

Diskrete Verteilungen — Rekapitulation

H: Hypergeometrische — ziehen ohne Rücklegen

$$X \sim H(k; n, N, K)$$

B: Binomial — Wieviele Treffer aus N?

$$X \sim B(N, p)$$

G: Geometrische — Wieviele Versuche bis Erfolg?

$$X \sim G(p)$$

NB: Negative Binomial – Wieviele Versuche zum r. Erfolg?

$$X \sim NB(r, p)$$

P: Poisson — Wieviele Ereignisse?

$$X \sim P(\lambda)$$

B, G, NB, P — alle nehmen Unabhängigkeit an.

und Spezialfälle wie Benford's Gesetz:

Die Verteilung der ersten Ziffer in Zahlen ist

$$P(\text{erste Ziffer} = n) = \log_{10} \left(1 + \frac{1}{n} \right) \quad n = 1, \dots, 9$$

Mathematische Resultate für diskrete Verteilungen

Für eine diskrete Zufallsvariable X mit Verteilung

$$P(X = x) = p(x) \quad x = 0, 1 \dots$$

gilt

$$\sum_{x=0}^{\infty} p(x) = 1$$

$$E[X] = \sum_{x=0}^{\infty} xp(x)$$

$$V[X] = \sum_{x=0}^{\infty} (x - E[X])^2 p(x)$$

z.B. $X \sim B(N, p)$ (Binomial Verteilung)

$$\sum_{x=0}^N \binom{N}{x} p^x (1-p)^{N-x} = 1$$

$$E[X] = \sum_{x=0}^N x \binom{N}{x} p^x (1-p)^{N-x}$$

$$= Np \sum_{y=0}^{N-1} \binom{N-1}{y} p^y (1-p)^{N-1-y} = Np$$

Beispiele von Anwendungen diskreter Verteilungen

1. 10% der Bevölkerung bringen Leistungen besser als der Wert L . Wieviele Leute muss man testen, bis man 20 hat?
2. Die Krankheit ALS hat eine Prävalenzrate von 3 bis 8 pro 100.000. Wäre es überraschend, 20 Patienten in Augsburg zu finden?
3. Nehmen wir an, dass 25% der Bevölkerung SPD wählen wollen. Was ist die Verteilung von SPD Wählern in einer Stichprobe der Größe 10? 100? 1000?
4. Für ein Fernsehprogramm soll ein Wähler von jeder Partei beim Einkaufen in der Stadt befragt werden. Wenn die Linken 5,5% Unterstützung haben, was ist die Verteilung der Anzahl Leute, die angesprochen werden muss, bis ein Wähler der Linken gefunden wird?
5. Im Durchschnitt gibt es ein Erdbeben der Stärke 8 oder höher jedes Jahr. Wie könnte man die Verteilung der Anzahl Erdbeben im nächsten Jahr modellieren?

3.8.3 Die Poisson Verteilung als eine Approximation für die Binomialverteilung

$$X \sim B(N, p)$$

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

Sei $Np = \lambda$, dann ist $X \sim B(N, \lambda/N)$ und

$$\begin{aligned} P(X = k) &= \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \\ &= \left(1 * \frac{N-1}{N} \dots \frac{N-k+1}{N}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \end{aligned}$$

Für $N \rightarrow \infty$ und λ konstant gilt

$$P(X = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

eine Poisson Verteilung.

Man findet auch

”Für $N \rightarrow \infty, p \rightarrow 0$ unter der Bedingung, dass $Np \rightarrow \lambda$, ist $P(Np)$ eine gute Approximation für $B(N, p)$.”

3.8.4 Die Poisson Verteilung ist die Zufallsverteilung

Seien

$$p_1(i) = P(i \text{ Ereignisse im Zeitintervall } t_1)$$

$$p_2(j) = P(j \text{ Ereignisse im Zeitintervall } t_2)$$

wo $t_1 \cap t_2 = \emptyset$ und

$$p_0(n) = P(n \text{ Ereignisse im Zeitintervall } t_0 = t_1 \cup t_2)$$

Dann

$$p_0(n) = \sum_{i=0}^n p_1(i)p_2(n-i)$$

Wir möchten, dass die Verteilungen $\{p_k(i)\}$ alle dieselbe Form haben. Wir nehmen an, dass die Ereignisse rein zufällig und unabhängig von einander sind.

\Rightarrow Die Wahrscheinlichkeit ein Ereignis in t_1 fällt, gegeben, dass es in $t_0 = t_1 \cup t_2$ fällt, ist

$$\alpha = \frac{t_1}{t_0}$$

Dann gilt

$$p_1(i)p_2(n-i) = \binom{n}{i} \alpha^i (1-\alpha)^{n-i} p_0(n)$$

Insbesondere haben wir

$$p_1(n)p_2(0) = \alpha^n p_0(n) \quad (\Rightarrow p_2(0) > 0)$$

und

$$p_1(n-1)p_2(1) = n\alpha^{n-1}(1-\alpha)p_0(n)$$

So dass

$$\frac{p_1(n)}{p_1(n-1)} = \frac{p_2(1)}{p_2(0)} \left(\frac{\alpha}{1-\alpha} \right) \frac{1}{n}$$

Für $n = 1$

$$\frac{p_1(1)}{p_1(0)} \frac{1}{\alpha} = \frac{p_2(1)}{p_2(0)} \frac{1}{1-\alpha} = c$$

$$\Rightarrow \frac{p_1(n)}{p_1(n-1)} = \frac{\alpha c}{n}$$

und so

$$p_1(n) = \frac{(\alpha c)^n}{n!} p_1(0)$$

eine Poisson Verteilung mit $\lambda = \alpha c$

3.8.5 Noch zwei Poisson Beispiele

1. Geburtstage

Gegeben 100 Studenten, wieviele haben heute Geburtstag? Als Approximation nehmen wir:

$$X \sim P\left(\frac{100}{365}\right)$$

$$P(\text{mindestens 1 Geburtstag}) \approx 1 - 0.76$$

2. Lotto

Wieviel Gewinner im Lotto?

$$P(\text{Größter Gewinn}) = \frac{1}{\binom{49}{6}} * 0.1$$

Gegeben 100.000.000 Tipps, die alle zufällig ausgewählt werden, modellieren wir die Anzahl Gewinner mit $X \sim P(0.715112)$

X	0	1	2	≥ 3
$P(X)$	0.489	0.350	0.125	0.036

3.9 R und diskrete Verteilungen

Für jede Verteilung gibt es vier Kommandos. z.B. für die Binomial Verteilung $X \sim B(N, p)$ haben wir

dbinom(x, N, p) $P(X = x)$

pbinom(y, N, p) $P(X \leq y)$

qbinom(s, N, p) $P(X \leq z) = s$

rbinom(n, N, p) eine Zufallsstichprobe der Größe n

Bei einer Poisson Verteilung hat man

dpois(x, λ), ppois(y, λ), qpois(z, λ), rpois(n, λ)

Für die Negativ Binomial Verteilung hat man

dnbinom(x, N, p) usw

Für die Hypergeometrische hat man

dhyper(k, K, N - K, n) usw