Zufällige Graphen

Christoph Schumacher

30. April 2020

RGCN I, Kap. 1.1, 1.2, 2.3: Stochastische Ordnung

Zusammenfassung

Größenverzerrte ZVn und stochastische Ordnung

Inhaltsverzeichnis

Fragen

1. Warum heißt $X^\star$ *size biased* version of X? — Alternative Beschreibung der Verteilung von $D^\star$: Wähle einen Knoten $U^\star$ zufällig aus allen Knoten in $[n]$ aus, aber nicht gleichverteilt, sondern *size biased*, genauer: proportional zu seinem Knotengrad:
$$\mathbb{P}(U^\star = u) = \frac{d_u}{\sum_{v\in[n]} d_v}. \tag{1}$$

Dann haben $d_{U^\star}$ und $D_n^\star$ dieselbe Verteilung:

$$\begin{aligned}
\mathbb{P}(d_{U^\star} = k) &= \sum_{u\in[n], d_u=k} \mathbb{P}(U^\star = u) && (2)\\
&= \sum_{u\in[n], d_u=k} \frac{k}{\sum_{v\in[n]} d_v} && (3)\\
&= \frac{k}{\frac{1}{n}\sum_{v\in[n]} d_v} \cdot \frac{1}{n} \sum_{u\in[n], d_u=k} 1 && (4)\\
&= \frac{k}{\mathbb{E}[D_n]} \mathbb{P}(D_n = k) && (5)\\
&\overset{(1.2.2)}{=} \mathbb{P}(D_n^\star = k). && (6)
\end{aligned}$$

1 Erdős-Rényi-Graph (1.1)

Motivation: Warum zufällige Graphen? — Effektive Beschreibung großer Netzwerke

Einfachstes Beispiel: Erdős-Rényi-Graph.

2 Gradfolge, typischer Grad (1.2)

- Gradfolge: $\mathbf{d} = (d_v)_{v\in[n]}$, wobei $[n] = \{1, \ldots, n\} = V(G)$.
- $U : \Omega \to V(G) = [n]$ gleichverteilter zufälliger Knoten,
- Typischer Grad $D_n := d_U$,
- Verteilung von D_n = Empirische Gradverteilung des Graphen = Histogram der Gradfolge $\mathbf{d}$

3 Größenverzerrte ZVn

Größenverzerrte Gradverteilung $D_n^\star$ = Grad eines Endes einer zufällig ausgewählten Kante

Warum gilt Gleichung (1.2.3)? — Zwei Möglichkeiten:

$$\mathbb{E}[D_n] = \mathbb{E}[d_U] = \sum_{u\in[n]} d_u \mathbb{P}(U = u) = \sum_{u\in[n]} d_u \frac{1}{n} = \frac{1}{n} \sum_{u\in[n]} d_u \tag{7}$$

oder

$$\mathbb{E}[D_n] = \sum_{k=0}^{n-1} k\mathbb{P}(D_n = k) = \sum_{k=0}^{n-1} k \frac{|\{u \in [n] \mid d_u = k\}|}{n} \tag{8}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1} k \sum_{u\in[n]} \mathbf{1}_{\{k\}}(d_u) = \frac{1}{n}\sum_{u\in[n]}\sum_{k=0}^{n-1} k\mathbf{1}_{\{k\}}(d_u) = \frac{1}{n}\sum_{u\in[n]} d_u \tag{9}$$

Beweis von Theorem 1.1. Wir rechnen mit der Gleichverteilung auf der Menge

$$E' := \big\{(v,w) \mid \{v,w\} \in E(G)\big\} \tag{10}$$

aller gerichteten Kanten. Der erste Knoten der Kante ist der, dessen Grad wir für $D_n^\star$ nehmen. Für die Gleichverteilung teilen wir Günstige durch Mögliche:

- Günstige: Die Anzahl aller von einem Knoten mit Grad k ausgehender gerichteter Kanten ist

$$k \cdot |\{v \in V(G) \mid d_v = k\}| = kn\mathbb{P}(D_n = k). \tag{11}$$

- Mögliche: Die Anzahl aller gerichteter Kanten ist

$$|E'| = 2|E(G)| = \sum_{v\in[n]} d_v = n\mathbb{E}[D_n]. \tag{12}$$

Zusammengesetzt:

$$\mathbb{P}(D_n^\star = k) = \frac{k\cdot|\{v \in V(G) \mid d_v = k\}|}{|E'|} = \frac{kn\mathbb{P}(D_n = k)}{n\mathbb{E}[D_n]} = \frac{k}{\mathbb{E}[D_n]}\mathbb{P}(D_n = k) \tag{13}$$

□

Warum ist (1.2.2) und (1.2.4) dasselbe? Aus (1.2.4) folgt (1.2.2):

$$\mathbb{P}(D_n^\star = k) = \mathbb{P}(D_n^\star \le k) - \mathbb{P}(D_n^\star \le k-1) \tag{14}$$

$$\overset{(1.2.4)}{=} \frac{\mathbb{E}[D_n\mathbf{1}_{\{D_n\le k\}}]}{\mathbb{E}[D_n]} - \frac{\mathbb{E}[D_n\mathbf{1}_{\{D_n\le k-1\}}]}{\mathbb{E}[D_n]} \tag{15}$$

$$= \frac{\mathbb{E}[D_n\mathbf{1}_{\{D_n=k\}}]}{\mathbb{E}[D_n]} \tag{16}$$

$$= \frac{k\mathbb{E}[\mathbf{1}_{\{D_n=k\}}]}{\mathbb{E}[D_n]} \tag{17}$$

$$= \frac{k\mathbb{P}(D_n = k)}{\mathbb{E}[D_n]} \tag{18}$$

Und umgekehrt:

$$\mathbb{P}(D_n^\star \le x) = \sum_{k=0}^{x} \mathbb{P}(D_n^\star = k) \tag{19}$$

$$\overset{(1.2.2)}{=} \sum_{k=0}^{x} k \frac{\mathbb{P}(D_n = k)}{\mathbb{E}[D_n]} \tag{20}$$

$$= \sum_{k=0}^{x} \frac{\mathbb{E}[D_n \mathbf{1}_{\{D_n=k\}}]}{\mathbb{E}[D_n]} \tag{21}$$

$$= \frac{\mathbb{E}[D_n \mathbf{1}_{\{D_n \leq x\}}]}{\mathbb{E}[D_n]} \tag{22}$$

Zum Beweis von Theorem 1.2: $D_n = X_n$, Zwischenschritt zu Gleichung (1.2.7): Erst gleichverteilt zufällig einen Knoten wählen, dann aus dessen Nachbarn zufällig und gleichverteilt einen Nachbarn wählen:

$$\mathbb{P}(X_n = k, Y_n = l) = \frac{1}{n} \sum_{v \in [n]} \frac{1}{d_v} \sum_{v \in N_u} \mathbf{1}_{\{d_u=k, d_v=l\}} = \frac{1}{n} \sum_{(u,v) \in E'} \mathbf{1}_{\{d_u=k, d_v=l\}} \frac{1}{d_v} \tag{23}$$

Gleichung (1.2.10) ist äquivalent zur zweiten binomischen Formel:

$$1 \leq \frac{1}{2} \left(\frac{x}{y} + \frac{y}{x} \right) \iff 2xy \leq x^2 + y^2 \iff 0 \leq x^2 - 2xy + y^2 = (x+y)^2 \tag{24}$$

4 Typische Weglänge

Keine Fragen.

5 Stochastische Ordnung (2.3)

Definition. Die ZV X ist *stochastisch kleiner* als die ZV Y, in Zeichen $X \preceq Y$, wenn

$$\forall x \in \mathbb{R}\colon F_X(x) = \mathbb{P}(X \leq x) \geq F_Y(x) \tag{25}$$

Kopplungsformulierung:

$$X \preceq Y \iff \exists \text{Kopplung } (\hat{X}, \hat{Y}) \text{ von } X, Y : \hat{X} \leq \hat{Y} \quad \text{fast sicher}$$

Im Beweis:

$$F^{-1}(u) := \inf\{x \in \mathbb{R} \mid F(x) \geq u\} \tag{26}$$

heißt *Quantiltransformation* (für die rechtsstetige Verteilungsfunktion F).

Rechnung zu (2.3.8):

$$F_X^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_X(x) \geq u\} \leq \inf\{x \in \mathbb{R} \mid F_Y(x) \geq u\} = F_Y^{-1}(u), \tag{27}$$

denn $\{x \in \mathbb{R} \mid F_X(x) \geq u\} \supseteq \{x \in \mathbb{R} \mid F_Y(x) \geq u\}$, denn $u \leq F_Y(x) \overset{(2.3.7)}{\Longrightarrow} u \leq F_X(x)$.

Beispiele:

- Binomialverteilung: $m \leq n \implies \mathrm{Bin}(m,p) \preceq \mathrm{Bin}(n,p)$ (eigentlich $X \preceq Y$ mit $X \sim \mathrm{Bin}(m,p), Y \sim \mathrm{Bin}(n,p)$)
- **Aufgabe 2.17**:
- Poissonverteilung: $\lambda \leq \mu \implies \mathrm{Poi}(\lambda) \preceq \mathrm{Poi}(\mu)$
- Geometrische Verteilung: $p \leq p' \in [0,1] \implies \mathrm{Geo}(p) \succeq \mathrm{Geo}(p')$.
  - Variante 1: $F_{\mathrm{Geo}(p)}(x) = 1-(1-p)^{\lfloor x \rfloor}$ (bzw. $F_{\mathrm{Geo}(p)}(x) = 1-(1-p)^{\lfloor x \rfloor - 1}$) ist für alle x monoton wachsend in p.
  - Variante 2: Wir brauchen ein Lemma.

    Lemma (Minimum geometrisch verteilter Zufallsgrößen)**.** *Seien X, Z geometrisch verteilt mit Parametern $p, q \in [0,1]$. Dann ist $Y := \min\{X, Z\}$ ebenfalls geometrisch verteilt mit Parameter $1-(1-p)(1-q)$.*

    Beweis. Wir stellen $X := \inf\{j \in \mathbb{N} \mid X_j = 1\}$ mit unabhängigen $X_j \sim \mathrm{Bin}(1,p)$ und $Z := \inf\{j \in \mathbb{N} \mid Z_j = 1\}$ mit unabhängigen $Z_j \sim \mathrm{Bin}(1,q)$ dar. Dann ist $Y = \inf\{j \in \mathbb{N} \mid X_j = 1 \text{ oder } Z_j = 1\}$ ebenfalls geometrisch verteilt mit Parameter

    $$\begin{aligned} \mathbb{P}(X_1 = 1 \text{ oder } Z_1 = 1) &= 1 - \mathbb{P}(X_1 = 0, Z_1 = 0) && (28) \\ &= 1 - \mathbb{P}(X_1 = 0)\mathbb{P}(Z_1 = 0) && (29) \\ &= 1 - (1-p)(1-q) = p + q - pq. && \square \end{aligned}$$

    Nun zur Aussage: Seien $X := \hat{X} \sim \mathrm{Geo}(p)$, $Z \sim \mathrm{Geo}(\frac{p'-p}{1-p})$ unabhängig. Dann ist

    $$\hat{Y} := \min\{X, Z\} \sim \mathrm{Geo}\big(1 - (1-p)(1 - \tfrac{p'-p}{1-p})\big) = \mathrm{Geo}(p'),$$

    so dass $(\hat{X}, \hat{Y})$ eine Kopplung von X und Y ist, und per Konstruktion haben wir $\hat{X} \geq \hat{Y}$ sichergestellt.
- Größenverzerrte ZVn: Der Beweis von Lemma 2.14 funktioniert auch, wenn f und g beide monoton fallend sind. Wenn f steigt und g fällt oder umgekehrt, dann dreht sich die Ungleichung um.

6 Erwartungswerte

In Theorem 2.15 Für positive ZVn X, Y mit $X \preceq Y$ kann man auch direkt mit der Definition argumentieren:

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x) \,\mathrm{d}x \tag{30}$$

$$
\begin{aligned}
&= \int_0^\infty (1 - \mathbb{P}(X \leq x))\,\mathrm{d}x && (31)\\
&\leq \int_0^\infty (1 - \mathbb{P}(Y \leq x))\,\mathrm{d}x && (32)\\
&= \int_0^\infty \mathbb{P}(Y > x)\,\mathrm{d}x && (33)\\
&= \mathbb{E}[Y]. && (34)
\end{aligned}
$$

7 Monotonie

Alles klar.