ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DMATH
Department of Mathematics

Diskrete Wahrscheinlichkeitsverteilungen

für D-UWIS, D-ERDW, D-USYS und D-HEST – SS15

Warum braucht es Wahrscheinlichkeitsverteilungen?

Essentially,
all models are wrong,
but some are useful.

- George E.P. Box

"*Übliche*" Verteilung für eine Aufgabe → Eigenschaften in Büchern/Software vorhanden → Typische Probleme einfach lösbar

Verteilungszoo – Diskrete W'keitsverteilungen

Lernziele heute

- Diskrete Verteilungen
- Parameterschätzung

Hausaufgaben

- Skript: Kapitel 3.1 – 3.2.1 lesen
- Serie 4 lösen
- Quiz 4 bearbeiten
- etutoR 3 anschauen

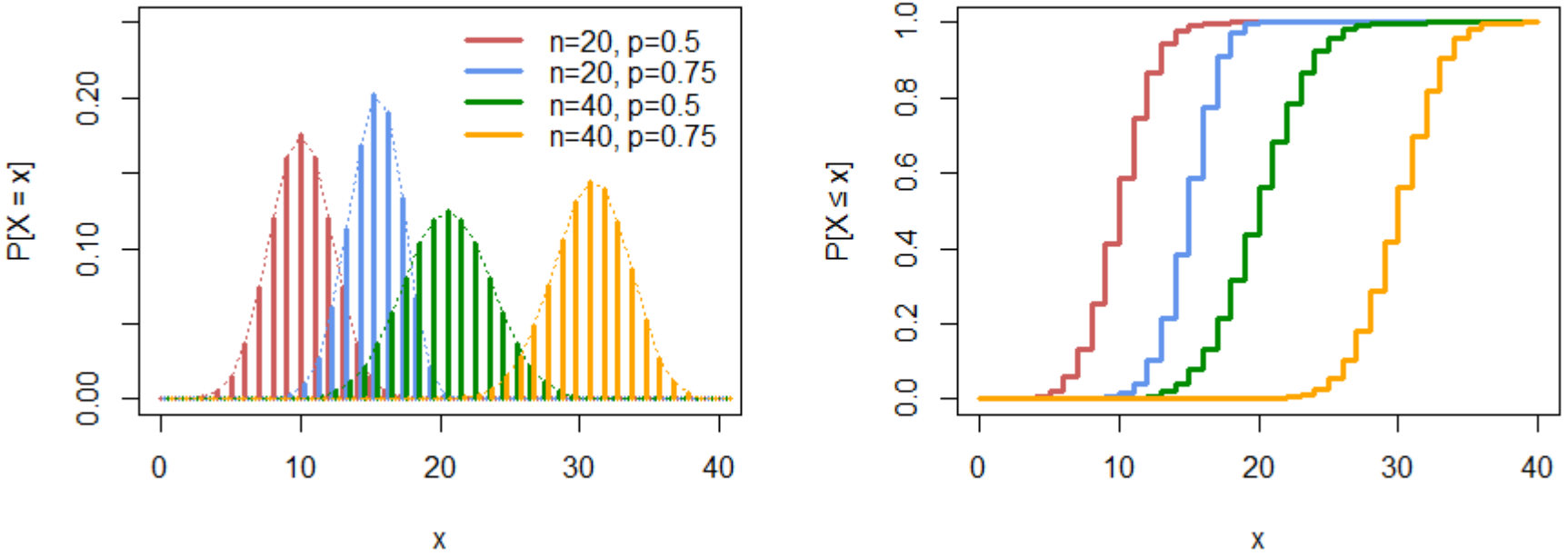
Binomialverteilung – $Bin(n, \pi)$

- Situation
  - Kaufe n Lose in einer Tombola
  - Alle Lose haben die gleiche Gewinnwahrscheinlichkeit
  - Lose sind unabhängig voneinander
- Zufallsvariable X: Anzahl Gewinne unter n Losen

➢ $X \sim Bin(n, \pi)$

- Binomialkoeffizient: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$
- $P[X = x] = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x \in \{0, 1, 2, \dots, n\}$
- $E(X) = n \cdot \pi, Var(X) = n \cdot \pi \cdot (1-\pi)$

Binomialverteilung – $Bin(n, \pi)$

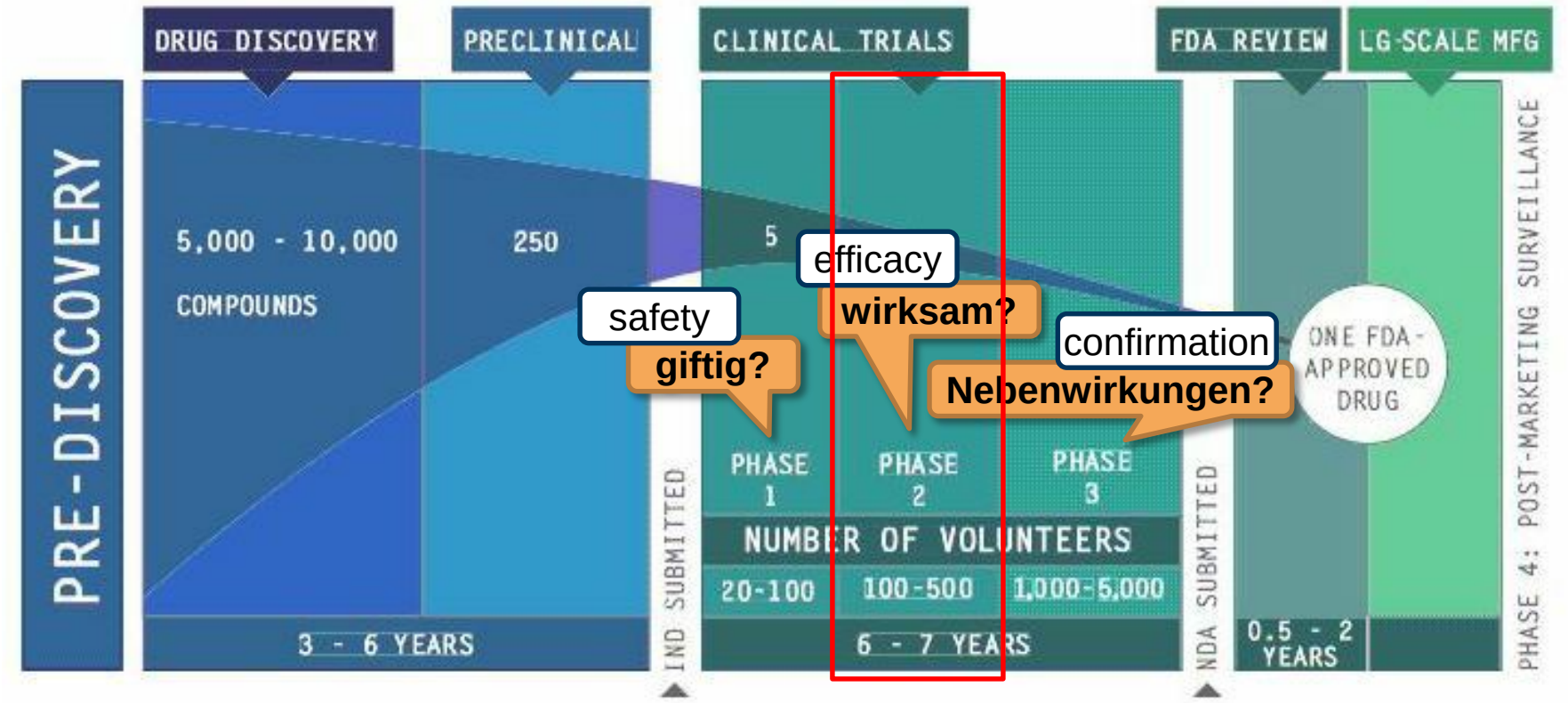


- Beispiel: **LD10**, $X \sim Bin(20, 0.1)$

$$P[X = 2] = \binom{20}{2} \cdot 0.1^2 \cdot 0.9^{18} \approx 0.285$$

$$E(X) = n \cdot \pi = 20 \cdot 0.1 = 2$$

Beispiel «Klinische Studie»



- **Lose**: Alle denkbaren Patienten
- n **gezogene Lose**: Patienten in der Studie
- **Gewinn**: Patient wird gesund
- π: Anteil aller denkbaren Patienten, die gesund werden

Beispiel «Clinical Trial – Phase 2»

- Hersteller behauptet: Medikament wirkt in 80% der Fällen
- Phase 2 Studie: von 100 Patienten werden nur 73 gesund
  - Ist das, bei einer Heilungsw'keit 80%, plausibel?
- X: Anzahl geheilter Patienten
- Falls Hersteller recht hat:

$$X \sim Bin(n = 100, \pi = 0.8)$$

- Wie testen wir die Behauptung «$\pi = 0.8$»?

- Versuch 1: $P[X = 73] = 0.022$

Überzeugt?

Exkurs: Größe einer Stichprobe

- $X \sim Bin(n = 100, \pi = 0.8)$
- Angenommen, wir haben genau $n \cdot \pi = 80$ Genesungen gesehen; wir sollten dem Hersteller also glauben

	n=100	n=1000	n=10'000	n=100'000
$P(X = n\pi)$	0.10	0.03	0.01	0.003

  - $P[X = 73]$ ist keine gute Kennzahl, weil die W'keit für jede beliebige Zahl klein wird, wenn man nur genug Beobachtungen hat!

	n=100	n=1000	n=10'000	n=100'000
$P(X \leq n\pi)$	0.54	0.51	0.504	0.501

  - $P[X \leq 73]$ ist eine gute Kennzahl; sie ist unabhängig von der Stichprobengrösse und leichter zu interpretieren.

⇨ P-Wert – W'keit für eine Beobachtung oder etwas noch Extremeres

Beispiel «Clinical Trial – Phase 2»

- Hersteller behauptet: Medikament wirkt in 80% der Fällen
- Phase 2 Studie: von 100 Patienten werden nur 73 gesund
  - Ist das, bei einer Heilungsw'keit 80%, plausibel?
- X: Anzahl geheilter Patienten
- Falls Hersteller recht hat:

$$X \sim Bin(n = 100, \pi = 0.8)$$

- Wie testen wir die Behauptung «$\pi = 0.8$»?

- Versuch 2: $P[X \leq 73] = 0.056$

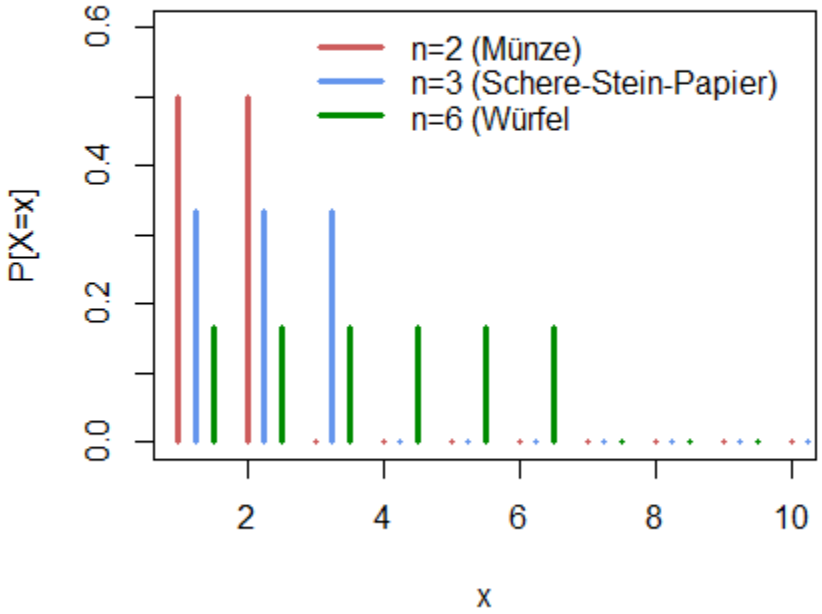
Uniforme Verteilung - $Unif(n)$

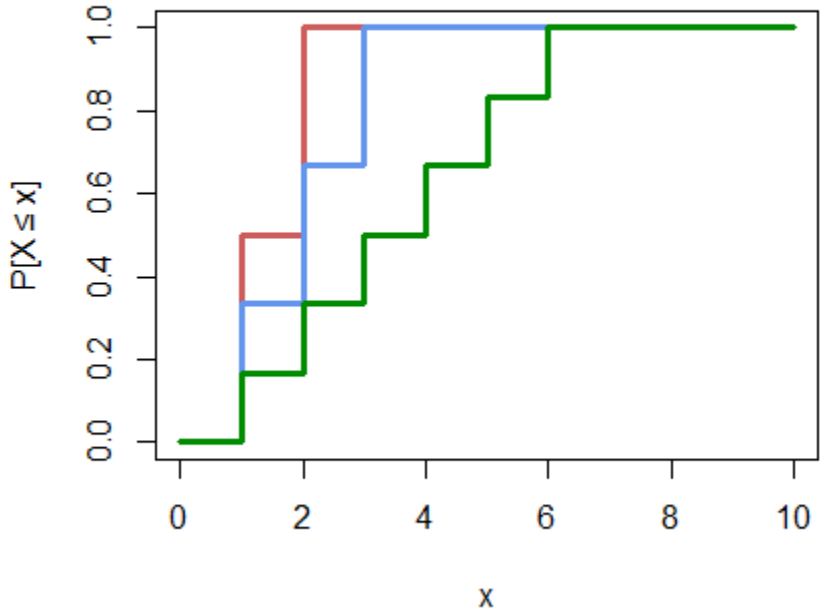
- Situation
  - Ziehe eine Zahl aus {1, 2, 3, …, n}
  - Alle Zahlen haben die gleiche Wahrscheinlichkeit
- Zufallsvariable X: Gezogene Zahl

➢ $X \sim Unif(n)$

- $P[X = x] = \frac{1}{n}, \quad x \in \{1, 2, \dots, n\}$
- $E(X) = \frac{n+1}{2}, Var(X) = \frac{(n+1)(n+2)}{12}$

Uniforme Verteilung - $Unif(n)$





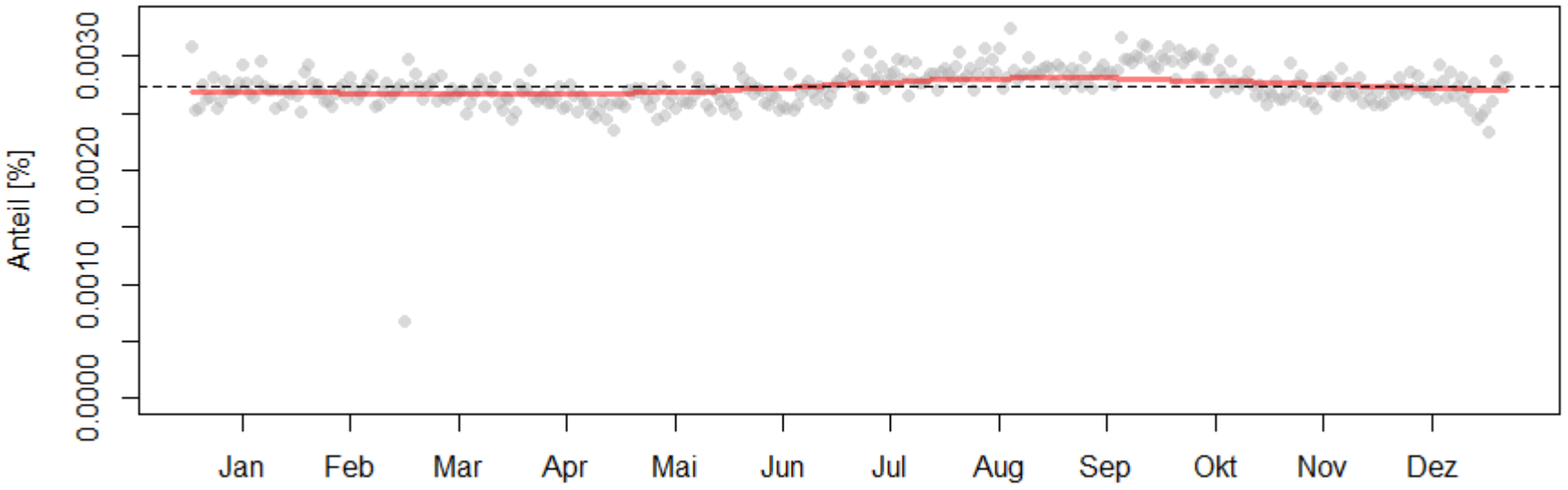
- Beispiel: **Würfel**, $X \sim Unif(6)$

$$P[X = x] = \frac{1}{6}$$

$$E(X) = \frac{6+1}{2} = 3.5$$

Sind Geburtstage uniform verteilt?

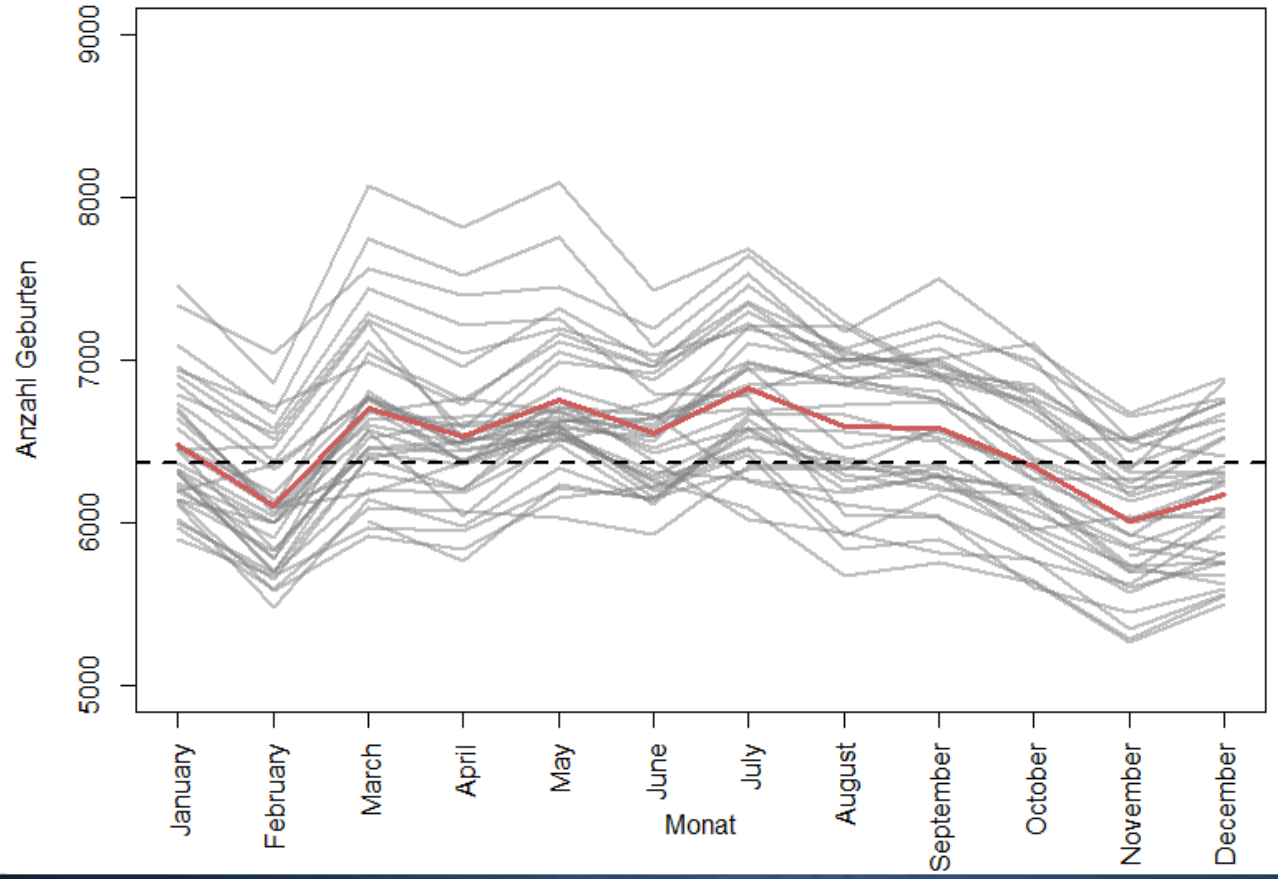
- Geburtstage aus einer Lebensversicherung 1981 – 1994



- In grober Näherung schon!

Sind Geburtstage uniform verteilt?

- Was wenn aus dem gleichen Land (Schweiz, 1973-2013)



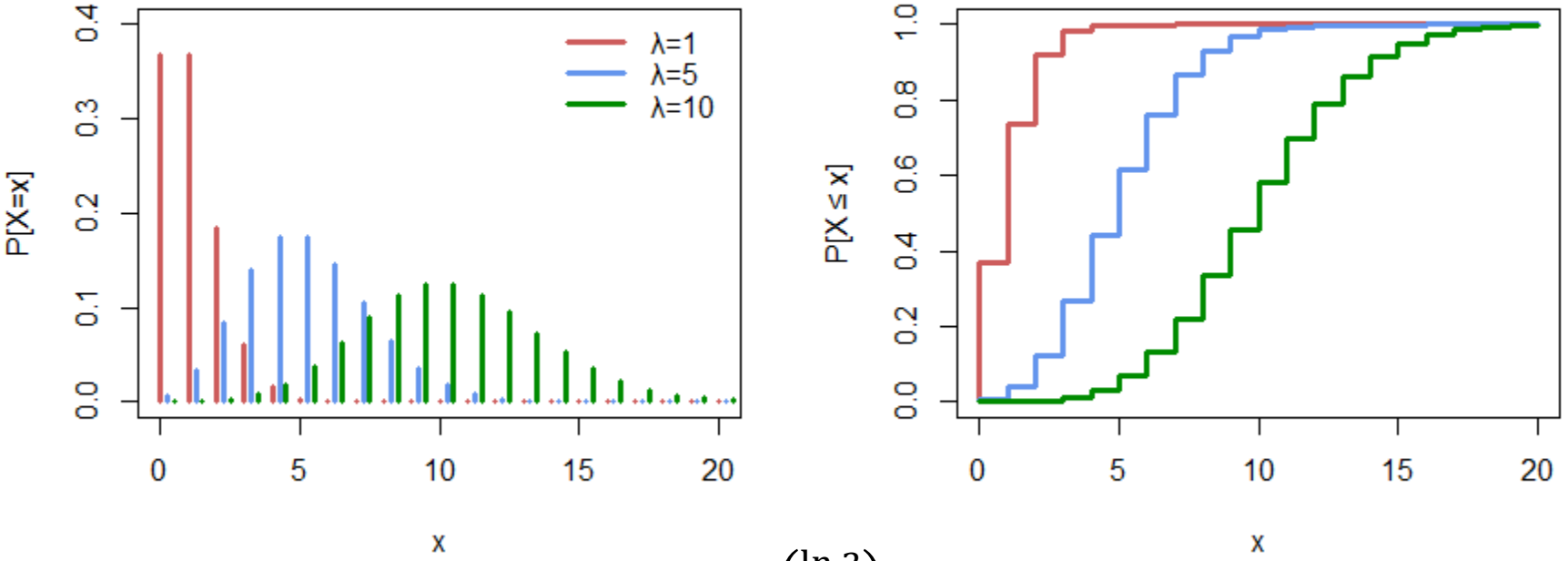
Poissonverteilung - $Pois(\lambda)$

- Situation
  - Seltene Ereignisse werden in einem vorgegebenen Zeitraum gezählt
- Zufallsvariable X: Anzahl beobachteter Ereignisse

➢ $X \sim Pois(\lambda)$

- $P[X = x] = \frac{\lambda^x}{x!}\exp(-\lambda)\,, \quad x \in \{0,1,\dots,\infty\}$
- $E(X) = \lambda, Var(X) = \lambda$

Poissonverteilung - $Pois(\lambda)$



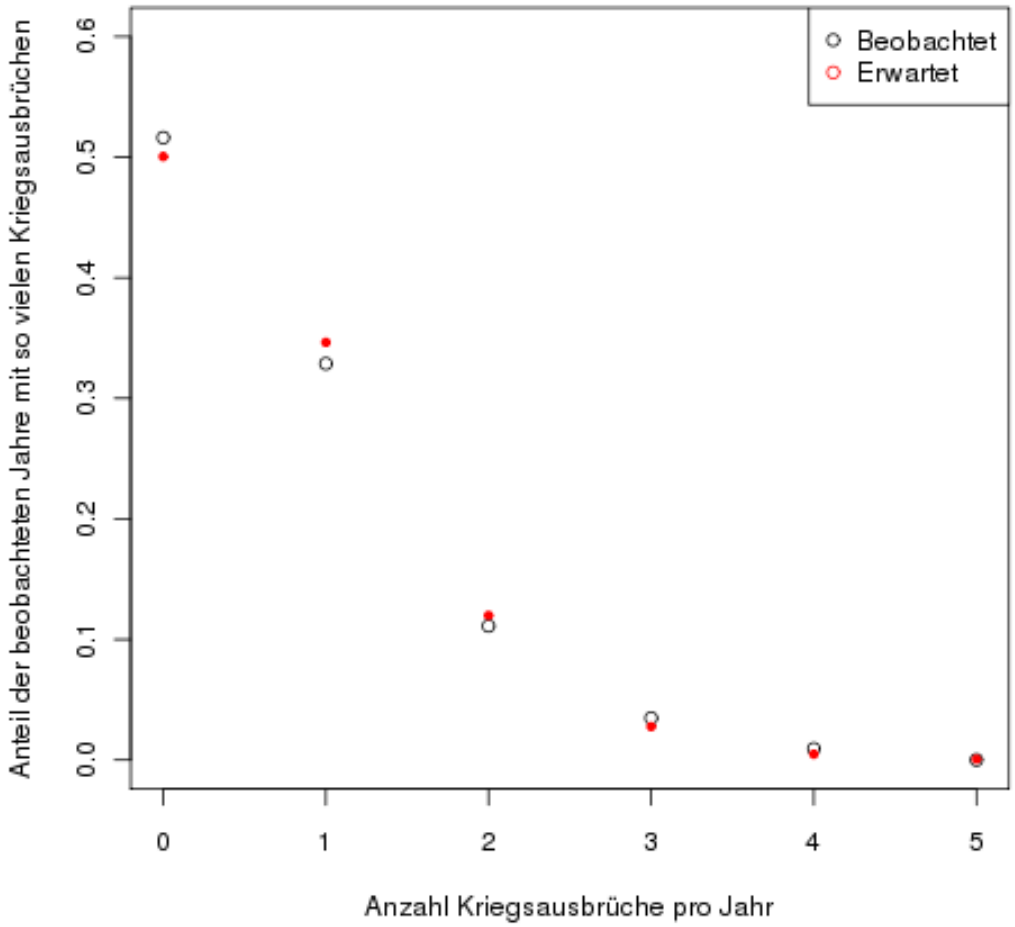
- Beispiel: **Cäsium-137**, $X \sim Pois\left(\frac{\ln 2}{27}\right) \approx Pois(0.026)$
  - ...Rate pro Jahr, d.h. $8.2 \times 10^{-10} \frac{1}{s}$

$$P[X = 1] = \frac{0.026^1}{1!} e^{-0.026} \approx 0.025$$

$$E(X) = \lambda = 0.026$$

  - $1\mu g$ ^{137}Cs enthält 10^{15} Nuklei, d.h. $\mu = N \cdot p = 8.2 \times 10^5$ Zerfalle pro Sekunde

Beispiel «Anzahl Kriege p.a. poissonverteilt?» (1500 – 1930, weltweit)



Exkurs: Besonderheit der Poissonverteilung

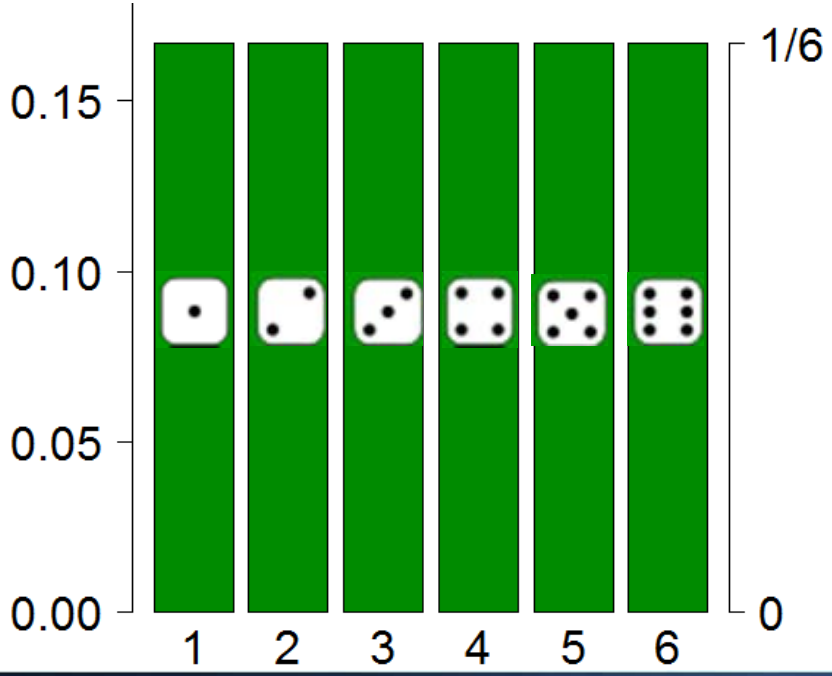
- Angenommen:
  - $X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2)$
  - X, Y sind unabhängig
- Bilde neue Zufallsvariable: $Z = X + Y$

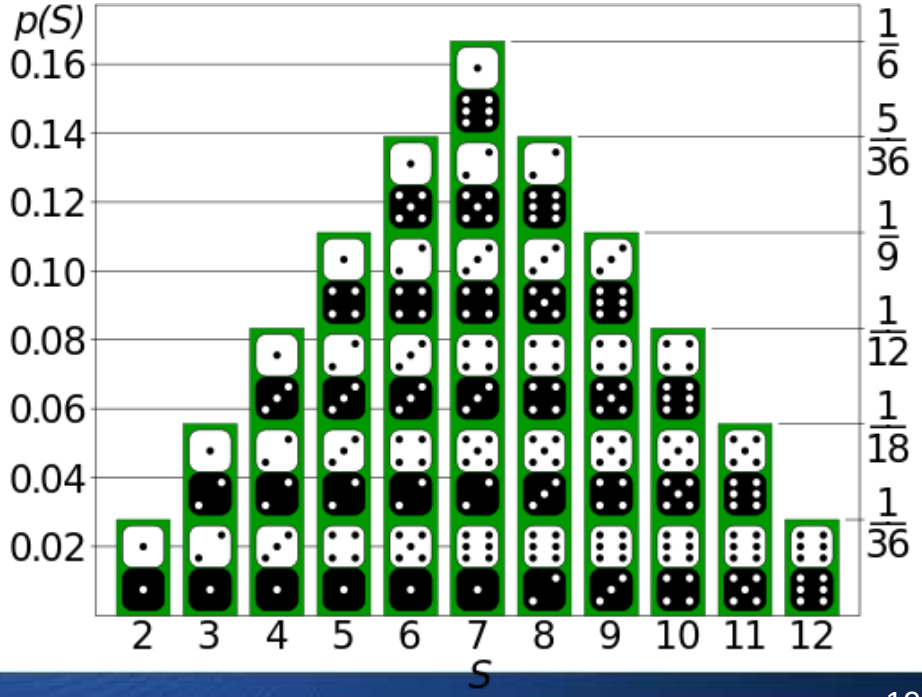
$$\Rightarrow Z \sim Pois(\lambda_1 + \lambda_2)$$

Das gilt normalerweise **nicht**!

Normalerweise: Summe von zwei Verteilungen gibt eine neue Verteilung

- Bsp: $X \sim Unif(\{1,2,3,4,5,6\}), Y \sim Unif(\{1,2,3,4,5,6\})$
  X, Y sind unabhängig
- $S = X + Y$ ist nicht uniform verteilt (Augensumme 2 ist selten, Augensumme 7 ist häufig)





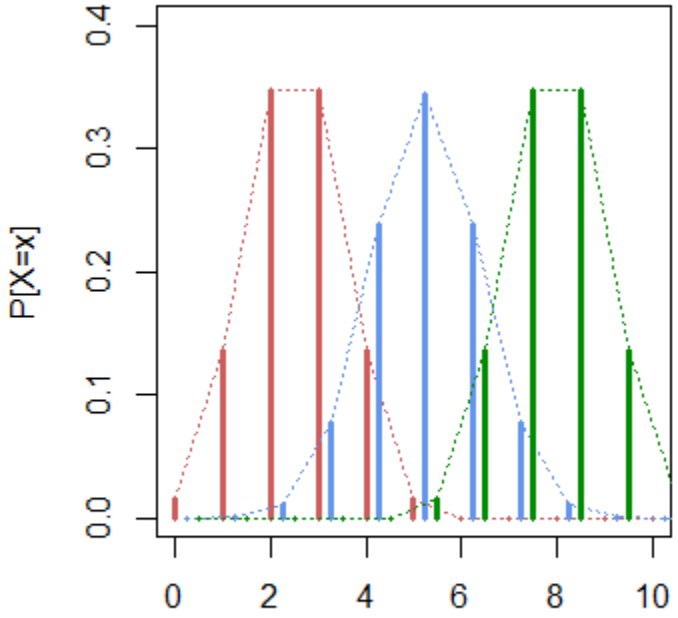
Hypergeometrische Verteilung - $Hyper(N, n, m)$

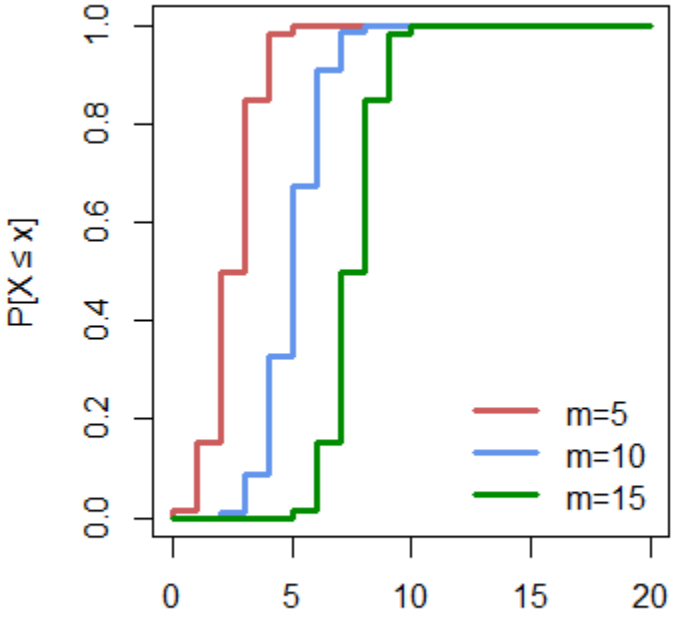
- Situation
  - Urne mit N Kugeln, m davon weiss und $N - m$ schwarz
  - Ziehe n Kugeln (ohne zurücklegen)
  - Wieviele Kugeln sind weiss?
- Zufallsvariable X: gezogene weisse Kugeln

➢ $X \sim Hyper(N, n, m)$

- $P[X = x] = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \quad x \in \{0,1, \dots, \min(m, n)\}$
- $E(X) = \frac{(n \cdot m)}{N}$, $Var(X)$ ziemlich kompliziert siehe

Hypergeometrische Verteilung - $Hyper(N, n, m)$





- Beispiel: **Urnenmodell**, $X \sim Hyper(20, 3, 7)$
  - 20 Kugeln, davon 7 markiert, 3 werden ohne zurücklegen gezogen

$$P[X = 1] = \frac{\binom{7}{1}\binom{13}{2}}{\binom{20}{3}} \approx 0.48, \qquad E(X) = \frac{3 \cdot 7}{20} = 1.05$$

Beispiel «Clinical Trial – Phase 3»

- Doppel-blinde, randomisierte Studie

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	**25**	**20**	**45**

- Falls Medikament keine Wirkung hat: Es gibt 24 Personen, bei denen unabhängig von der Gruppenzuteilung fest steht, dass sie gesund werden

Beispiel «Clinical Trial – Phase 3»

- ZV X: Anzahl geheilter Patienten in Medikamentengruppe
- unter $\mathcal{H}_0$ (keine Wirkung): $X \sim Hyper(N = 45, m = 24, n = 25)$

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	**25**	**20**	**45**

- Ist es dann plausibel 15 geheilte Patienten in der Medikamentengruppe zu beobachten?

$$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.76 = \boxed{0.24}$$

- Wenn nicht wirksam, durchaus möglich 15 oder mehr…

Momentenmethode

- Beispiel «Bachforellenzucht»
  - in 100 zufällig ausgewählten Bächen werden Brütlinge ausgesetzt
  - nach knapp einem Jahr werden in 54 Standorten grosse Jungtiere gefunden
  - Wie gross ist die Wahrscheinlichkeit, dass eine solche Wildzucht erfolgreich ist?

- X: Anzahl Jungtiere, die überleben

$$X \sim Bin(n = 100, \pi = ?)$$

  - Beobachtung: $x = 54$

- Momentenmethode um π zu schätzen:

$$E(X) = n \cdot \pi; \;\; E(X) \approx x = 54 \rightarrow x \approx n \cdot \pi \rightarrow \pi \approx \frac{x}{n} = 0.54$$

Momentenmethode

- Beispiel «Capture-Recapture»
  - Wie gross ist eine Population, von der wir ***gar nichts*** weiter wissen?
- Lincoln-Peterson Methode:
  - Fange m zufällige Tiere, markiere, lasse wieder laufen
  - Fange n zufällige Tiere
  - ZV X: Anzahl markierter Tiere im zweiten Fang
- $X \sim Hyper(N, n, m)$, wobei N die Grösse der Pop. ist; x markierte Tiere im zweiten Fang
- Idee: «Erwartungswert ≈ Beobachtung»
  - $E(X) = \frac{n \cdot m}{N} \approx x \rightarrow N \approx \frac{n \cdot m}{x}$
- Ungenau, aber OK für richtige Grössenordnung

engl. **grösste Wahrscheinlichkeit**

Maximum-Likelihood Methode

- Bsp: $n=600$ Personen erhalten neues Medikament; $x=30$ haben als Nebenwirkung Kopfschmerzen
- Wie gross ist der Anteil Personen mit diesen Nebenwirkungen in der Gesamtbevölkerung (>600)?

- Binomialverteilung:
  - X: Anzahl Personen mit Kopfschmerzen
  - $X \sim Bin(n = 600, \pi)$
  - $P[X = 30] = \binom{600}{30} \pi^{30} (1-\pi)^{570}$

- Maximum-Likelihood Estimate (MLE) $\hat{\pi}$ für π, ist der Wert, der $P[X = 30]$ maximiert.

Maximum-Likelihood Methode

- mit dem Computer:
  - berechne $P[X = 30]$ für verschiedene Werte von π:

π	…	0.03	0.04	0.05	0.06	0.07	…
$P[X = 30]$		0.002	0.036	0.075	0.042	0.010	

Maximum
$\hat{\pi} \approx 0.05$

- analytisch:
  - $P[X = x] = \binom{n}{x}\pi^x(1-\pi)^{n-x} =: f(\pi)$ «likelihood»
  - Analysis: Finde π, sodass $f(\pi)$ maximal ist (siehe Skript S. 25)
- Ergebnis: $\hat{\pi} = \frac{x}{n} = \frac{30}{600} = 0.05$

Zusammenfassung

- Diskrete Verteilungen:
  - Binomial, Uniform, Poisson, Hypergeometrisch
- Parameterschätzung:
  - Momentenmethode, Maximum-Likelihood Estimation

Hausaufgaben

- Skript: Kapitel 3.1 – 3.2.1 lesen
- Serie 4 lösen
- Quiz 4 bearbeiten
- etutoR 3 anschauen