Zufällige Graphen

Christoph Schumacher

27. April 2020

RGCN I, Kap. 2.2: Kopplungen

Zusammenfassung

Ziel: Quantitative Abschätzung der Konvergenz

$$\mathrm{Bin}(n, \lambda/n) \xrightarrow[n\to\infty]{d} \mathrm{Poi}(\lambda) \tag{2.2.1}$$

mit Kopplungsbeweis

Inhaltsverzeichnis

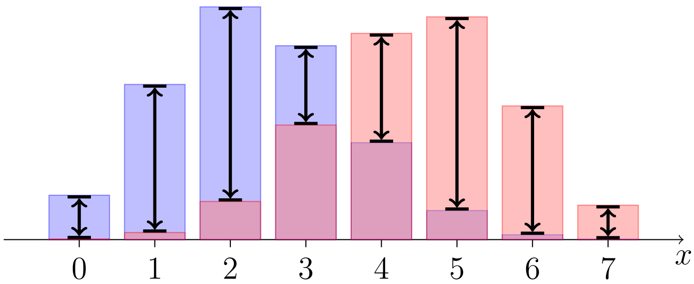
Fragen

1. Warum kopiert man die ZVn auf einen Wahrscheinlichkeitsraum? — Um sie direkt vergleichen zu können, und nicht nur ihre Verteilungen.

2. Wofür sind Kopplungen gut? — Wir können $\hat{X}$ und $\hat{Y}$ abhängig wählen. Wenn wir „viel“ Abhängigkeit hinkriegen, dann sind die Verteilungen von X und Y ähnlich bzw. nah zusammen. Umgekehrt: wenn die Verteilungen von X

und Y ähnlich sind, dann können wir eine Kopplung finden, bei der die Werte der ZVn mit großer Wahrscheinlichkeit gleich sind. Die Ähnlichkeit der Verteilungen spiegelt sich dann in den Werten wieder und machen die gekoppelten Zufallsvariablen vergleichbar.

3. Wie stellt man sich $d_{\mathrm{TV}}(p, q) = \frac{1}{2}\sum_x |p_x - q_x|$ vor? —



Die Pfeile ↧ zeigen $|p_x - q_x|$ im Histogramm. Diese Größen summiert man alle, halbiert die Summe, weil zwei Wahrscheinlichkeitsverteilungen beitragen, und das ergibt $d_{\mathrm{TV}}(p, q)$. Beachte: In der Rechnung ergibt sich, dass die blauen Flächen zusammen so groß sind wie die roten Flächen, siehe (7). Der Überlapp $p_x \wedge q_x$ ist das, was p und q gemeinsam haben, vergleiche (2.2.11).

4. Was bedeutet (2.2.9)? — Theorem 2.9 gibt eine äquivalente Beschreibung von d_{TV}:

$$d_{\mathrm{TV}}(\mu, \nu) = \inf\{\mathbb{P}(\hat{X} \neq \hat{Y}) \mid (\hat{X}, \hat{Y}) \text{ Kopplung mit } \hat{X} \sim \mu,\ \hat{Y} \sim \nu\} \qquad (1)$$

5. Impliziert Theorem 2.10 schon (2.2.1)? Oder ist das sogar eine quantitative Fassung davon? — Ja, Konvergenz in Verteilung ist in Def. 2.1(a) definiert. Für alle $x \in \mathbb{R}$ (wir müssen noch nicht einmal Menge $\mathbb{N}$ der Unstetigkeitsstellen der Verteilungsfunktion der Poissonverteilung ausnehmen) gilt

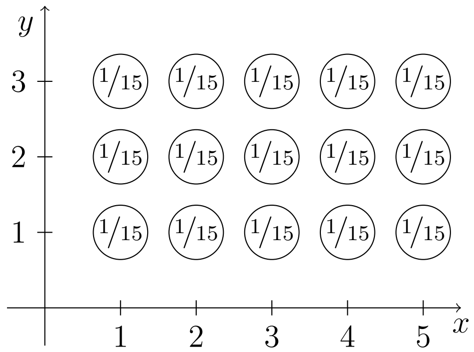
$$|\mathbb{P}(X_n \leq x) - \mathbb{P}(X \leq x)| \leq d_{\mathrm{TV}}(\mathrm{Bin}(n, \lambda/n), \mathrm{Poi}(\lambda)) \leq \frac{\lambda^2}{n} \xrightarrow{n\to\infty} 0, \qquad (2)$$

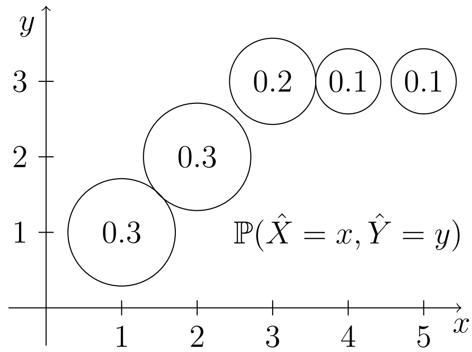
wobei wir die Definition (2.2.4) von d_{TV} und (39) aus Aufgabe 2.15 verwendet haben.

1 Kopplungen

Eine *Kopplung* von zwei ZVn X und Y ist ein zufälliger Vektor $(\hat{X}, \hat{Y})$ mit *Randverteilungen* $\mathcal{L}(X) = \mathcal{L}(\hat{X})$ und $\mathcal{L}(Y) = \mathcal{L}(\hat{Y})$.

Skizze zu *Kopplung* zweier diskreter ZVn X und Y:





Randverteilungen rechts:

$$\mathbb{P}(X = x) = \begin{cases} 0.3 & (x = 1) \\ 0.3 & (x = 2) \\ 0.2 & (x = 3) \\ 0.1 & (x = 4) \\ 0.1 & (x = 5) \end{cases} \qquad \mathbb{P}(Y = y) = \begin{cases} 0.3 & (y = 1) \\ 0.3 & (y = 2) \\ 0.4 & (y = 3) \end{cases} \tag{3}$$

Man kann auch gleich Kopplungen von beliebig vielen ZVn definieren. (In Kap. 4.1.1 wird eine überabzählbare Kopplung benutzt.)

Definition (Kopplung von ZVn)**.** Sei M eine Menge und X_m, $m \in M$, Zufallsvariablen. Die Familie $(\hat{X}_m)_{m \in M}$ von Zufallsvariablen heißt *Kopplung* der Zufallsvariablen X_m, $m \in M$, wenn alle $\hat{X}_m$ auf demselben Wahrscheinlichkeitsraum definiert sind und $\mathcal{L}(X_m) = \mathcal{L}(\hat{X}_m)$ für alle $m \in M$.

2 Totalvariation

Totalvariation = Metrik auf Wahrscheinlichkeitsverteilungen!

Notation:

- Messraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,
- Wahrscheinlichkeitsmaße μ, ν auf $\mathbb{R}$.

Definition:

$$d_{\mathrm{TV}}(\mu, \nu) := \sup_{A \in \mathcal{B}(\mathbb{R})} |\mu(A) - \nu(A)| \tag{2.2.4}$$

Jetzt $\Omega := \mathcal{X} \cup \mathcal{Y} \subseteq \mathbb{R}$ diskret, $p, q \colon \Omega \to [0, 1]$ Gewichtsfunktionen von μ bzw. ν, Behauptung: (**Aufgabe 2.13**)

$$d_{\mathrm{TV}}(p, q) = \frac{1}{2} \sum_{x \in \Omega} |p_x - q_x| \tag{2.2.6}$$

Beweis. Verwende
$$A_+ := \{x \in \Omega \mid p_x \geq q_x\}, \tag{4}$$
denn das könnte das Supremum aus (2.2.4) realisieren. Wie groß wird die Differenz denn?

$$|\mu(A_+) - \nu(A_+)| = \left|\sum_{x\in A_+} p_x - \sum_{x\in A_+} q_x\right| = \left|\sum_{x\in A_+} (p_x - q_x)\right| = \sum_{x\in A_+} (p_x - q_x), \tag{5}$$
$$|\mu(A_+^c) - \nu(A_+^c)| = \left|\sum_{x\in A_+^c} p_x - \sum_{x\in A_+^c} q_x\right| = \left|\sum_{x\in A_+^c} (p_x - q_x)\right| = \sum_{x\in A_+^c} |p_x - q_x|. \tag{6}$$

Außerdem gilt

$$|\mu(A_+^c) - \nu(A_+^c)| = |1 - \mu(A_+) - (1 - \nu(A_+)| = |\mu(A_+) - \nu(A_+)|. \tag{7}$$

Die beiden Summen aus (5) und (6) ergänzen sich. Wir werden also A_+ unter der Summe los, indem wir den Durchschnitt bilden:

$$|\mu(A_+) - \nu(A_+)| = \frac{1}{2}\big(|\mu(A_+) - \nu(A_+)| + |\mu(A_+^c) - \nu(A_+^c)|\big) = \frac{1}{2}\sum_{x\in\Omega}|p_x - q_x|. \tag{8}$$

Jetzt müssen wir nur noch prüfen, dass A_+ tatsächlich das Supremum realisiert. Für alle $A \subseteq \Omega$ können wir rechnen:

$$\mu(A) - \nu(A) = \mu(A\cap A_+) - \nu(A\cap A_+) + \underbrace{\mu(A\setminus A_+^c) - \nu(A\setminus A_+^c)}_{\leq 0} \tag{9}$$
$$\leq \mu(A\cap A_+) - \nu(A\cap A_+) \leq \mu(A_+) - \nu(A_+), \tag{10}$$

und andersherum auch:

$$\mu(A) - \nu(A) = \underbrace{\mu(A\cap A_+) - \nu(A\cap A_+)}_{\geq 0} + \mu(A\setminus A_+^c) - \nu(A\setminus A_+^c) \tag{11}$$
$$\geq \mu(A\cap A_+^c) - \nu(A\cap A_+^c) \geq \mu(A_+^c) - \nu(A_+^c). \tag{12}$$

Jetzt Beträge nehmen und fertig! □

Dieselbe Behauptung gibt es jetzt auch mit Integralen und Dichten statt Summen und Gewichtsfunktionen.

Seien jetzt $f, g\colon \mathbb{R} \to [0,1]$ Lebesguedichten von μ bzw. ν. Behauptung:

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\int_{\mathbb{R}} |f(x) - g(x)|\,\mathrm{d}x \tag{2.2.8}$$

Beweis. Verwende $A_+ := \{x \in \Omega \mid f(x) \geq g(x)\} \in \mathcal{A}$, denn f und g sind messbar. Rechne wie oben:

$$|\mu(A_+) - \nu(A_+)| = \left| \int_{A_+} f(x)\,\mathrm{d}x - \int_{A_+} g(x)\,\mathrm{d}x \right| = \int_{A_+} (f(x) - g(x))\,\mathrm{d}x, \tag{13}$$

$$|\mu(A_+^c) - \nu(A_+^c)| = \left| \int_{A_+^c} f(x)\,\mathrm{d}x - \int_{A_+^c} g(x)\,\mathrm{d}x \right| = \int_{A_+^c} |f(x) - g(x)|\,\mathrm{d}x. \tag{14}$$

Gleichung (7) ist genau so immer noch korrekt. Die beiden Integrale aus (13) und (14) ergänzen sich:

$$|\mu(A_+) - \nu(A_+)| = \frac{1}{2}\big(|\mu(A_+) - \nu(A_+)| + |\mu(A_+^c) - \nu(A_+^c)|\big) = \frac{1}{2}\int_{\mathbb{R}} |f(x) - g(x)|\,\mathrm{d}x. \tag{15}$$

Dass A_+ tatsächlich das Supremum realisiert, steht in (9)–(12). □

3 Maximale Kopplung

Theorem 2.9 sagt: X, Y ZVn mit Gewichtsfunktionen p, q. Die Kopplung mit der größten Wahrscheinlichkeit, dass $\hat{X} = \hat{Y}$ ist, hat $\mathbb{P}(\hat{X} \neq \hat{Y}) = d_{\mathrm{TV}}(q, p)$

Rechnungen zum Beweis:

- Behauptung: (**Aufgabe 2.14**)

$$\sum_{x\in\Omega} (p_x - (p_x \wedge q_x)) = \sum_{x\in\Omega} (q_x - (p_x \wedge q_x)) = d_{\mathrm{TV}}(p, q) \tag{2.2.13}$$

Beweis. Es gilt

$$\sum_{x\in\Omega} (p_x - (p_x \wedge q_x)) \stackrel{(4)}{=} \sum_{x\in A_+} (p_x - q_x) = d_{\mathrm{TV}}(p, q), \tag{16}$$

wegen (5), (8) und (2.2.6). Das mit q geht genauso. □

- Behauptung:

$$\mathbb{P}(\hat{X} = x) = p_x, \qquad \mathbb{P}(\hat{Y} = y) = q_y. \tag{2.2.14}$$

Beweis. Für alle $x \in \Omega$ gilt

$$(p_x - (p_x \wedge q_x))(q_x - (p_x \wedge q_x)) = 0, \tag{17}$$

denn einer der Faktoren ist gleich Null. Damit:

$$
\begin{aligned}
\mathbb{P}(\hat{X} = x) &= \sum_{y \in \Omega} \mathbb{P}(\hat{X} = x, \hat{Y} = y) && (18)\\
&= (p_x \wedge q_x) + \sum_{y \in \Omega \setminus \{x\}} \frac{(p_x - (p_x \wedge q_x))(q_y - (p_y \wedge q_y))}{d_{\mathrm{TV}}(p, q)} && (19)\\
&\overset{(17)}{=} (p_x \wedge q_x) + \sum_{y \in \Omega} \frac{(p_x - (p_x \wedge q_x))(q_y - (p_y \wedge q_y))}{d_{\mathrm{TV}}(p, q)} && (20)\\
&= (p_x \wedge q_x) + \frac{p_x - (p_x \wedge q_x)}{d_{\mathrm{TV}}(p, q)} \underbrace{\sum_{y \in \Omega} (q_y - (p_y \wedge q_y))}_{= d_{\mathrm{TV}}(p,q)} && (21)\\
&= (p_x \wedge q_x) + p_x - (p_x \wedge q_x) = p_x. && (22)
\end{aligned}
$$

Das mit $\hat{Y}$ geht dann genauso. □

- Behauptung:

$$
\mathbb{P}(\hat{X} = \hat{Y} = x) = p_x \wedge q_x, \tag{2.2.11}
$$

$$
\mathbb{P}(\hat{X} = x, \hat{Y} = y) = \frac{(p_x - (p_x \wedge q_x))(q_y - (p_y \wedge q_y))}{d_{\mathrm{TV}}(p, q)}, \qquad x \neq y. \tag{2.2.12}
$$

definiert eine Wahrscheinlichkeitsverteilung. Alle Terme sind nicht negativ. Wir prüfen die Normierung:

Beweis.

$$
\sum_{x \in \Omega} \sum_{y \in \Omega} \mathbb{P}(\hat{X} = x, \hat{Y} = y) \overset{(2.2.14)}{=} \sum_{x \in \Omega} p_x = 1. \tag{23}
$$

□

- (2.2.15) verwendet (17).
- Behauptung:

$$
1 - \sum_{x \in \Omega} (p_x \wedge q_x) = d_{\mathrm{TV}}(p, q). \tag{2.2.20}
$$

Beweis.

$$
1 - \sum_{x \in \Omega} (p_x \wedge q_x) = \sum_{x \in \Omega} (p_x - (p_x \wedge q_x)) \overset{(2.2.13)}{=} d_{\mathrm{TV}}(p, q). \tag{24}
$$

□

4 Anwendung bei Binomial- und Poissonverteilung

Theorem 2.10 gibt eine Abschätzung für die beste Kopplung einer Bionomial- und und Poissonverteilung.

Behauptung:

$$\mathbb{P}(\hat{X} \neq \hat{Y}) \leq \frac{\lambda^2}{n}. \tag{2.2.22}$$

Beweis. Wähle $p_i := \frac{\lambda}{n}$, dann ist

$$\sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} \frac{\lambda^2}{n^2} = n\frac{\lambda^2}{n^2} = \frac{\lambda^2}{n}. \tag{25}$$

□

Gleichung (2.2.23) bedeutet

$$p_{i,x} = \begin{cases} 1 - p_i & (x = 0), \\ p_i & (x = 1), \\ 0 & (x \geq 2). \end{cases} \tag{26}$$

Zu (2.2.25):

$$x = 0: \qquad \min\{p_{i,0}, q_{i,0}\} = \min\{1 - p_i, \mathrm{e}^{-p_i}\} = 1 - p_i \tag{27}$$

$$x = 1: \qquad \min\{p_{i,1}, q_{i,1}\} = \min\{p_i, p_i\mathrm{e}^{-p_i}\} = p_i \tag{28}$$

Boole's inequality ist Subadditivität.

5 Kommentar zu Konvergenz in TV-Abstand und punktweiser Konvergenz der Gewichtsfunktionen

Der Text verweist auf **Aufgabe 2.16**. Seien $p^{(n)} = (p_x^{(n)})_{x\in\mathcal{X}}$, $n \in \mathbb{N}$, Gewichtsfunktionen, so dass $p_x := \lim_{n\to\infty} p_x^{(n)}$ existiert und $p := (p_x)_x$ wieder eine Gewichtsfunktion ist. Behauptung:

$$\lim_{n\to\infty} d_{\mathrm{TV}}(p^{(n)}, p) = 0. \tag{29}$$

Beweis. Mal sehen, ob wir das einfach ausrechnen können.

$$d_{\mathrm{TV}}(p^{(n)}, p) = \frac{1}{2} \sum_{x\in\mathcal{X}} |p_x^{(n)} - p_x| = \dots \tag{30}$$

Jetzt müssten wir den Limes und die Reihe vertauschen. Das heißt aber, dass wir zwei Limiten vertauschen müssen, und dafür braucht man etwas Gleichmäßigkeit.

Die kommt bestimmt aus der Tatsache, dass p sich zu 1 summiert. Vielleicht können wir die Summe abbrechen und den Fehler abschätzen.

Sei $\varepsilon > 0$. Wähle $A \subseteq X$ endlich und so groß, dass $\sum_{x\in A} p_x \geq 1 - \varepsilon$. Dann ist

$$
\begin{aligned}
d_{\mathrm{TV}}(p^{(n)}, p) &= \frac{1}{2}\sum_{x\in\mathcal{X}} |p_x^{(n)} - p_x| && (30)\\
&\leq \frac{1}{2}\sum_{x\in A} |p_x^{(n)} - p_x| + \frac{1}{2}\sum_{x\in\mathcal{X}\setminus A} (p_x^{(n)} + p_x) = \dots && (31)
\end{aligned}
$$

Um den Fehler abzuschätzen, wählen wir $N \in \mathbb{N}$ so groß, dass für alle $n \geq N$

$$
|p_x^{(n)} - p_x| < \frac{\varepsilon}{|A|} \tag{32}
$$

erfüllt ist. Das geht, weil $\lim_n p_x^{(n)} = p_x$ und A endlich ist. Dann erhalten wir als Fehlerabschätzung mit der Dreiecksungleichung

$$
\begin{aligned}
\sum_{x\in\mathcal{X}\setminus A} (p_x^{(n)} + p_x) &= \sum_{x\in\mathcal{X}\setminus A} p_x^{(n)} + \sum_{x\in\mathcal{X}\setminus A} p_x && (33)\\
&= 1 - \sum_{x\in A} \underbrace{p_x^{(n)}}_{\geq p_x - |p_x^{(n)} - p_x|} + 1 - \underbrace{\sum_{x\in A} p_x}_{\geq 1-\varepsilon} && (34)\\
&\leq 1 - \sum_{x\in A} (p_x - |p_x^{(n)} - p_x|) + 1 - (1-\varepsilon) && (35)\\
&\leq 1 - \sum_{x\in A} p_x + \sum_{x\in A} |p_x^{(n)} - p_x| + \varepsilon && (36)\\
&\leq 3\varepsilon. && (37)
\end{aligned}
$$

Jetzt zusammensetzen: Für alle $n \geq N$ gilt

$$
\begin{aligned}
d_{\mathrm{TV}}(p^{(n)}, p) &= \frac{1}{2}\sum_{x\in\mathcal{X}} |p_x^{(n)} - p_x| && (30)\\
&\leq \frac{1}{2}\sum_{x\in A} |p_x^{(n)} - p_x| + \frac{1}{2}\sum_{x\in\mathcal{X}\setminus A} (p_x^{(n)} + p_x) && (31)\\
&\leq \frac{1}{2}\varepsilon + \frac{3}{2}\varepsilon = 2\varepsilon. && (38)
\end{aligned}
$$

Und weil $\varepsilon > 0$ beliebig war, haben wir damit (29) bewiesen. □

Aufgabe 2.15: Voraussetzungen:

- $\lambda > 0$, $n \in \mathbb{N}$
- $X \sim \mathrm{Bin}(n, \lambda/n)$, $f_i := \mathbb{P}(X = i)$,
- $Y \sim \mathrm{Poi}(\lambda)$, $g_i := \mathbb{P}(Y = i)$.

Behauptung:

$$d_{\mathrm{TV}}(f, g) \leq \frac{\lambda^2}{n} \tag{39}$$

und $\forall i \in \mathbb{N}$:

$$|\mathbb{P}(X = i) - \mathbb{P}(Y = i)| \leq \lambda^2/n. \tag{40}$$

Beweis. Mit Theorem 2.10 finden wir eine Kopplung $(\hat{X}, \hat{Y})$ von X und Y, so dass folgendes gilt:

$$d_{\mathrm{TV}}(f, g) \overset{\text{Thm. 2.9}}{\leq} \mathbb{P}(\hat{X} \neq \hat{Y}) \overset{\text{Thm. 2.10}}{\leq} \frac{\lambda^2}{n}.$$

Außerdem ist

$$|\mathbb{P}(X = i) - \mathbb{P}(Y = i)| \leq \sup_A |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| = d_{\mathrm{TV}}(f, g) \leq \lambda^2/n. \quad \square$$