Wenn der Vektor X eine Dichte hat, dann ist

$$F(x_1,\dots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1,\dots,t_n) dt_n \dots dt_1 .$$

Wie lautet jeweils die gemeinsame PMF in den beiden obigen Szenarien für $x = (0, 1, 0, 1)$?

Umgekehrt, lassen sich aus der gemeinsamen Verteilung die Randverteilungen bestimmen.

Marginale und gemeinsame Verteilung

Wir betrachten zur Vereinfachung Zufallsvektoren (X, Y) der Dimension 2. Die folgenden Konzepte lassen sich leicht generalisieren.

- Die Verteilung von Y ist gegeben durch

$$F^Y(y) = \lim_{x\to\infty} F(x, y) = \lim_{x\to\infty} P(X \leq x, Y \leq y).$$

- Die PMF von Y ist gegeben durch

$$p(y) := P(Y = y) = \sum_x P(X = x, Y = k) = \sum_x p(x, y).$$

- Die Dichte von Y ist gegeben durch

$$f^Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Marginale und gemeinsame Verteilung

Beispiel (Rand-PMF)

Wir haben eine Urne mit einer schwarzen, einer roten und einer blauen Kugeln. Wir ziehen 5 mal mit Zurücklegen. Sei (X_1, X_2, X_3) die Anzahl der gezogenen Kugeln der einzelnen Farben. Dann haben wir

$$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = \binom{5}{k_1, k_2, k_3} \frac{1}{3^5}.$$

Es folgt, dass

$$P(X_1 = 2) = \frac{1}{3^5} \sum_{0 \le k_2, k_3 \le 5} \binom{5}{2, k_2, k_3},$$

wobei $\binom{n}{k_1, k_2, k_3} = 0$ falls $k_1 + k_2 + k_3 \neq n$.

Marginale und gemeinsame Verteilung

Beispiel (Randdichte/Randverteilung)

Angenommen (X, Y) ist ein Zufallsvektor mit Dichtefunktion

$$f(x, y) = (1 + x + y)/2, \quad 0 \leq x, y \leq 1.$$

Beachte Konvention: $f(x, y) = 0$ wenn $0 \leq x, y \leq 1$ nicht gilt.
Wir müssen immer nur über den Bereich $f(x, y) \neq 0$ integrieren.

Die Dichtefunktion von X ist dann

$$f^X(x) = \int_0^1 f(x, y) dy = \frac{x}{2} + \frac{3}{4}.$$

Die Verteilungsfunktion von X ist dann

$$F^X(x) = \int_0^x f^X(t) dt = \frac{x^2}{4} + \frac{3x}{4}.$$

Marginale und gemeinsame Verteilung

Sei $X_0 = 1$ wenn es heute regnet und 0 sonst. Sei $X_1 = 1$ wenn es morgen regnet und 0 sonst. Der Vektor $X = (X_0, X_1)$ hat PMF

$$p_{00} = 1/3, \quad p_{01} = 1/2, \quad p_{10} = 1/12, \quad p_{11} = 1/12.$$

wobei $p_{ij} = P(X_0 = i, X_1 = j)$. Wie groß ist die Wahrscheinlichkeit, dass es morgen regnet?

Marginale und gemeinsame Verteilung

Sei X_0 die heutige Regenmenge (l/m^2) in Graz und Sei X_1 morgige Regenmenge in Graz. Der Vektor $X = (X_0, X_1)$ habe die Dichtefunktion

$$f(x, y) = \frac{1}{6} e^{-1/2x - 1/3y}, \quad x, y \geq 0.$$

Weise nach, dass es sich hier um eine Dichtefunktion handelt. Wie groß ist die Wahrscheinlichkeit dass es entweder heute, oder morgen mehr als $5l/m^2$ regnet?