Wdh: Wa. verteilung

- Zufallsvariable $X : \Omega \to \mathbb{R}$

  $X(\omega) : \omega \mapsto x$

- Wa. verteilung gibt Wa. für $X = x$: $P(X = x)$

  Bsp: Münze; Kopf → +3 SFr ; Zahl → −2 SFr

Tabelle

x	-2	$+3$
$P(X=x)$	$0{,}5$	$0{,}5$

$$\sum_{\text{alle } x} P(X=x) = 1$$

Bsp: Wa. verteilung durch Funktion beschreiben

Gewinnspiel: Zahl x aus $\{1,2,3,\dots,99,100\}$ mit Wa. $\frac{x}{5050}$

Gewinn = x

$$P(X=x) = \frac{x}{5050} \; ; \; \sum_{\text{alle } x} P(X=x) = 1$$

- Berühmte / Nützliche Wa. verteilungen:
  - Uniform $\{1,2,\dots,n\}$: $P(X=x) = \frac{1}{n}$ „gleiche Wa."
  - Binomial (n,p) : $P(X=x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$ „Lose"
  - Hypergeometrisch (N,n,m) : $P(X=x) = \dfrac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$ „Urne"
  - Poisson (λ) : $P(X=x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$ „Unfälle"

- Kennzahlen:

$$E[X] = \sum_x x \cdot P(X=x)$$

$$Var(X) = E\left[(X-E[X])^2\right] = \sum_x (x-E[X])^2 \cdot P(X=x)$$

$$\sigma_X = \sqrt{Var(X)}$$

Beispiele: Uniforme Verteilung

1) Geburtstag einer zufälligen Person ist x. Tag im Jahr

X: Nummer des Tages im Jahr, an dem die Person Geburtstag hat

$P[X=x] = 1/365$

Wa. am 1.1. oder 2.1. Geburtstag zu haben:

$P[X=1 \cup X=2] = P[X=1] + P[X=2] = \frac{1}{365} + \frac{1}{365} = \frac{2}{365}$

Beispiele: Binomialverteilung

1) Geburtstagslotto: 18 Tage angegeben; wer an diesen Tagen Geburtstag hat gewinnt

X: Anzahl Gewinner bei 200 Studenten

$X \sim Bin\left(n=200;\ \pi = \frac{18}{365} \approx 0{,}05\right)$

$P(X=0) = \binom{n}{0} \pi^0 (1-\pi)^{n-0} = 0{,}95^{200} \approx 3{,}5 \cdot 10^{-5}$

$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) \approx 0{,}002$

$P(X \le k) \stackrel{!}{=} 0{,}95 \Rightarrow k \approx 15$

2) LD50: „Lethal Dose 50%"
Wenn man diese Dosis einnimmt, stirbt man mit 50% Wa.

z.B. 10 Tiere; Was ist die Wa., dass alle Tiere überleben?

$P(X=0) = \binom{n}{0} \pi^0 (1-\pi)^{n-0} = 0{,}5^{10} \approx 0{,}001$

Beispiele: Hypergeometrische Verteilung

1) Angenommen, Medikament wirkt nicht

N Patienten = Kugeln

m Personen, die von selbst gesund werden = markierte Kugeln

n Personen bekommen Med. = gezogene Kugeln

z.B: $N=30$, $m=10$, $n=15$

Es werden 9 Personen in der Medikamentengruppe gesund.

Wie wahrscheinlich ist das? X: Anz. ges. Pers. in Med. gruppe

$X \sim \text{Hyper}(30, 15, 10)$

$$P[X=9] = \frac{\binom{10}{9}\binom{20}{6}}{\binom{30}{15}} \approx 0{,}0025$$

Wie wahrscheinlich ist es, MINDESTENS 9 Genesungen in der Medikamentengruppe zu beobachten?

$$P[X \geq 9] = P[X=9] + P[X=10] \approx 0{,}0026$$

mehr als 10 markierte Bälle gibt es nicht

Unter der Annahme, dass das Medikament nicht wirkt, ist die Beobachtung sehr unwahrscheinlich

$\Downarrow$

Es gibt guten Grund zur Annahme, dass das Medikament wirkt

↳ siehe: Fisher's Exact Test

Beispiele: Poissonverteilung

1) Rettungsdienst Zürich
Kann bis zu 5 Unfälle pro Stunde versorgen (tagsüber)
Letztes Jahr: Im Mittel 0,7 Unfälle pro Stunde

X: Anzahl Unfälle in einer Stunde

$X \sim \text{Pois}(0{,}7)$

Wie gross ist die Wa., dass man eine Std frei hat?

$$P[X=0] = \frac{0{,}7^0}{0!} e^{-0{,}7} = e^{-0{,}7} \approx 0{,}5$$

Wie gross ist die Wa., dass in einer Stunde mind. ein Unfall nicht versorgt werden kann?

$$P[X>5] = 1 - P[X \leq 5] =$$

$$= 1 - \left(P[X=0] + P[X=1] + \ldots + P[X=5]\right) =$$

$$= 1 - \left(\frac{0{,}7^0}{0!} e^{-0{,}7} + \frac{0{,}7^1}{1!} e^{-0{,}7} + \ldots + \frac{0{,}7^5}{5!} e^{-0{,}7}\right) \approx$$

$$\approx 1 - 0{,}9999 = 0{,}0001$$

2) Anzahl bewaffneter Konflikte pro Jahr zwischen 1500 A.D. und 1931 A.D.

Momentenmethode

Bsp: Grösse einer Bevölkerung abschätzen

Variante 1: Struktur bekannt

X: Anz. Studenten, die im März Geburtstag haben

$X \sim Bin(n, \pi)$ mit $\pi \approx 1/12$ und n gesucht

$$E[X] = n \cdot p \Rightarrow n = \frac{E[X]}{p}$$

Schätzung für $E[X]$: Anz. beobachteter Studenten mit Geb. im März

$\Rightarrow \widehat{E[X]}$

$\Rightarrow$ Schätzung für n: $\hat{n} = \dfrac{\widehat{E[X]}}{p}$

Variante 2: Keine Information über Struktur

$\rightarrow$ Capture: Sammle m Personen und markiere

$\rightarrow$ Recapture: Sammle unabhängig davon n Personen

(Personen aus „Capture" können auch vorkommen)

Wieviele markierte Personen in Recapture?

X: Anz. markierter Personen in Recapture

$X \sim Hyper(N, n, m)$; n, m bekannt, N unbekannt

$$E[X] = \frac{n \cdot m}{N} \Rightarrow N = \frac{n \cdot m}{E[X]}$$

Schätzung $\widehat{E[X]}$ = Anz. beobachteter, markierter Pers in Recap.

$\Rightarrow$ Schätzung $\hat{N} = \dfrac{n \cdot m}{\widehat{E[X]}}$

Maximum – Likelihood Methode

Bsp: $n = 600$ Personen erhalten neues Medikament

$x = 30$ haben als Nebenwirkung Kopfschmerzen

Wie gross ist wohl der Anteil Personen mit diesen Nebenwirkungen in der Gesamtbevölkerung (>600

Variante 1: Computer

π	...	0,03	0,04	0,05	0,06	0,07	...
$P(X=30)$		0,002	0,036	0,075	0,042	0,010	

$\Rightarrow \hat{\pi} = 0{,}05$

Variante 2: Analytisch

$$P(X=x) = \binom{n}{x} \pi^x (1-\pi)^{n-x} =: f(\pi) \quad (\text{„likelihood“})$$

Analysis: $f(\pi)$ max $\Leftrightarrow$ $\log(f(\pi))$ max

$$\log P(X=x) = \log\left[\binom{n}{x}\pi^x(1-\pi)^{n-x}\right] = \quad (\text{„log-likelihood“})$$

$$= \log\binom{n}{x} + x\log\pi + (n-x)\log(1-\pi) =: \text{A}$$

$$\frac{d}{d\pi} P(X=x) = 0 \Leftrightarrow \frac{d}{d\pi}\log\big(P(X=x)\big) = 0$$

$$\frac{d}{d\pi}\log(P(X=x)) = \frac{d}{d\pi}\text{A} = 0 + \frac{x}{\pi} + \frac{n-x}{1-\pi}(-1) =$$

$$= \frac{x}{\pi} - \frac{n-x}{1-\pi} \overset{!}{=} 0$$

$$\Rightarrow \frac{x}{\pi} = \frac{n-x}{1-\pi} \Rightarrow x - x\pi = n\pi - x\pi$$

$$\Rightarrow \pi = \frac{x}{n} = \frac{30}{600} = 5\%$$