Wahrscheinlichkeitstheorie für Informatikstudien (Kapitel #3)

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.

Wichtige Verteilungen und Zufallsvariablen

Kurze Zusammenfassung

Hypergeometrische Verteilung:

Wir ziehen n mal ohne Zurücklegen aus einer Urne mit N_1 weißen und $N - N_1$ schwarzen Kugeln. Die PMF lautet:

$$p_k := P(|\text{weiß}| = k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}, \quad k \in \{0, \ldots, n\}.$$

Binomialverteilung:

Wir ziehen n mal mit Zurücklegen aus einer Urne mit N_1 weißen und $N - N_1$ schwarzen Kugeln. Sei $p = N_1/N$. Die PMF lautet:

$$p_k := P(|\text{weiß}| = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \ldots, n\}.$$

Kurze Zusammenfassung

Beachte: wir hätten gleich den kleineren Raum $\Omega = \{0, \ldots, n\}$ nehmen können und direkt darauf $P(k) := p_k$ definieren.

$\Longrightarrow$ Dann ist aber die Motivation unklar.

Wir haben gezeigt, dass p_k im hypergeometrischen Modell, sowie im Binomialmodell annähernd gleich sind, wenn die Anzahl der Kugeln in der Urne gg. unendlich geht.

D.h. die Binomialverteilung kann als ein Grenzfall der hypergeometrischen Verteilung gesehen werden, in welchem die Anzahl der Kugeln in der Urne sehr groß ist, aber das Verhältnis der Farben konstant gehalten wird.

Die Poisson-Verteilung

Die Poisson Verteilung ist wiederum ein Grenzfall der Binomialverteilung.

Nehmen wir ein Experiment mit Binomialverteilung, wobei $p \to 0$ und $n \to \infty$ mit $np \to \lambda > 0$. Wie kann man die PMF

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

numerisch gut approximieren?

Beispiel (Lotto)

Frage: Wenn n Lottoscheine verkauft wurden, wie groß ist die Wahrscheinlichkeit dass k Scheine darunter sind, in denen alle 6 Zahlen richtig sind?

Beachte: hier ist n typischerweise: *sehr* groß und $p = \binom{45}{6}^{-1} \approx 1.228 \times 10^{-7}$ *sehr* klein.

Die Poisson-Verteilung

Unter obigen Annahmen, haben wir für jedes fixe ℓ, dass $(n-\ell)p \to \lambda$. Für jedes fixe k gilt $(1-p)^{-k} \to 1$. Daher folgt, dass

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n \cdots (n-k+1)}{k!} p^k (1-p)^{n-k}$$

$$= \underbrace{\frac{np \times (n-1)p \times \cdots \times (n-k+1)p}{k!}}_{\to \frac{\lambda^k}{k!}} \underbrace{\left(1-\frac{pn}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{(1-p)^{-k}}_{\to 1}$$

$$\to \frac{\lambda^k e^{-\lambda}}{k!} =: p_k.$$

Die Poisson-Verteilung

Die PMF $p_k = \frac{\lambda^k e^{-\lambda}}{k!}$, $k = 0, 1, \ldots$ heißt Poisson-Verteilung.

Beachte, dass p_k wirklich eine PMF auf $\{0, 1, 2, \ldots\}$ definiert, also $p_k \geq 0$ und

$$\sum_{k \geq 0} p_k = 1.$$

Wir können nun mit der Poisson-Verteilung ein neues probabilistisches Modell definieren, in welchem wir nicht mehr direkt auf das motivierende Urnenmodell zurückgreifen müssen.

Die obige Herleitung macht klar, warum das Poisson-Modell oft benutzt wird um seltene Ereignisse zu beschreiben.

Die Poisson-Verteilung

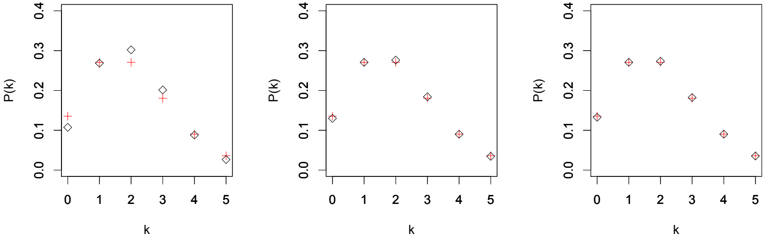

Abbildung: Poisson Approximation: $n = 10$, $n = 50$ und $n = 100$. Rote Kreuze sind die PMF einer Poisson-Verteilung mit $\lambda = 2$ und die Rauten sind die PMF einer Binomialverteilung mit Parametern n und $p = 2/n$.

Die Poisson-Verteilung

Beispiel (Lotto)

Angenommen 20 Mio. Lottoscheine wurden verkauft. Wie groß ist die Wahrscheinlichkeit mindestens 2 Gewinner zu haben? Wir verwenden die Poisson Approximation und erhalten $np \approx 2.46$.

Beispiel (Fußball)

Angenommen die Anzahl der Tore pro Spiel eines Clubs folgt einer Poisson-Verteilung mit Parameter $\lambda = 2$. Wie groß ist die Wahrscheinlichkeit, dass die Mannschaft in mindestens einem Spiel (in einer Saison mit 50 Spielen) mehr als 4 Tore schießt?

Das nächste diskrete Wahrscheinlichkeitsmodell ist durch folgendes Beispiel motiviert.

Beispiel (Auf einen Erfolg warten)

Angenommen wir spielen " Mensch ärgere Dich nicht". Damit wir anfangen können, müssen wir eine 6 würfeln. Wie groß ist die Wahrscheinlichkeit, dass wir mehr als 20 mal würfeln müssen um anfangen zu können?

Wie lautet ein passendes Urnenmodell?

Geometrische Verteilung

Angenommen wir ziehen (mit Zurücklegen) aus einer Urne mit N_1 roten und $N - N_1$ blauen Kugeln. Sei $p = N_1/N$. Wie groß ist die Wahrscheinlichkeit, dass wir k Versuche brauchen um eine rote Kugel zu ziehen?

Um wieder ein Laplacemodell zu benutzen, brauchen wir zunächst einen endlichen Stichprobenraum, also setzen wir $\Omega = \{(\omega_1, \ldots, \omega_n) \in \{1, \ldots, N\}^n\}$. Die Zahlen 1 bis N_1 bezeichne die roten Kugeln und der Rest die blauen Kugeln.

Wir haben $|\Omega| = N^n$ und aus Symmetriegründen erscheint jede Sequenz mit derselben Wahrscheinlichkeit N^{-n}.

Geometrische Verteilung

Das Ereignis $A_k = \{\text{erste rote Kugel nach } k \text{ Versuchen}\}$ (sei hier $1 \leq k \leq n$) impliziert, dass die ersten $k-1$ Zahlen in der Folge alle $> N_1$ sind und dass die k-te Zahl $\leq N_1$ ist. Es gibt keine Einschränkung für die anderen Zahlen. Dies ergibt $(N - N_1)^{k-1} \times N_1 \times N^{n-k}$ mögliche Kombinationen.

Daraus folgern wir, dass

$$P(A_k) = (1-p)^{k-1} \times p \quad \text{für } 1 \leq k \leq n$$

und

$$P((A_1 \cup \ldots \cup A_n)^c) = (1-p)^n.$$

Die PMF mit $p_k = (1-p)^{k-1}p$, $k = 1, 2, 3, \ldots$ nennen wir die geometrische Verteilung.

Geometrische Verteilung

Beispiel (Mensch ärgere Dich nicht)

Wir spielen Mensch ärgere Dich nicht". Wie groß ist die Wahrscheinlichkeit, dass wir mehr als 20x würfeln müssen um anfangen zu können?

Überblick

Diskrete Gleichverteilung auf Ω:

$$p_k = 1/|\Omega|, \quad k \in \Omega.$$

Hypergeometrische Verteilung:

$$p_k = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}, \quad k \in \{0, \ldots, n\}.$$

Binomialverteilung:

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, \ldots, n\}.$$

Überblick

Poisson-Verteilung:

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \geq 0.$$

Geometrische Verteilung:

$$p_k = p(1-p)^{k-1}, \quad k \geq 1.$$