Kapitel U

Kombinatorik und Näherungen

Stochastik ist die Kunst,
mit unsicheren Ereignissen sicher zu rechnen.

Sprichwort

On voit, par cet Essai, que la théorie des probabilités
n'est, au fond, que le bon sens réduit au calcul.

Pierre-Simon de Laplace (1749–1827)

Inhalt dieses Kapitels U

Motivation und Zielsetzung

Viele Zufallsexperimente beschreiben wir durch **Standardmodelle**:

- Produktraum bei unabhängig durchgeführten Experimenten,
- Ziehung von Losen (mit/ohne Zurücklegen, mit/ohne Reihenfolge),
- Schubfachmodelle (mit geeigneten zusätzlichen Bedingungen).

Hier hilft geschicktes systematisches Abzählen, also **Kombinatorik**.
Die entstehenden Wahrscheinlichkeitsverteilungen sind zwar vielfältig, haben aber Gemeinsamkeiten, die Sie kennen und nutzen sollten.
Drei besonders wichtige werden wir in diesem Kapitel behandeln:

- die hypergeometrische Verteilung $H(N, K, n)$ für Stichproben,
- die Binomialverteilung $B(n, t)$ für Trefferanzahlen,
- die Poisson–Verteilung $P(\lambda)$ als Näherung.

Viele Rechnungen vereinfachen wir durch hilfreiche **Näherungen**:

- Für große Stichproben ($N \to \infty$) gilt $H(N, K, n) \approx B(n, K/N)$.
- Gesetz der kleinen Zahlen: $B(n, t) \approx P(nt)$ für kleines t.
- Lokaler Grenzwertsatz: $B(n, t) \approx N(\mu, \sigma^2)$, siehe Kapitel V.

Vorgehensweise

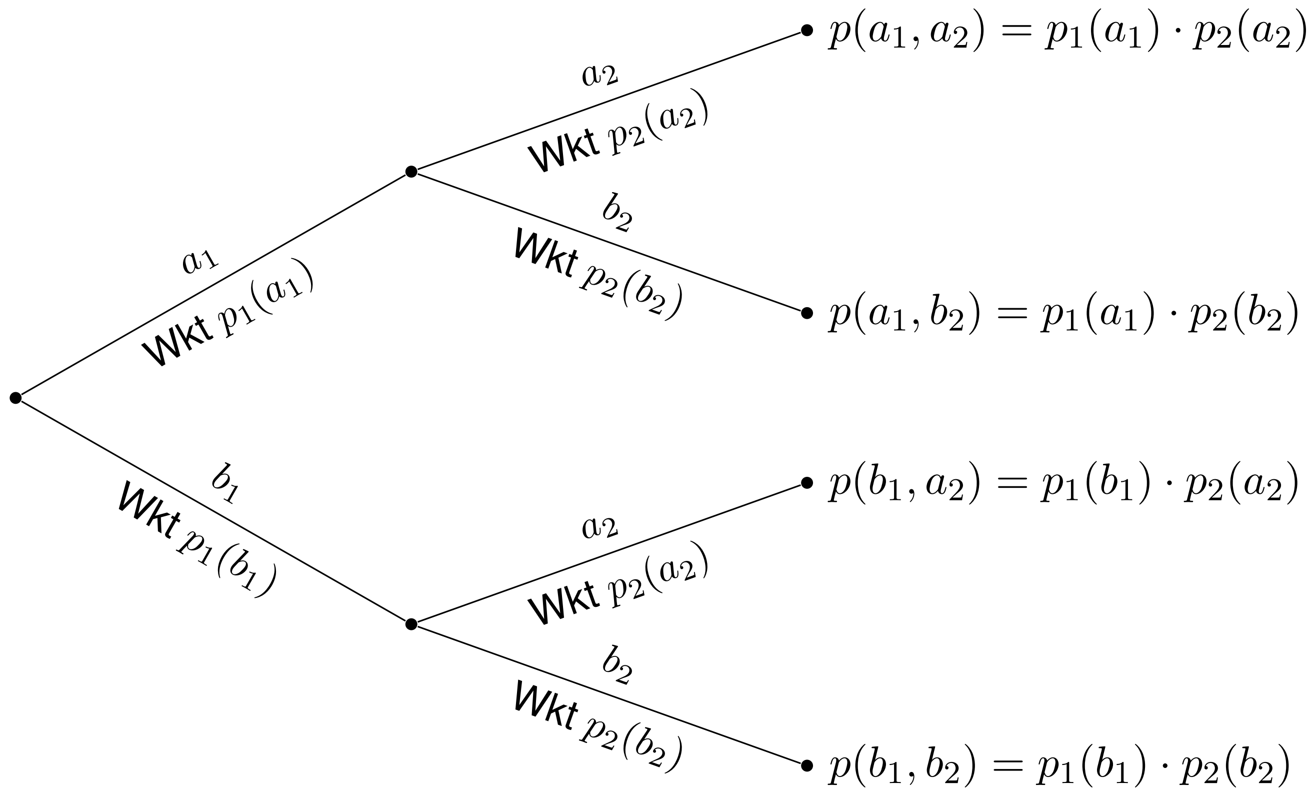
Abschätzung von Risiken ist eine typische und wichtige Anwendung der Wahrscheinlichkeitsrechnung in den Ingenieurwissenschaften.
Ich beginne daher mit Produktexperimenten als Modell unabhängiger Ereignisse und ersten Anwendungen auf Ausfallwahrscheinlichkeiten.
Nach Einführung der Urnenmodelle kommen wir zu Stichproben und damit zu hypergeometrischen, Binomial- und Poisson–Verteilungen.

Auf Grundlage eines stochastischen Modells können Sie die gesuchten Wkten exakt berechnen, d.h. aus den angenommenen Daten ableiten.
Oft ist die exakte Berechnung zu aufwändig, und Sie begnügen sich mit einer guten Näherung. Nützliche Tricks sollten Sie kennen und nutzen!

Grenzwertsätze formuliere ich möglichst mit expliziten Fehlerschranken.
Dies ermöglicht, die üblichen Approximationen nicht nur blind oder naiv nach Gefühl anzuwenden, sondern bequem ihre Güte abzuschätzen.
Ob neben der Berechnung der Approximation auch eine Abschätzung des Fehlers notwendig ist, hängt vom geforderten Sicherheitsniveau ab.
Die Werkzeuge hierzu sind jedenfalls vorhanden und nicht schwierig.

Baummodell für unabhängige Experimente

Durchführung von zwei **unabhängigen Experimenten**:
$(\Omega_1 = \{a_1, b_1, \dots\}, \mathbf{P}_1)$ und $(\Omega_2 = \{a_2, b_2, \dots\}, \mathbf{P}_2)$.



Produkt von Wahrscheinlichkeitsräumen

Definition U1A: Produktwahrscheinlichkeit, zunächst diskret

Seien $(\Omega_1, \mathbf{P}_1)$ und $(\Omega_2, \mathbf{P}_2)$ diskrete Wahrscheinlichkeitsräume.
Der Produktraum ist das **kartesische Produkt**

$$\Omega = \Omega_1 \times \Omega_2 = \{ (\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \; \omega_2 \in \Omega_2 \}.$$

Die **Produktwahrscheinlichkeit** $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2 : \Omega \to \mathbb{R}$ ist

$$\mathbf{P}\big(\{(\omega_1, \omega_2)\}\big) = \mathbf{P}_1\big(\{\omega_1\}\big) \cdot \mathbf{P}_2\big(\{\omega_2\}\big).$$

Per Additivität setzt sich dies auf alle Ereignisse $A \subseteq \Omega$ fort zur Formel

$$\mathbf{P}(A) = \sum_{(\omega_1, \omega_2) \in A} \mathbf{P}_1\big(\{\omega_1\}\big) \cdot \mathbf{P}_2\big(\{\omega_2\}\big).$$

Der **Produktraum** $(\Omega, \mathbf{P}) = (\Omega_1, \mathbf{P}_1) \times (\Omega_2, \mathbf{P}_2) = (\Omega_1 \times \Omega_2, \mathbf{P}_1 \otimes \mathbf{P}_2)$ ist selbst ein diskreter WRaum, denn $\mathbf{P}$ ist additiv und erfüllt $\mathbf{P}(\Omega) = 1$.

Beispiel: Zweimaliges unabhängiges Würfeln, siehe T124.

Ausfallwahrscheinlichkeit bei 2 Teilen

Aufgabe: Eine Maschine bestehe aus zwei unabhängigen Teilen. Ihre Ausfallwkten seien $p_1 = 0.1 = 10\%$ und $p_2 = 0.2 = 20\%$.

(1) Konstruieren Sie als Modell hierfür explizit einen WRaum $(\Omega, \mathbf{P})$.

(2) Mit welcher Wahrscheinlichkeit funktionieren beide? fällt eins aus?

Lösung: (1) Stochastisches Modell $(\Omega_1, \mathbf{P}_1)$ für Teil T_1 allein:

$$\Omega_1 = \{0,1\} \quad \text{mit} \quad 0 = \text{„}T_1 \text{ fällt aus“}, \qquad 1 = \text{„}T_1 \text{ funktioniert“}$$
$$\mathbf{P}_1(\{0\}) = p_1 = 0.1, \quad \mathbf{P}_1(\{1\}) = \overline{p_1} = 0.9$$

Stochastisches Modell $(\Omega_2, \mathbf{P}_2)$ für Teil T_2 allein:

$$\Omega_2 = \{0,1\} \quad \text{mit} \quad 0 = \text{„}T_2 \text{ fällt aus“}, \qquad 1 = \text{„}T_2 \text{ funktioniert“}$$
$$\mathbf{P}_2(\{0\}) = p_2 = 0.2, \quad \mathbf{P}_2(\{1\}) = \overline{p_2} = 0.8$$

Wegen Unabhängigkeit ist der Produktraum hier das richtige Modell: Wir betrachten also den Produktraum $(\Omega, \mathbf{P}) = (\Omega_1 \times \Omega_2, \mathbf{P}_1 \otimes \mathbf{P}_2)$, also $\Omega = \{00, 01, 10, 11\}$ mit Produktwahrscheinlichkeit $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$.

Ausfallwahrscheinlichkeit bei 2 Teilen

(2) Wegen der vorausgesetzten Unabhängigkeit der beiden Teile ist die Produktwahrscheinlichkeit hier das richtige Modell:

T_1 fällt aus und T_2 fällt aus:	$\mathbf{P}(\{00\}) = p_1 \cdot p_2 = 0.1 \cdot 0.2 = 0.02$
T_1 fällt aus und T_2 funktioniert	$\mathbf{P}(\{01\}) = p_1 \cdot \overline{p_2} = 0.1 \cdot 0.8 = 0.08$
T_1 funktioniert und T_2 fällt aus:	$\mathbf{P}(\{10\}) = \overline{p_1} \cdot p_2 = 0.9 \cdot 0.2 = 0.18$
T_1 funktioniert und T_2 funktioniert:	$\mathbf{P}(\{11\}) = \overline{p_1} \cdot \overline{p_2} = 0.9 \cdot 0.8 = 0.72$

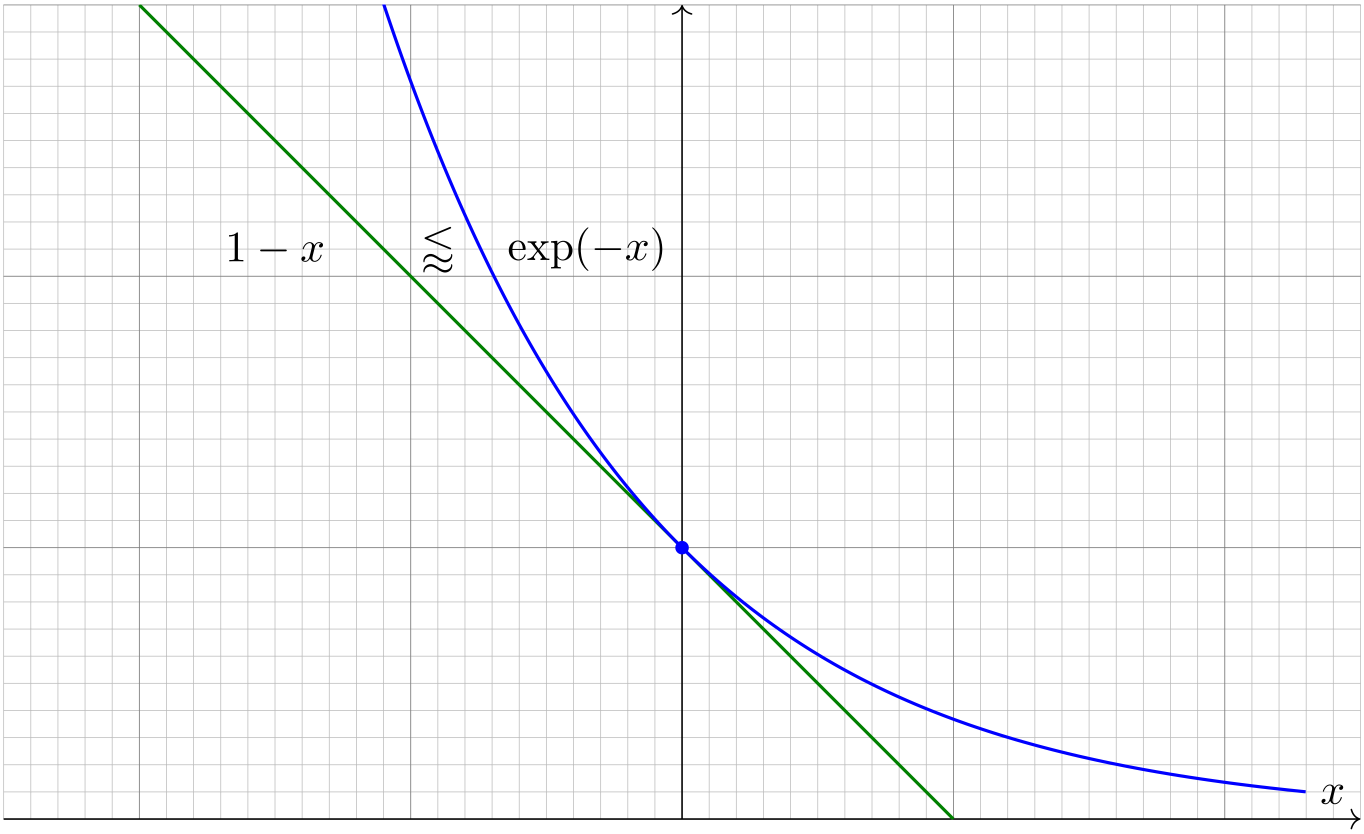
Mit Wahrscheinlichkeit 0.72 funktionieren beide Teile.
Mit Wahrscheinlichkeit 0.28 fällt (mindestens) eins aus.

☺ Das explizite Modell (1) ist detailliert; das kann helfen. Oft genügt die implizite Schreibweise wie in (2); sie ist schnell und bequem.

☺ Es liegt immer ein Modell zu Grunde; im Zweifel führt man es aus. Zur Illustration notiere ich hier (übertrieben) sorgfältig jeden Schritt.

⚠ Verwechseln Sie nicht die gesuchte Wkt und ihre Gegenwkt! Solche Aufgaben sind leicht, wenn Sie sorgfältig arbeiten.

Erinnerung an die Taylor–Entwicklung $\exp(-x) \gtrapprox 1 - x$



Für alle x gilt $1 - x \le \exp(-x)$, und für $x \approx 0$ zudem $1 - x \approx \exp(-x)$.
Wir schreiben dies im Folgenden kurz und prägnant $1 - x \lessapprox \exp(-x)$.

Erinnerung an die Taylor–Entwicklung $\exp(-x) \gtrapprox 1 - x$

Dieses graphische Argument können und wollen wir präzisieren:

Lemma U1B: Exponentialnäherung

(1) Für alle $x \in \mathbb{R}$ gelten die Ungleichungen

$$\exp(x) \ge 1 + x \qquad \text{und} \qquad \exp(-x) \ge 1 - x.$$

(2) Für x um 0 sind dies die Taylor–Näherung bis zur ersten Ordnung:

$$\exp(x) \gtrapprox 1 + x \qquad \text{und} \qquad \exp(-x) \gtrapprox 1 - x.$$

Aufgabe: Beweisen Sie diese Ungleichungen durch Kurvendiskussion.

Lösung: (1) Für $f : \mathbb{R} \to \mathbb{R} : x \mapsto \exp(x)$ gilt $f(0) = 1$ und $f' = f$.
Die Tangente im Punkt $x = 0$ ist somit $x \mapsto f(0) + f'(0)\,x = 1 + x$.
Wegen $f'' = f' = f > 0$ ist f konvex, liegt also oberhalb der Tangente.
Das beweist die Ungleichung $\exp(x) \ge 1 + x$, ebenso $\exp(-x) \ge 1 - x$.

(2) Anschaulich: Für $x \approx 0$ gilt $\exp(x) \approx 1 + x$, ebenso $\exp(-x) \approx 1 - x$.
Der Fehler ist höherer Ordnung, kurz $O(x^2)$, genauer siehe Restterm!
Wir schreiben dies im Folgenden kurz und prägnant $1 - x \lessapprox \exp(-x)$.

Ausfallwahrscheinlichkeit bei n Teilen

Ein Projekt habe n **single points of failure** / Teilschritte $T_1, \dots, T_n$. Zum Gesamterfolg muss jeder dieser Teilschritte erfolgreich sein.

Eine Maschine bestehe aus n **unabhängigen** Teilen $T_1, \dots, T_n$. Jedes Teil T_k hat eine gewisse Ausfallwahrscheinlichkeit $p_k \in [0, 1]$.

Aufgabe: Mit welcher Wahrscheinlichkeit funktionieren alle Teile? Mit welcher Wahrscheinlichkeit fällt (mindestens) ein Teil aus?

Lösung: Für diese Wkten gilt exakt bzw. näherungsweise:

$$\begin{aligned} \mathbf{P}(\text{Alle Teile funktionieren}) &= (1-p_1)\cdots(1-p_n) \\ &\lessapprox \mathrm{e}^{-p_1}\cdots\mathrm{e}^{-p_n} = \mathrm{e}^{-(p_1+\dots+p_n)}. \\ \mathbf{P}(\text{Mindestens eins fällt aus}) &= 1-(1-p_1)\cdots(1-p_n) \\ &\gtrapprox 1-\mathrm{e}^{-(p_1+\dots+p_n)}. \end{aligned}$$

☺ Diese Näherung ist praktisch für $p_1, \dots, p_n$ klein und n groß: Statt einem $n$fachen Produkt genügt uns eine eine $n$fache Summe! Dies ist die Poisson–Wkt $P(\lambda)(0)$ mit $\lambda = p_1 + \dots + p_n$, siehe U3C.

Ausfallwahrscheinlichkeit bei n Teilen

☺ Ein wichtiger Spezialfall sind **unabhängige identische** Teile.

Satz U1C: Ausfallwahrscheinlichkeiten

Von n unabhängigen Teilen habe jedes Ausfallwkt $p \in [0, 1]$. Dann gilt:

$$\boxed{\begin{aligned} \mathbf{P}(\text{Alle Teile funktionieren}) &= (1-p)^n &&\lessapprox \mathrm{e}^{-np} \\ \mathbf{P}(\text{Mindestens eins fällt aus}) &= 1-(1-p)^n &&\gtrapprox 1-\mathrm{e}^{-np} \end{aligned}}$$

Aufgabe: Bei einer Ausfallwahrscheinlichkeit $p = 0.01$, mit welcher Wahrscheinlichkeit funktionieren alle Bauteile?

Bei 50 Bauteilen? Wahrscheinlichkeit $0.99^{50} \approx \mathrm{e}^{-0.5} \approx 0.61$

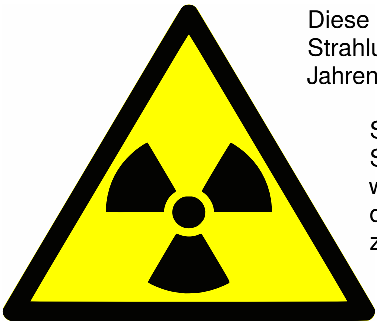
Bei 100 Bauteilen? Wahrscheinlichkeit $0.99^{100} \approx \mathrm{e}^{-1} \approx 0.37$

Bei 200 Bauteilen? Wahrscheinlichkeit $0.99^{200} \approx \mathrm{e}^{-2} \approx 0.14$

☺ Für den Ausfall von $1, 2, 3, \dots$ Teilen gilt exakt die Binomialverteilung $B(n, p)$ und als praktische Näherung die Poisson–Verteilung $P(np)$. Poissons Gesetz der kleinen Zahlen (U3E) liefert eine Fehlerschranke.

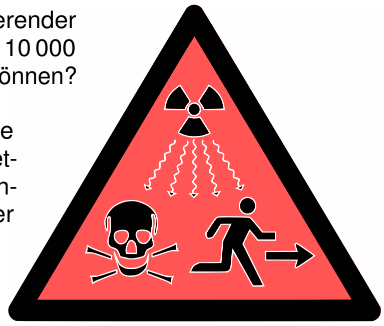
Wahrscheinlichkeit eines Reaktorunfalls

Wo steckt der Fehler? „Die Klausur bestehe ich mit 40% Wkt, bei drei Prüfungsversuchen also mit 120%iger Sicherheit.“



Diese Symbole warnen vor ionisierender Strahlung. Wird man sie auch in 10 000 Jahren noch richtig interpretieren können?

Schon heute werden diese Symbole falsch verstanden, etwa als „Propeller“ oder „Wundermedizin, die Tote wieder zum Leben erweckt.“



Aus der Presse: „In $14\,000$ Reaktorjahren sind 4 Unfälle der höchsten Stufe aufgetreten. Demnach tritt bei 143 europäischen Reaktoren in 30 Jahren ein solcher Unfall mit Wahrscheinlichkeit von über 100% auf.“

Dieses katastrophale Argument verläuft mutmaßlich so:

Wahrscheinlichkeit $p = \frac{4}{14000} \approx 0.029\%$ pro Reaktorjahr.

Betrachtete Laufzeit $n = 143 \cdot 30 = 4290$ Reaktorjahre.

Falsche Rechnung: $np = \frac{4}{14000} \cdot 4290 \approx 1.23$, also 123% Wkt.

Wahrscheinlichkeit eines Reaktorunfalls

Nach der Nuklearkatastrophe von Fukushima im März 2011 wurde über Risiken diskutiert. Der Physiker Bernard Laponche und der Ingenieur Benjamin Dessus schrieben in der französischen Tageszeitung *La Libération* zur Wahrscheinlichkeit von Reaktorunfällen folgendes:

Le parc actuel de réacteurs des centrales nucléaires cumule 14000 années-réacteur, ce qui correspond à environ 450 réacteurs fonctionnant durant trente et un ans. [. . .] Sur ce parc, cinq réacteurs ont connu un accident grave (un à Three Mile Island, un à Tchernobyl et trois à Fukushima), dont quatre sont des accidents majeurs (Tchernobyl et Fukushima). [. . .] La France compte actuellement 58 réacteurs en fonctionnement et l'Union européenne un parc de 143 réacteurs. Sur la base du constat des accidents majeurs survenus ces trente dernières années, la probabilité d'occurrence d'un accident majeur sur ces parcs serait donc de 50% pour la France et de plus de 100% pour l'UE. (*Accident nucléaire: une certitude statistique*, La Libération, 3.6.2011)

Diese kreative Wahrscheinlichkeitsrechnung wollen wir kritisch prüfen. Nicht alles, was Sie in der Zeitung oder im Internet oder im Lehrbuch lesen, ist wahr und gut und nützlich. Bleiben Sie (selbst)kritisch!

Wahrscheinlichkeit eines Reaktorunfalls

Aufgabe: (1) Was ist hier falsch? (2) Welche Rechnung wäre richtig?

Lösung: (1) Eine Wkt von „über 100%“ ist blühender Unsinn! (Kompetenz stärkt Argumente, Inkompetenz schwächt sie.)

☺ Wir haben oben bereits die richtige Formel abgeleitet! Das ist nicht weiter schwer, und soviel Sorgfalt muss sein.

(2) Modell: Unfälle ereignen sich unabhängig mit Wkt p pro Raktorjahr. Wahrscheinlichkeit, in n Reaktorjahren **keinen Unfall** zu erleben:

$$(1-p)^n \lessapprox \mathrm{e}^{-np}$$

Wahrscheinlichkeit, **mindestens einen Unfall** zu erleben:

$$1-(1-p)^n \gtrapprox 1-\mathrm{e}^{-np}$$

Hier gilt $np = 1.23$ und $\mathrm{e}^{-1.23} \lessapprox 29.3\%$, also $1-\mathrm{e}^{-1.23} \gtrapprox 70.7\%$. Die Wkt mindestens eines Unfalls beträgt also etwas über 70%.

☺ Hier ist $\mu = np$ die erwartete Anzahl der Unfälle. Die Näherung $\mathrm{e}^{-np} = P(\mu)(0)$ gehört zur Poisson–Verteilung (Satz U3E).

Wahrscheinlichkeit eines Reaktorunfalls

⚠ Auch die korrekte Rechnung ergibt noch eine alarmierend hohe Wahrscheinlichkeit. Wir können nun darüber nachdenken, ob die Daten verlässlich und die Annahmen überzeugend sind… Aber wenigstens stimmt jetzt die Rechnung: Soviel Handwerk muss man beherrschen!

Ob die Argumente *sachlich* stichhaltig sind, ist eine sehr schwierige Frage. Für eine gesicherte Auswertung ist die Fallzahl sehr gering. Die statistische Analyse vergangener Unfälle und die daraus abgeleitete Schätzung zukünftiger Unfälle sind daher nur schwer möglich. Zum Beispiel sind die drei Unfälle von Fukushima nicht unabhängig, ihre Zählweise also diskussionswürdig. Die Sachlage ist komplex, die hier vorgestellten Zahlen sind nur ein grober erster Anhaltspunkt.

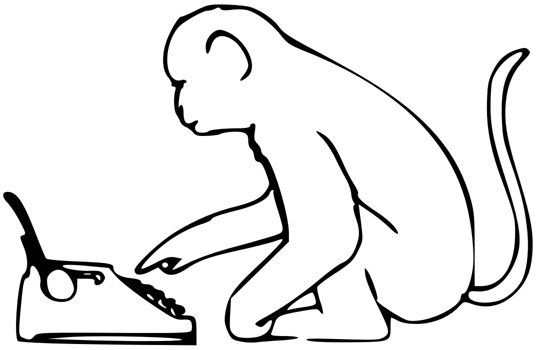
Klar ist immerhin, dass die Unfallwahrscheinlichkeit nicht verschwindend gering ist, und schon gar nicht gleich Null. Zur Vereinfachung schlage ich hier vor, für unsere Beispielrechnung die Annahmen der Autoren zunächst zu akzeptieren und auf dieser Grundlage lediglich zu prüfen, ob die Argumente *rechnerisch* richtig sind. Zur Betonung halten wir daher nochmals fest:

Eine Wahrscheinlichkeit von „über 100%“ ist blühender Unsinn!

Der Fehler liegt in der allzu naiven Rechnung: Bei einer Unfallwahrscheinlichkeit von p pro Reaktorjahr und einer Laufzeit von n Reaktorjahren beträgt die Gesamtwahrscheinlichkeit nicht np. Das fällt spätestens bei einer (groben) Plausibilitätsprüfung auf! Wie jede ernsthafte Frage verdient die Reaktorsicherheit Umsicht und Sorgfalt und insbesondere korrekte Rechnungen!

Der endlos tippende Affe

Ein Affe tippt zufällig aber endlos lange auf einer Schreibmaschine. Wird er dabei irgendwann Shakespeares *Hamlet* schreiben? Ja! Müssen wir darauf sehr lange warten? Ja, aber sicher!



Bildquelle: wikipedia.org

Satz U1D: vom endlos tippenden Affen, quantitative Fassung

Sei T ein vorgegebener Text der Länge L über einem Alphabet A. Ein zufälliger Text Z der Länge nL enthält T mit Wahrscheinlichkeit

$$\mathbf{P}(Z \text{ enthält } T) \;\geq\; 1 - (1 - |A|^{-L})^n \nearrow 1 \quad \text{für } n \to \infty.$$

Für $p \in [0,1[$ gilt $\mathbf{P}(Z \text{ enthält } T) \geq p$ für $n \geq -\ln(1-p) \cdot |A|^L$.

Der Satz vom endlos tippenden Affen

Aufgabe: (1) Erklären Sie diese Abschätzung und den Grenzwert.

(2) Rechnen Sie ein Zahlenbeispiel für $|A| = 26$ und $L = 130\,000$.

(3) Ist dies als Strategie für Klausuren zu empfehlen?

(4) Die extrem lange Wartezeit wird manchmal als Argument gegen die Evolution durch zufällige Mutationen verwendet. Ist das gerechtfertigt?

Die Formulierung des „endlos tippenden Affen“ ist provokant-anschaulich, drückt aber präzise einen mathematischen Satz aus! Das betrachtete Alphabet A habe $|A|$ Zeichen. Zum Beispiel im lateinischen Alphabet 26 Klein- und 26 Großbuchstaben, dazu Ziffern, Satzzeichen, etc. Bei Gleichverteilung wird jedes Zeichen mit positiver Wahrscheinlichkeit $a = 1/|A|$ getippt.

Für das wiederholte Tippen gehen wir von einem Produktexperiment aus. Unser fleißiger Affe produziert so eine *unendliche* Folge zufälliger Zeichen. Darin findet sich jede *endliche* Folge mit Wahrscheinlichkeit 1. Neben *Hamlet* wird er also auch *Harry Potter* Band 1–7 tippen, ebenso Darwins *Über die Entstehung der Arten*, die gesamte Nationalbibliothek, und überhaupt alle bisher veröffentlichten Texte. Die zu erwartende Wartezeit ist allerdings *sehr sehr laaang*…

Zugrundeliegendes Modell: Auf dem Alphabet A betrachten wir eine WVerteilung $q : A \to [0,1]$ mit $a = \min q > 0$, sodass jedes Zeichen mit positiver Wahrscheinlichkeit $\geq a$ getippt wird. Bei Gleichverteilung hätten wir $a = 1/|A|$; die folgende Abschätzung gilt etwas allgemeiner. Bei k–facher unabhängiger Wiederholung des zufälligen Tippvorgangs betrachten wir als stochastisches Modell den Produktraum A^k mit der Produktwahrscheinlichkeit.

Der Satz vom endlos tippenden Affen

(1) Beweis der Abschätzung: Die Wkt, bei L zufälligen Zeichen den Text T zu tippen, ist $\geq a^L > 0$. (Sie ist zugegeben sehr klein, aber nicht Null). Die Misserfolgswkt ist $\leq 1 - a^L < 1$ (nahe Eins, aber nicht gleich Eins).

Wenn der Affe $2L$ Zeichen tippt, so ist die Misserfolgswkt $\leq (1 - a^L)^2$.
Wenn der Affe $3L$ Zeichen tippt, so ist die Misserfolgswkt $\leq (1 - a^L)^3$.
Wenn der Affe nL Zeichen tippt, so ist die Misserfolgswkt $\leq (1 - a^L)^n$.
Wegen $(1 - a^L) < 1$ konvergiert $(1 - a^L)^n \searrow 0$, langsam aber sicher!

Für hinreichend großes n ist die Misserfolgswkt also beliebig nahe an 0. Umgekehrt liegt daher die Erfolgswkt bei $\geq 1 - (1 - a^L)^n \nearrow 1$. Genauer: Für Erfolgswkt $\geq p$, also Misserfolgswkt $\leq 1 - p$, genügt demnach

$$(1 - a^L)^n \leq 1 - p.$$

Nach n aufgelöst bedeutet das: $n \ln(1 - a^L) \leq \ln(1 - p)$.
Wegen $\ln(1 - t) \lessapprox -t$ genügt es, die einfachere Ungleichung $n(-a^L) \leq \ln(1 - p)$ zu erfüllen, also $n \geq -\ln(1 - p)/a^L$.
Für $a = 1/|A|$ erhalten wir die Formel des Satzes.

Der Satz vom endlos tippenden Affen

(2) Zahlenbeispiel: Die Abschätzung ist zwar recht grob, aber illustrativ. Wir betrachten $|A| = 26$, also nur Großbuchstaben ohne Satzzeichen. Die Textlänge $L = 130\,000$ entspricht in etwa die Länge von Hamlet. Für Erfolgswkt $> 99.99\%$ genügt ein zufälliger Text der Länge

$$nL \quad \geq \quad 10 \cdot 26^{130\,000} \cdot 130000 \quad \approx \quad 10^{184\,000}.$$

☺ Das scheinbare Paradox ist damit aufgelöst und numerisch illustriert: Die Schwierigkeit liegt in der harmlosen Formulierung „endlos lange“. Hamlet kommt irgendwann, aber es wird dauern… *sehr sehr laaange*!

(3) Nein, als Strategie für Klausuren ist das nicht zu empfehlen: Niemand will, niemand kann so lange warten. Nicht nachmachen!

☺ Die Erfolgswkt geht gegen 100% durch gründliche Vorbereitung und nicht durch endloses Wiederholen. Das kann sich jeder ausrechnen. Diese Strategie ist wesentlich sinnvoller und effizienter.

(4) Wesentliche Evolutionsfaktoren sind neben zufälliger Mutation auch Rekombination und Selektion. Das ändert das Modell grundlegend!

Geburtstagsparadox / Kollisionswahrscheinlichkeit

Wie groß ist die Wahrscheinlichkeit Q, dass unter $k = 25$ zufälligen Personen mindestens zwei am gleichen Tag des Jahres geboren sind?

Annahmen: Die Geburtstage sind gleichverteilt auf $n = 365$ Tage und untereinander unabhängig (insbesondere keine Zwillinge, etc).

Aufgabe: Berechnen Sie die Wkt Q exakt und näherungsweise.

Komplement: Alle k Personen haben verschiedene Geburtstage. Die Wahrscheinlichkeit $P = 1 - Q$ ist leicht auszurechnen:

$$P = \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$$

$$= \prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right) \lessapprox \prod_{j=0}^{k-1}\exp\left(-\frac{j}{n}\right) = \exp\left(-\sum_{j=0}^{k-1}\frac{j}{n}\right) = \exp\left(-\frac{k(k-1)}{2n}\right)$$

Für $k = 25$ und $n = 365$ gilt $P \lessapprox \mathrm{e}^{-0.821} \lessapprox 0.44$ und somit $Q \gtrapprox 0.56$.

🙂 Intuitiv halten viele eine Kollision für unwahrscheinlich. Tatsächlich liegt die Wkt bei über 50%, daher trägt dieses überraschende Ergebnis den Namen „Geburtstagsparadox“. Probieren Sie es selbst einmal aus!

Geburtstagssatz / Kollisionswahrscheinlichkeit

Satz U1E: Geburtstagssatz / Kollisionswahrscheinlichkeit

Aus n Möglichkeiten wird k mal zufällig ausgewählt (wobei $1 \le k \le n$). Die Wahrscheinlichkeit $P_{n,k}$, dabei k verschiedene auszuwählen, ist

$$\boxed{P_{n,k} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \lessapprox \exp\left(-\frac{k(k-1)}{2n}\right).}$$

Die Wahrscheinlichkeit $Q_{n,k}$ mindestens einer Kollision ist demnach

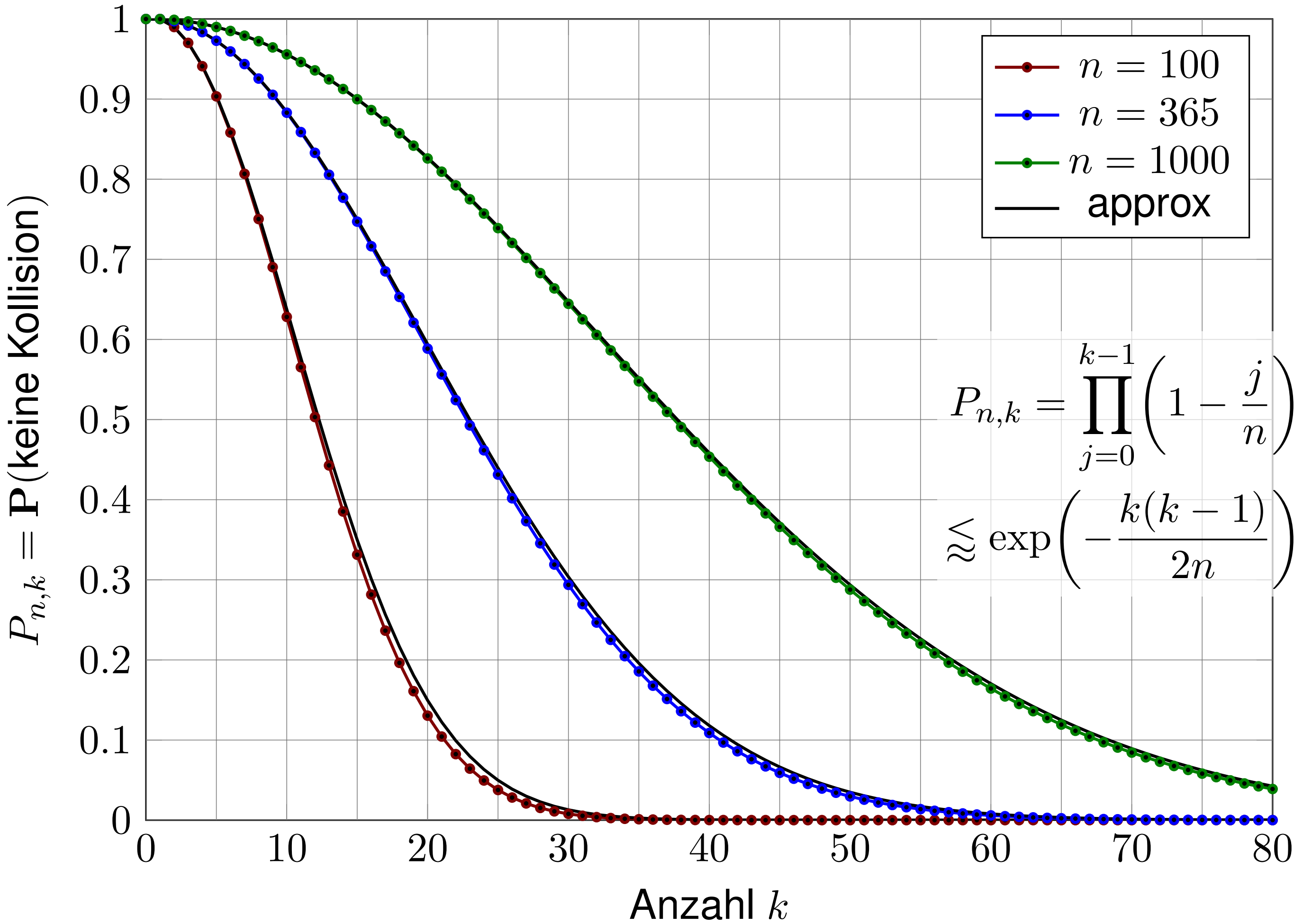
$$\boxed{Q_{n,k} = 1 - P_{n,k} \gtrapprox 1 - \exp\left(-\frac{k(k-1)}{2n}\right).}$$

Aus n Möglichkeiten k mal zufällig auswählen soll heißen *gleichverteilt* und *unabhängig*. Die genannte Ungleichung „$\le$“ bzw. „$\ge$“ gilt hier immer. Die Näherung „$\approx$“ ist gut falls n groß ist und k klein. Wie die folgenden Graphen zeigen, ist selbst für große k die Abweichung meist akzeptabel.

🙂 Der Graph ist die rechte Hälfte der Gaußschen Glockenkurve $\mathrm{e}^{-x^2/2}$.

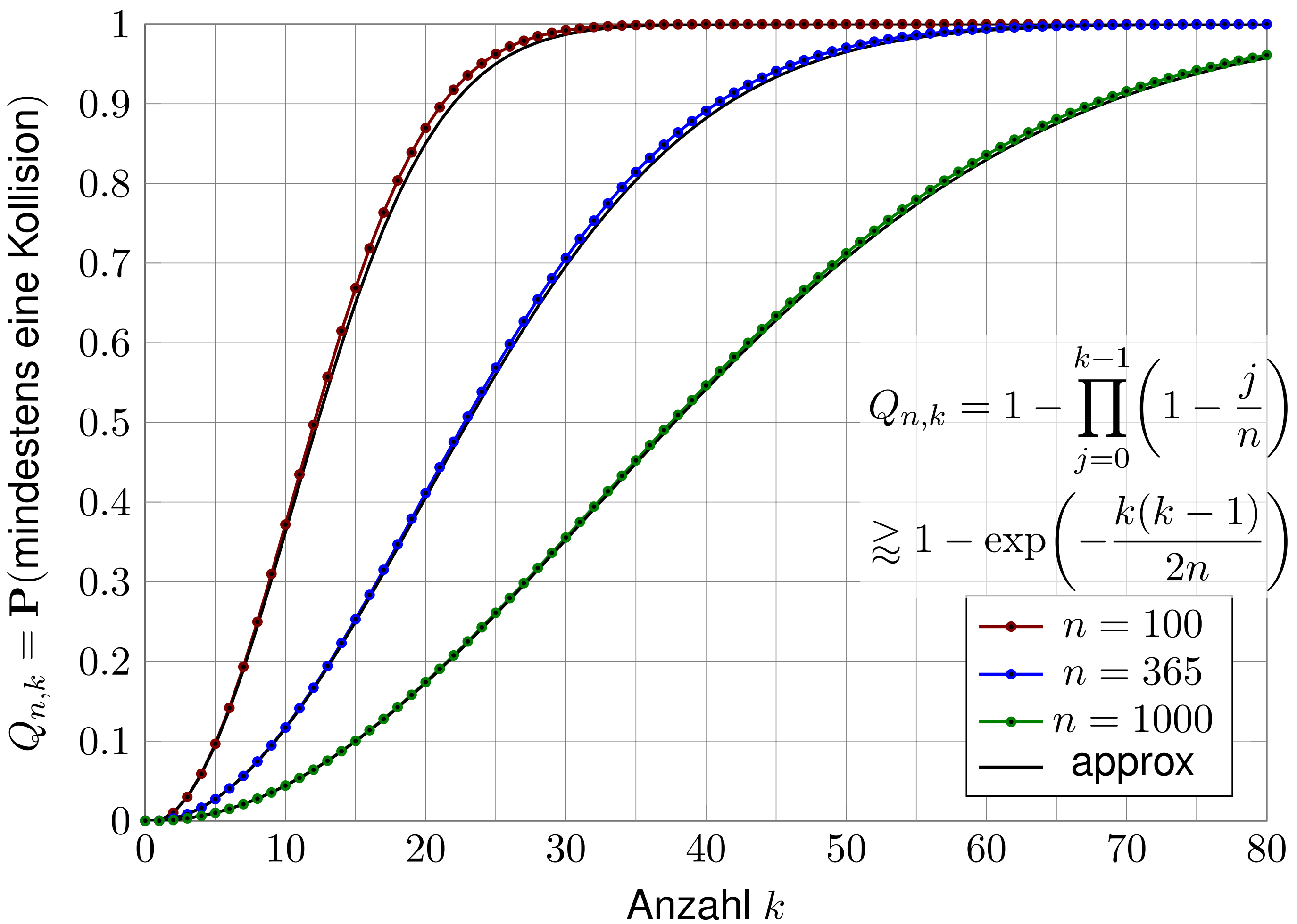
Geburtstagssatz / Kollisionswahrscheinlichkeit

U119



Geburtstagssatz / Kollisionswahrscheinlichkeit

U120



Kombinatorische Abzählformeln

Viele Zufallsexperimente beschreiben wir als **Laplace–Modell** durch einen endlichen Wahrscheinlichkeitsraum $(\Omega, \mathbf{P})$ mit Gleichverteilung:

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl aller Ergebnisse in } \Omega} = \frac{\text{günstige Ergebnisse}}{\text{mögliche Ergebnisse}}$$

☺ Es genügt hier also, Elemente zu zählen. Das ist meist einfach.

⚠ In realistischen Anwendungen können Ω und A sehr groß sein.

Beispiel: Beim Lotto „6 aus 49“ gibt es ca. 14 Millionen Möglichkeiten. Mengen solcher Größe möchten Sie nicht mehr per Hand abzählen.

Sind die betrachteten Mengen noch etwas größer, so versagen selbst Computer beim expliziten Aufzählen. Wir benötigen daher effiziente Techniken zum geschickten, systematischen Abzählen...

☺ Hierzu nutzen wir die folgenden kombinatorischen Abzählformeln!

☺ Für große Werte nutzen wir zudem geeignete Näherungsformeln.

📖 M. Aigner: *Diskrete Mathematik.* Vieweg 2006

Kombinatorische Abzählformeln

Anzahl aller **Tupel** der Länge k von n Elementen:

$$n^k = n \cdot n \cdots n \qquad (k \text{ Faktoren})$$

Anzahl aller **Tupel ohne Wiederholungen**:

$$n^{\underline{k}} = n(n-1)(n-2)\cdots(n-k+1) \qquad (k \text{ Faktoren})$$

Anzahl aller **Permutationen** von n Elementen:

$$n! = n^{\underline{n}} = n(n-1)(n-2)\cdots 1 \qquad (n \text{ Faktoren})$$

Anzahl aller **Auswahlen** von k aus n Elementen

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 1} = \frac{n^{\underline{k}}}{k!}$$

Für große n und kleine k haben wir folgende Näherungsformeln:

$$\boxed{n^{\underline{k}}/n^k \lessapprox \mathrm{e}^{-k(k-1)/(2n)} \quad \text{und} \quad n!/n^n \sim \sqrt{2\pi n}\,\mathrm{e}^{-n}}$$

📖 Siehe Kimmerle–Stroppel, Lineare Algebra, §1.3, zur Wiederholung von **Fakultät** und **Binomialkoeffizienten**. Das Produkt $n^{\underline{k}} = n(n-1)(n-2)\cdots(n-k+1) = n!/(n-k)!$ heißt **fallende Fakultät** von n mit k Faktoren. Sie entsteht hier als nützliche Abzählformel.

Tupel

Aus $\{0,1\}$ lassen sich folgende 2–Tupel (Paare) bilden:

$$\{0,1\}^2 = \{\ (0,0),\ (0,1),\ (1,0),\ (1,1)\ \}$$

Aus $\{0,1\}$ lassen sich folgende 3–Tupel (Tripel) bilden:

$$\begin{aligned}\{0,1\}^3 = \{\ &(0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\\ &(1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\ \}\end{aligned}$$

Aus $\{0,1\}$ lassen sich folgende 4–Tupel (Quadrupel) bilden:

$$\begin{aligned}\{0,1\}^4 = \{\ &(0,0,0,0),\ (0,0,0,1),\ (0,0,1,0),\ (0,0,1,1),\\ &(0,1,0,0),\ (0,1,0,1),\ (0,1,1,0),\ (0,1,1,1),\\ &(1,0,0,0),\ (1,0,0,1),\ (1,0,1,0),\ (1,0,1,1),\\ &(1,1,0,0),\ (1,1,0,1),\ (1,1,1,0),\ (1,1,1,1)\ \}\end{aligned}$$

Das entspricht dem n–maligen Münzwurf (mit $0 =$ Kopf, $1 =$ Zahl). Beim n–maligen Würfeln betrachten wir analog $\Omega = \{1,2,3,4,5,6\}^n$ als Ergebnismenge; den Fall $n = 2$ haben wir auf Seite T124 aufgezählt.

Tupel

Für alle $n, k \in \mathbb{N}$ haben wir die **Potenz**

$$\boxed{n^k = \text{Anzahl aller **Tupel** der Länge } k \text{ von } n \text{ Elementen.}}$$

Rekursionsformel: $n^0 = 1$ und $n^{k+1} = n^k \cdot n$.

Ausgeschrieben: $n^k = n \cdot n \cdot n \cdots n$ mit k Faktoren.

Zudem betrachten wir die **fallende Fakultät**

$$\boxed{n^{\underline{k}} = \text{Anzahl aller **Tupel ohne Wiederholungen**.}}$$

Rekursionsformel: $n^{\underline{0}} = 1$ und $n^{\underline{k+1}} = n^{\underline{k}} \cdot (n-k)$.

Ausgeschrieben: $n^{\underline{k}} = n(n-1)(n-2)\cdots(n-k+1)$.

Nach Geburtstagssatz U1E gilt $n^{\underline{k}}/n^k \lessapprox \exp\big(-k(k-1)/(2n)\big)$.

Für kleine k und große n haben wir also die Näherungsformel

$$\boxed{n^{\underline{k}} \lessapprox n^k \, \mathrm{e}^{\frac{-k(k-1)}{2n}}.}$$

Permutationen

Permutationen der Elemente $1, 2$:

$$(1,2),\ (2,1)$$

Permutationen der Elemente $1, 2, 3$:

$$(1,2,3),\ (1,3,2),\ (2,1,3),\ (2,3,1),\ (3,1,2),\ (3,2,1)$$

Permutationen der Elemente $1, 2, 3, 4$:

$$\begin{aligned}
&(1,2,3,4),\ (1,2,4,3),\ (1,3,2,4),\ (1,3,4,2),\ (1,4,2,3),\ (1,4,3,2),\\
&(2,1,3,4),\ (2,1,4,3),\ (2,3,1,4),\ (2,3,4,1),\ (2,4,1,3),\ (2,4,3,1),\\
&(3,1,2,4),\ (3,1,4,2),\ (3,2,1,4),\ (3,2,4,1),\ (3,4,1,2),\ (3,4,2,1),\\
&(4,1,2,3),\ (4,1,3,2),\ (4,2,1,3),\ (4,2,3,1),\ (4,3,1,2),\ (4,3,2,1)
\end{aligned}$$

Eine Permutation der Menge $\{1, \dots, n\}$ ist ein n–Tupel, in dem jedes Element $1, \dots, n$ genau einmal vorkommt. Hierzu gibt es $n! := n^{\underline{n}} = n(n-1)(n-2)\cdots 1$ Möglichkeiten.

Obige Näherungsformel liefert die Ungleichung $n! \le n^n\,\mathrm{e}^{\frac{1-n}{2}}$, ist hier aber unnötig grob. Für die Fakultät geben wir die ersten Werte und als Näherung die Stirling–Formel.

Permutationen

Für alle $n \in \mathbb{N}$ haben wir die **Fakultät**

$$\boxed{n! = \text{Anzahl aller } \textbf{Permutationen} \text{ von } n \text{ Elementen.}}$$

Rekursionsformel: $0! = 1$ und $(n+1)! = n! \cdot (n+1)$.

Ausgeschrieben: $n! = 1 \cdot 2 \cdot 3 \cdots n$. Die ersten Werte sind:

$0! = 1$	$7! = 5\,040$	$14! = 87\,178\,291\,200$
$1! = 1$	$8! = 40\,320$	$15! = 1\,307\,674\,368\,000$
$2! = 2$	$9! = 362\,880$	$16! = 20\,922\,789\,888\,000$
$3! = 6$	$10! = 3\,628\,800$	$17! = 355\,687\,428\,096\,000$
$4! = 24$	$11! = 39\,916\,800$	$18! = 6\,402\,373\,705\,728\,000$
$5! = 120$	$12! = 479\,001\,600$	$19! = 121\,645\,100\,408\,832\,000$
$6! = 720$	$13! = 6\,227\,020\,800$	$20! = 2\,432\,902\,008\,176\,640\,000$

Dank Stirling–Formel B3D haben wir für große n die Näherung

$$\boxed{n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n}$$

Potenzmenge

Die **Potenzmenge** $\mathfrak{P}(\Omega)$ ist die Menge aller Teilmengen von Ω:

$$\begin{aligned}
\mathfrak{P}(\emptyset) &= \{\ \emptyset\ \} \\
\mathfrak{P}(\{1\}) &= \{\ \emptyset,\ \{1\}\ \} \\
\mathfrak{P}(\{1,2\}) &= \{\ \emptyset,\ \{1\},\{2\},\ \{1,2\}\ \} \\
\mathfrak{P}(\{1,2,3\}) &= \{\ \emptyset,\ \{1\},\{2\},\{3\},\ \{1,2\},\{1,3\},\{2,3\},\ \{1,2,3\}\ \} \\
\mathfrak{P}(\{1,2,3,4\}) &= \{\ \emptyset,\ \{1\},\{2\},\{3\},\{4\}, \\
&\qquad \{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}, \\
&\qquad \{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\},\ \{1,2,3,4\}\ \}
\end{aligned}$$

Für die Anzahl aller Teilmengen gilt

$$\boxed{|\mathfrak{P}(\Omega)| = 2^{|\Omega|}.}$$

Beweis: In einer n–elementigen Menge Ω entsteht jede Teilmenge $A \subseteq \Omega$ durch n unabhängige Wahlen: Jedes Element $\omega \in \Omega$ liegt entweder in A oder nicht, und diese Wahlen bestimmen A. Vornehm ausgedrückt: Die Indikatorfunktion stiftet eine Bijektion $\mathfrak{P}(\Omega) \to \{0,1\}^\Omega : A \mapsto \mathbf{I}_A$, ihre Umkehrfunktion ist demnach $\{0,1\}^\Omega \to \mathfrak{P}(\Omega) : f \mapsto \operatorname{supp}(f) = \{\, \omega \in \Omega \mid f(\omega) = 1 \,\}$.

Binomialkoeffizienten

Für $n, k \in \mathbb{N}$ haben wir den **Binomialkoeffizienten**

$$\boxed{\binom{n}{k} = \text{Anzahl aller **Auswahlen** von } k \text{ aus } n \text{ Elementen.}}$$

Es gilt $\binom{n}{0} = \binom{n}{n} = 1$ und $\binom{n}{k} = 0$ für $k > n$ sowie $\binom{n}{k} = \binom{n}{n-k}$.
Für alle $n, k \ge 1$ gelingt die Berechnung durch die Rekursionsformel

$$\boxed{\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.}$$

Beweis: Entweder wird das Element n ausgewählt, dann gibt es genau $\binom{n-1}{k-1}$ Möglichkeiten, noch $k-1$ Elemente aus den verbleibenden $n-1$ auszuwählen. Oder das Element n wird nicht ausgewählt, dann gibt es noch genau $\binom{n-1}{k}$ Möglichkeiten, k Elemente aus den verbleibenden $n-1$ auszuwählen. Beide Fälle sind disjunkt, ihre Summe ist die behauptete Rekursionsformel.

Zusammen mit der vorigen Formel $|\mathfrak{P}(\Omega)| = 2^{|\Omega|}$ erhalten wir mühelos:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Binomialkoeffizienten

Hieraus erhalten wir **Pascals Dreieck für Binomialkoeffizienten**:

$\binom{n}{k}$	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

Für festes k und $n \to \infty$ (also eine Spalte in obiger Tabelle) gilt

$$\boxed{\binom{n}{k} \quad \sim \quad \frac{n^k}{k!} \exp\left(-\frac{k(k-1)}{2n}\right) \quad \sim \quad \frac{n^k}{k!}.}$$

Binomischer Lehrsatz

Für je zwei Elemente a, b in einem Ring mit $ab = ba$ gilt:

$$\begin{aligned}
(a+b)^0 &= 1 \\
(a+b)^1 &= 1a + 1b \\
(a+b)^2 &= 1a^2 + 2ab + 1b^2 \\
(a+b)^3 &= 1a^3 + 3a^2b + 3ab^2 + 1b^3 \\
(a+b)^4 &= 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4
\end{aligned}$$

Für alle $n \in \mathbb{N}$ gilt (per vollständiger Induktion):

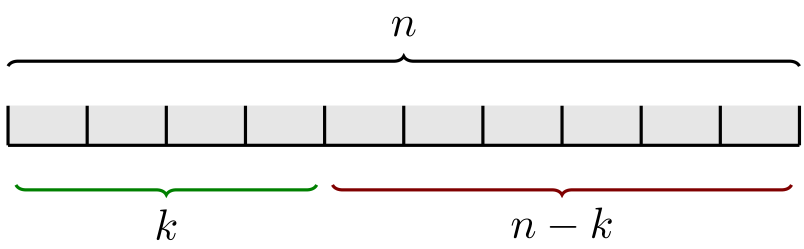
$$\boxed{(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k}$$

Speziell für $a = b = 1$ erhalten wir damit erneut:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Binomialkoeffizienten

Auf wie viele Arten können wir aus n Objekten k auswählen?



Die Gesamtzahl der Möglichkeiten ist:

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 1} = \frac{n^{\underline{k}}}{k!}}$$

Kombinatorische Herleitung: Wir wählen k aus n Elementen. Hierzu können wir alle n Elemente auf $n!$ Arten anordnen. Bei jeder Anordnung wählen wir die ersten k Elemente aus. Die ersten k dürfen wir dabei auf $k!$ Arten umordnen. Die letzten $n-k$ dürfen wir auf $(n-k)!$ Arten umordnen. Die Auswahl ändert sich hierdurch nicht. Dies liefert die obige Formel.

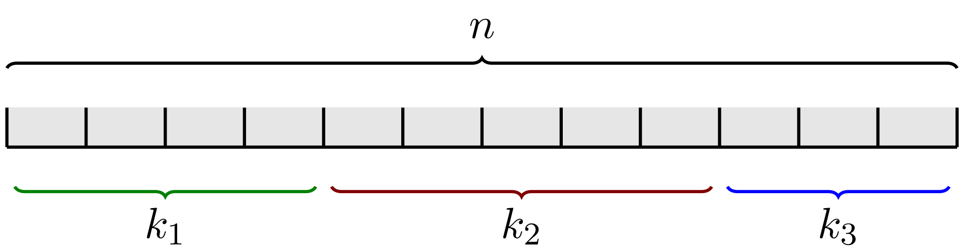
Hieraus erhalten wir die folgenden groben Schranken:

$$\frac{n^k}{k^k} \quad\le\quad \binom{n}{k} \quad\le\quad \frac{n^k}{k!}\exp\left(-\frac{k(k-1)}{2n}\right) \quad\le\quad \frac{n^k}{k!}$$

Multinomialkoeffizienten

Der Binomialkoeffizient $\binom{n}{k}$ zählt, auf wie viele Arten wir n Objekte in zwei Gruppen der Größe k und $n-k$ aufteilen können. Etwas allgemeiner können wir folgendes zählen:

Auf wie viele Arten können wir n Objekte in ℓ Gruppen zu $k_1, \dots, k_\ell$ Elementen aufteilen? Hierbei sei $k_1, \dots, k_\ell \ge 0$ und $k_1 + \dots + k_\ell = n$.



Die Gesamtzahl der Möglichkeiten ist:

$$\boxed{\binom{n}{k_1, k_2, \dots, k_\ell} := \frac{n!}{k_1! \cdot k_2! \cdots k_\ell!}}$$

Kombinatorische Herleitung: Wir können die n Elemente auf $n!$ Arten anordnen. Die ersten k_1 bilden die erste Gruppe, die nächsten k_2 bilden die zweite Gruppe, usw. Innerhalb jeder Gruppe können wir beliebig umordnen. Dies liefert die obige Formel.

Anwendungsbeispiele

Aufgabe: Wie viele Kartenverteilungen gibt es beim Skat?

Lösung: Von 32 Karten bekommt jeder der drei Spieler 10 Karten:

$$\binom{32}{10,10,10,2} = \frac{32!}{10!\,10!\,10!\,2!} \approx 2.75 \cdot 10^{15}$$

Aufgabe: Auf wie viele Weisen lassen sich 12 (verschiedenfarbige) Bonbons gerecht unter zwei / drei / vier / zwölf Kindern aufteilen?

Lösung: Die Anzahl der möglichen Aufteilungen berechnen wir zu:

$$\begin{aligned}
\binom{12}{6,6} &= \frac{12!}{6!\,6!} &&= 924 \\
\binom{12}{4,4,4} &= \frac{12!}{4!\,4!\,4!} &&= 34650 \\
\binom{12}{3,3,3,3} &= \frac{12!}{3!\,3!\,3!\,3!} &&= 369600 \\
\binom{12}{1,1,1,1,1,1,1,1,1,1,1,1} &= 12! &&= 479001600
\end{aligned}$$

Anwendungsbeispiele

Aufgabe: Wie wahrscheinlich ist es, bei 10 Münzwürfen genau 7 mal Kopf und 3 mal Zahl zu erhalten?

$$\begin{aligned}
\Omega &= \{0,1\}^{10}, & |\Omega| &= 2^{10} &&= 1024 \\
A &= \left\{ \begin{array}{l} \text{Ergebnisse } \omega \in \Omega \text{ mit} \\ \omega_1 + \dots + \omega_{10} = 3 \end{array} \right\}, & |A| &= \binom{10}{7,3} &&= 120 \\
& & \mathbf{P}(A) &= |A|/|\Omega| &&\approx 0.117
\end{aligned}$$

Aufgabe: Wie wahrscheinlich ist es, bei 10maligem Würfeln, die Augenzahlen $1,2,3,4,5,6$ mit Häufigkeit $1,2,3,3,0,1$ zu erhalten?

$$\begin{aligned}
\Omega &= \{1,2,3,4,5,6\}^{10}, & |\Omega| &= 6^{10} &&= 60\,466\,176 \\
A &= \left\{ \begin{array}{l} \text{Ergebnisse } \omega \in \Omega \text{ mit} \\ \text{diesen Häufigkeiten} \end{array} \right\}, & |A| &= \binom{10}{1,2,3,3,0,1} &&= 50\,400 \\
& & \mathbf{P}(A) &= |A|/|\Omega| &&\approx 0.00083
\end{aligned}$$

Hier wurde ein Laplace–Experiment betrachtet, d.h. alle Ergebnisse sind gleich wahrscheinlich. Es genügt daher, die Anzahl $|A|$ der gesuchten Ergebnisse durch die Gesamtzahl $|\Omega|$ aller Ergebnisse zu teilen. Diese Anzahlen können wir nun leicht bestimmen.

Zerlegungen

Die Menge $\{1\}$ erlaubt nur eine Zerlegung, nämlich $\{\{1\}\}$.
Die Menge $\{1,2\}$ erlaubt zwei Zerlegungen: $\{\{1,2\}\}$ und $\{\{1\},\{2\}\}$.
Die Menge $\{1,2,3\}$ erlaubt die folgenden fünf Zerlegungen:

$$\{\{1,2,3\}\},\ \{\{1\},\{2,3\}\},\ \{\{1,2\},\{3\}\},\ \{\{1,3\},\{2\}\},\ \{\{1\},\{2\},\{3\}\}.$$

Die Menge $\{1,2,3,4\}$ erlaubt die folgenden fünfzehn Zerlegungen:

$$\{\{1,2,3,4\}\},$$

$$\{\{1,2\},\{3,4\}\},\ \{\{1,3\},\{2,4\}\},\ \{\{1,4\},\{2,3\}\},$$
$$\{\{1\},\{2,3,4\}\},\ \{\{2\},\{1,3,4\}\},\ \{\{3\},\{1,2,4\}\},\ \{\{4\},\{1,2,3\}\},$$

$$\{\{1,2\},\{3\},\{4\}\},\ \{\{1,3\},\{2\},\{4\}\},\ \{\{1,4\},\{2\},\{3\}\},$$
$$\{\{2,3\},\{1\},\{4\}\},\ \{\{2,4\},\{1\},\{3\}\},\ \{\{3,4\},\{1\},\{2\}\},$$

$$\{\{1\},\{2\},\{3\},\{4\}\}.$$

Zerlegungszahlen

Für $n,k\in\mathbb{N}$ betrachten wir die Zerlegungszahl

$$\left\{{n \atop k}\right\} = \text{Anzahl aller } \textbf{Zerlegungen} \text{ von } n \text{ Elementen in } k \text{ nicht-leere Gruppen (beliebiger Größe).}$$

Offenbar gilt $\left\{{0 \atop 0}\right\} = 1$ und $\left\{{n \atop k}\right\} = 0$ für $k > n$.
Für $n \ge 1$ gilt $\left\{{n \atop 0}\right\} = 0$, $\left\{{n \atop 1}\right\} = 1$, $\left\{{n \atop n}\right\} = 1$.
Für $n \ge 2$ gilt zudem $\left\{{n \atop n-1}\right\} = \binom{n}{2}$, $\left\{{n \atop 2}\right\} = 2^{n-1}-1$. (Warum?)
Für alle $n,k\ge 1$ gelingt die Berechnung durch die Rekursionsformel

$$\boxed{\left\{{n \atop k}\right\} = \left\{{n-1 \atop k-1}\right\} + k\left\{{n-1 \atop k}\right\}.}$$

Beweis: Entweder das Element n bildet alleine eine Gruppe, dann gibt es genau $\left\{{n-1 \atop k-1}\right\}$ Möglichkeiten, die verbleibenden $n-1$ Elemente in $k-1$ Gruppen zu zerlegen.
Oder aber das Element n ist Teil einer größeren Gruppe, dann gibt es hierfür genau $\left\{{n-1 \atop k}\right\}$ Zerlegungen und jeweils k Wahlen, das Element n zu einer bestehenden Gruppe hinzuzufügen.

Die Zerlegungszahlen sind ebenso interessant und natürlich wie die Binomialkoeffizienten, wenn auch weniger prominent. Für $\left\{{n \atop k}\right\}$ gibt es leider keine so schöne Formel wie für $\binom{n}{k}$.

Hieraus erhalten wir folgendes **Dreieck für Zerlegungszahlen**:

$\left\{{n \atop k}\right\}$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	1	3	1							
4	1	7	6	1						
5	1	15	25	10	1					
6	1	31	90	65	15	1				
7	1	63	301	350	140	21	1			
8	1	127	966	1701	1050	266	28	1		
9	1	255	3025	7770	6951	2646	462	36	1	
10	1	511	9330	34105	42525	22827	5880	750	45	1

Diese Werte heißen auch **Stirling–Zahlen** (zweiter Art).

Für festes k und $n \to \infty$ (also eine Spalte in obiger Tabelle) gilt

$$\left\{{n \atop k}\right\} \sim \frac{k^n}{k!} \qquad \text{im Gegensatz zu} \qquad \binom{n}{k} \sim \frac{n^k}{k!}.$$

Aufgabe: (Aus dem Mathekalender 2012) Sie backen $k = 100$ Kekse. Sie geben n Chocolate Chips zum Teig, verteilen diese gründlich zufällig und teilen den Teig in 100 Kekse auf. Wieviele Chips brauchen Sie, damit mit 90% Sicherheit jeder Keks mindestens einen Chip enthält?

Lösung: Die Wkt, dass kein Chip im ersten Keks landet, ist $(99/100)^n$. Die Wkt für mindestens einen im ersten Keks ist $1 - (99/100)^n$. Für $A_i = \{\text{mindestens ein Chip in Keks } i\}$ gilt $\mathbf{P}(A_i) = 1 - 0.99^n$.

(1) Näherung: Zur Vereinfachung rechnen wir, als wären die Ereignisse A_i unabhängig. (Das ist genau genommen nicht richtig, siehe T224, erweist sich aber als erstaunlich gute Näherung, siehe unten.)

Die gewünschte Bedingung vereinfacht sich dann zu:

$$f(n) := (1 - 0.99^n)^{100} \quad \overset{!}{\geq} \quad 0.9$$

Wir erhalten somit (dank Taschenrechner):

$$n \geq \frac{\ln(1 - 0.9^{1/100})}{\ln 0.99} \approx 682.17$$

Anwendungsbeispiel: Schokokekse

(2) Exakt gibt es $k^n = 100^n$ Verteilungen der n Chips auf $k = 100$ Kekse. Darunter sind genau $k!\left\{{n \atop k}\right\}$ mit mindestens einem Chip in jedem Keks.

Anders gesagt: Es gibt k^n Abbildungen $f:\{1,\dots,n\} \to \{1,\dots,k\}$, davon sind genau $k!\left\{{n \atop k}\right\}$ surjektiv. (Schubfachmodell, siehe U222)

Die gewünschte Bedingung ist demnach:

$$g(n) := \frac{k!}{k^n}\left\{{n \atop k}\right\} \quad \overset{!}{\geq} \quad 0.9$$

Ausrechnen für $n = 100, 101, 102, \dots$ mit Hilfe eines Computers liefert:

$$g(682) = 0.899499\dots, \qquad g(683) = 0.900455\dots$$

☺ Sie brauchen also tatsächlich 683 Chocolate Chips! Das rechtfertigt nachträglich die naive Näherung in der obigen vereinfachten Rechnung.

Das erklärt noch nicht, warum diese Näherung so gut funktioniert, oder in welchen anderen Situationen Sie diesen Trick anwenden können. Hierzu wäre eine allgemeine Fehlerabschätzung notwendig.

Zusammenfassung: Abzählformeln

Viele Zufallsexperimente beschreiben wir als **Laplace–Modell** durch einen endlichen Wahrscheinlichkeitsraum $(\Omega, \mathbf{P})$ mit Gleichverteilung:

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl aller Ergebnisse in } \Omega} = \frac{\text{günstige Ergebnisse}}{\text{mögliche Ergebnisse}}$$

☺ Zum systematischen Abzählen nutzen wir folgende Funktionen:

n^k = Anzahl aller **Tupel** der Länge k von n Elementen,
$n^{\underline{k}}$ = Anzahl aller **Tupel ohne Wiederholungen**,
$n!$ = Anzahl aller **Permutationen** von n Elementen,
$\binom{n}{k}$ = Anzahl aller **Auswahlen** von k aus n Elementen,
$\left\{{n \atop k}\right\}$ = Anzahl aller **Zerlegungen** von n Elementen in k Gruppen.

Diese Grundfunktionen gehören zum kombinatorischen Handwerk, da solche und ähnliche Abzählprobleme sehr häufig vorkommen.

☺ Zur exakten Berechnung haben wir die obigen Rekursionsformeln.

☺ Für große Werte nutzen wir zudem geeignete Näherungsformeln.

Schubfachmodelle im Überblick

Auf wie viele Arten können wir k Murmeln auf n Zellen verteilen?

Beim Verteilen gibt es drei Varianten:

- **beliebig viele Murmeln pro Zelle**
- **höchstens eine Murmel pro Zelle**
- **mindestens eine Murmel pro Zelle**

Auch zur Unterscheidung der Murmeln gibt es zwei Varianten:

- **mit Unterscheidung**: Die Murmeln sind nummeriert $1, \dots, k$.
- **ohne Unterscheidung**: Die Murmeln sind ununterscheidbar.

Die Zuordnung der Murmeln zu den Zellen entspricht einer Abbildung

$$f : \{1, \dots, k\} \to \{1, \dots, n\}.$$

Ununterscheidbare Murmeln dürfen wir umsortieren und erhalten so

$$f(1) \le f(2) \le \dots \le f(k).$$

Wir zählen nun beliebige / injektive / surjektive Abbildungen.

Schubfachmodelle im Überblick

Auf wie viele Arten kann man k Objekte auf n Fächer verteilen?

Schubfachmodell	unterscheidbare Objekte	ununterscheidbare Objekte
beliebig viele Objekte pro Fach	n^k	$\binom{n+k-1}{k}$
höchstens eins pro Fach (injektiv)	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$
mindestens eins pro Fach (surjektiv)	$n!\left\{ {k \atop n} \right\}$	$\binom{k-1}{n-1}$

Unterscheidbare Objekte denken wir uns mit $1, \dots, k$ nummeriert. Bei ununterscheidbaren Objekten dürfen wir Objekte untereinander vertauschen: Das bedeutet, wir betrachten Anordnungen dann als gleich, wenn sie sich nur durch die Nummerierung der Objekte unterscheiden. (Für Ausführungen siehe `en.wikipedia.org/wiki/Twelvefold_way`.)

Schubfachmodelle: unterscheidbare Objekte

☺ Die Tabelle ist dank unserer Vorarbeit leicht zu verstehen.

(1) Wir verteilen die Objekte $1, \dots, k$ auf die Fächer $1, \dots, n$.
Wir betrachten also Abbildungen $f : \{1, \dots, k\} \to \{1, \dots, n\}$.

Beliebige Abbildungen: n^k Möglichkeiten.

Wenn wir k Objekte beliebig auf n Fächer verteilen dürfen, dann haben wir für jedes Objekt n Wahlen. Diese sind untereinander unabhängig. Insgesamt erhalten wir n^k Möglichkeiten.

Injektive Abbildungen: $n^{\underline{k}} = \frac{n!}{(n-k)!}$ Möglichkeiten.

Wenn wir in jedes Fach höchstens ein Objekt legen dürfen, so muss $k \le n$ gelten. Für das erste Objekt haben wir n Möglichkeiten, für das zweite nur noch $n-1$, für das dritte nur noch $n-2$, usw. Insgesamt erhalten wir also $n \cdot (n-1) \cdots (n-k+1) = n^{\underline{k}} = \frac{n!}{(n-k)!}$ Möglichkeiten.

Surjektive Abbildungen: $n! \left\{ {k \atop n} \right\}$ Möglichkeiten.

Wenn wir in jedes Fach mindestens ein Objekt legen müssen, so muss $k \ge n$ gelten.
Wir zerlegen zunächst alle k Objekte in n Gruppen. Dies gelingt auf $\left\{ {k \atop n} \right\}$ Möglichkeiten.
Diese n Gruppen können wir nun auf $n!$ Weisen n den Fächern zuordnen.

Schubfachmodelle: ununterscheidbare Objekte

(2) Bei ununterscheidbaren Objekten dürfen wir Objekte vertauschen:
Nach Umordnung können wir $f(1) \le f(2) \le \dots \le f(k)$ annehmen.

Injektive Abbildungen: $f(1) < f(2) < \dots < f(k)$

Dann ist f bereits durch $\{f(1), f(2), \dots, f(k)\}$ bestimmt.

Hier gibt es genau $\binom{n}{k}$ Möglichkeiten, diese Teilmenge zu wählen.

Surjektive Abbildungen: $f(1) \le f(2) \le \dots \le f(k)$.

Für genau $n-1$ der $k-1$ Vergleiche gilt „$<$“, sonst „$=$“.

Hierzu gibt es genau $\binom{k-1}{n-1} = \binom{k-1}{k-n}$ Möglichkeiten.

Beliebige Abbildungen: $f(1) \le f(2) \le \dots \le f(k)$.

Definiere $g : \{1, \dots, k\} \to \{1, \dots, n+k-1\}$ durch $g(x) = f(x) + x - 1$.

Es gilt dann $g(1) < g(2) < \dots < g(k)$. Dies stiftet eine Bijektion $f \leftrightarrow g$.

Für g gibt es genau $\binom{n+k-1}{k}$ Möglichkeiten; somit auch für f.

Urnenmodelle im Überblick

Viele Zufallsexperimente lassen sich als **Urnenmodelle** sehr effizient beschreiben.

Aus einer Urne mit n durchnummerierten Kugeln ziehen wir k Kugeln (oder Lose):

- Ziehung mit oder ohne Zurücklegen,
- Ergebnis mit oder ohne Reihenfolge.

Die Gesamtzahl der möglichen Ergebnisse berechnet sich wie folgt:

Urnenmodell	mit Reihenfolge		ohne Reihenfolge	
ohne Zurücklegen	$\frac{n!}{(n-k)!}$	Laplace	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$	Laplace
mit Zurücklegen	n^k	Laplace	$\binom{n+k-1}{k}$	nicht Laplace!

Urnenmodelle im Überblick

Beim Ziehen der Losnummern gibt es zwei Varianten:

- **mit Zurücklegen** (mZ) — Nach jedem Zug wird die Nummer der gezogenen Kugel notiert und die Kugel zurückgelegt.
- **ohne Zurücklegen** (oZ) — Kugeln werden nicht zurückgelegt. Jede Zahl kann daher höchstens einmal auftreten.

Das Endergebnis können wir auf zwei Weisen zusammenfassen:

- **mit** Berücksichtigung der **Ziehungsreihenfolge** (mR)
- **ohne** Berücksichtigung der **Ziehungsreihenfolge** (oR)

☺ Dies entspricht obigen Schubfachmodellen: Jede Ziehung ist eine Abbildung $f : \{1, \dots, k\} \to \{1, \dots, n\}$. Ohne Zurücklegen ist sie injektiv. Ohne Reihenfolge dürfen wir umsortieren zu $f(1) \le f(2) \le \dots \le f(k)$.

☺ Jede Ziehung oZmR, oZoR, mZmR ist **ein Laplace–Experiment**: Die Wahrscheinlichkeit ist hier **gleichverteilt**, also $p(\omega) = 1/|\Omega|$.

⚠ Die Ziehung mZoR hingegen ist **kein Laplace–Experiment**: Die Wahrscheinlichkeit ist **nicht gleichverteilt**! (Olympialotterie 1971)

Beispiel: Zweimaliges Würfeln

Zweimal Würfeln ergibt die Gleichverteilung auf $6^2 = 36$ Ergebnissen:

$$\Omega = \left\{ \begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix} \right\}$$

Das entspricht dem Urnenmodell „2 aus 6“ mZmR und ist laplacesch. Umsortiert (also mZoR) bleiben nur noch $\binom{6+2-1}{2} = 21$ Ergebnisse:

$$\Omega^* = \left\{ \begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ & & (3,3), & (3,4), & (3,5), & (3,6), \\ & & & (4,4), & (4,5), & (4,6), \\ & & & & (5,5), & (5,6), \\ & & & & & (6,6) \end{matrix} \right\}$$

Diese sind nicht gleichwahrscheinlich: $p(1,1) = \frac{1}{36}$ aber $p(1,2) = \frac{2}{36}$.

Beispiel: Große Straße beim Kniffel

Aufgabe: Wie wahrscheinlich ist eine große Straße beim Kniffel?

Beim Kniffel (oder Yahtzee) würfelt man mit fünf Würfeln. Eine große Straße besteht aus fünf aufeinanderfolgende Zahlen. Statt „einmal 5 Würfel“ kann man „5 mal einen Würfel“ werfen.

Lösung: Ergebnismenge $\Omega = \{1,2,3,4,5,6\}^5$ mit Gleichverteilung. In diesem Modell ist das Ereignis „große Straße“ die Teilmenge

$$S = \Big\{ \omega \in \Omega \;\Big|\; \{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\} = \{1,2,3,4,5\} \text{ oder } \{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\} = \{2,3,4,5,6\} \Big\}.$$

Anzahl $|S| = 2 \cdot 5! = 240$. Die Wahrscheinlichkeit ist demnach

$$\mathbf{P}(S) = \frac{|S|}{|\Omega|} = \frac{2 \cdot 5!}{6^5} = \frac{240}{7776} = \frac{5}{162} \approx 0.0308$$

Urnenmodell ohne Zurücklegen mit Reihenfolge

Urnenmodell „k aus n“ ohne Zurücklegen mit Berücksichtigung der Ziehungsreihenfolge. Die Ergebnismenge ist hier:

$$\Omega = \left\{ \omega \in \{1, \dots, n\}^k \;\middle|\; \omega_i \neq \omega_j \text{ für } i \neq j \right\}$$

Anzahl der Elemente:

$$|\Omega| = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

🙂 Alle Ergebnisse sind gleich wahrscheinlich (Laplace–Experiment). Die Elementarwahrscheinlichkeiten sind demnach:

$$p(\omega) = \frac{1}{|\Omega|} = \frac{(n-k)!}{n!}$$

Beispiel: Anzahl der Permutationen einer Menge mit n Elementen. Das Urnenmodell Ziehung „n aus n“ oZmR ergibt $n!$ Permutationen.

Urnenmodell ohne Zurücklegen ohne Reihenfolge

Urnenmodell „k aus n“ ohne Zurücklegen ohne Berücksichtigung der Ziehungsreihenfolge. Äquivalent hierzu: aufsteigende Anordnung, also

$$\begin{aligned}
\Omega &= \left\{ \omega \subseteq \{1, \dots, n\} \;\middle|\; |\omega| = k \right\} \\
&= \left\{ \{\omega_1, \omega_2, \dots, \omega_k\} \subseteq \{1, \dots, n\} \;\middle|\; \omega_1 < \omega_2 < \dots < \omega_k \right\} \\
&\cong \left\{ (\omega_1, \omega_2, \dots, \omega_k) \in \{1, \dots, n\}^k \;\middle|\; \omega_1 < \omega_2 < \dots < \omega_k \right\}
\end{aligned}$$

Anzahl der Elemente:

$$|\Omega| = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

🙂 Alle Ergebnisse sind gleich wahrscheinlich (Laplace–Experiment). Die Elementarwahrscheinlichkeiten sind demnach:

$$p(\omega) = \frac{k!(n-k)!}{n!}$$

Beispiel: Lotto entspricht dem Urnenmodell „6 aus 49“ oZoR
Dies ergibt $\binom{49}{6} = 13\,983\,816$ mögliche Ergebnisse.

Urnenmodell mit Zurücklegen mit Reihenfolge

Urnenmodell „k aus n“ mit Zurücklegen mit Berücksichtigung der Ziehungsreihenfolge. Die Ergebnismenge besteht aus allen k–Tupeln:

$$\Omega = \{1, \dots, n\}^k = \left\{ (\omega_1, \omega_2, \dots, \omega_k) \;\middle|\; \omega_1, \omega_2, \dots, \omega_k \in \{1, \dots, n\} \right\}$$

Die Anzahl der Elemente ist demnach die Potenz:

$$|\Omega| = \underbrace{n \cdot n \cdot \dots \cdot n}_{k \text{ mal}} = n^k$$

🙂 Alle Ergebnisse sind gleich wahrscheinlich (Laplace–Experiment). Die Elementarwahrscheinlichkeiten sind demnach:

$$p(\omega) = \frac{1}{n^k}$$

Beispiel: Anzahl der Ergebnisse bei k–maligem Würfeln.
Darstellung als Urnenmodell: Ziehung „k aus 6“ mZmR.
Dies ergibt 6^k mögliche Ergebnisse.

Urnenmodell mit Zurücklegen ohne Reihenfolge

Mögliche Wahl der Ergebnismenge (aufsteigende Reihenfolge):

$$\Omega = \left\{ \omega = (\omega_1, \dots, \omega_k) \in \{1, \dots, n\}^k \;\middle|\; \omega_1 \le \omega_2 \le \dots \le \omega_k \right\}$$

Die Anzahl der Elemente ist zunächst nicht direkt ersichtlich...
Genial-einfacher Rechentrick: Wir wählen eine geschickte Bijektion!

$$\Phi : \Omega \to \Omega' = \left\{ \omega' \in \{1, \dots, n+k-1\}^k \;\middle|\; \omega'_1 < \omega'_2 < \dots < \omega'_k \right\}$$
$$\omega \leftrightarrow \omega' \quad \text{mit} \quad \omega'_j = \omega_j + j - 1 \quad \text{für alle} \quad j = 1, \dots, k$$

Beispiele dieser Bijektion zur Illustration:

$$\Phi(1,1,1,1,1) = (1,2,3,4,5)$$
$$\Phi(1,1,2,4,4) = (1,2,4,7,8)$$

Diese raffinierte Sichtweise löst das Abzählproblem:

$$|\Omega| = |\Omega'| = \binom{n+k-1}{k}$$

⚠ Hier sind nicht alle Ergebnisse gleich wahrscheinlich!

Beispiel: Buben beim Skat

Aufgabe: Wie welcher Wkt erhalten Sie $0, 1, 2, 3, 4$ Buben beim Skat?

Lösung: Jeder Spieler erhält zufällig 10 der insgesamt 32 Karten. Das Kartenspiel enthält 4 Buben und 28 weitere Karten.

$$\begin{aligned}
\mathbf{P}(4 \text{ Buben}) &= \binom{4}{4}\binom{28}{6} \Big/ \binom{32}{10} = \frac{21}{3596} \approx 0.00583 \\
\mathbf{P}(3 \text{ Buben}) &= \binom{4}{3}\binom{28}{7} \Big/ \binom{32}{10} = \frac{66}{899} \approx 0.0734 \\
\mathbf{P}(2 \text{ Buben}) &= \binom{4}{2}\binom{28}{8} \Big/ \binom{32}{10} = \frac{2079}{7192} \approx 0.289 \\
\mathbf{P}(1 \text{ Bube}) &= \binom{4}{1}\binom{28}{9} \Big/ \binom{32}{10} = \frac{385}{899} \approx 0.428 \\
\mathbf{P}(0 \text{ Buben}) &= \binom{4}{0}\binom{28}{10} \Big/ \binom{32}{10} = \frac{1463}{7192} \approx 0.203
\end{aligned}$$

Hypergeometrische Verteilung

Aufgabe: In einer Urne liegen N Kugeln, davon genau K rote.
Wir wählen zufällig n der N Kugeln aus. (Ziehung oZoR)
Sei $A_k =$ „Es werden genau k der K roten Kugeln gezogen“.
Wie wahrscheinlich ist das Ereignis A_k?

Lösung: Die Wahrscheinlichkeit für das Ereignis A_k ist

$$\mathbf{P}(A_k) = \frac{|A_k|}{|\Omega|} = \binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n}.$$

Erläuterung: Das ist ein Laplace–Experiment mit $|\Omega| = \binom{N}{n}$.
Das Ereignis A_k setzt sich aus zwei unabhängigen Teilen zusammen:

- Auswahl von k aus K roten Kugeln: $\binom{K}{k}$ Möglichkeiten
- Auswahl von $n-k$ aus $N-K$ weiteren Kugeln: $\binom{N-K}{n-k}$

Also gilt $|A_k| = \binom{K}{k}\binom{N-K}{n-k}$. Dies ergibt die obige Wahrscheinlichkeit.

Hypergeometrische Verteilung

Satz U3A: hypergeometrische Verteilung

Wir betrachten eine **Stichprobe** ohne Zurücklegen ohne Reihenfolge: Gesamtgröße N, davon K mögliche Treffer, Stichprobengröße $n \le N$. Die Wahrscheinlichkeit $p(k)$ für k Treffer in der Stichprobe ist dann

$$p(k) = \binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n}.$$

Dies nennen wir die **hypergeometrische Verteilung**

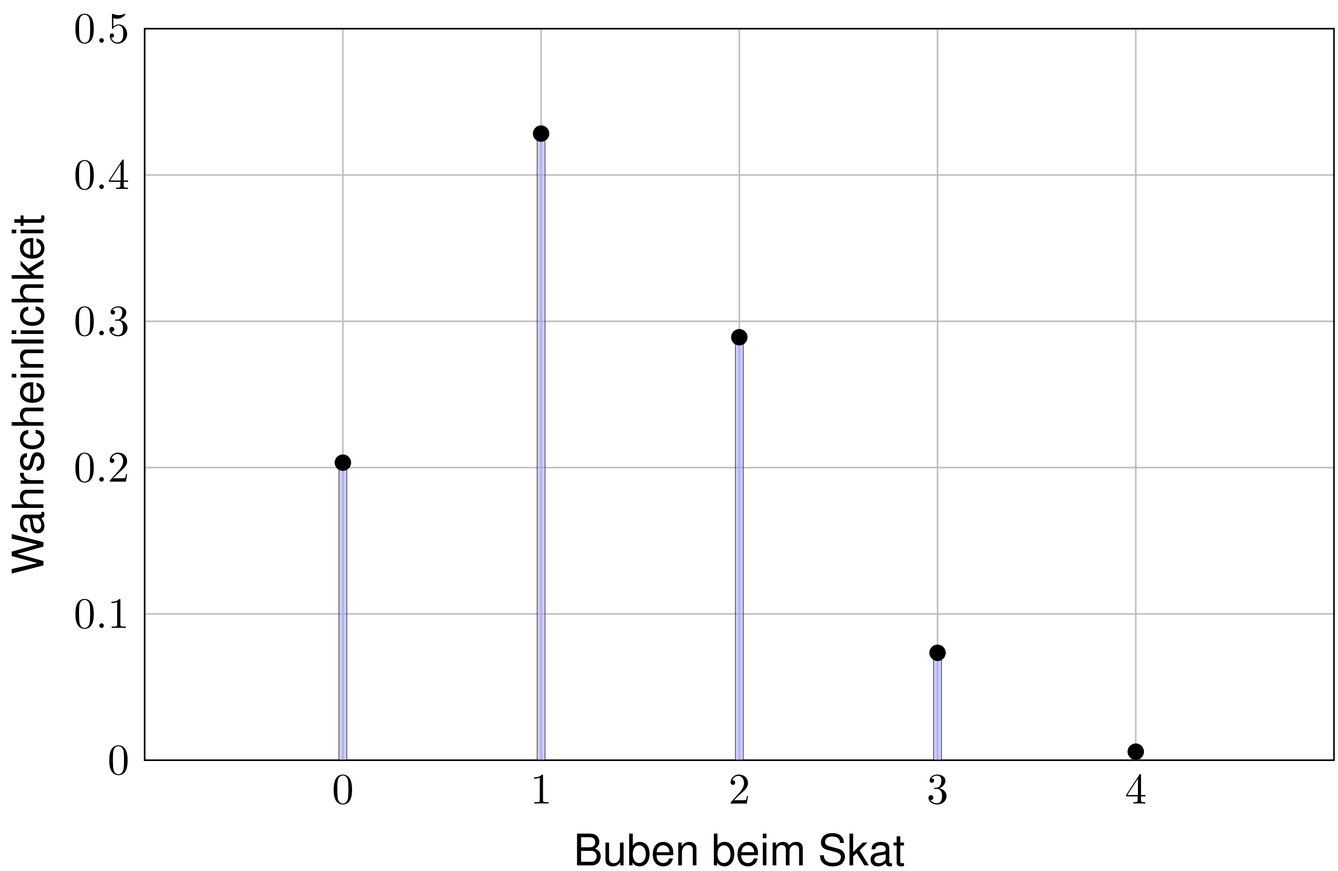
$$H(N,K,n) \,:\, \mathbb{N} \to [0,1] \,:\, k \mapsto p(k).$$

Dies ist eine WVerteilung, demnach gilt $p(0) + p(1) + \dots + p(n) = 1$.
Hieraus erhalten wir ohne weitere Mühe die **Vandermonde–Identität**:

$$\sum_{k=0}^{n} \binom{K}{k}\binom{N-K}{n-k} = \binom{N}{n}$$

Für $k < 0$ oder $k > n$ sowie $k > K$ oder $n-k > N-K$ gilt $p(k) = 0$.

Beispiel: $H(32,4,10)$–Verteilung beim Skat



☺ Nur in etwa jedem 200. Blatt erhält man alle vier Buben. Immerhin!

Beispiel: Lotto „6 aus 49“

Aufgabe: Gezogen werden 6 Zahlen aus 49.

(1) Wie viele Möglichkeiten gibt es insgesamt?

(2) Wie wahrscheinlich sind $k = 0, 1, 2, 3, 4, 5, 6$ Richtige?

Lösung: Hierzu muss man das Ziehungsverfahren präzisieren:

(1) Urnenmodell „6 aus 49“ ohne Zurücklegen ohne Reihenfolge.

Von insgesamt 49 Zahlen wählt der Spieler 6 aus.

Hierzu gibt es $\binom{49}{6} = 13\,983\,816$ Möglichkeiten.

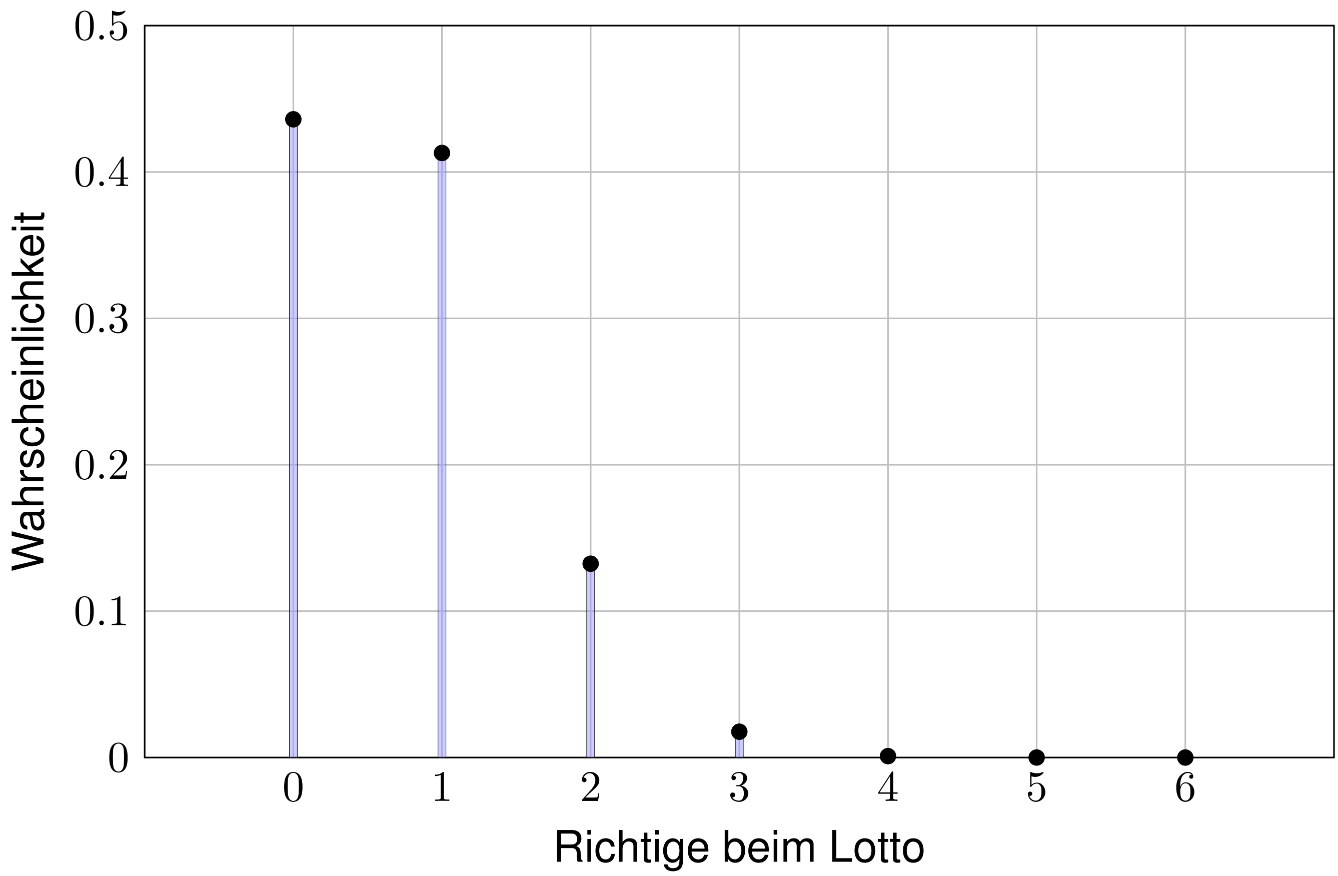
So sagt es auch die freundliche Stimme aus dem Radio bei der Werbung für Lotto 6 aus 49: „Gewinnchance 1 zu 14 Millionen. Spielteilnahme ab 18. Glücksspiel kann süchtig machen.“

(2) Dies wird durch die hypergeometrische Verteilung gegeben:

$$H(49,6,6)(k) = \binom{6}{k}\binom{43}{6-k} \Big/ \binom{49}{6}$$

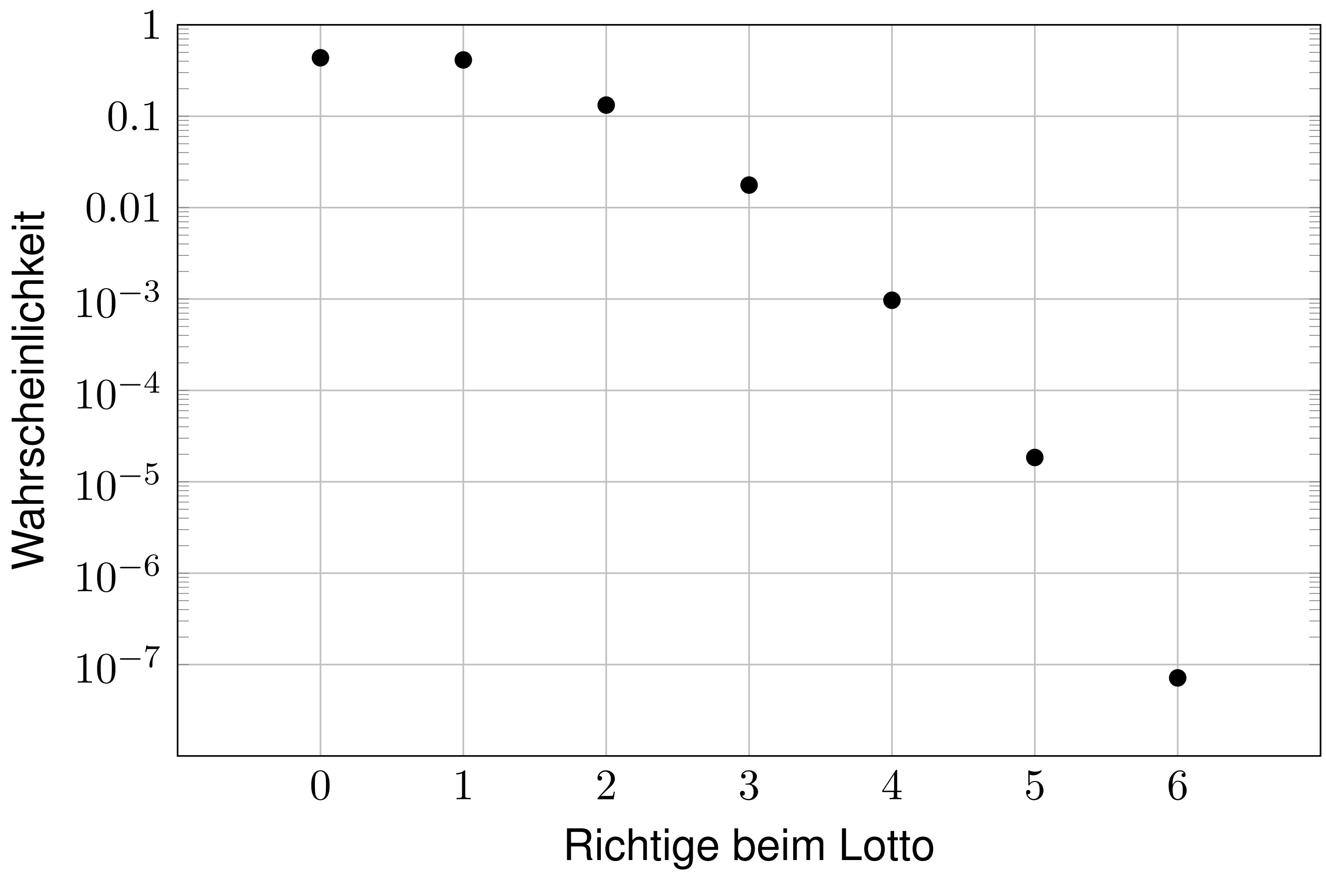
☺ Unsere passenden Werkzeuge erleichtern solche Abzählungen. Die folgenden Graphiken illustrieren die konkreten Werte.

Beispiel: $H(49,6,6)$–Verteilung beim Lotto



☹ Die Wkten für 4, 5, 6 Richtige sind in dieser Graphik kaum erkennbar.

Beispiel: $H(49,6,6)$–Verteilung beim Lotto



Eine logarithmische Skala ist hier wesentlich informativer!

Exkurs: Lohnt es sich, Lotto zu spielen?

Ich denke: Nein! Beim Lotto werden nur 50% der Einsätze als Gewinne ausgeschüttet: Ein zufälliger Tipp erhält also bei 1 Euro Einsatz im Durchschnitt nur 50 Cent Gewinn. Der Ziehungsmechanismus scheint recht zuverlässig, wie die eingangs angegebenen Statistiken andeuten. (Es ist wohl ein skurriler Zufall, dass ausgerechnet die 13 bislang die seltenste Zahl ist... Es ist ja auch nicht weiter verdächtig, dass die häufigste Zahl bislang die 38 ist.)

Selbst wenn die Gewinnwahrscheinlichkeiten gleichverteilt sind, die Gewinne sind es nicht! Durch eine geschickte Wahl des Tipps kann man den erwarteten Gewinn (geringfügig) erhöhen. Die Höhe der Auszahlung hängt nämlich stark davon ab, wie viele Spieler die gezogenen Zahlen getippt haben. Die Problematik wurde 1986 im französischen Parlament vom Abgeordneten Dr. Masson vorgetragen, der eine Offenlegung der Statistik forderte [Der Spiegel 19/1986]; die Abgeordneten konnten den Argumenten jedoch nicht folgen. Hier eine Kurzfassung:

Viele Spieler tippen Geburtstage (Monate $1,\dots,12$, Tage $1,\dots,31$) oder bevorzugen gewisse Anordnungen. Wird eine beliebte Kombination gezogen, so gibt es viele Gewinner, jede einzelne Gewinnausschüttung ist dann geringer. Hingegen kann man mit höheren Auszahlungen rechnen, wenn man Tipps wählt, die die anderen Spieler vernachlässigen. Mit einem gut gewählten Tipp kann man statt 50 Cent eine Erwartung von 60 bis 70 Cent erreichen [Krengel 2005, S. 11]. Ob mehr möglich ist, weiß ich nicht. 📖 Henze, Riedwyl: *How to win more.* A K Peters 1998.

Aber selbst mit einer ausgefuchsten Analyse des Spielverhaltens der anderen Spieler bleibt das Lottospielen vermutlich ein Verlustgeschäft. Nur eins ist sicher: Der Betreiber gewinnt immer. Boshaft nennt man solche staatlichen Lotterien deshalb auch eine „Steuer auf Dummheit“.

Die Binomialverteilung $B(n,t)$

Sie führen n **unabhängige Experimente** durch mit Trefferwkt $t \in [0,1]$.
Ergebnisraum $\Omega = \{0,1\}^n$, Produktraum über $\{1 = \text{Treffer}, 0 = \text{Niete}\}$.
Für $\omega \in \Omega$ mit k Treffern und $n-k$ Nieten ist die Wkt $t^k(1-t)^{n-k}$.
Es gibt genau $\binom{n}{k}$ solcher Ergebnisse $\omega \in \Omega$ mit k Treffern.

Satz U3B: Binomialverteilung

Die Wahrscheinlichkeit $p(k)$ für genau k Treffer ist gegeben durch

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

Dies nennen wir die **Binomialverteilung**

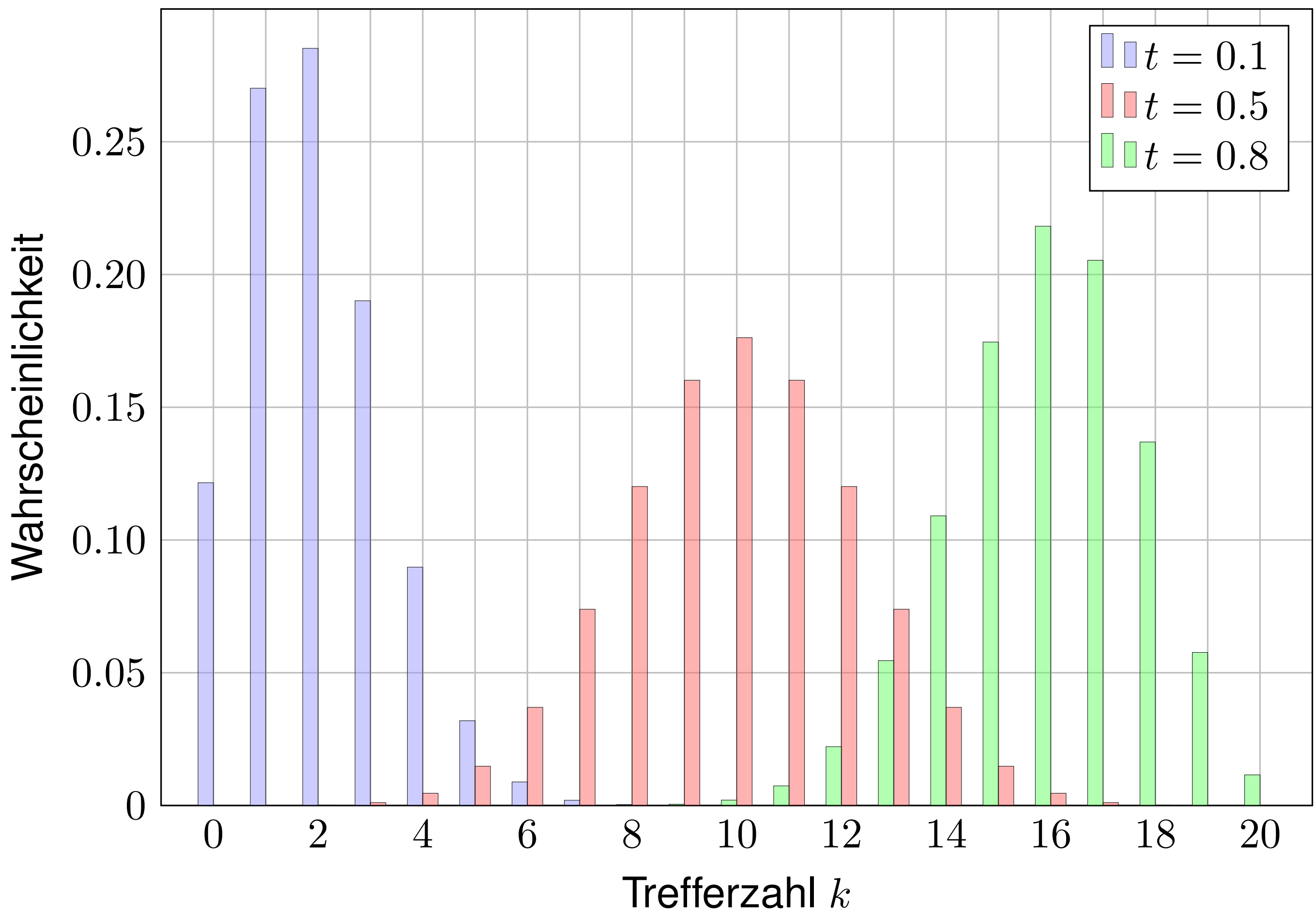
$$B(n,t) \;:\; \mathbb{N} \to [0,1] \;:\; k \mapsto p(k).$$

Aufgabe: Ist dies eine WVerteilung? **Lösung:** Ja, denn $p(k) \ge 0$ und

$$\sum_{k\in\mathbb{N}} p(k) \;=\; \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \;=\; \Big[t + (1-t)\Big]^n \;=\; 1.$$

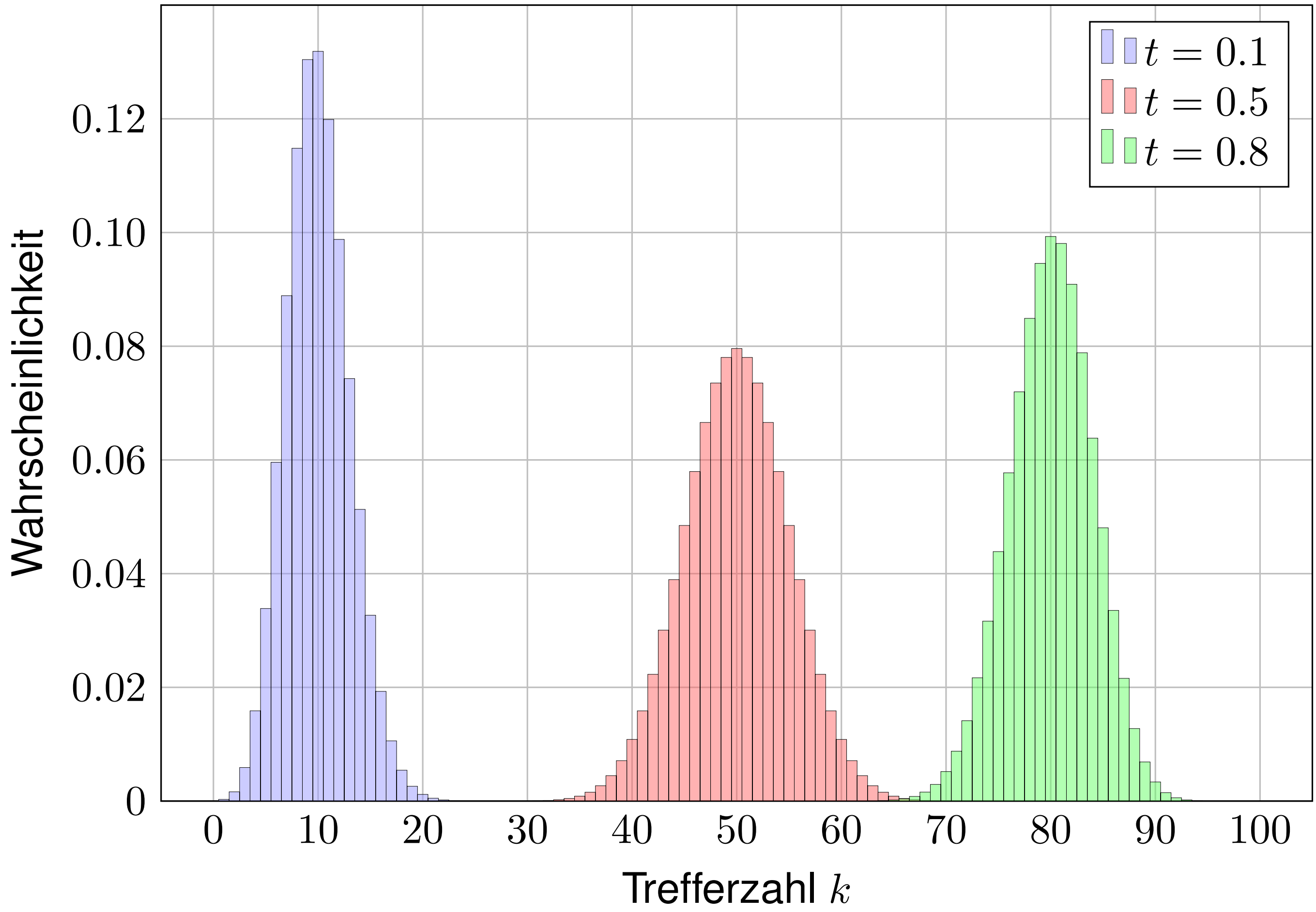
Der Träger von $B(n,t)$ ist $\{0,\dots,n\}$: Für $k<0$ oder $k>n$ gilt $p(k)=0$.

Beispiele zur Binomialverteilung: $B(20,t)$



Beispiele zur Binomialverteilung: $B(100, t)$

U311



Anwendung: Anzahl der Geburtstagskinder

U312

Aufgabe: Im Hörsaal sitzen 300 Personen (zufällig und unabhängig). Mit welcher Wahrscheinlichkeit haben $0, 1, 2, 3, \dots$ heute Geburtstag?

Lösung: Diese Anzahl $B(n, t)$–verteilt mit $n = 300$ und $t = 1/365$. Die Wahrscheinlichkeit für $0, 1, 2, 3, \dots$ Geburtstagskinder ist:

$$B(n,t)(0) = \binom{300}{0} \left(\frac{1}{365}\right)^0 \left(\frac{364}{365}\right)^{300} \approx 0.4391$$

$$B(n,t)(1) = \binom{300}{1} \left(\frac{1}{365}\right)^1 \left(\frac{364}{365}\right)^{299} \approx 0.3619$$

$$B(n,t)(2) = \binom{300}{2} \left(\frac{1}{365}\right)^2 \left(\frac{364}{365}\right)^{298} \approx 0.1486$$

$$B(n,t)(3) = \binom{300}{3} \left(\frac{1}{365}\right)^3 \left(\frac{364}{365}\right)^{297} \approx 0.0406$$

🙂 Für $B(n,t)(0)$ kennen wir bereits die Näherung $(1-t)^n \lessapprox e^{-nt}$. Mit $nt = 300/365 \approx 0.8219$ finden wir $B(n,t)(0) \approx e^{-0.819} \approx 0.4396$.

Von der Binomial- zur Poisson–Verteilung

Die Binomialverteilung $B(n,t)$ ist für große n mühsam zu berechnen.

Aufgabe: Wir suchen eine praktische Näherung für n groß und t klein. Sei $\lambda > 0$ gegeben und $t = \lambda/n$. Berechnen Sie für $n \to \infty$ den Limes der Binomialverteilungen $B(n, \lambda/n)$. **Lösung:** Geduldig ausrechnen:

$$\begin{aligned}
B\left(n, \tfrac{\lambda}{n}\right)(k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to \mathrm{e}^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \quad \to \quad \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}
\end{aligned}$$

Beispiel: Für $n = 300$ und $t = 1/365$ gilt $\lambda = nt = 300/365 \approx 0.8219$. Statt exakt $B(n,t)(k)$ berechnen wir bequemer $p(k) = \mathrm{e}^{-\lambda} \cdot \lambda^k/k!$:

$$p(0) \approx 0.4396, \quad p(1) \approx 0.3613, \quad p(2) \approx 0.1485, \quad p(3) \approx 0.0407.$$

Die Poisson–Verteilung $P(\lambda)$

Definition U3C: Poisson–Verteilung

Die **Poisson–Verteilung** $P(\lambda)$ für $\lambda \geq 0$ ist gegeben durch

$$P(\lambda) \; : \; \mathbb{N} \to [0,1] \; : \; k \mapsto p(k) = \mathrm{e}^{-\lambda} \cdot \lambda^k/k!.$$

Aufgabe: Ist dies eine WVerteilung? **Lösung:** Ja, denn $p(k) \geq 0$ und

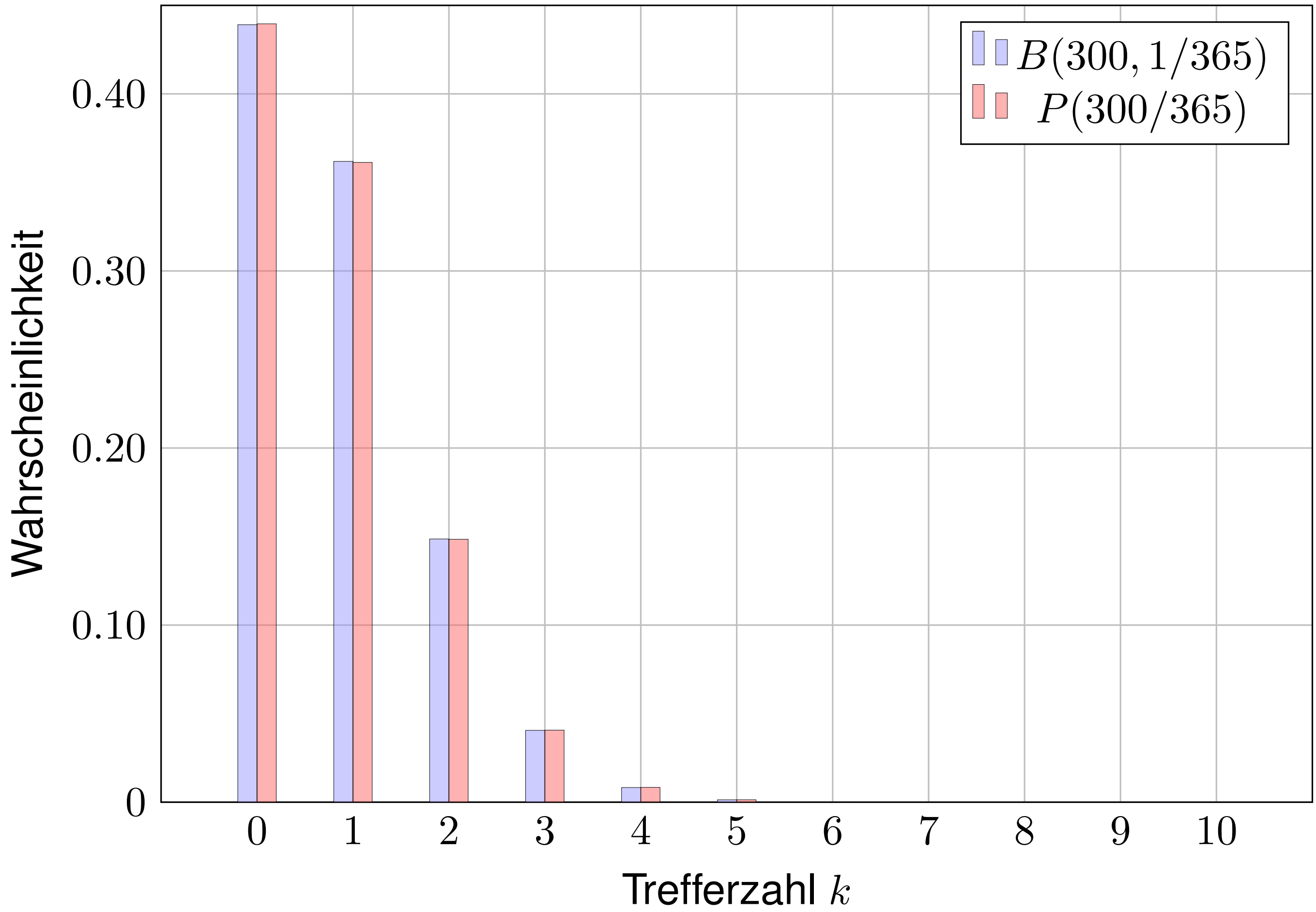
$$\sum_{k=0}^{\infty} p(k) = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \mathrm{e}^{\lambda} = 1.$$

🙂 So lässt sich die Poisson–Verteilung besonders leicht merken. Die Terme $\lambda^k/k!$ entsprechen der Exponentialreihe $\mathrm{e}^\lambda = \sum_{k=0}^\infty \lambda^k/k!$.

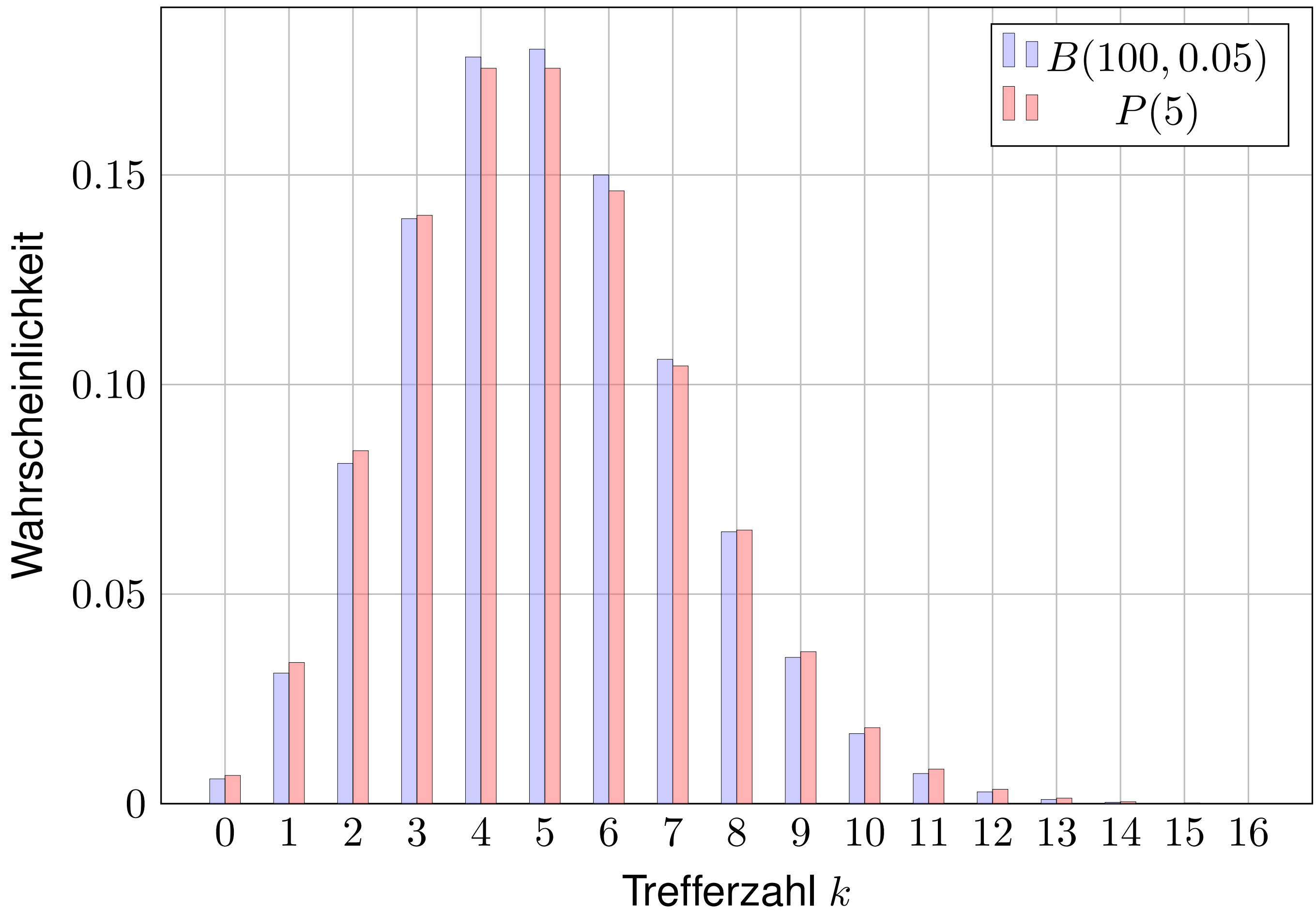
🙂 Der Grenzübergang $B(n, \lambda/n) \to P(\lambda)$ erklärt die Bedeutung: $P(\lambda)$ entspricht der Häufigkeit von unabhängigen Ereignissen geringer Wahrscheinlichkeit bei sehr vielen Wiederholungen.

Wie gut ist die Approximation von $B(n,t)$ durch $P(nt)$? Auch dies wollen wir nun quantitativ beantworten!

Vergleich von Binomial- und Poisson–Verteilung



Vergleich von Binomial- und Poisson–Verteilung



Abstand von Wahrscheinlichkeitsmaßen

Oft wollen oder müssen wir zwei WMaße $\mathbf{P}_0$ und $\mathbf{P}_1$ auf Ω vergleichen: Ist $\mathbf{P}_0$ mühsam aber $\mathbf{P}_1$ bequem, so wollen wir $\mathbf{P}_0$ durch $\mathbf{P}_1$ ersetzen.

Die beiden WMaße $\mathbf{P}_0$ und $\mathbf{P}_1$ ordnen jedem Ereignis $A \subseteq \Omega$ die Wkten $\mathbf{P}_0(A)$ bzw. $\mathbf{P}_1(A)$ zu. Der totale Abstand zwischen $\mathbf{P}_0$ und $\mathbf{P}_1$ ist das Supremum, die größtmögliche Abweichung, die hierbei auftreten kann:

Definition U3D: totaler Abstand zweier WMaße

Der **totale Abstand** von zwei diskreten WMaße $\mathbf{P}_0$ und $\mathbf{P}_1$ auf Ω ist:

$$\big\|\mathbf{P}_0 - \mathbf{P}_1\big\| \;:=\; \sup_{A\subseteq\Omega}\big|\mathbf{P}_0(A) - \mathbf{P}_1(A)\big| \;=\; \frac{1}{2}\sum_{\omega\in\Omega}\big|\mathbf{P}_0(\{\omega\}) - \mathbf{P}_1(\{\omega\})\big|$$

Diese Fehlerschranke nutzen wir bei näherungsweisen Rechnungen: Ist der Abstand klein genug, etwa $\|\mathbf{P}_0 - \mathbf{P}_1\| \le \varepsilon = 0.5 \cdot 10^{-3}$, so können wir $\mathbf{P}_0(A)$ durch $\mathbf{P}_1(A)$ ersetzen und machen dabei schlimmstenfalls einen Fehler von ε, das heißt, es gilt $\mathbf{P}_0(A) = \mathbf{P}_1(A) + \delta$ mit $|\delta| \le \varepsilon$.

Der totale Abstand ist (bis auf einen Faktor $1/2$) die Summe über alle punktweisen Abstände: Die linke Seite $\sup_{A\subseteq\Omega}$ ist leicht zu verstehen, die rechte Seite $\sum_{\omega\in\Omega}$ ist leicht zu berechnen. Diese Umformulierung ist oft leichter zugänglich: Es genügt, die absolute Differenz über alle Ergebnisse zu addieren. Die genaue Rechnung U318 beweist dies und erklärt den Faktor $1/2$.

Abstand von Wahrscheinlichkeitsmaßen

Beweis der Umformulierung: Wir betrachten $f:\Omega \to \mathbb{R}$ mit $f(\omega) = \mathbf{P}_0(\{\omega\}) - \mathbf{P}_1(\{\omega\})$ und beweisen $\|\mathbf{P}_0 - \mathbf{P}_1\| \le \frac{1}{2}\sum_{\omega\in\Omega}|f(\omega)|$ sowie $\|\mathbf{P}_0 - \mathbf{P}_1\| \ge \frac{1}{2}\sum_{\omega\in\Omega}|f(\omega)|$.

(1) Für jede beliebige Teilmenge $A \subseteq \Omega$ gilt:

$$\begin{aligned}
\mathbf{P}_0(A) - \mathbf{P}_1(A) &= \sum_{\omega\in A} f(\omega) \;\le\; \sum_{\omega\in A} f^+(\omega) \;\le\; \sum_{\omega\in\Omega} f^+(\omega) \\
&= \sum_{\omega\in\Omega} \frac{1}{2}\Big[|f(\omega)| + f(\omega)\Big] \;=\; \frac{1}{2}\sum_{\omega\in\Omega}\big|f(\omega)\big|
\end{aligned}$$

Die letzte Gleichung verdanken wir $\sum_{\omega\in\Omega} f(\omega) = \mathbf{P}_0(\Omega) - \mathbf{P}_1(\Omega) = 0$. Symmetrisch gilt $\mathbf{P}_1(A) - \mathbf{P}_0(A) \le \frac{1}{2}\sum_{\omega\in\Omega}\big|f(\omega)\big|$. Insgesamt folgt $|\mathbf{P}_0(A) - \mathbf{P}_1(A)| \le \frac{1}{2}\sum_{\omega\in\Omega}|f(\omega)|$.

Nach Definition ist $\|\mathbf{P}_0 - \mathbf{P}_1\|$ die kleinste obere Schranke von $|\mathbf{P}_0(A) - \mathbf{P}_1(A)|$ über alle $A \subseteq \Omega$, also $\|\mathbf{P}_0 - \mathbf{P}_1\| \le \frac{1}{2}\sum_{\omega\in\Omega}|f(\omega)|$. Wir zeigen nun die umgekehrte Ungleichung.

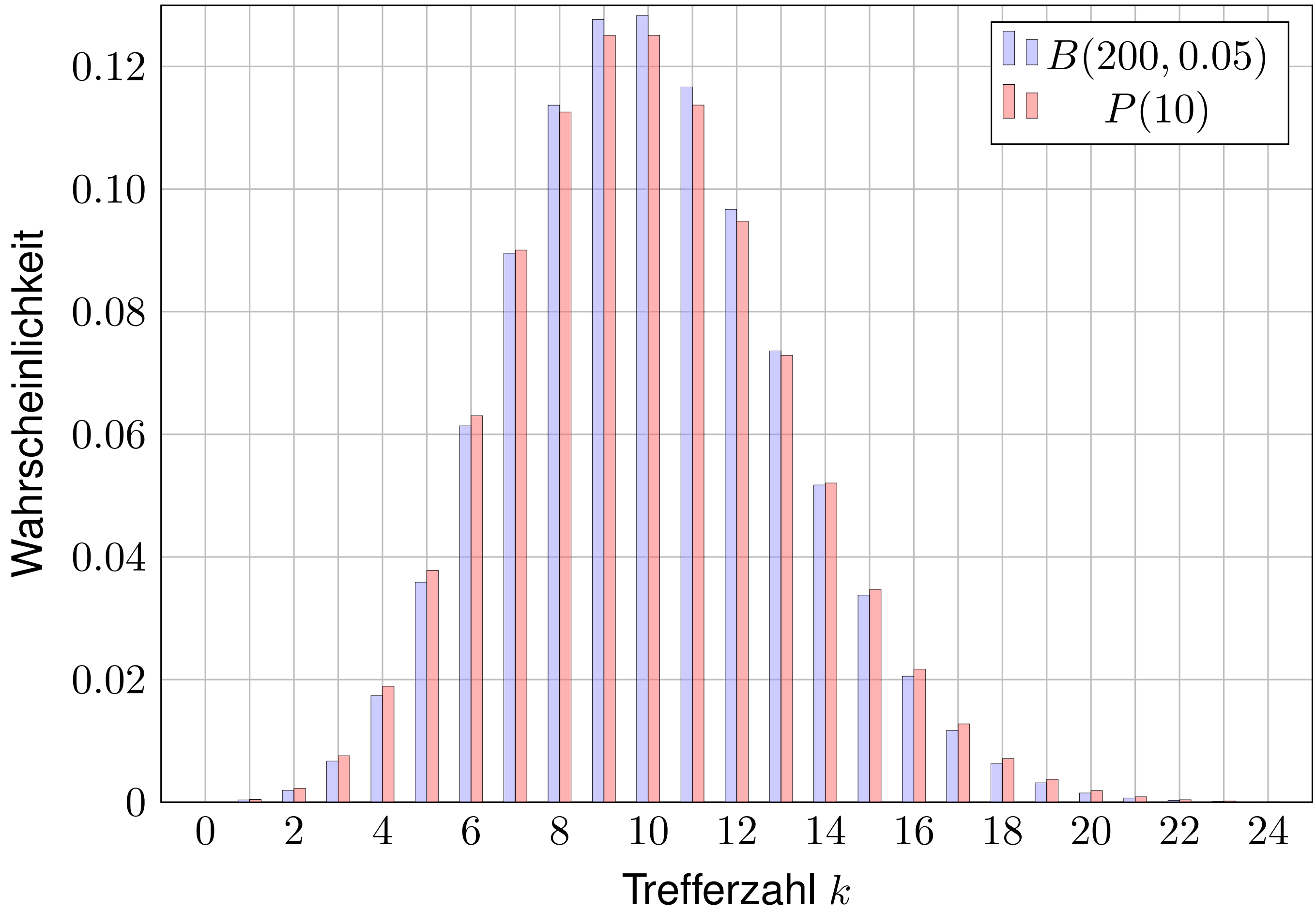
(2) Für die beiden Mengen $A = \{\,\omega\in\Omega \mid f(\omega) > 0\,\}$ und $B = \{\,\omega\in\Omega \mid f(\omega) < 0\,\}$ gilt

$$0 = \sum_{\omega\in\Omega} f(\omega) \;=\; \sum_{\omega\in A} f(\omega) + \sum_{\omega\in B} f(\omega) \qquad \text{und somit}$$

$$0 \le \sum_{\omega\in\Omega}\big|f(\omega)\big| = \sum_{\omega\in A} f(\omega) - \sum_{\omega\in B} f(\omega) = 2\big[\mathbf{P}_0(A) - \mathbf{P}_1(A)\big] \le 2\big\|\mathbf{P}_0 - \mathbf{P}_1\big\|.$$

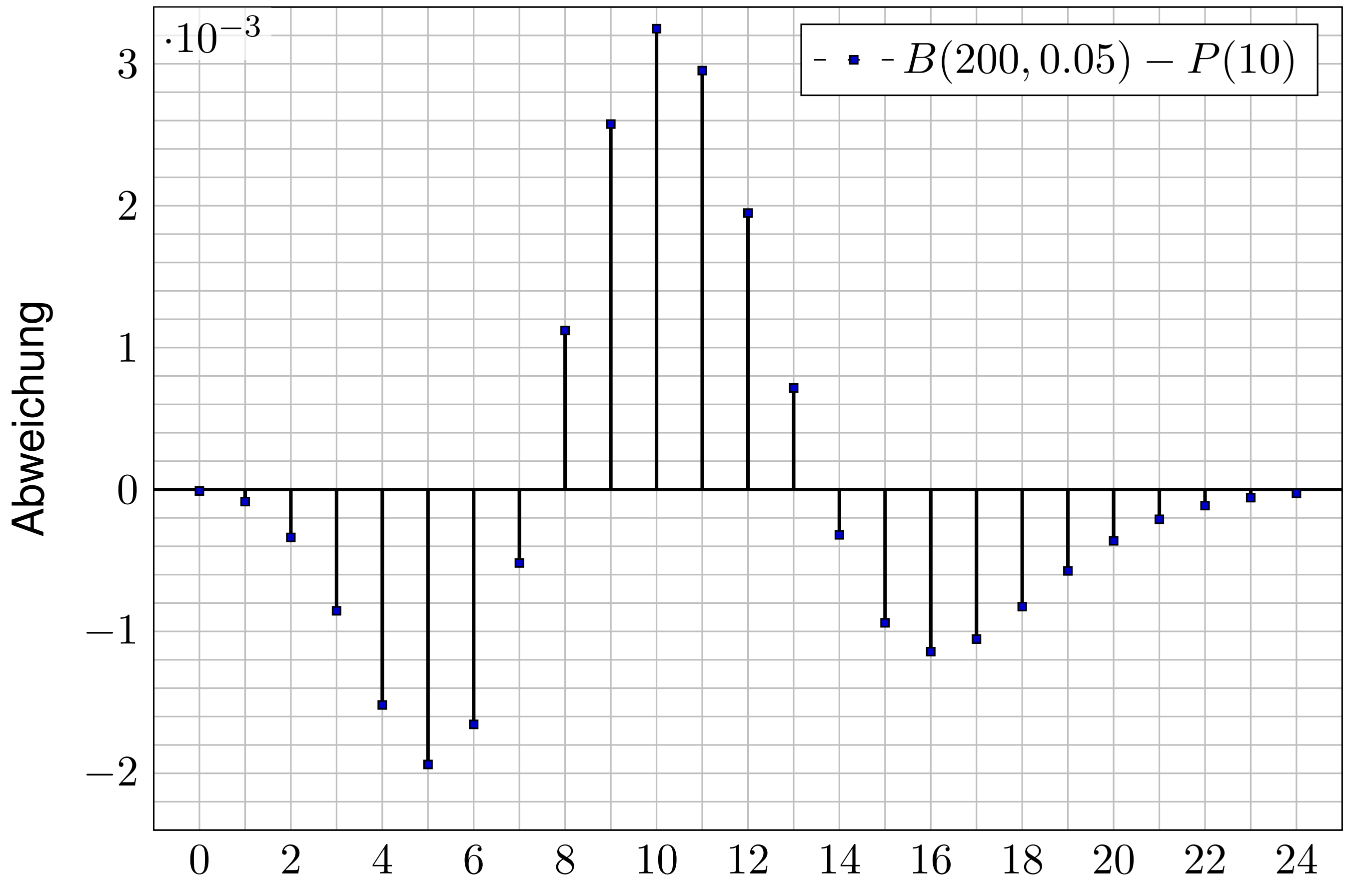
Dies zeigt $\|\mathbf{P}_0 - \mathbf{P}_1\| \ge \frac{1}{2}\sum_{\omega\in\Omega}|f(\omega)|$. Die Ungleichungen (1) und (2) beweisen Gleichheit. Unsere Rechnung zeigt zudem, dass das Supremum für A (und B) tatsächlich angenommen wird.

Vergleich von Binomial- und Poisson–Verteilung



Vergleich von Binomial- und Poisson–Verteilung

Wie gut ist diese Näherung? Punktweise bis auf 0.0033 genau!



Poissons Gesetz der kleinen Zahlen

Für große n ist die Binomialverteilung $B(n,t)$ mühsam. Bei kleinem t können wir sie durch die bequemere Poisson–Verteilung $P(nt)$ ersetzen.

Satz U3E: Poissons Gesetz der kleinen Zahlen

(1) Für $n \to \infty$ gilt die punktweise Konvergenz $B(n, \lambda/n) \to P(\lambda)$.

(2) Es gilt $B(n,t) \approx P(nt)$ mit Fehlerschranke für den totalen Abstand:

$$\boxed{\big\|B(n,t) - P(nt)\big\| \le nt^2 = \lambda^2/n \searrow 0}$$

Hierbei ist $\lambda = nt$ der Erwartungswert und Parameter für $P(\lambda)$.

☺ Dank dieses Satzes können wir den Fehler bequem abschätzen!

Z.B. liegen $B(200, 0.05)$ und $P(10)$ höchstens 0.5 auseinander, hingegen $B(2000, 0.005)$ und $P(10)$ noch höchstens 0.05, $B(20\,000, 0.0005)$ und $P(10)$ nur noch höchstens 0.005.

Gilt der Fehler als klein genug, so kann man die mühsame Binomialverteilung durch die wesentlich bequemere Poisson–Verteilung ersetzen. Den Grenzübergang haben wir oben schon ausgerechnet U313. Die Fehlerabschätzung des totalen Abstandes ist etwas trickreicher U435. Wir verallgemeinern dies in Satz U4C und rechnen es anschließend ausführlich und elegant nach.

Anwendungsbeispiel: Pixelfehler

Aufgabe: Bei der Herstellung eines Displays mit $n = 1920 \times 1080$ Pixeln ist jedes mit Wkt $t = 1/2\,764\,800$ defekt, alle unabhängig voneinander.

(1) Mit welcher Wkt p hat ein Display mehr als 2 Pixelfehler? (Garantie)

(2) Welche Verteilung gilt hier exakt? Welche gilt näherungsweise? Wie groß ist der Approximationsfehler, d.h. der totale Abstand?

Lösung: (2) Exakt $B(n,t)$, näherungsweise $P(\lambda)$ mit $\lambda = nt = 0.75$. Der totale Abstand $\|B(n,t) - P(\lambda)\| \le \lambda^2/n < 10^{-6}$ ist extrem klein.
☺ Die Näherung ist sehr genau und rechtfertigt die Rechnung in (1).

(1) Statt der exakten Verteilung nutzen wir die bequeme Näherung:

$$\begin{aligned} 1 - p &= \binom{n}{0} t^0 (1-t)^n + \binom{n}{1} t^1 (1-t)^{n-1} + \binom{n}{2} t^2 (1-t)^{n-2} \\ &\approx \mathrm{e}^{-\lambda} \left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} \right) \approx 0.95949 \end{aligned}$$

Garantiefall: Mehr als 2 fehlerhafte Pixel treten mit Wkt $p \approx 0.04051$ auf.
☺ Die Poisson–Näherung vereinfacht unsere Rechnung spürbar.

Vergleich von hypergeometrisch zu binomial

Stichprobe: Gesamtgröße N, davon K Treffer, Stichprobengröße n.
Die Trefferzahl k folgt hier der **hypergeometrischen Verteilung**

$$H(N,K,n)(k) = \binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n}.$$

Stichprobe mit Zurücklegen: Stichprobengröße n, Trefferwkt $t = \frac{K}{N}$.
Die Trefferzahl k folgt in diesem Fall der **Binomialverteilung**

$$B(n,t)(k) = \binom{n}{k} t^k (1-t)^{n-k}.$$

Ist N groß gegenüber n, so erwarten wir $H(N,K,n)(k) \approx B(n,t)(k)$.

Anschaulich: Zurücklegen oder nicht macht kaum einen Unterschied.

Diese Idee überprüfen wir anhand von Beispielen und Graphiken für endliches N sowie durch die Berechnung des Grenzwertes für $N \to \infty$.

Der abschließende Satz U3F gibt explizite handfeste Fehlerschranken: Der totale Abstand von $H(N,K,n)$ und $B(n,K/N)$ ist kleiner als n/N.

Vergleich von hypergeometrisch zu binomial

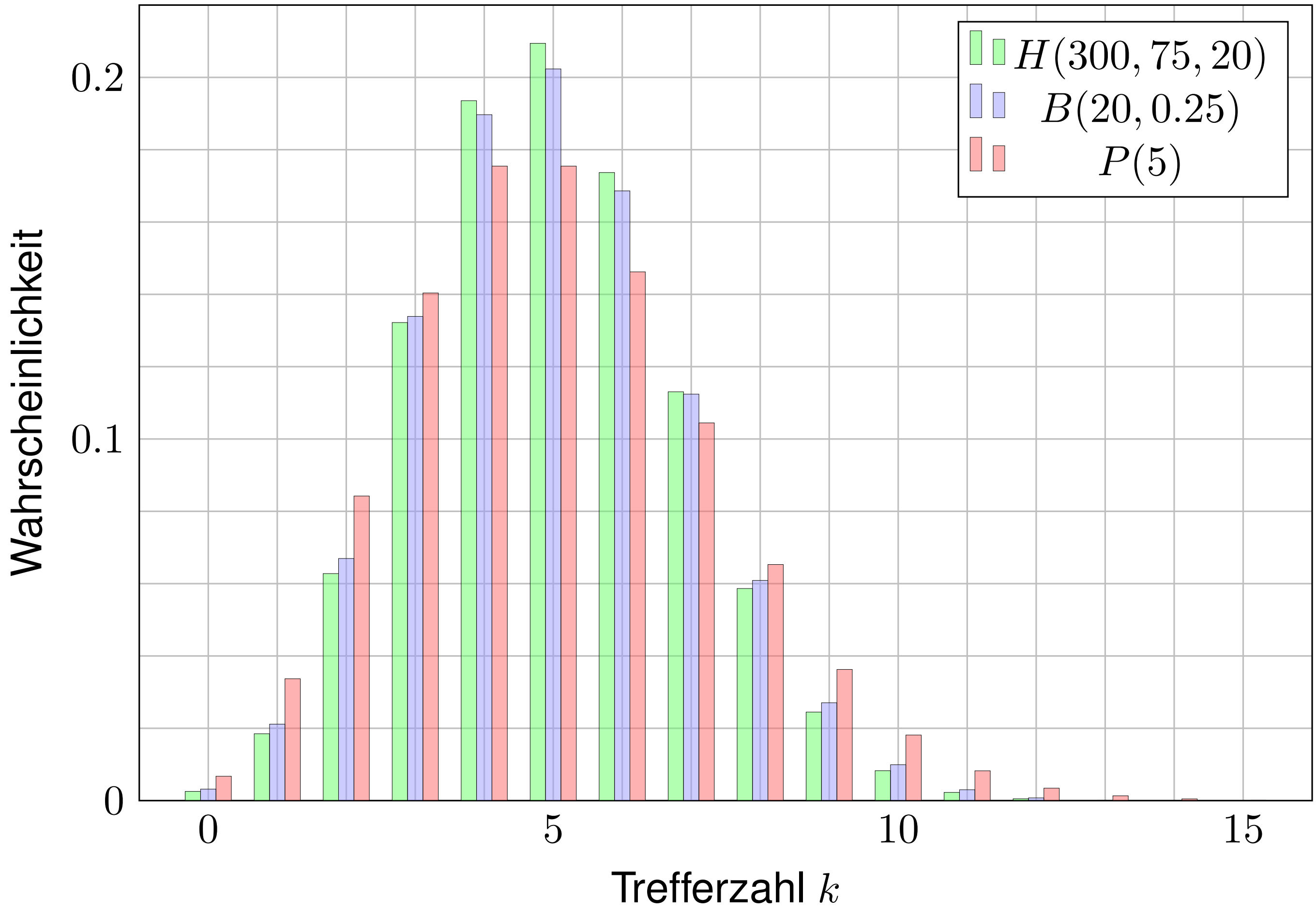
Aufgabe: Für $N \to \infty$ und $K/N \to t$ zeige man $H(N,K,n) \to B(n,t)$.

Lösung: Ausschreiben und geduldig vereinfachen:

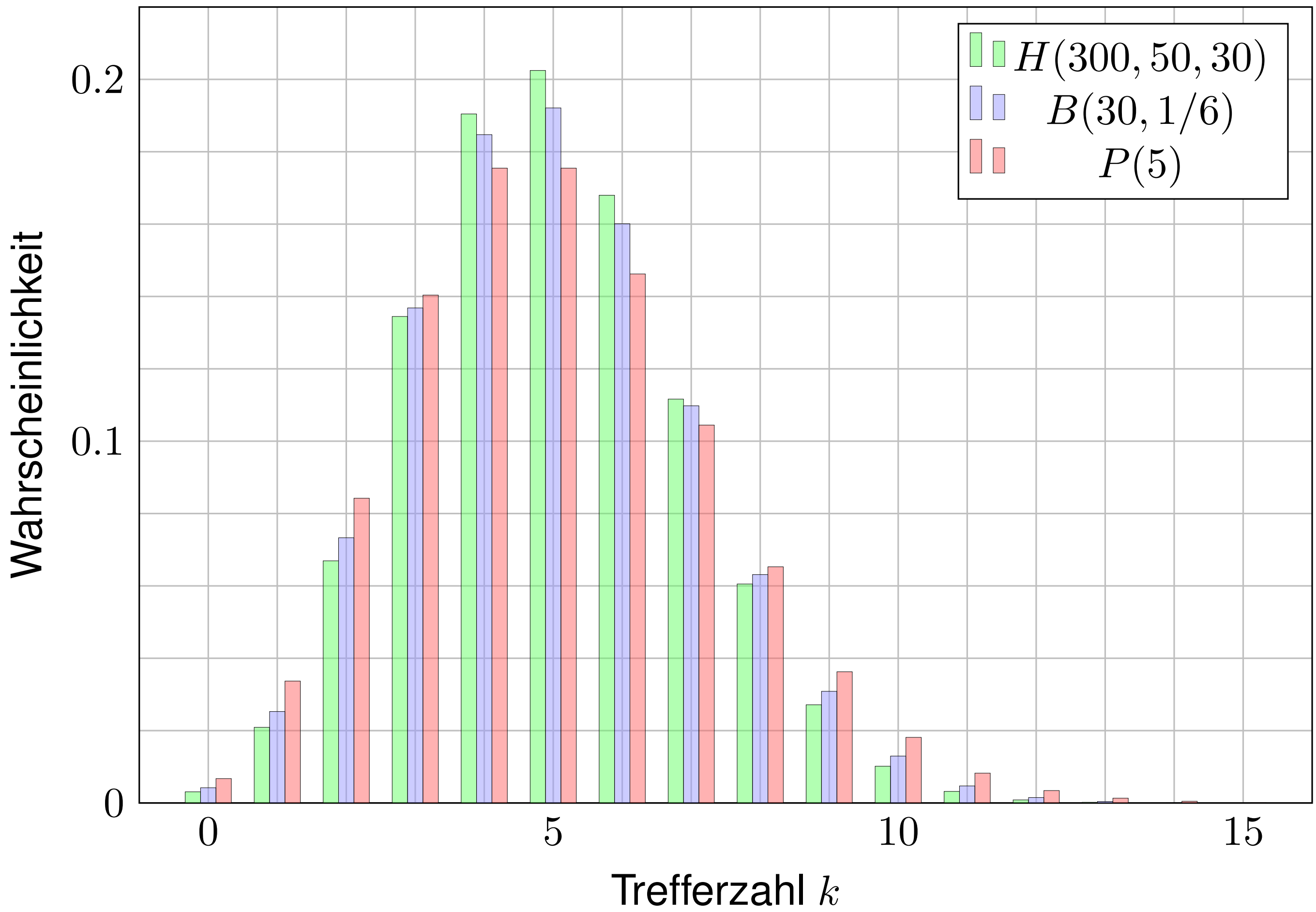
$$\begin{aligned}
\binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n} &= \frac{K^{\underline{k}}}{k!} \frac{(N-K)^{\underline{n-k}}}{(n-k)!} \frac{n!}{N^{\underline{n}}} \\
&= \binom{n}{k} \frac{K^{\underline{k}}}{N^{\underline{k}}} \frac{(N-K)^{\underline{n-k}}}{(N-k)^{\underline{n-k}}} \\
&\sim \binom{n}{k} \left(\frac{K}{N}\right)^k \left(\frac{N-K}{N}\right)^{n-k} \\
&\to \binom{n}{k} t^k (1-t)^{n-k}
\end{aligned}$$

Wie gut ist die Approximation von $H(N,K,n)$ durch $B(n,K/N)$?
Die Güte der Approximation ist für die Praxis wesentlich, leider lässt sie sich aus unserer einfachen Grenzwertrechnung noch nicht ablesen.
Der folgende Satz präzisiert dies durch eine explizite Fehlerschranke.

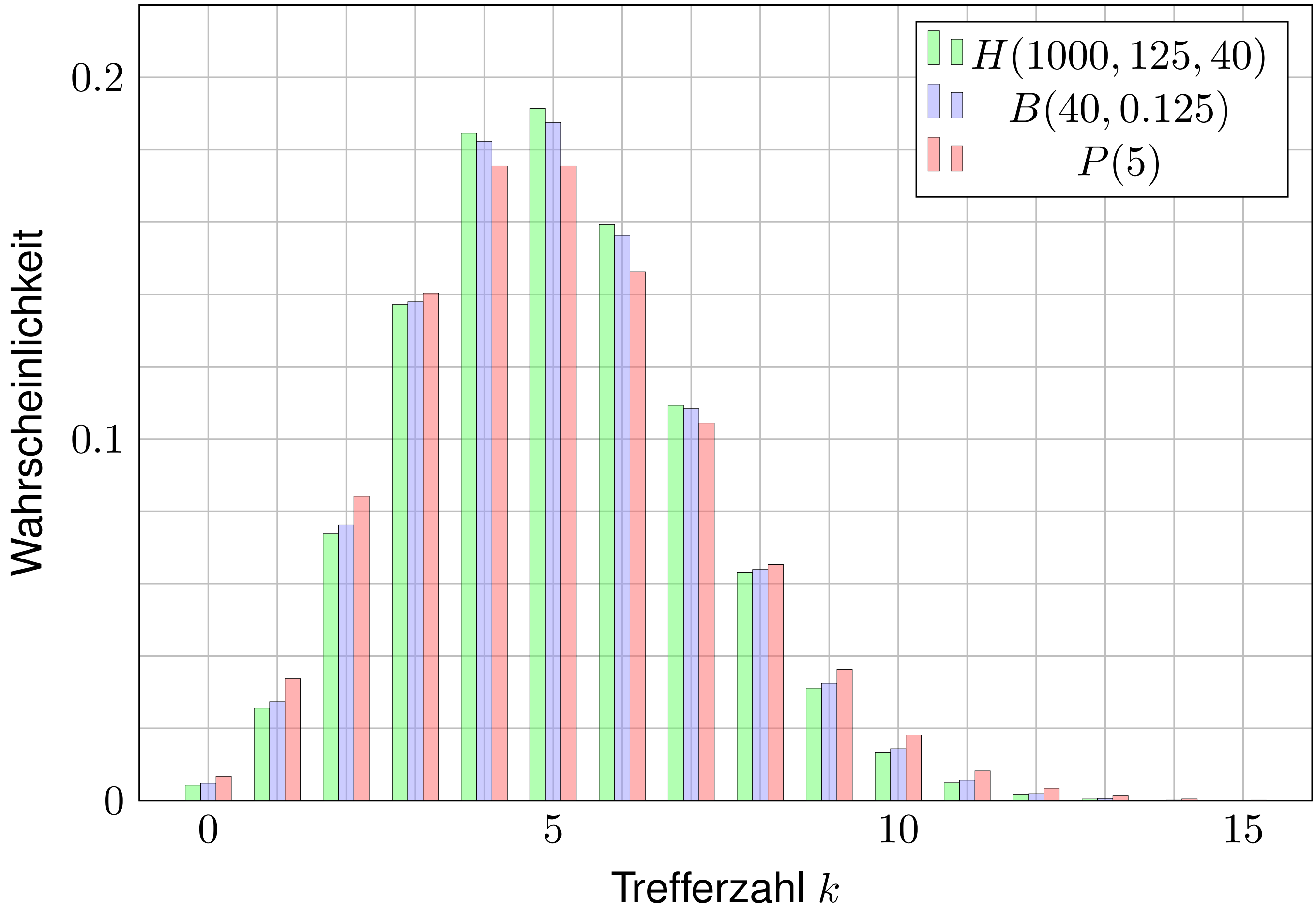
Vergleich: hypergeometrisch, binomial, Poisson



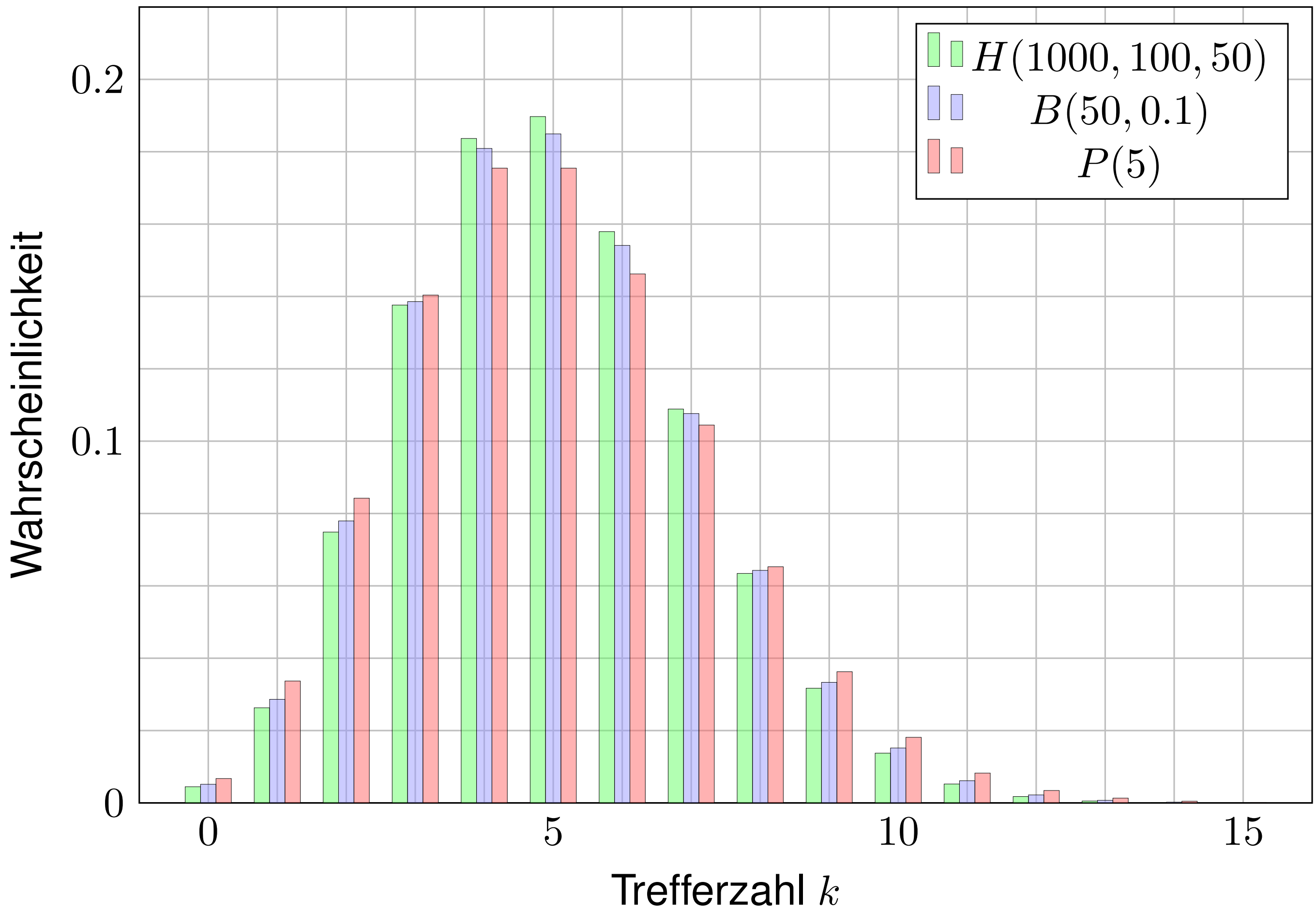
Vergleich: hypergeometrisch, binomial, Poisson



Vergleich: hypergeometrisch, binomial, Poisson



Vergleich: hypergeometrisch, binomial, Poisson



Abstand von hypergeometrisch zu binomial

Für große Stichproben möchten wir die hypergeometrische Verteilung $H(N,K,n)$ durch die Binomialverteilung $B(n,k/N)$ ersetzen.

Satz U3F: Abstand von hypergeometrisch zu binomial

(1) Für $N \to \infty$ und $K/N \to t$ gilt Konvergenz $H(N,K,n) \to B(n,t)$.

(2) Genauer gilt bei $K > n$ folgende Schranke für den totalen Abstand:

$$\boxed{\big\|H(N,K,n) - B(n,K/N)\big\| \ < \ n/N \quad \searrow \quad 0}$$

☺ Dank dieses Satzes können wir den Fehler bequem abschätzen.

Z.B. liegen $H(1000,100,50)$ und $B(50,0.1)$ höchstens 0.05 auseinander, hingegen $H(10000,1000,50)$ und $B(50,0.1)$ nur noch höchstens 0.005.

Innerhalb der genannten Fehlerschranken nutzen wir die Näherungen

$$\boxed{H(N,K,n) \approx B(n,K/N) \quad \text{und} \quad B(n,t) \approx P(nt).}$$

Die Grenzübergänge haben wir oben ausgerechnet. Die Fehlerabschätzung ist schwieriger und wird hier nicht ausgeführt. Ist der Fehler klein genug, so kann man die sehr mühsame hypergeometrische Verteilung durch die weniger mühsame Binomialverteilung ersetzen.

Anwendungsbeispiel zu Stichproben

In einer Population von $N = 500\,000$ Tieren sind $K = 25\,000$ erkrankt. Zwecks Untersuchung werden $n = 200$ Tiere zufällig ausgewählt.

Aufgabe: (1) Welche Verteilung hat die Anzahl k der kranken Tiere?
(2) Welche Verteilungen bieten sich als bequeme Näherungen an?
(3) Wie gut sind diese Näherungen? Sind sie bis auf $0.5 \cdot 10^{-3}$ genau?

Lösung: (1) Die Anzahl der kranken Tiere ist $H(N,K,n)$–verteilt.

(2) Es wäre bequemer, mit $B(n,K/N)$ oder $P(nK/N)$ zu rechnen.

(3) Beim Übergang von $H(N,K,n)$ zu $B(n,t)$ mit $t = K/N = 0.05$ machen wir einen totalen Fehler von $< n/N = 0.0004$. Das reicht.

Beim Übergang von $B(n,t)$ zu $P(\lambda)$ mit $\lambda = nt = 10$ machen wir einen totalen Fehler von $\le \lambda^2/n = 0.5$. Das scheint allzu grob: Vorsicht!

$$\begin{aligned}
H(N,K,n):&\quad 8 \mapsto 0.11373, && 9 \mapsto 0.12771, && 10 \mapsto 0.12838,\\
B(n,K/N):&\quad 8 \mapsto 0.11372, && 9 \mapsto 0.12769, && 10 \mapsto 0.12836,\\
P(nK/N):&\quad 8 \mapsto 0.11260, && 9 \mapsto 0.12511, && 10 \mapsto 0.12511
\end{aligned}$$

Die Poisson–Verteilung ist in diesen Punkten immerhin bis auf $0.5 \cdot 10^{-2}$ genau. Die Fehler in verschiedenen Punkten können sich jedoch addieren; das Supremum ist der totale Abstand U3D.

Anwendungsbeispiel zur Wareneingangsprüfung

Firma Schlaule kauft einen billigen Restposten von $50\,000$ Bauteilen. Der Lieferant sichert vertraglich zu, dass höchstens $2\,500$ defekt sind. Als Eingangsprüfung nehmen Sie eine Stichprobe von 100 Teilen und finden 6 defekte. Ihr Chef folgert: „Der Vertrag wurde nicht eingehalten."

Aufgabe: (1) Kann der Chef auch irren? Wenn ja, mit welcher Wkt? Er wird Ihre Rechnung anzweifeln; begründen Sie deshalb ausführlich! Welche Näherungen sind möglich, welche Fehlergrenzen gelten? W204

(2) Ihr Chef will nun den Lieferanten verklagen. Sollte sich der Vorwurf als falsch erweisen, wären die Kosten enorm. Die Anwältin rät daher zu einer erneuten Stichprobe und verlangt einen Test mit 97% Sicherheit. Bei welchen Stichprobenergebnissen können Sie dies garantieren?

Lösung: Wir betrachten zunächst den extremen Fall: $K = 2\,500$ defekte von $N = 50\,000$ Teilen, also $t = K/N = 5\%$. Eine zufällige Stichprobe von $n = 100$ ohne Zurücklegen enthält k defekte mit Wkt $H(N,K,n)(k)$.

Näherung $B(n,t)$ mit Fehler $< n/N = 0.002$: Das scheint akzeptabel.

Näherung $P(5)$ mit Fehler $\le nt^2 = 0.25$: Hier ist Vorsicht geboten!

Anwendungsbeispiel zur Wareneingangsprüfung

Wir berechnen die drei Verteilungen, die jeweils zweite Zeile kumuliert.

k	0	1	2	3	4	5	6	7	8	9
$H(N,K,n)$	.0059	.0311	.0811	.1395	.1782	.1802	.1502	.1061	.0648	.0348
Kumul	.0059	.0370	.1180	.2576	.4358	.6160	.7662	.8723	.9371	.9719
$B(n,t)$	.0059	.0312	.0812	.1396	.1781	.1800	.1500	.1060	.0649	.0349
Kumul	.0059	.0371	.1183	.2578	.4360	.6160	.7660	.8720	.9369	.9718
$P(\lambda)$	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653	.0363
Kumul	.0067	.0404	.1247	.2650	.4405	.6160	.7622	.8666	.9319	.9682

Eine Tabellenkalkulation hilft, z.B. *LibreOffice*. Hier weicht $B(n,t)$ kaum von $H(N,K,n)$ ab, $P(\lambda)$ zwar etwas mehr, aber insgesamt noch wenig.

(1) Angenommen, höchstens $2\,500$ der $50\,000$ Teile sind defekt.
Stichprobe ≤ 5 mit Wkt $\gtrapprox 61\%$, Stichprobe ≥ 6 mit Wkt $\lessapprox 39\%$.
Bei bis zu 38% aller Stichproben würde Ihr Chef demnach irren.

(2) Stichprobe ≤ 9 mit Wkt $\gtrapprox 97\%$, Stichprobe ≥ 10 mit Wkt $\lessapprox 3\%$.
Das Risiko einer irrtümlichen Klage wäre demnach kleiner als 3%.
Im Falle $K < 2\,500$ verstärken sich die Ungleichungen noch.

Ausfallwahrscheinlichkeit

Von n **unabhängigen** Teilen habe jedes Ausfallwkt p. Dann gilt:

$$\begin{aligned}
\mathbf{P}(\text{Alle Teile funktionieren}) &= (1-p)^n &\lessapprox \mathrm{e}^{-np}\\
\mathbf{P}(\text{Mindestens eins fällt aus}) &= 1-(1-p)^n &\gtrapprox 1-\mathrm{e}^{-np}
\end{aligned}$$

Taylor–Entwicklung: Es gilt $\exp(x) \gtrapprox 1+x$ und $\exp(-x) \gtrapprox 1-x$.

Genauer: Für alle $x \in \mathbb{R}$ gilt $\exp(x) \ge 1+x$ und $\exp(-x) \ge 1-x$, und für $x \approx 0$ gilt zudem $\exp(x) \approx 1+x$ und $\exp(-x) \approx 1-x$.

Eine Maschine bestehe aus vielen **unabhängigen** Teilen $T_1, \dots, T_n$. Jedes Teil T_k hat eine gewisse Ausfallwahrscheinlichkeit $p_k \in [0,1]$.

$$\begin{aligned}
\mathbf{P}(\text{Alle Teile funktionieren}) &= (1-p_1)\cdots(1-p_n)\\
&\lessapprox \mathrm{e}^{-p_1}\cdots\mathrm{e}^{-p_n} = \mathrm{e}^{-(p_1+\dots+p_n)}.\\
\mathbf{P}(\text{Mindestens eins fällt aus}) &= 1-(1-p_1)\cdots(1-p_n)\\
&\gtrapprox 1-\mathrm{e}^{-(p_1+\dots+p_n)}.
\end{aligned}$$

☺ Die Näherung ist praktisch für $p_1, \dots, p_n$ klein und n groß. Dies ist die Poisson–Verteilung $P(\lambda)$ mit $\lambda = p_1 + \dots + p_n$. U314

Kollisionswahrscheinlichkeit

Aus n Möglichkeiten wird k mal zufällig ausgewählt (wobei $1 \le k \le n$). Die Wahrscheinlichkeit $P_{n,k}$, dabei k verschiedene auszuwählen, ist

$$P_{n,k} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \lessapprox \exp\left(-\frac{k(k-1)}{2n}\right).$$

☺ Der Graph ist die rechte Hälfte der Gaußschen Glockenkurve $\mathrm{e}^{-x^2/2}$. Die Wahrscheinlichkeit $Q_{n,k}$ mindestens einer Kollision ist demnach

$$Q_{n,k} = 1 - P_{n,k} \gtrapprox 1 - \exp\left(-\frac{k(k-1)}{2n}\right).$$

Geburtstagsparadox: Mit welcher Wkt Q sind unter $k = 25$ zufälligen Personen mindestens zwei am gleichen Tag des Jahres geboren?

Lösung: Für $k = 25$ und $n = 365$ gilt $P \lessapprox \mathrm{e}^{-0.821} \lessapprox 0.44$, also $Q \gtrapprox 0.56$.

☺ Intuitiv hält man eine Kollision hier für unwahrscheinlich. Tatsächlich liegt die Wkt bei über 50%, daher heißt dieses überraschende Ergebnis auch „Geburtstagsparadox“. Probieren Sie es selbst einmal aus!

Kombinatorik / Schubfachmodelle

Auf wie viele Arten kann man k Objekte auf n Fächer verteilen?

Schubfachmodell	unterscheidbare Objekte	ununterscheidbare Objekte
beliebig viele Objekte pro Fach	n^k	$\binom{n+k-1}{k}$
höchstens eins pro Fach (injektiv)	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$
mindestens eins pro Fach (surjektiv)	$n!\left\{ {k \atop n} \right\}$	$\binom{k-1}{n-1}$

Unterscheidbare Objekte denken wir uns mit $1, \dots, k$ nummeriert.
Bei ununterscheidbaren Objekten dürfen wir Objekte untereinander vertauschen: Das bedeutet, wir betrachten Anordnungen dann als gleich, wenn sie sich nur durch die Nummerierung der Objekte unterscheiden.
(Für Ausführungen siehe `en.wikipedia.org/wiki/Twelvefold_way`.)

Kombinatorik / Urnenmodelle

Aus n durchnummerierten Kugeln ziehen wir k Kugeln (oder Lose):
Ziehung mit / ohne Zurücklegen, Ergebnis mit / ohne Reihenfolge.
Die Gesamtzahl der möglichen Ergebnisse berechnet sich wie folgt:

Urnenmodell	mit Reihenfolge		ohne Reihenfolge	
ohne Zurücklegen	$\frac{n!}{(n-k)!}$	Laplace	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$	Laplace
mit Zurücklegen	n^k	Laplace	$\binom{n+k-1}{k}$	nicht Laplace!

Beispiel: In einer Urne liegen N Kugeln, davon sind genau K rot.
Wir ziehen zufällig n der N Kugeln (oZoR). Welche Wkt hat das Ereignis $A_k =$ „Es werden genau k der K roten Kugeln gezogen“? **Lösung:**

$$\mathbf{P}(A_k) = \frac{|A_k|}{|\Omega|} = \binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n}$$

Abstand von Wahrscheinlichkeitsmaßen

Oft wollen oder müssen wir zwei WMaße $\mathbf{P}_0$ und $\mathbf{P}_1$ auf Ω vergleichen: Ist $\mathbf{P}_0$ mühsam aber $\mathbf{P}_1$ bequem, so wollen wir $\mathbf{P}_0$ durch $\mathbf{P}_1$ ersetzen.

Die beiden WMaße $\mathbf{P}_0$ und $\mathbf{P}_1$ ordnen jedem Ereignis $A \subseteq \Omega$ die Wkten $\mathbf{P}_0(A)$ bzw. $\mathbf{P}_1(A)$ zu. Der **totale Abstand** zwischen $\mathbf{P}_0$ und $\mathbf{P}_1$ ist das Supremum, die größtmögliche Abweichung, die hierbei auftreten kann:

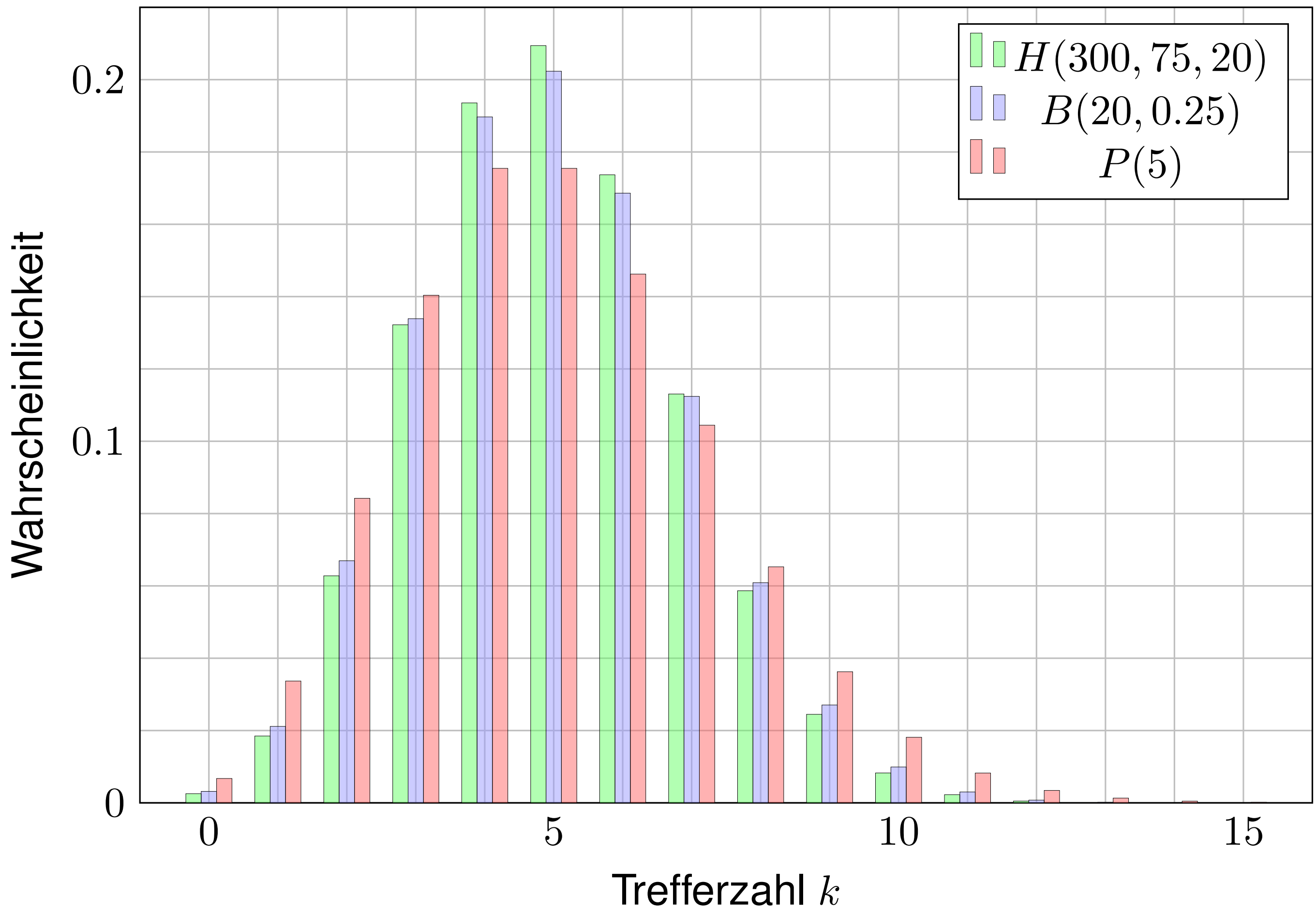
$$\boxed{\left\|\mathbf{P}_0 - \mathbf{P}_1\right\| := \sup_{A\subseteq\Omega}\left|\mathbf{P}_0(A) - \mathbf{P}_1(A)\right| = \frac{1}{2}\sum_{\omega\in\Omega}\left|\mathbf{P}_0(\{\omega\}) - \mathbf{P}_1(\{\omega\})\right|}$$

Diese Fehlerschranke nutzen wir bei näherungsweisen Rechnungen: Ist der Abstand klein genug, etwa $\|\mathbf{P}_0 - \mathbf{P}_1\| \le \varepsilon = 0.5 \cdot 10^{-3}$, so können wir $\mathbf{P}_0(A)$ durch $\mathbf{P}_1(A)$ ersetzen und machen dabei schlimmstenfalls einen Fehler von ε, das heißt, es gilt $\mathbf{P}_0(A) = \mathbf{P}_1(A) + \delta$ mit $|\delta| \le \varepsilon$.

Der totale Abstand ist (bis auf einen Faktor $1/2$) die Summe über alle punktweisen Abstände: Die linke Seite $\sup_{A\subseteq\Omega}$ ist leicht zu verstehen, die rechte Seite $\sum_{\omega\in\Omega}$ ist leicht zu berechnen. Diese Umformulierung ist oft leichter zugänglich: Es genügt, die absolute Differenz über alle Ergebnisse zu addieren. Die genaue Rechnung U318 beweist dies und erklärt den Faktor $1/2$.

Die Rechnung zeigt zudem, dass das Supremum angenommen wird, also ein Maximum ist. Für die Menge $A = \{\, \omega \in \Omega \mid \mathbf{P}_0(\{\omega\}) > \mathbf{P}_1(\{\omega\}) \,\}$ gilt $\|\mathbf{P}_0 - \mathbf{P}_1\| = \left|\mathbf{P}_0(A) - \mathbf{P}_1(A)\right|$.

Hypergeometrisch, binomial, Poisson



Hypergeometrisch, binomial, Poisson

Stichprobe: Gesamtgröße N, davon K Treffer, Stichprobengröße n.

Die Trefferzahl k folgt der **hypergeometrischen Verteilung**

$$\boxed{H(N,K,n)(k) = \binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n}.}$$

Ein Experiment mit Trefferwkt t mit n–mal unabhängig wiederholt, z.B. Stichprobengröße n mit Zurücklegen, Trefferwkt $t = K/N$.

Die Trefferzahl k folgt hier der **Binomialverteilung**

$$\boxed{B(n,t)(k) = \binom{n}{k} t^k (1-t)^{n-k}.}$$

Damit sind $H(N,K,n)$ und $B(n,t)$ WVerteilungen auf $\{0,\dots,n\} \subset \mathbb{N}$.

Für $N \to \infty$ und $K/N \to t$ gilt die Konvergenz $H(N,K,n) \to B(n,t)$. Genauer gilt für $K > n$ folgende Abschätzung des **totalen Abstands**

$$\boxed{\big\| H(N,K,n) - B(n,K/N) \big\| < n/N.}$$

Hypergeometrisch, binomial, Poisson

Die **Poisson–Verteilung** $P(\lambda)$ zum Parameter $\lambda \ge 0$ ist gegeben durch

$$\boxed{P(\lambda)(k) = \mathrm{e}^{-\lambda}\, \lambda^k / k!.}$$

Damit ist $P(\lambda)$ eine Wahrscheinlichkeitsverteilung auf $\mathbb{N} = \{0,1,2,\dots\}$. Das Gesetz der kleinen Zahlen besagt $B(n,\lambda/n) \to P(\lambda)$ für $n \to \infty$.

Genauer gilt folgende Abschätzung des **totalen Abstands**:

$$\boxed{\big\| B(n,\lambda/n) - P(\lambda) \big\| \le \lambda^2/n.}$$

Für Trefferwahrscheinlichkeit $t \in [0,1]$ und $\lambda = nt$ erhalten wir

$$\boxed{\big\| B(n,t) - P(nt) \big\| \le nt^2.}$$

🙂 Je nach geforderter Genauigkeit können wir so bequem von hypergeometrisch über binomial zu Poisson übergehen.

🙂 Das verallgemeinert und präzisiert die Näherungsformel U108 für Ausfallwahrscheinlichkeiten bei n gleichen unabhängigen Bauteilen.

Verständnisfragen: unabhängige Experimente

Versuchen Sie, folgende Fragen frei aber genau zu beantworten, etwa so, wie Sie dies einer Kommiliton:in / Kolleg:in erklären wollen.

Wie beschreibt man zwei unabhängige Experimente in einem gemeinsamen Modell, d.h. auf einem Wahrscheinlichkeitsraum? Wie berechnet man z.B. die Sicherheit von drei Bauteilen mit Ausfallwkten p_1, p_2, p_3? Warum ist die Unabhängigkeit wesentlich?

Wie wahrscheinlich ist es, dass von n unabhängigen Teilen mit gleicher Ausfallwkt p mindestens eins ausfällt? Was gilt exakt? Welche Näherung ist für große n nützlich? Wie kommt sie zustande? Inwiefern ist dies ein Spezialfall der Poisson–Verteilung?

Was besagt das Geburtstagsparadox? Was ist daran paradox? Wie berechnet man allgemein die Wkt einer Kollision? exakt? Welche Näherung ist hierzu nützlich? Wie kommt sie zustande? Welche Rolle spielt hierbei die Gaußsche Glockenkurve $\exp(-x^2)$?

Verständnisfragen: Abzählformeln

Welche Abzählformeln gelten bei der Ziehung aus einer Urne mit / ohne Zurücklegen sowie mit / ohne Beachtung der Ziehungsreihenfolge? Welche davon sind Laplace-Experimente, mit gleichverteilter Wkt?

Drei wichtige Beispiele: Wie definiert man folgende Verteilungen?

- die hypergeometrische Verteilung $H(N, K, n)$
- die Binomialverteilung $B(n, t)$
- die Poisson–Verteilung $P(\lambda)$

Nennen Sie Anwendungen, in denen obige Verteilungen auftreten; inwiefern sind diese Modelle jeweils exakt oder nur angenähert?

Wie definiert man allgemein den totalen Abstand zwischen zwei diskreten Wahrscheinlichkeitsmaßen $\mathbf{P}_0, \mathbf{P}_1$ auf einer Menge Ω? Hierzu gibt es zwei äquivalente Formeln, als Supremum über alle Mengen (Ereignisse) und als Summe über alle Punkte (Ergebnisse). Wie lauten diese beiden Formeln und was bedeuten sie anschaulich?

Verständnisfragen: Gesetz der kleinen Zahlen

Dies führt zum ersten wichtigen Grenzwertsatz dieses Kapitels:

$$B(n,t) \approx P(\lambda)$$

In welchem Sinne nähert sich die Binomial- der Poisson–Verteilung? Nennen / erfinden Sie mindestens zehn konkrete Anwendungsbeispiele, in denen diese Näherung praktisch nützlich ist, sinnvoll und hilfreich.

Wie lautet und was besagt Poissons Gesetz der kleinen Zahlen? Wie bestimmt man hierzu die nötige Kenngröße λ? In welchen Situationen kann man diese Näherung anwenden? Welche Voraussetzungen braucht man und welche Schlussfolgerung gewinnt man?

Wie lautet die Fehlerschranke beim Gesetz der kleinen Zahlen? Wie verhält sie sich für $n \to \infty$? Wie muss t angepasst werden? Wozu nützt diese Schranke in konkreten Anwendungen, bei festem n? Wann darf man sie getrost ignorieren? meistens? manchmal? nie?

Verständnisfragen: Trefferwkt bei Stichproben

Schließlich zum ergänzenden Grenzwertsatz für Stichproben:

$$H(N,K,n) \approx B(n,t)$$

In welchem Sinne nähert sich die hypergeom. der Binomialverteilung? Nennen / erfinden Sie mindestens zehn konkrete Anwendungsbeispiele, in denen diese Näherung praktisch nützlich ist, sinnvoll und hilfreich.

In der Praxis betrachtet man häufig eine Stichprobe der Größe n aus einer Population N, entweder mit oder ohne Zurücklegen. Wie lauten jeweils die exakten Wkten für die Trefferzahlen? Welches der beiden Modelle ist bequemer zu berechnen?

Warum sind beide Modelle für großes N nahezu gleich? Wie weit liegen sie höchstens auseinander? Wie verhält sich ihr Abstand für $N \to \infty$? Wozu nützt diese Schranke in konkreten Anwendungen, bei festem n? Wann darf man sie getrost ignorieren? meistens? manchmal? nie?

Kollision bei Losen

Aufgabe: Eine Tombola enthält 850 durchnummerierte Lose. Sie ziehen zufällig und unabhängig 51 mal ein Los *mit Zurücklegen*. Wie groß ist die Wkt P, dabei 51 *verschiedene* Lose zu ziehen?

Lösung: Wir nutzen den Geburtstagssatz U1E. Für die Wahrscheinlichkeit gilt mit $n = 850$ und $k = 51$:

$$\begin{aligned}
P &= \Big(1 - \frac{0}{n}\Big)\Big(1 - \frac{1}{n}\Big)\Big(1 - \frac{2}{n}\Big)\cdots\Big(1 - \frac{k-1}{n}\Big) \\
&\lessapprox \exp\Big(-\frac{0}{n}\Big)\exp\Big(-\frac{1}{n}\Big)\exp\Big(-\frac{2}{n}\Big)\cdots\exp\Big(-\frac{k-1}{n}\Big) \\
&= \exp\Big(-\frac{0+1+2+\dots+(k-1)}{n}\Big) \\
&= \exp\Big(-\frac{k(k-1)}{2n}\Big) \quad = \quad \exp(-1.5) \quad \lessapprox \quad 0.223.
\end{aligned}$$

⚠ Die Wkt $< 23\%$ ist viel geringer als zunächst vermutet: In 850 scheint für 51 verschiedene Lose genug Platz. Wie beim Geburtstagsparadox sind Kollisionen mit $> 77\%$ viel wahrscheinlicher als man naiv erwartet.

Kollision beim Lotto

Aufgabe: (1) Wie viele Ergebnisse gibt es beim Lotto „6 aus 49“? (2) Mit welcher Wahrscheinlichkeit Q tritt bei 5000 unabhängigen Ziehungen mindestens ein Ziehungsergebnis doppelt auf?

Lösung: (1) Wir nutzen unsere kombinatorischen Abzählformeln:

Beim Lotto „6 aus 49“ sind $\binom{49}{6} = 13\,983\,816$ Ergebnisse möglich.

(2) Wir nutzen den Geburtstagssatz U1E wie in der vorigen Aufgabe:

Mit $n = \binom{49}{6}$ und $k = 5000$ gilt $k(k-1)/(2n) \gtrapprox 25/28 \gtrapprox 0.89$.

Die Wahrscheinlichkeit, keine Kollision zu erleben, ist demnach:

$$P \quad \lessapprox \quad \exp\Big(-\frac{k(k-1)}{2n}\Big) \quad \lessapprox \quad \exp(-0.89) \quad \lessapprox \quad 0.411$$

⚠ Die Wahrscheinlichkeit, dass mindestens ein Ziehungsergebnis doppelt vorkommt, ist demnach $Q = 1 - P \gtrapprox 0.589$, also fast 60%.

Das ist viel wahrscheinlicher, als man zunächst vermuten würde. Im deutschen Lotto „6 aus 49“ wurden bei den ersten 5000 Ziehungen (von 1955 bis bis 2011) tatsächlich zweimal die gleichen sechs Zahlen gezogen: Sowohl am 20.12.1986 als auch am 21.06.1995 waren es die Zahlen 15, 25, 27, 30, 42, 48. Ein solches Vorkommnis ist nach unserer Rechnung durchaus plausibel.

Mehrfache Hauptgewinne beim Lotto

Aufgabe: Beim Lotto beträgt die Wkt für 6 Richtige etwa 1:14 Mio. An einem typischen Samstag werden etwa 52 Mio Tipps abgegeben.

(1) Wie groß ist die Wkt für $k = 0, 1, 2, 3, 4, 5, \ldots$ Tipps mit 6 Richtigen?
(2) Welche vereinfachenden Annahmen möchten / müssen Sie machen?
(3) Welche Approximation ist geeignet? Wir groß ist dabei der Fehler?

Lösung: $n = 52\,000\,000$ Versuche mit Trefferwkt $t = 1/14\,000\,000$.
(1) Die Trefferzahl ist binomialverteilt: $B(n,t)(k) = \binom{n}{k} t^k (1-t)^{n-k}$
(2) Wir nehmen hier die Tipps als gleichverteilt und unabhängig an.
(3) Eine Näherung ist die Poisson–Verteilung $P(\mu)$ mit $\mu = nt \approx 3.714$:

$$B(n,t)(k) = \binom{n}{k} t^k (1-t)^{n-k} \quad \approx \quad P(\mu)(k) = \mathrm{e}^{-\mu} \frac{\mu^k}{k!}$$

☺ Der totale Abstand $\|B(n,t) - P(\mu)\| \le nt^2 < 3 \cdot 10^{-7}$ ist extrem klein.

k	0	1	2	3	4	5	6	7	8	9
$P(\mu)(k)$	.024	.091	.168	.208	.193	.144	.089	.047	.022	.009

Übertrag und Jackpot beim Lotto

Aufgabe: Beim Lotto gibt es ca. 140 Mio Möglichkeiten mit Superzahl. Samstags werden etwa 52 Mio Tipps abgegeben, Mittwochs nur 26 Mio.

(1) Wir groß ist jeweils die Wkt, dass der Jackpot nicht geknackt wird?
(2) Eine Woche lang? zwei? drei? vier? ... (3) Nennen Sie explizit Ihre vereinfachenden Annahmen und die daraus abgeleiteten Modelle.

Lösung: (3) $M = 10 \cdot \binom{49}{6} \approx 14 \cdot 10^7$ Möglichkeiten. Ziehungen sind unabhängig und gleichverteilt, also Trefferwkt $t = 1/M$. Wir nehmen hierzu vereinfachend die Tipps als unabhängig und gleichverteilt an.

(1) Bei n Tipps mit Trefferwkt t wird der Jackpot nicht geknackt mit Wkt

$$q \;=\; B(n,t)(0) \;=\; (1-t)^n \;=\; \left(1 - \frac{nt}{n}\right)^n \;\approx\; \mathrm{e}^{-nt} \;=\; P(nt)(0)$$

Mittwochs $n_M \cdot t = 26/140 \approx 0.1857$, also $q_M = \mathrm{e}^{-0.1857} \approx 0.8305$
Samstags $n_S \cdot t = 52/140 \approx 0.3714$, also $q_S = q_M^2 = \mathrm{e}^{-0.3714} \approx 0.6897$

(2) Jackpot wird in einer Woche nicht geknackt: $q = \mathrm{e}^{-0.5571} \approx 0.5728$.
Mit Wkt $q^{k-1}(1-q)$ wird der Jackpot erst in der k–ten Woche geknackt.

☺ Das ist die geometrische Verteilung T4A: Die Wkt fällt exponentiell.

Symmetrie der hypergeometrischen Verteilung

Aufgabe: Beweisen Sie die folgende Symmetrieformel:

$$\boxed{H(N,K,n)(k) = H(N,n,K)(k)}$$

(1) Schreiben Sie beide Seiten explizit aus und vergleichen Sie.

(2) Sie wählen aus der Gesamtmenge $\Omega = \{1,2,\dots,N\}$ zufällig eine Teilmenge $A \subseteq \Omega$ mit $|A| = K$ Elementen und eine Teilmenge $B \subseteq \Omega$ mit $|B| = n$ Elementen. Mit welcher Wahrscheinlichkeit gilt $|A \cap B| = k$?

Lösung: (1) Wir setzen die Definition ein und vergleichen:

$$H(N,K,n)(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{K!\,(N-K)!\,n!(N-n)!}{k!(K-k)!\,(n-k)!(N-K-n+k)!\,N!}$$

$$H(N,n,K)(k) = \frac{\binom{n}{k}\binom{N-n}{K-k}}{\binom{N}{K}} = \frac{n!\,(N-n)!\,K!(N-K)!}{k!(n-k)!\,(K-k)!(N-n-K+k)!\,N!}$$

(2) Wenn wir erst A fixieren, finden wir $H(N,K,n)(k)$. Wenn wir erst B fixieren, finden wir $H(N,n,K)(k)$. Beide Wkten sind jedoch dieselben!

Diskrete Verteilungen und ihre Kenngrößen

Satz U4A: diskrete Verteilungen und ihre Kenngrößen

$X \sim$ Verteilung	Ω	$p(k) = \mathbf{P}(\{k\})$	$\mathbf{E}$	$\mathbf{V}$
Geometrisch $G(q)$	$\mathbb{N}_{\ge 1}$	$(1-q)q^{k-1}$	$\frac{1}{1-q}$	$\frac{q}{(1-q)^2}$
Hyperg. $H(N,K,n)$	$\{0,\dots,n\}$	$\binom{K}{k}\binom{N-K}{n-k}/\binom{N}{n}$	$n\frac{K}{N}$	$n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$
Binomial $B(n,t)$	$\{0,\dots,n\}$	$\binom{n}{k}t^k(1-t)^{n-k}$	nt	$nt(1-t)$
Poisson $P(\lambda)$	$\mathbb{N}$	$\mathrm{e}^{-\lambda}\,\lambda^k/k!$	λ	λ

Aufgabe: Rechnen Sie dies für die geometrische Verteilung nach.
Lösung: Die Rechnung kennen Sie aus dem vorigen Kapitel, Satz T4A.

Für die folgenden Rechnungen nutzen wir insbesondere die Identität

$$k\binom{K}{k} = \frac{K\cdot(K-1)\cdots(K-k+1)}{(k-1)\cdots 1} = K\binom{K-1}{k-1}.$$

Erwartung der hypergeometrischen Verteilung

Aufgabe: Sei $X \sim H(N, K, n)$ hypergeometrisch verteilt. Zeigen Sie:

$$\boxed{\mathbf{E}[X] = n\frac{K}{N} \quad \text{und} \quad \mathbf{V}[X] = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}}$$

Diese Formeln sind symmetrisch in n und K, ganz wie es sein muss!

Lösung: (1) Wir berechnen zunächst den Erwartungswert:

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{k=0}^{n} k\,\mathbf{P}(X{=}k) = \sum_{k=0}^{n} k\binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n} \\
&= \sum_{k=0}^{n} K\binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)} \Big/ \frac{N}{n}\binom{N-1}{n-1} \\
&= \frac{nK}{N} \underbrace{\sum_{k=0}^{n} \binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)} \Big/ \binom{N-1}{n-1}}_{=1}
\end{aligned}$$

☺ Die letzte Summe kumuliert die $H(N{-}1, K{-}1, n{-}1)$–Verteilung.

Varianz der hypergeometrischen Verteilung

(2) Ebenso berechnen wir $\mathbf{E}[X^2]$, dazu zunächst $\mathbf{E}[X(X-1)]$:

$$\begin{aligned}
\mathbf{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{K}{k}\binom{N-K}{n-k} \Big/ \binom{N}{n} \\
&= \sum_{k=0}^{n} K(K-1)\binom{K-2}{k-2}\binom{(N-2)-(K-2)}{(n-2)-(k-2)} \Big/ \frac{N}{n}\frac{N-1}{n-1}\binom{N-2}{n-2} \\
&= \frac{n(n-1)K(K-1)}{N(N-1)} \underbrace{\sum_{k=0}^{n} \binom{K-2}{k-2}\binom{(N-2)-(K-2)}{(n-2)-(k-2)} \Big/ \binom{N-2}{n-2}}_{=1}
\end{aligned}$$

Hieraus erhalten wir schließlich die ersehnte Varianz:

$$\begin{aligned}
\mathbf{V}[X] &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \mathbf{E}[X(X-1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 \\
&= \frac{n(n-1)K(K-1)}{N(N-1)} + \frac{nK}{N} - \frac{n^2K^2}{N^2} = \frac{nK(N-K)(N-n)}{N^2(N-1)}
\end{aligned}$$

☺ Die Rechnung ist länglich, aber jeder einzelne Schritt ist leicht.

Erwartung der Binomialverteilung

Aufgabe: Sei $X \sim B(n,t)$ binomialverteilt. Zeigen Sie:

$$\boxed{\mathbf{E}[X] = nt \quad \text{und} \quad \mathbf{V}[X] = nt(1-t)}$$

Lösung: Klar als Summe unabhängiger Zufallsvariablen, siehe T333! Alternativ gelingt die Rechnung auch direkt: (1) Die Erwartung ist:

$$\begin{aligned}
\mathbf{E}\big[X\big] &= \sum_{k=0}^{n} k\,\mathbf{P}(X{=}k) = \sum_{k=1}^{n} k \binom{n}{k} t^k (1-t)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} t^k (1-t)^{n-k} \\
&= nt \sum_{k=1}^{n} \binom{n-1}{k-1} t^{k-1} (1-t)^{(n-1)-(k-1)} \\
&= nt \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} t^j (1-t)^{(n-1)-j}}_{=1}
\end{aligned}$$

☺ Die letzte Summe kumuliert alle Wkten der $B(n{-}1,t)$–Verteilung.

Varianz der Binomialverteilung

(2) Ebenso berechnen wir $\mathbf{E}[X^2]$, dazu zunächst $\mathbf{E}[X(X-1)]$:

$$\begin{aligned}
\mathbf{E}\big[X(X-1)\big] &= \sum_{k=0}^{n} k(k-1) \binom{n}{k} t^k (1-t)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1) \binom{n-2}{k-2} t^k (1-t)^{n-k} \\
&= n(n-1)t^2 \sum_{k=2}^{n} \binom{n-2}{k-2} t^{k-2} (1-t)^{(n-2)-(k-2)} \\
&= n(n-1)t^2 \underbrace{\sum_{j=0}^{n-2} \binom{n-2}{j} t^j (1-t)^{(n-2)-j}}_{=1}
\end{aligned}$$

Hieraus erhalten wir schließlich die ersehnte Varianz:

$$\begin{aligned}
\mathbf{V}\big[X\big] &= \mathbf{E}\big[X^2\big] - \mathbf{E}\big[X\big]^2 = \mathbf{E}\big[X(X-1)\big] + \mathbf{E}\big[X\big] - \mathbf{E}\big[X\big]^2 \\
&= n(n-1)t^2 + nt - (nt)^2 = nt(1-t)
\end{aligned}$$

☺ Die Rechnung ist länglich, aber jeder einzelne Schritt ist leicht.

Illustration zur Binomialverteilung

Ein Experiment mit Trefferwkt 0.5 wird 100 mal unabhängig wiederholt.
Aufgabe: Was ist hier die Verteilung? Erwartung? Varianz? Streuung?
Lösung: $B(100, 0.5)$, Erwartung $\mu{=}50$, Varianz $\sigma^2{=}25$, Streuung $\sigma{=}5$.

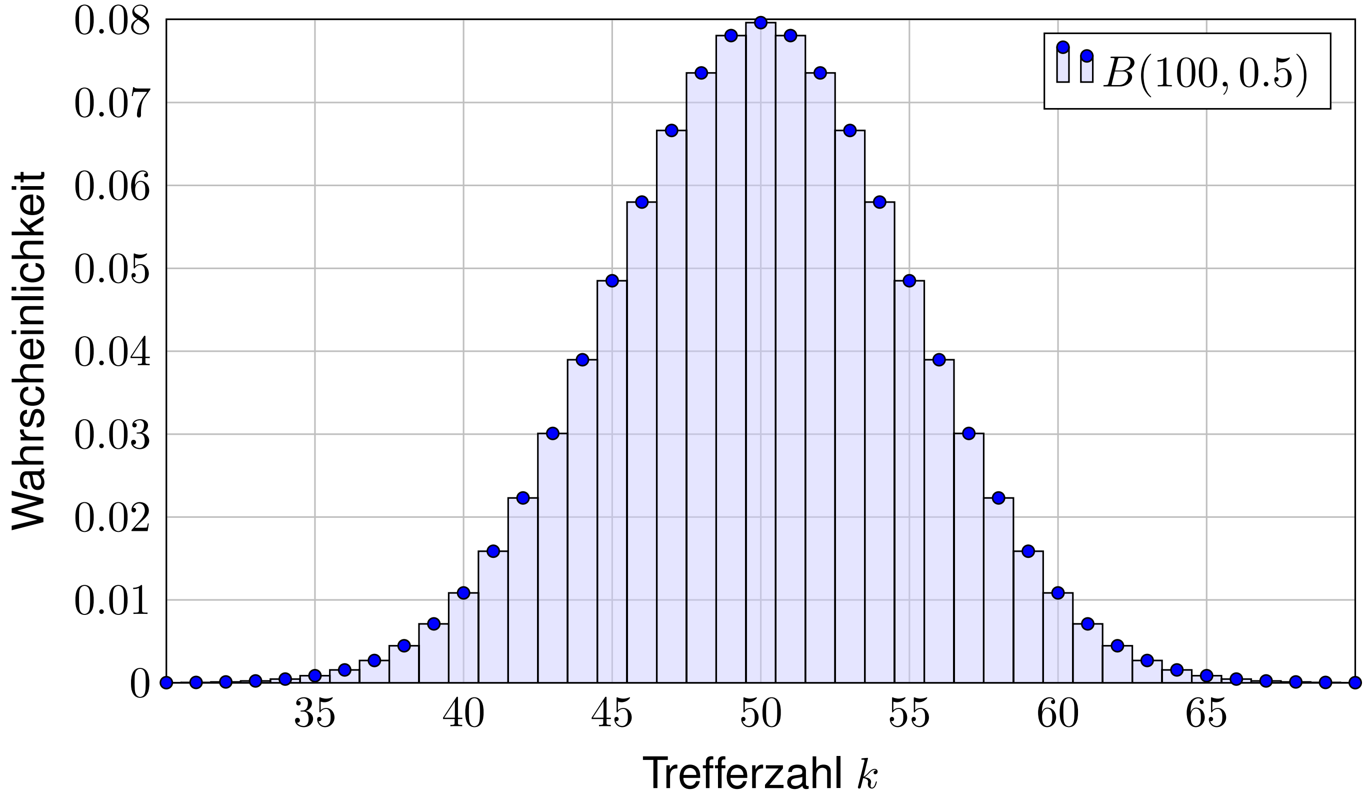
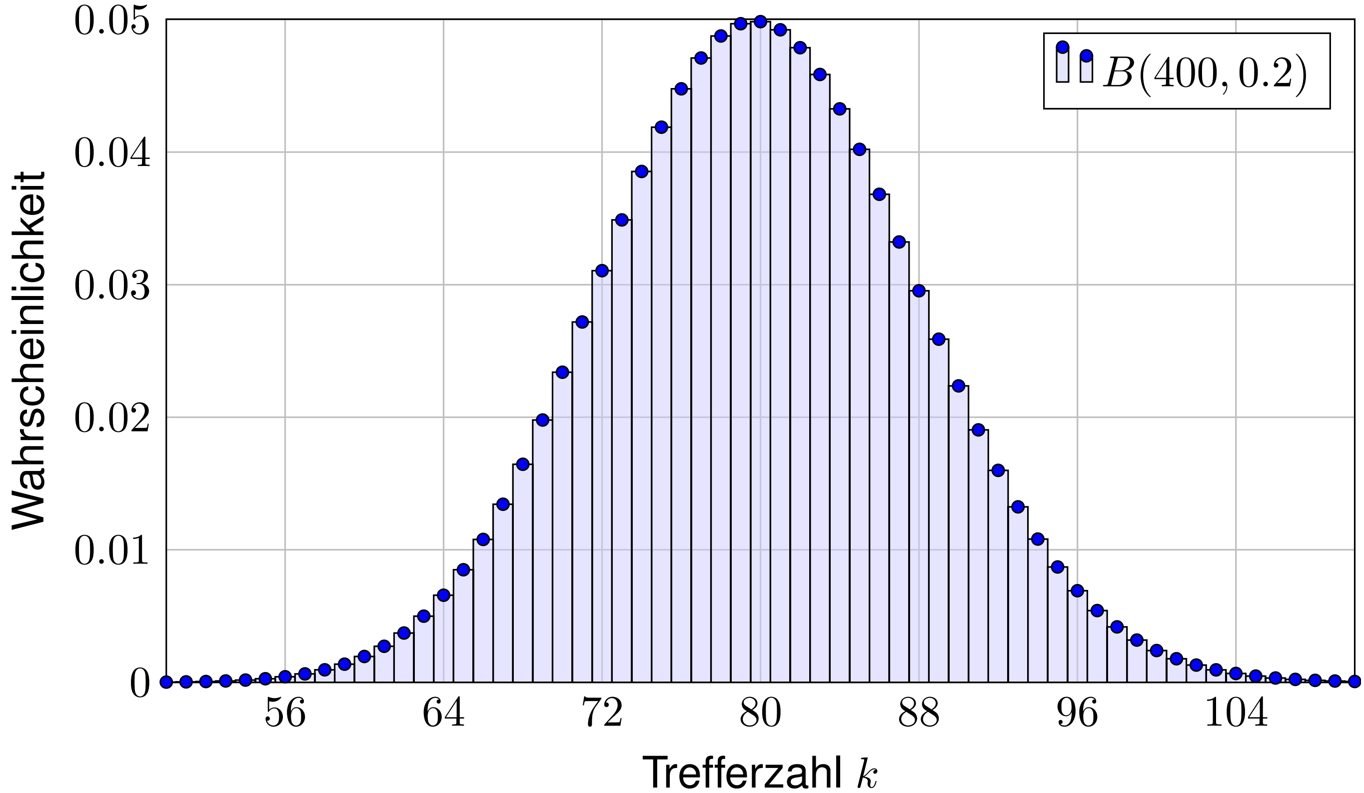


Illustration zur Binomialverteilung

Ein Experiment mit Trefferwkt 0.2 wird 400 mal unabhängig wiederholt.
Aufgabe: Was ist hier die Verteilung? Erwartung? Varianz? Streuung?
Lösung: $B(400, 0.2)$, Erwartung $\mu{=}80$, Varianz $\sigma^2{=}64$, Streuung $\sigma{=}8$.



Erwartung der Poisson–Verteilung

Aufgabe: Sei $X \sim P(\lambda)$ Poisson–verteilt. Zeigen Sie:

$$\boxed{\mathbf{E}[X] = \lambda \quad \text{und} \quad \mathbf{V}[X] = \lambda}$$

Beweis: (1) Die Erwartung der Poisson–Verteilung $P(\lambda)$ ist

$$\begin{aligned}
\mathbf{E}\big[X\big] &= \sum_{k=0}^{\infty} k\,\mathbf{P}(X{=}k) &&= \sum_{k=0}^{\infty} k\,\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} \\
&= \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!}\,\mathrm{e}^{-\lambda} &&= \lambda \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\,\mathrm{e}^{-\lambda}}_{=1} &&= \lambda.
\end{aligned}$$

🙂 Die letzte Summe kumuliert alle Wkten der $P(\lambda)$–Verteilung.

🙂 Die Verteilungen werden immer einfacher von hypergeometrisch über binomial zu Poisson, und das spüren wir auch in den Rechnungen.

Varianz der Poisson–Verteilung

(2) Ebenso berechnen wir $\mathbf{E}[X^2]$, dazu zunächst $\mathbf{E}[X(X-1)]$:

$$\begin{aligned}
\mathbf{E}\big[X(X-1)\big] &= \sum_{k=0}^{\infty} k(k-1)\,\mathbf{P}(X{=}k) &&= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda} \\
&= \sum_{k=2}^{\infty} \lambda^2 \frac{\lambda^{k-2}}{(k-2)!}\,\mathrm{e}^{-\lambda} &&= \lambda^2 \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\,\mathrm{e}^{-\lambda}}_{=1} &&= \lambda^2.
\end{aligned}$$

Hieraus erhalten wir schließlich die ersehnte Varianz:

$$\begin{aligned}
\mathbf{V}\big[X\big] &= \mathbf{E}\big[X^2\big] - \mathbf{E}\big[X\big]^2 \\
&= \mathbf{E}\big[X(X-1)\big] + \mathbf{E}\big[X\big] - \mathbf{E}\big[X\big]^2 \\
&= \lambda^2 + \lambda - \lambda^2 \quad = \quad \lambda
\end{aligned}$$

🙂 Die Rechnung ist länglich, aber jeder einzelne Schritt ist leicht.

Illustration zur Poisson–Verteilung

Die Verteilung $P(4)$ hat Erwartung $\mu = 4$ und Streuung $\sigma = 2$.

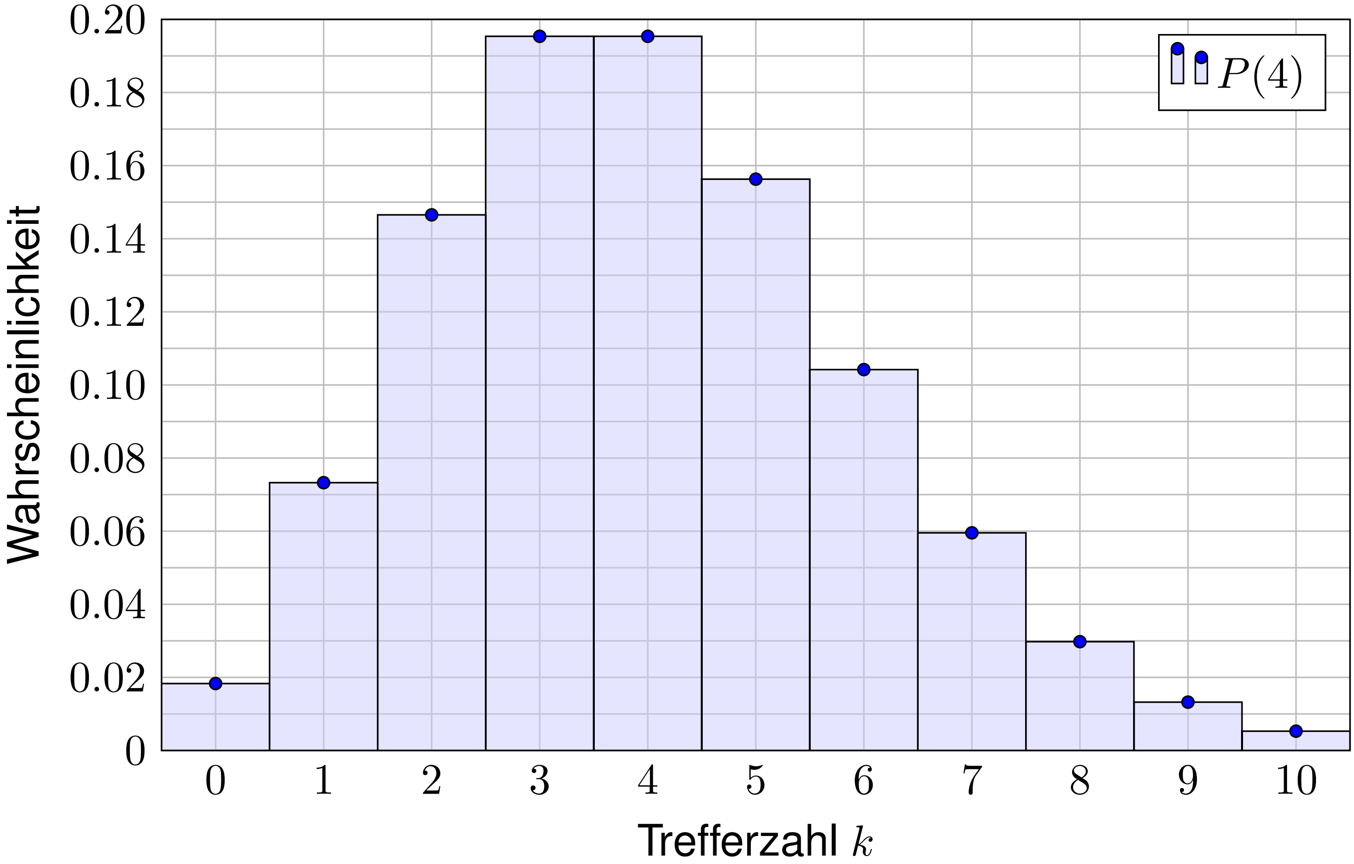
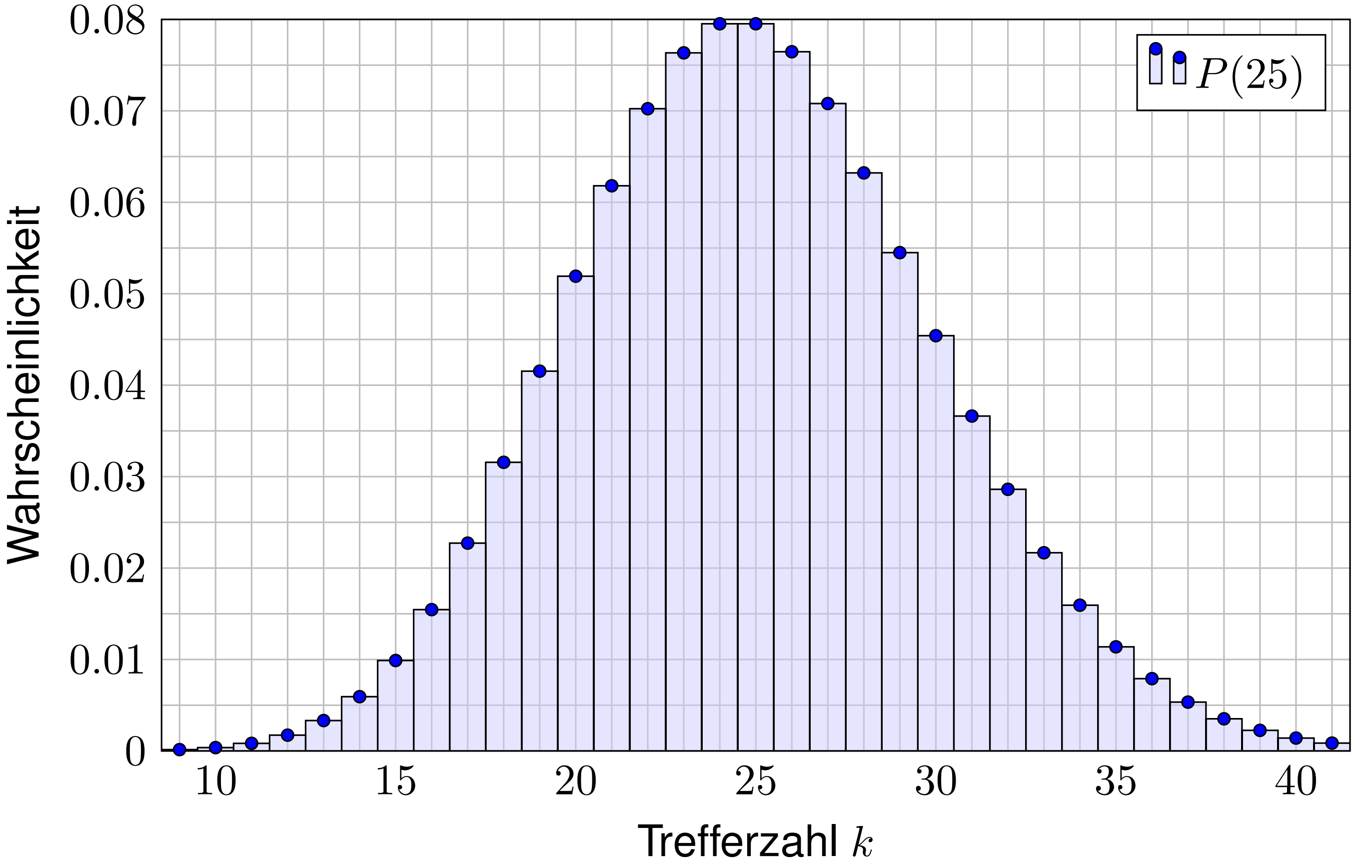


Illustration zur Poisson–Verteilung

Die Verteilung $P(25)$ hat Erwartung $\mu = 25$ und Streuung $\sigma = 5$.



Atomzerfall: Vergleich von Messung und Theorie

Ein klassisches Beispiel für eine Poisson–verteilte Zufallsvariable X ist die Anzahl der pro Sekunde emittierten α–Partikel einer radioaktiven Substanz: Die Probe enthält sehr viele Atome, aber jedes Atom zerfällt mit sehr geringer Wahrscheinlichkeit – identisch verteilt und unabhängig von allen anderen. (Bei einer Kettenreaktion sieht es anders aus! U461)

Messungen an einer Probe Americum 241 (Halbwertszeit 458 Jahre) ergaben für 12169 Intervalle von 1–Sekunden–Länge die folgende Verteilung für die Anzahl k der Zerfälle (nach Berkson 1966):

k	0	1	2	3	4	5	Σ
n	5267	4436	1800	534	111	21	12169

Aufgabe: (1) Berechnen Sie den Mittelwert dieser Messdaten.
(2) Welche Poisson–Verteilung passen Sie diesem Datensatz an?
(3) Was sind die zugehörigen theoretischen (abs./rel.) Häufigkeiten?
(4) Was ist der totale Abstand zwischen empirischer und theoretischer Verteilung? Bestätigen oder widerlegen diese Daten die anfängliche Annahme, dass die Zahl der Zerfälle einer Poisson–Verteilung folgt?

Atomzerfall: Vergleich von Messung und Theorie

(1) Der arithmetische Mittelwert μ der gemessenen Zerfallszahlen ist:

$$\frac{0\cdot 5267 + 1\cdot 4436 + 2\cdot 1800 + 3\cdot 534 + 4\cdot 111 + 5\cdot 21}{12169} = \frac{10187}{12169} \approx 0.83713$$

(2) Die Poisson–Verteilung ist gegeben durch $P(\lambda)(k) = \mathrm{e}^{-\lambda}\lambda^k/k!$ und hat Erwartungswert λ (Satz U4A). Daher bietet sich die Wahl $\lambda = \mu$ an.

(3) Die Verteilung $P(\mu)$ gibt uns folgende theoretische Häufigkeiten:

k	0	1	2	3	4	5	≥ 6	Summe
n	5267	4436	1800	534	111	21	0	12169
$\mathrm{e}^{-\mu}\mu^k/k! \cdot \Sigma$	5268.6	4410.5	1846.1	515.1	107.8	18.0	2.9	12169
n/Σ	.43282	.36453	.14792	.04388	.00912	.00173	.00000	1
$\mathrm{e}^{-\mu}\mu^k/k!$	.43295	.36244	.15170	.04233	.00886	.00148	.00023	1

Die erste Zeile zeigt die Anzahl k der Zerfälle. Die zweite und dritte Zeile zeigen die absolute Häufigkeit empirisch bzw. theoretisch. Die vierte und fünfte Zeile zeigen die relative Häufigkeit empirisch bzw. theoretisch.

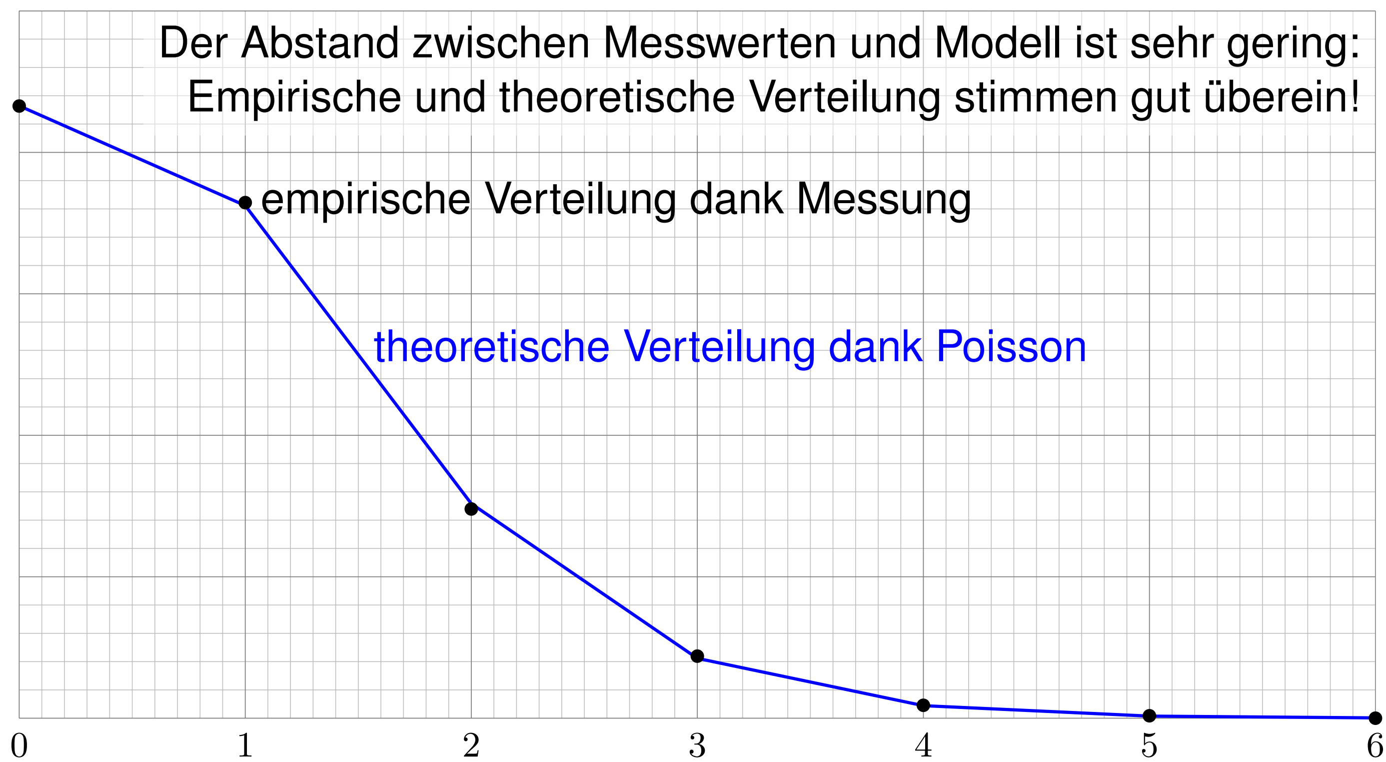
Atomzerfall: Vergleich von Messung und Theorie

(4) Die empirische Verteilung ist $P:\mathbb{N}\to[0,1]$ mit $P(k)=n(k)/12169$. Die theoretische Verteilung ist $P(\mu):\mathbb{N}\to[0,1]$ mit $P(\mu)(k)=\mathrm{e}^{-\mu}\mu^k/k!$. Ihr totaler Abstand ist (gemäß Definition U3D) gegeben durch

$$
\begin{aligned}
\|P-P(\mu)\| &= \frac{1}{2}\sum_{k=0}^{\infty}\left|\frac{n(k)}{12169}-P(\mu)(k)\right| \\
&= \frac{1}{2}\cdot\frac{1.6+25.5+46.1+18.9+3.2+3.0+2.9}{12169} \\
&= \frac{1}{2}\cdot\frac{101.2}{12169} \quad\lessapprox\quad 0.00416
\end{aligned}
$$

Für den totalen Abstand U3D haben wir hier die bequemere zweite Formel verwendet: Für jedes Ergebnis $k=0,1,2,3,4,5,6,\dots$ berechnen wir die Differenz der Wkten $|P(k)-P(\mu)(k)|$, summieren auf und teilen durch Zwei. Alternativ können wir auch leicht die erste Formel für das Supremum anwenden: Der Abstand $|P(A)-P(\mu)(A)|$ wird maximal für $A\subseteq\mathbb{N}$, wenn A alle Ergebnisse k mit $P(k)>P(\mu)(k)$ enthält, hier also $A=\{1,3,4,5\}$. Der Abstand wird ebenso maximal für das Komplement $\overline{A}=\{0,2,\geq 6\}$. Summiert man die Abweichungen für jedes der Elemente einer dieser Mengen, dann erhält man ebenfalls den totalen Abstand. Diese Aufgabe bietet so die Gelegenheit, den totalen Abstand anhand eines konkreten Beispiels zu begreifen.

Atomzerfall: Vergleich von Messung und Theorie



Eine perfekte Übereinstimmung ist hier sicher nicht zu erwarten, im Gegenteil wäre sie eher verdächtig: Die Messung ist zufälligen Schwankungen unterworfen! Eine Wiederholung der Messung würde nicht genau dieselben Werte ergeben, sondern nur ähnliche Werte innerhalb der üblichen Schwankungsbreite. Glücklicherweise hilft uns auch hier das Gesetz der großen Zahlen!

Der totale Abstand ist eine Metrik.

Für $u \in \mathbb{R}^n$ nutzen wir die euklidische **Norm** $\|u\| := \sqrt{u_1^2 + \dots + u_n^2}$.
Diese erfreut sich folgender Eigenschaften für alle $u, v \in V$ und $\lambda \in \mathbb{R}$:

(N1) $\|v\| \ge 0$, und $\|v\| > 0$ für $v \ne 0$ (positive Definitheit)
(N2) $\|\lambda \cdot v\| = |\lambda| \cdot \|v\|$ (Homogenität über $\mathbb{R}$)
(N3) $\|u + v\| \le \|u\| + \|v\|$ (Dreiecksungleichung)

Der damit definierte euklidische **Abstand** $d(u, v) := |u - v|$ erfüllt dann:

(M1) $d(u, u) = 0$, und $d(u, v) > 0$ für $u \ne v$ (Definitheit)
(M2) $d(u, v) = d(v, u)$ (Symmetrie)
(M3) $d(u, w) \le d(u, v) + d(v, w)$. (Dreiecksungleichung)

Definition U4B: Metrik und metrischer Raum

Eine **Metrik** auf einer Menge X ist eine Abstandsfunktion

$$d : X \times X \to [0, \infty] : (u, v) \mapsto d(u, v),$$

die die obigen Bedingungen (M1,2,3) für alle $u, v, w \in X$ erfüllt.
Das Paar (X, d) nennen wir dann einen **metrischen Raum**.
Im Raum (X, d) ist **Konvergenz** $v_n \to v$ definiert durch $d(v_n, v) \to 0$.

Der totale Abstand ist eine Metrik.

Aufgabe: Sei Ω eine Menge und hierüber sei $\mathscr{M}(\Omega)$ die Menge aller diskreten WMaße (T1D). Ist der totale Abstand (U3D) eine Metrik?

$$\|-\| : \mathscr{M}(\Omega) \times \mathscr{M}(\Omega) \to [0, 1] : (\mathbf{P}, \mathbf{Q}) \mapsto \|\mathbf{P} - \mathbf{Q}\|$$

Lösung: Wir haben die drei Eigenschaften (M1,2,3) zu prüfen.
Hierzu seien $\mathbf{P}, \mathbf{Q}, \mathbf{R} \in \mathscr{M}(\Omega)$ beliebige diskrete WMaße auf Ω.

(M1) Für $\mathbf{P} = \mathbf{Q}$ gilt offensichtlich $\|\mathbf{P} - \mathbf{Q}\| = 0$. Für $\mathbf{P} \ne \mathbf{Q}$ hingegen gibt es $\omega \in \Omega$ mit $\mathbf{P}(\{\omega\}) \ne \mathbf{Q}(\{\omega\})$, und daraus folgt $\|\mathbf{P} - \mathbf{Q}\| > 0$.

(M2) Die Symmetrie $\big\|\mathbf{P} - \mathbf{Q}\big\| = \big\|\mathbf{Q} - \mathbf{P}\big\|$ ist klar nach Definition.

(M3) Die Dreiecksungleichung für $\|-\|$ rechnen wir sorgfältig nach:

$$\begin{aligned}
\big\|\mathbf{P} - \mathbf{R}\big\| &= \frac{1}{2} \sum_{\omega \in \Omega} \big|\mathbf{P}(\{\omega\}) - \mathbf{Q}(\{\omega\}) + \mathbf{Q}(\{\omega\}) - \mathbf{R}(\{\omega\})\big| \\
&\le \frac{1}{2} \sum_{\omega \in \Omega} \big|\mathbf{P}(\{\omega\}) - \mathbf{Q}(\{\omega\})\big| + \big|\mathbf{Q}(\{\omega\}) - \mathbf{R}(\{\omega\})\big| \\
&= \big\|\mathbf{P} - \mathbf{Q}\big\| + \big\|\mathbf{Q} - \mathbf{R}\big\|
\end{aligned}$$

Diese Eigenschaften von $\|-\|$ folgen aus denen des Betrages $|-|$.

Poissons Gesetz der kleinen Zahlen

Poissons Gesetz der kleinen Zahlen U3E sagt $\|B(n,t) - P(nt)\| \le nt^2$. Wir kennen bereits zahlreiche Beispiele und effiziente Anwendungen. Wir wollen dies allgemein und flexibel formulieren und dann beweisen:

Satz U4C: Poissons Gesetz der kleinen Zahlen

Seien $X_1, \dots, X_n : \Omega \to \{0,1\}$ unabhängig mit $P(X_k{=}1) = t_k \in [0,1]$, also jeweils $B(1,t_k)$–verteilt mit $\mathbf{E}(X_k) = t_k$ und $\mathbf{V}(X_k) = t_k(1-t_k)$.

Die Summe $S = X_1 + \dots + X_n$ hat den Erwartungswert $\mu = t_1 + \dots + t_n$ und dank Unabhängigkeit die Varianz $\sigma^2 = t_k(1-t_k) + \dots + t_n(1-t_n)$.

Für kleine t_k ist S annähernd Poisson–verteilt, $\mathbf{P}_S \approx P(\mu)$. Genauer:

$$\boxed{\big\|\mathbf{P}_S - P(\mu)\big\| \quad \le \quad t_1^2 + \dots + t_n^2}$$

Spezialfall: Für $t_1 = \dots = t_n = t$ ist S exakt $B(n,t)$–verteilt, und somit

$$\boxed{\big\|B(n,t) - P(\mu)\big\| \quad \le \quad nt^2 = \mu^2/n.}$$

☺ Vergleichen Sie dies später mit dem zentralen Grenzwertsatz W1D.

Poissons Gesetz der kleinen Zahlen

☺ Mit den Techniken dieses Kapitels können wir diese allgemeine Fassung beweisen und die explizite Fehlerschranke nachrechnen:

Aufgabe: (0) Wiederholen Sie den totalen Abstand diskreter WMaße.

(1) Sind $X \sim P(\lambda)$ und $Y \sim P(\mu)$ unabhängig und Poisson–verteilt, dann ist ihre Summe $X + Y \sim P(\lambda + \mu)$ ebenfalls Poisson–verteilt.

(2) Berechnen Sie den Abstand der WMaße $B(1,t)$ und $P(t)$. Vergröbern Sie die exakte Rechnung zu $\big\|B(1,t) - P(t)\big\| \le t^2$.

(3) Bei Produktmaßen addieren sich die totalen Abstände gemäß der einfachen Formel $\big\|\mathbf{P} \otimes \mathbf{Q} - \mathbf{P}' \otimes \mathbf{Q}'\big\| \le \big\|\mathbf{P} - \mathbf{P}'\big\| + \big\|\mathbf{Q} - \mathbf{Q}'\big\|$.

(4) Seien $\mathbf{P}, \mathbf{P}'$ WMaße auf Ω und $S : \Omega \to \mathbb{N}$ eine Zufallsvariable. Für den Abstand der Bildmaße gilt dann $\|\mathbf{P}_S - \mathbf{P}'_S\| \le \|\mathbf{P} - \mathbf{P}'\|$.

(5) Folgern Sie Poissons Gesetz der kleinen Zahlen (Satz U4C). Dies beinhaltet Satz U3E als Spezialfall $\|B(n,t) - P(nt)\| \le nt^2$.

☺ Die Rechnung ist länglich, aber jeder einzelne Schritt ist leicht. Die explizite Fehlerschranke ist, wie wir wissen, oft sehr hilfreich.

Lösung: (0) Wir beginnen mit der Erinnerung zur Abstandsmessung. Für diskrete WMaße $\mathbf{P}_0, \mathbf{P}_1$ nutzen wir den totalen Abstand U3D:

$$\big\|\mathbf{P}_0 - \mathbf{P}_1\big\| := \sup_{A\subseteq\Omega}\big|\mathbf{P}_0(A) - \mathbf{P}_1(A)\big| = \frac{1}{2}\sum_{\omega\in\Omega}\big|\mathbf{P}_0(\{\omega\}) - \mathbf{P}_1(\{\omega\})\big|$$

(1) Wir berechnen die Wkten geduldig und sorgsam:

$$\begin{aligned}
\mathbf{P}\big[X+Y=n\big] &\overset{\text{Zerlegung}}{=} \sum_{k+\ell=n}\mathbf{P}\big[X{=}k, Y{=}\ell\big] \\
\overset{\text{Unabh'keit}}{=} \sum_{k+\ell=n}\mathbf{P}\big[X{=}k\big]\cdot\mathbf{P}\big[Y{=}\ell\big] &\overset{\text{Poisson}}{=} \sum_{k+\ell=n}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\cdot\mathrm{e}^{-\mu}\frac{\mu^\ell}{\ell!} \\
= \frac{1}{n!}\mathrm{e}^{-(\lambda+\mu)}\underbrace{\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k}}_{\text{binomische Formel!}} &= \underbrace{\frac{(\lambda+\mu)^n}{n!}\mathrm{e}^{-(\lambda+\mu)}}_{\text{erneut Poisson!}}
\end{aligned}$$

☺ Dies beweist die ersehnte Faltungsformel $P(\lambda) * P(\mu) = P(\lambda+\mu)$. Diese bemerkenswerte Eigenschaft beweisen wir erneut auf Seite U445, dort mit erzeugenden Funktionen: genauso leicht und noch eleganter.

(2) Auf der Menge $\mathbb{N}$ vergleichen wir die Binomialverteilung $B(1,t)$ mit Trefferwkt $t\in[0,1]$ und die Poisson–Verteilung $P(t)$ zum Parameter t:

$$\begin{aligned}
2\big\|B(1,t) - P(t)\big\| &= \textstyle\sum_{k\in\mathbb{N}}\big|B(1,t)(k) - P(t)(k)\big| \\
&\overset{\text{Def}}{=} \left|(1-t) - \frac{t^0}{0!}\mathrm{e}^{-t}\right| + \left|t - \frac{t^1}{1!}\mathrm{e}^{-t}\right| + \sum_{k=2}^{\infty}\frac{t^k}{k!}\mathrm{e}^{-t} \\
&\overset{\text{U1B}}{=} (\mathrm{e}^{-t} - 1 + t) + t(1-\mathrm{e}^{-t}) + (1 - \mathrm{e}^{-t} - t\,\mathrm{e}^{-t}) \\
&= 2t(1-\mathrm{e}^{-t}) \overset{\text{U1B}}{\leq} 2t^2
\end{aligned}$$

Hierbei nutzen wir zweimal die Ungleichung $1-t\leq\mathrm{e}^{-t}$, siehe U1B.

☺ Wir erhalten die ersehnte Abschätzung des totalen Abstands:

$$\boxed{\big\|B(1,t) - P(t)\big\| \leq t^2}$$

Dies ist Satz U4C für den einfachsten Fall $n=1$. Wir übertragen die Ungleichung nun von $X_1,\dots,X_n$ auf ihre Summe $S = X_1+\dots+X_n$.

Poissons Gesetz der kleinen Zahlen

(3) Gegeben sind WMaße $\mathbf{P}, \mathbf{P}'$ auf der Menge A sowie $\mathbf{Q}, \mathbf{Q}'$ auf B. Auf $\Omega = A \times B$ vergleichen wir die Produktmaße $\mathbf{P} \otimes \mathbf{Q}$ und $\mathbf{P}' \otimes \mathbf{Q}'$:

$$
\begin{aligned}
2\big\|\mathbf{P} \otimes \mathbf{Q} - \mathbf{P}' \otimes \mathbf{Q}'\big\| &= \sum_{(a,b)\in A\times B} \big|p(a)q(b) - p'(a)q'(b)\big| \\
&= \sum_{(a,b)\in A\times B} \big|p(a)q(b) - p'(a)q(b) + p'(a)q(b) - p'(a)q'(b)\big| \\
&\le \sum_{(a,b)\in A\times B} \big|p(a) - p'(a)\big| \cdot q(b) + p'(a) \cdot \big|q(b) - q'(b)\big| \\
&= \sum_{a\in A}\sum_{b\in B} q(b) \cdot \big|p(a) - p'(a)\big| + \sum_{b\in B}\sum_{a\in A} p'(a) \cdot \big|q(b) - q'(b)\big| \\
&= \sum_{a\in A}\big|p(a) - p'(a)\big| + \sum_{b\in B}\big|q(b) - q'(b)\big| = 2\|\mathbf{P} - \mathbf{P}'\| + 2\|\mathbf{Q} - \mathbf{Q}'\|
\end{aligned}
$$

☺ Wir erhalten folgende Ungleichung für den totalen Abstand:

$$
\boxed{\big\|\mathbf{P} \otimes \mathbf{Q} - \mathbf{P}' \otimes \mathbf{Q}'\big\| \quad \le \quad \|\mathbf{P} - \mathbf{P}'\| + \|\mathbf{Q} - \mathbf{Q}'\|}
$$

Poissons Gesetz der kleinen Zahlen

(4) Es gilt $|\mathbf{P}_S(A) - \mathbf{P}'_S(A)| = |\mathbf{P}(S^{-1}(A)) - \mathbf{P}'(S^{-1}(A))| \le \|\mathbf{P} - \mathbf{P}'\|$ für jede Teilmenge $A \subseteq \mathbb{N}$. Hieraus folgt die nützliche Ungleichung

$$
\boxed{\|\mathbf{P}_S - \mathbf{P}'_S\| \quad \le \quad \|\mathbf{P} - \mathbf{P}'\|.}
$$

(5) Dank (2,3) haben die Produktmaße $\mathbf{P} = B(1,t_1) \otimes \cdots \otimes B(1,t_n)$ und $\mathbf{P}' = P(t_1) \otimes \cdots \otimes P(t_n)$ auf der Ergebnismenge $\mathbb{N}^n$ den totalen Abstand

$$
\begin{aligned}
\big\|\mathbf{P} - \mathbf{P}'\big\| &= \big\| B(1,t_1) \otimes \cdots \otimes B(1,t_n) \ - \ P(t_1) \otimes \cdots \otimes P(t_n) \big\| \\
&\le \big\|B(1,t_1) - P(t_1)\big\| + \cdots + \big\|B(1,t_n) - P(t_n)\big\| \ \le \ t_1^2 + \cdots + t_n^2.
\end{aligned}
$$

Unter $\mathbf{P}$ ist die Treffersumme $S = X_1 + \cdots + X_n$ verteilt gemäß $\mathbf{P}_S$. Unter $\mathbf{P}'$ ist das Bildmaß $\mathbf{P}'_S = P(\mu)$ die Poisson–Verteilung dank (1). Dank (4) haben beide Bildmaße auf $\mathbb{N}$ den totalen Abstand

$$
\boxed{\big\|\mathbf{P}_S - P(\mu)\big\| \quad \le \quad t_1^2 + \cdots + t_n^2.}
$$

☺ Das ist Poissons Gesetz der kleinen Zahlen gemäß Satz U4C. Speziell für $t_1 = \cdots = t_n = t$ ist S exakt $B(n,t)$–verteilt, also gilt:

$$
\boxed{\big\|B(n,t) - P(\mu)\big\| \quad \le \quad nt^2 = \mu^2/n}
$$

Erzeugende Funktion: einführendes Beispiel

Aufgabe: (1) Sei $X \sim B(n,t)$. Berechnen Sie $G(z) := \mathbf{E}\big[z^X\big]$ für $z \in \mathbb{C}$.
(2) Berechnen Sie aus G umgekehrt $G^{(k)}(0)/k!$ für $k = 0, 1, 2, \dots$

Lösung: (1) Für $X \sim B(n,t)$ binomialverteilt wissen wir

$$\mathbf{P}(X{=}k) \;=\; \binom{n}{k} t^k (1-t)^{n-k}.$$

Als Funktion $G(z) = \mathbf{E}\big[z^X\big]$ erhalten wir ein hübsches Polynom:

$$G(z) \;=\; \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \cdot z^k \;=\; (tz+1-t)^n$$

(2) Es gilt $G^{(k)}(z) = n(n-1)\cdots(n-k+1)\, t^k\, (tz+1-t)^{n-k}$, also

$$\frac{1}{k!} G^{(k)}(0) \;=\; \binom{n}{k} t^k (1-t)^{n-k}.$$

☺ Das geht ganz leicht und das Ergebnis ist noch dazu sehr einfach! Diesen Trick entwickeln wir nun zu einer allgemeinen Rechenmethode.

Erzeugende Funktion: Definition

Definition U4D: erzeugende Funktion einer WVerteilung auf $\mathbb{Z}$

Zur diskreten WVerteilung $g : \mathbb{Z} \to [0,1]$ ist die **erzeugende Funktion**

$$\boxed{G(z) \;:=\; \sum\nolimits_{k\in\mathbb{Z}} g(k) z^k \;=\; \dots + g(-1)z^{-1} + g(0)z^0 + g(1)z^1 + \dots.}$$

Ist die Zufallsvariable $X : \Omega \to \mathbb{Z}$ verteilt gemäß $\mathbf{P}(X{=}k) = g(k)$, so gilt

$$\boxed{G(z) \;=\; \sum\nolimits_{k\in\mathbb{Z}} \mathbf{P}(X{=}k)\, z^k \;=\; \mathbf{E}\big[z^X\big] \qquad \text{für } z \in K(0,\sigma,\rho) \subseteq \mathbb{C}.}$$

Aus G rekonstruieren wir die Verteilung g durch die Wegintegrale (F2Q)

$$\boxed{\mathbf{P}(X{=}k) \;=\; g(k) \;=\; \frac{1}{2\pi\mathrm{i}} \oint_{\partial B(0,1)} \frac{G(z)}{z^{k+1}}\,\mathrm{d}z \qquad \text{für } k \in \mathbb{Z}.}$$

Im Falle einer Potenzreihe, also $X : \Omega \to \mathbb{N}$, genügen die Ableitungen

$$\boxed{\mathbf{P}(X{=}k) \;=\; g(k) \;=\; \frac{1}{k!} G^{(k)}(0) \qquad \text{für } k \in \mathbb{N}.}$$

Erzeugende Funktion: Faltung wird zum Produkt.

Erzeugende Funktionen bieten uns sehr effiziente Rechenregeln:

Satz U4E: Faltung wird zum Produkt erzeugender Funktionen.

Für die Summe $X+Y$ unabhängiger Zufallsvariablen $X,Y:\Omega\to\mathbb{Z}$ gilt

$$\boxed{\begin{aligned} G_{X+Y}(z) &= \mathbf{E}\big[z^{X+Y}\big] = \mathbf{E}\big[z^X\cdot z^Y\big] \\ &= \mathbf{E}\big[z^X\big]\cdot\mathbf{E}\big[z^Y\big] = G_X(z)\cdot G_Y(z). \end{aligned}}$$

Aufgabe: Seien $X\sim B(m,t)$ und $Y\sim B(n,t)$ unabhängig, $m,n\in\mathbb{N}$. Ist die Summe $S=X+Y$ binomialverteilt? Mit welchen Parametern?

Lösung: Die erzeugenden Funktionen haben wir oben ausgerechnet:

$$G_X(z) = (tz+1-t)^m \quad \text{und} \quad G_Y(z) = (tz+1-t)^n$$

Für die Summe $S=X+Y$ erhalten wir somit ohne weitere Mühen

$$G_S(z) = G_X(z)\cdot G_Y(z) = (tz+1-t)^{m+n}.$$

Somit ist die Summe $S\sim B(m{+}n,t)$ tatsächlich binomialverteilt.

Wir erhalten die Faltungsformel $B(m,t)*B(n,t)=B(m{+}n,t)$.

Erzeugende Funktion: Faltung wird zum Produkt.

Glücklicherweise lässt sich die Verteilung von S aus G_S rekonstruieren! Die erzeugende Funktion G_S bestimmt die gesamte Verteilung von S.

Alternativ können wir die Verteilung von $S=X+Y$ direkt ausrechnen durch die entsprechende Reihe, ganz ohne erzeugende Funktionen:

$$\begin{aligned} \mathbf{P}\big[X+Y=s\big] &\overset{\text{Zerlegung}}{=} \sum_{k+\ell=s} \mathbf{P}\big[X{=}k, Y{=}\ell\big] \\ &\overset{\text{Unabh'keit}}{=} \sum_{k+\ell=s} \mathbf{P}\big[X{=}k\big]\cdot\mathbf{P}\big[Y{=}\ell\big] \\ &\overset{\text{binom'vert}}{=} \sum_{k+\ell=s} \binom{m}{k} t^k(1-t)^{m-k}\cdot\binom{n}{\ell}t^\ell(1-t)^{n-\ell} \\ &= t^s(1-t)^{m+n-s}\underbrace{\sum_{k=0}^{s}\binom{m}{k}\binom{n}{s-k}}_{=\binom{m+n}{s} \text{ dank Satz U3A}} \end{aligned}$$

Dies zeigt $S\sim B(m{+}n,t)$ durch direktes Aufsummieren. Unsere erzeugenden Funktionen kodieren dieselbe Information noch effizienter!

Erzeugende Funktion: Poisson–Verteilung

Aufgabe: (1) Sei $X \sim P(\lambda)$ mit $\lambda \geq 0$. Berechnen Sie $G(z) = \mathbf{E}\left[z^X\right]$.
(2) Seien $X \sim P(\lambda)$ und $Y \sim P(\mu)$ unabhängig mit Parametern $\lambda, \mu \geq 0$. Ist die Summe $S = X + Y$ Poisson–verteilt? Mit welchem Parameter?

Lösung: (1) Die erzeugende Funktion ist hier denkbar einfach:

$$G(z) = \sum_{k=0}^{\infty} \mathbf{P}(X{=}k) \cdot z^k = \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \cdot z^k = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = \mathrm{e}^{\lambda(z-1)}$$

Der Exponentialreihe sei Dank! Der Konvergenzradius ist hier ∞.

(2) Für die Summe $S = X + Y$ erhalten wir somit ohne weitere Mühen

$$G_S(z) = G_X(z) \cdot G_Y(z) = \mathrm{e}^{\lambda(z-1)} \cdot \mathrm{e}^{\mu(z-1)} = \mathrm{e}^{(\lambda+\mu)(z-1)}.$$

Somit ist die Summe $S \sim P(\lambda + \mu)$ tatsächlich Poisson–verteilt.

Wir erhalten die schöne Faltungsformel $P(\lambda) * P(\mu) = P(\lambda + \mu)$.

Auch hier stellt sich die Frage: Lässt sich die Verteilung von S aus der Funktion G_S rekonstruieren? Zum Glück wissen wir $g(k) = G^{(k)}(0)/k!$. Die erzeugende Funktion G_S bestimmt so die gesamte Verteilung von S.

Erzeugende Funktion: Poisson–Verteilung

Alternativ können wir auch hier die Verteilung von S direkt ausrechnen:

$$\begin{aligned}
\mathbf{P}\left[X + Y = n\right] &\overset{\text{Zerlegung}}{=} \sum_{k+\ell=n} \mathbf{P}\left[X{=}k, Y{=}\ell\right] \\
&\overset{\text{Unabh'keit}}{=} \sum_{k+\ell=n} \mathbf{P}\left[X{=}k\right] \cdot \mathbf{P}\left[Y{=}\ell\right] \\
&\overset{\text{Poisson}}{=} \sum_{k+\ell=n} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \cdot \mathrm{e}^{-\mu} \frac{\mu^\ell}{\ell!} \\
&= \frac{1}{n!} \mathrm{e}^{-(\lambda+\mu)} \underbrace{\sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k}}_{\text{binomische Formel!}} = \frac{(\lambda+\mu)^n}{n!} \mathrm{e}^{-(\lambda+\mu)}
\end{aligned}$$

Dies zeigt $S \sim P(\lambda + \mu)$ durch direktes Aufsummieren. Unsere erzeugenden Funktionen kodieren dieselbe Information noch effizienter!

Für viele unserer wichtigsten WVerteilungen ist die erzeugende Funktion einfach gebaut. Das ist ein Glücksfall, den wir gerne nutzen.

Für die hypergeometrische Verteilung gilt dies leider nicht.

Erzeugende Funktion: geometrische Verteilung

Aufgabe: Berechnen Sie die erzeugende Funktion $G(z) = \mathbf{E}\big[z^X\big]$ für eine geometrisch verteilte Zufallsvariable $X \sim \mathrm{Geom}(q)$ mit $0 \le q < 1$. Welchen Konvergenzradius hat die Potenzreihe $G(z)$ in diesem Fall?

Lösung: Wir kennen die geometrische Verteilung aus T4A: Für $n \in \mathbb{N}_{\ge 1}$ gilt $\mathbf{P}(X{=}n) = (1-q)q^{n-1}$. Die erzeugende Funktion ist demnach:

$$\begin{aligned} G(z) &= \sum_{k\in\mathbb{Z}} \mathbf{P}(X{=}k)\cdot z^k &&= \sum_{n=1}^{\infty}(1-q)q^{n-1}z^n \\ &= (1-q)z\sum_{j=0}^{\infty}(qz)^j &&= \frac{(1-q)z}{1-qz} \quad \text{für } |z| < 1/q \end{aligned}$$

Die rationale Funktion G hat ihren einzigen Pol im Punkt $z = 1/q$. Der Konvergenzradius ist demnach $\rho = 1/q > 1$, beliebig dicht an 1!

⚠ Es kann also vorkommen, dass die erzeugende Funktion $G(z)$ nicht für alle $z \in \mathbb{C}$ definiert ist. Wegen $G(1) = 1$ ist die Reihe aber zumindest für alle $|z| \le 1$ absolut konvergent. Das genügt für unsere Anwendungen.

Erzeugende Funktion: Konvergenzradius

Aufgabe: Sei $X : \Omega \to \mathbb{N}$ eine ZVariable mit Verteilung $\mathbf{P}(X{=}k) = g(k)$ und erzeugender Funktion $G(z) = \mathbf{E}\big[z^X\big] = \sum_{k=0}^{\infty} g(k)z^k$ für $z \in \mathbb{C}$.

(1) Berechnen Sie $G(1)$. Ist somit der Konvergenzradius ≥ 1? Wie rekonstruieren Sie die Koeffizienten $g(k)$ aus der Funktion $G(z)$?

(2) Was gilt entsprechend für den allgemeinen Fall $X : \Omega \to \mathbb{Z}$?

Lösung: (1) Als WVerteilung gilt $g(k) \ge 0$ und $G(1) = \sum_{k=0}^{\infty} g(k) = 1$. Dank Majoranten-Kriterium konvergiert $G(z)$ absolut für alle $|z| \le 1$. Der Konvergenzradius ρ erfüllt also $\rho \ge 1$, und wir erhalten die Funktion

$$G \,:\, \overline{B}(0,1) = \big\{\, z \in \mathbb{C} \;\big|\; |z| \le 1 \,\big\} \to \mathbb{C}.$$

Innerhalb des Konvergenzkreises ist G beliebig oft differenzierbar. Diese Potenzreihe können wir termweise differenzieren und erhalten $G^{(k)}(z) = k!\,g(k)z^0 + \dots$, also $g(k) = G^{(k)}(0)/k!$ wie in U4D angegeben.

☺ Die WVerteilung g und die Funktion G bestimmen sich gegenseitig.

(2) Allgemein für $X : \Omega \to \mathbb{Z}$ ist $G(z)$ eine Laurent-Reihe. Sie konvergiert mindestens auf der Kreislinie $|z| = 1$, allgemein dem Kreisring $K(0, \sigma, \rho)$. Der Nullpunkt ist eine Singularität, wir nutzen daher Wegintegrale U4D.

Satz U4F: Erwartung und Varianz aus erzeugender Funktion

Sei $X:\Omega\to\mathbb{Z}$ eine Zufallsvariable mit Verteilung $\mathbf{P}(X{=}k)=g(k)$ und erzeugender Funktion $G(z)=\mathbf{E}\big[z^X\big]=\sum_{k\in\mathbb{Z}}g(k)z^k$. Dann gilt

$$\boxed{G(1)=1,\quad G'(1)=\mathbf{E}\big[X\big],\quad G''(1)=\mathbf{E}\big[X(X-1)\big],}$$

falls das erste bzw. zweite Moment existiert. Allgemein erhalten wir so jedes **faktorielle Moment** $G^{(n)}(1)=\mathbf{E}\big[X(X-1)\cdots(X-n+1)\big]$, falls die zugehörige Reihe konvergiert. Speziell für die Varianz folgt daraus:

$$\boxed{\begin{aligned}\mathbf{V}(X)=\mathbf{E}[X^2]-\mathbf{E}[X]^2&=\mathbf{E}[X(X-1)]+\mathbf{E}[X]-\mathbf{E}[X]^2\\&=G''(1)+G'(1)-G'(1)^2\end{aligned}}$$

Aufgabe: Rechnen Sie dies nach. **Lösung:** Die n–te Ableitung ist

$$G^{(n)}(z)=\sum_{k\in\mathbb{Z}}k(k-1)\cdots(k-n+1)\,g(k)\,z^{k-n}.$$

Einsetzen von $z=1$ liefert die angegebenen Formeln, Konvergenz vorausgesetzt. Dieser Trick beschert uns die wichtigsten Kenngrößen.

Aufgabe: Berechnen Sie auf diese Weise erneut Erwartung und Varianz (1) der Binomialverteilung $B(n,t)$ und (2) der Poisson–Verteilung $P(\lambda)$ sowie (3) der geometrischen Verteilung $\mathrm{Geom}(q)$.

Lösung: (1) Für $X\sim B(n,t)$ haben wir $G(z)=(tz+1-t)^n$. Damit gilt:

$$\begin{aligned}&G'(z)=nt\,(tz+1-t)^{n-1}\quad\Longrightarrow\quad\mathbf{E}(X)=G'(1)=nt\\&G''(z)=n(n-1)\,t^2\,(tz+1-t)^{n-2}\quad\Longrightarrow\\&\mathbf{V}(X)=G''(1)+G'(1)-G'(1)^2=nt(1-t)\end{aligned}$$

(2) Für $X\sim P(\lambda)$ haben wir $G(z)=\mathrm{e}^{\lambda(z-1)}$. Daraus folgt:

$$\begin{aligned}&G'(z)=\lambda\,\mathrm{e}^{\lambda(z-1)}\quad\Longrightarrow\quad\mathbf{E}(X)=G'(1)=\lambda\\&G''(z)=\lambda^2\,\mathrm{e}^{\lambda(z-1)}\quad\Longrightarrow\\&\mathbf{V}(X)=G''(1)+G'(1)-G'(1)^2=\lambda^2+\lambda-\lambda^2=\lambda\end{aligned}$$

☺ Das entspricht den zuvor berechneten Werten, siehe U421, U425: Die direkte Berechnung führt ebenso zum Ziel, erzeugende Funktionen organisieren die Rechnung jedoch besonders geschickt und effizient.

Erzeugende Funktion: Erwartung und Varianz

Dieser Trick funktioniert ebenso für die geometrische Verteilung:

(3) Auch für $X \sim \operatorname{Geom}(q)$ kennen wir die erzeugende Funktion

$$G(z) \quad := \quad \sum_{k=1}^{\infty}(1-q)q^{k-1}z^k \quad = \quad \frac{(1-q)z}{1-qz}.$$

Es gilt $G(1)=1$. Die erste Ableitung liefert uns die Erwartung:

$$G'(z) = \frac{1-q}{(1-qz)^2} \quad \Longrightarrow \quad \mathbf{E}(X) = G'(1) = \frac{1}{1-q}$$

Die zweite Ableitung liefert uns die Varianz:

$$G''(z) = \frac{2q(1-q)}{(1-qz)^3} \quad \Longrightarrow$$

$$\mathbf{V}(X) = G''(1) + G'(1) - G'(1)^2 = \frac{q}{(1-q)^2}$$

☺ Das entspricht den zuvor direkt berechneten Werten aus Satz T4A. Auch dort war die Berechnung der erforderlichen Potenzreihen trickreich. Erzeugende Funktionen organisieren die Rechnung besonders effizient.

Faltung geometrischer Verteilungen

Aufgabe: Seien $X \sim \operatorname{Geom}(q)$ und $Y \sim \operatorname{Geom}(p)$ unabhängig mit $0 \le q, p < 1$. Ist die Summe $S = X + Y$ ebenfalls geometrisch verteilt?

Lösung: Die erzeugenden Funktionen haben wir oben ausgerechnet:

$$G_X(z) = \frac{(1-q)z}{1-qz} \quad \text{und} \quad G_Y(z) = \frac{(1-p)z}{1-pz}$$

Für die Summe $S = X + Y$ erhalten wir dank Unabhängigkeit sofort

$$G_S(z) = G_X(z) \cdot G_Y(z) = \frac{(1-q)(1-p)z^2}{(1-qz)(1-pz)}.$$

Wäre S geometrisch verteilt, also $S \sim \operatorname{Geom}(r)$ mit $0 \le r < 1$, so wäre

$$G_S(z) = \frac{(1-r)z}{1-rz}.$$

Der Vergleich zeigt: S ist nicht geometrisch verteilt. (Warum nicht?)

☺ Die Familie der geometrisch verteilten Zufallsvariablen ist also nicht stabil unter Faltung: Das sehen wir leicht dank erzeugender Funktionen! Die Faltungen untersuchen wir in der nächsten Aufgabe und Satz U4G.

Die negative Binomialverteilung $NB(r,p)$

Wir wiederholen ein Zufallsexperiment mit Trefferwahrscheinlichkeit p und Gegenwkt $q = 1 - p$. Sei X_r der Zeitpunkt des r–ten Treffers.

Aufgabe: (1) Berechnen Sie erneut $\mathbf{P}[X_1{=}n]$ und allgemein $\mathbf{P}[X_r{=}n]$. Konstruieren Sie hierzu einen geeigneten WRaum als Grundlage.

(2) Berechnen Sie die erzeugende Funktion $G_r(z) = \sum_{n\in\mathbb{N}} \mathbf{P}[X_r{=}n] z^n$. Was erkennen Sie daran? Folgern Sie hieraus Erwartung und Varianz.

Lösung: (1) Wir finden die Wahrscheinlichkeiten

$$\mathbf{P}\big[X_1{=}n\big] = pq^{n-1} \qquad \text{und} \qquad \mathbf{P}\big[X_r{=}n\big] = \binom{n-1}{r-1} p^r q^{n-r}.$$

Wir nennen dies die **negative Binomialverteilung**, kurz $X_r \sim NB(r,p)$.

Genauer: Ein Einzelexperiment liefert das Ergebnis 0 (Niete) mit Wkt q oder 1 (Treffer) mit Wkt p. Als WRaum $(\Omega, \mathscr{A}, \mathbf{P})$ nutzen wir die Menge Ω aller 0–1–Folgen $\omega = (\omega_1, \omega_2, \omega_3, \dots)$ und dem WMaß erzeugt durch $\mathbf{P}[\{(\omega_1,\dots,\omega_n,*)\}] = p^r q^{n-r}$, wobei $r = \omega_1 + \dots + \omega_n$ die Trefferzahl ist. Die Bedingung $X_r(\omega) = n$ bedeutet $\omega_n = 1$ und $\omega_1 + \dots + \omega_{n-1} = r - 1$. Hierzu gibt es $\binom{n-1}{r-1}$ disjunkte Möglichkeiten, jede davon mit Wkt $p^r q^{n-r}$.

Die negative Binomialverteilung $NB(r,p)$

(2) Für $X_1 \sim NB(1,p) = \mathrm{Geom}(q)$ kennen wir die erzeugende Funktion:

$$G_1(z) = \sum_{k\in\mathbb{Z}} \mathbf{P}[X_1{=}k] \cdot z^k = \sum_{n=1}^{\infty} pq^{n-1} z^n = pz \sum_{j=0}^{\infty} (qz)^j = \frac{pz}{1-qz}$$

Für den allgemeinen Fall $X_r \sim NB(r,p)$ finden wir entsprechend:

$$\begin{aligned} G_r(z) &= \sum_{k\in\mathbb{Z}} \mathbf{P}[X_r{=}k] \cdot z^k = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r q^{n-r} z^n \\ &= (pz)^r \sum_{j=0}^{\infty} \binom{j+r-1}{r-1} (qz)^j = (pz)^r \Bigg[\sum_{i=0}^{\infty} (qz)^i\Bigg]^r = \left(\frac{pz}{1-qz}\right)^r \end{aligned}$$

☺ Dies ist das r–fache Produkt von $G_1(z)$. Wir erhalten die Faltungsformeln $NB(r,p) = \mathrm{Geom}(q)^{*r}$ und $NB(r,p) * NB(s,p) = NB(r{+}s,p)$.

☺ Anschaulich: Die Wartezeit $X_r = W_1 + \dots + W_r$ auf den r–ten Treffer ist die Summe der Wartezeiten $W_1, \dots, W_r \sim \mathrm{Geom}(q)$ auf den jeweils nächsten Treffer. Diese sind ohne Gedächtnis, also unabhängig!

Wir fassen unsere Ergebnisse in folgendem schönen Satz zusammen.

Satz U4G: negative Binomialverteilung

Wir wiederholen ein Zufallsexperiment mit Trefferwahrscheinlichkeit p und Gegenwkt $q = 1 - p$. Für den Zeitpunkt X_r des r–ten Treffers gilt

$$\mathbf{P}\big[X_1{=}n\big] = pq^{n-1} \qquad \text{und} \qquad \mathbf{P}\big[X_r{=}n\big] = \binom{n-1}{r-1} p^r q^{n-r}.$$

Wir nennen dies die **negative Binomialverteilung**, kurz $X_r \sim NB(r,p)$. Im Spezialfall $r = 1$ ist dies die geometrische Verteilung, $X_1 \sim \operatorname{Geom}(q)$.

Für die erzeugende Funktion $G_r(z) = \sum_{n\in\mathbb{N}} \mathbf{P}[X_r{=}n] \cdot z^n$ gilt

$$G_1(z) = \frac{pz}{1-qz} \qquad \text{und} \qquad G_r(z) = \left(\frac{pz}{1-qz}\right)^r$$

Das zeigt $NB(r,p) = \operatorname{Geom}(q)^{*r}$ und $NB(r,p) * NB(s,p) = NB(r{+}s,p)$. Die mittlere Wartezeit bis zum r–ten Treffer und ihre Varianz sind also

$$\mathbf{E}[X_r] = r/p \qquad \text{und} \qquad \mathbf{V}[X_r] = rq/p^2.$$

Aufgabe: Sie werfen unabhängig wiederholt einen fairen sechsseitigen Würfel bis Sie im n–ten Versuch das hundertste Mal eine „6“ erhalten.

(1) Bestimmen Sie die Wahrscheinlichkeit $p(n)$ als Funktion von n sowie Erwartung und Varianz der Wartezeit bis zum hundertsten Treffer.

(2) Mit welcher Wkt gelingen 100 Treffer mit höchstens 710 Würfen? Welche Abschätzung liefert Chebychev? der LGS/ZGS? Was gilt exakt?

Lösung: (1) Die Wkt ist negativ-binomial-verteilt mit Parametern $p = 1/6$ und $r = 100$, explizit $p(n) = NB(100, 1/6)(n) = \binom{n-1}{99}(1/6)^{100}(5/6)^{n-100}$. Dank Satz U4G gilt $\mu = r/p = 600$ und $\sigma^2 = rq/p^2 = 3000$ also $\sigma \approx 55$.

(2) Chebychev liefert eine grobe aber bequeme erste Abschätzung:

$$\mathbf{P}\big[X_r > \mu + 2\sigma\big] \le \frac{1}{1+2^2} = 20\% \quad \text{also} \quad \mathbf{P}\big[X_r \le \mu + 2\sigma\big] \ge 80\%$$

Später erhalten wir noch wesentlich bessere Näherungen durch den lokalen Grenzwertsatz V3A und den zentralen Grenzwertsatz W1D.

Die exakte Wkt ist $\sum_{n=100}^{710} p(n) = \sum_{k=100}^{710} \binom{710}{k}(1/6)^k(5/6)^{710-k}$: mühsam! Die Summe $0.97309\ldots$ lassen wir durch einen Computer berechnen.

Erzeugende Funktion löst Rekursionsgleichung!

Die erzeugende Funktion $G_X(z) = \mathbf{E}[z^X]$ ist eine konzise Schreibweise für die Verteilung einer Zufallsvariablen $X : \Omega \to \mathbb{Z}$. Die Methode ist sehr effizient, wenn wir $\mathbf{E}[X] = G'(1)$ berechnen wollen, aber die explizite Verteilung von X kompliziert ist. Hier ist ein frappierendes Beispiel:

Aufgabe: Seien $X_1, X_2, \ldots \sim B(1,p)$ unabhängige Experimente und T die Wartezeit zum ersten Erfolgslauf der Länge r, ausgeschrieben:

$$T(\omega) := \inf\left\{\, k \ge r \;\middle|\; X_k(\omega) = X_{k-1}(\omega) = \cdots = X_{k-r+1}(\omega) = 1 \,\right\}$$

Berechnen Sie rekursiv $\mathbf{P}[T{=}n]$ und daraus die Erwartung $\mathbf{E}[T]$.

Lösung: (1) Für $n < r$ gilt $\mathbf{P}[T{=}n] = 0$. Für $n = r$ gilt $\mathbf{P}[T{=}r] = p^r$. Für $n > r$ zerlegen wir: An einer Stelle $k \in \{1, 2, \ldots, r\}$ muss die erste 0 auftreten, danach noch Wartezeit $n - k$. Somit gilt die Rekursionsformel

$$\mathbf{P}[T{=}n] \;=\; \sum_{k=1}^{r} p^{k-1}(1-p)\,\mathbf{P}[T{=}n-k].$$

☺ Ein Computer kann damit alle Werte $\mathbf{P}[T{=}n]$ rekursiv berechnen. Geschlossene Formel ist schwierig... gelingt aber als erz. Funktion!

Erzeugende Funktion löst Rekursionsgleichung!

(2) Sei $G(z) := \mathbf{E}[z^T] = \sum_{n=1}^{\infty} z^n\,\mathbf{P}[T{=}n]$. Einsetzen von (1) ergibt:

$$\begin{aligned}
G(z) &\overset{\text{Rek}}{\underset{(1)}{=}} z^r p^r + \sum_{n>r} z^n \sum_{k=1}^{r} p^{k-1}(1-p)\,\mathbf{P}[T = n-k] \\
&\overset{\text{Distr}}{=} (zp)^r + z(1-p) \sum_{k=1}^{r} (zp)^{k-1} \sum_{n>r} z^{n-k}\,\mathbf{P}[T = n-k] \\
&\overset{\text{Def}}{=} (zp)^r + z(1-p)\,\frac{1-(zp)^r}{1-zp}\,G(z)
\end{aligned}$$

☺ Die erzeugende Funktion löst wunderbar die Rekursion! Wir finden:

$$G(z) = \frac{z^r p^r (1-zp)}{1 - z + z^{r+1} p^r (1-p)}$$

Wir erhalten so explizit die gesuchte Funktion G. Bei Bedarf rekonstruieren wir daraus all ihre Koeffizienten $\mathbf{P}[T{=}n]$ durch Partialbruchzerlegung und Entwicklung. Im Spezialfall $r = 1$ ist dies die vertraute geometrische Verteilung, siehe Satz T4A und die erz. Funktion U447.

☺ Für den Erwartungswert $\mathbf{E}[T]$ genügt uns einmaliges Ableiten:

$$\mathbf{E}[T] \overset{\text{U4F}}{=} G'(1) = \ldots = \frac{1-p^r}{p^r(1-p)}$$

Zufallssummen zufälliger Länge

Wir nutzen oft Summen $S_n = \sum_{k=1}^{n} X_k$ unabhängiger Zufallsvariablen. Manchmal ist selbst die Länge n dieser Summe zufällig! Ausführlich:

Satz U4H: Zufallssummen zufälliger Länge

Seien $T, X_1, X_2, \ldots : \Omega \to \mathbb{N}$ unabhängige Zufallsvariablen. Wir setzen

$$S := \sum_{k=1}^{T} X_k, \quad \text{also} \quad S : \Omega \to \mathbb{N} : \omega \mapsto S(\omega) := \sum_{k=1}^{T(\omega)} X_k(\omega).$$

Alle X_k seien gleichverteilt. Dann gilt $G_S = G_T \circ G_X$. Dank Kettenregel folgt daraus $\mathbf{E}(S) = \mathbf{E}(T)\mathbf{E}(X)$ und $\mathbf{V}(S) = \mathbf{V}(T)\mathbf{E}(X)^2 + \mathbf{E}(T)\mathbf{V}(X)$.

Nachrechnen: Dank U4E wissen wir $G_{X_1+\dots+X_n}(z) = G_X(z)^n$, also:

$$G_S(z) \stackrel{\text{Def}}{=} \sum_{k=0}^{\infty} z^k \, \mathbf{P}[S{=}k] \;\underset{\text{Unabh'keit}}{\stackrel{\text{Zerlegung}}{=}}\; \sum_{k=0}^{\infty} z^k \sum_{n=0}^{\infty} \mathbf{P}[T{=}n] \, \mathbf{P}[S_n{=}k]$$

$$\stackrel{\text{Distr}}{=} \sum_{n=0}^{\infty} \mathbf{P}[T{=}n] \sum_{k=0}^{\infty} z^k \, \mathbf{P}[S_n{=}k] \stackrel{\text{U4E}}{=} \sum_{n=0}^{\infty} \mathbf{P}[T{=}n] \, G_X(z)^n \stackrel{\text{Def}}{=} G_T(G_X(z))$$

Zufallssummen zufälliger Länge

Erwartung und Varianz finden wir durch Ableiten dank Kettenregel:

$$\begin{aligned}
\mathbf{E}[S] &\stackrel{\text{U4F}}{=} G_S'(1) = G_T'(G_X(1))\, G_X'(1) = G_T'(1)\, G_X'(1) = \mathbf{E}[T]\, \mathbf{E}[X] \\
\mathbf{E}[S(S-1)] &\stackrel{\text{U4F}}{=} G_S''(1) = G_T''(G_X(1))\, G_X'(1)^2 + G_T'(G_X(1))\, G_X''(1) \\
\mathbf{V}[S] &\stackrel{\text{U4F}}{=} G_S''(1) + G_S'(1) - G_S'(1)^2 = \ldots = \mathbf{V}[T]\, \mathbf{E}[X]^2 + \mathbf{E}[T]\, \mathbf{V}[X]
\end{aligned}$$

🙂 Erzeugende Funktionen bieten eine bequem-konzise Schreibweise.

Aufgabe: Eine radioaktive Probe emittiert $T \sim P(\lambda)$ Teilchen/Zeit. U429
Im Messgerät wird jedes mit Wkt p registriert, unabhängig voneinander. Berechnen Sie Verteilung, Erwartung, Varianz der registrierten Teilchen.

Lösung: Wir haben $X_1, X_2, \ldots \sim B(1,p)$, also $G_X(z) = (zp - p + 1)$. Für $T \sim P(\lambda)$ wissen wir $G_T(z) = \mathrm{e}^{\lambda(z-1)}$. Für die Summe $S = \sum_{k=1}^{T} X_k$ folgt dank U4H dann $G_S(z) = G_T(G_X(z)) = \mathrm{e}^{\lambda p(z-1)}$, also $S \sim P(\lambda p)$.

Wir finden die Erwartung $\mathbf{E}(S) = \lambda p = \mathbf{E}(T)\mathbf{E}(X)$ und die Varianz $\mathbf{V}(S) = \lambda p = \mathbf{V}(T)\mathbf{E}(X)^2 + \mathbf{E}(T)\mathbf{V}(X)$, so wie es sein muss.

🙂 Leistungsstarke Theorie ermöglicht effiziente Berechnung.

Wir beginnen mit einem Individuum. In seiner kurzen Lebenszeit erzeugt es X Nachkommen, mit $\mathbf{P}[X{=}k] = p_k$ und $G(z) := \mathbf{E}[z^X] = \sum_{k=0}^{\infty} p_k z^k$. Alle Nachkommen vermehren sich ebenso, unabhängig voneinander. Die Zufallsvariable Z_n sei die Anzahl der Individuuen in Generation n.

Francis Galton (1822–1911) und Henry William Watson (1827–1903) untersuchten um 1873 das Aussterben aristokratischer Familienamen. Gleiches gilt für bedrohte Tierarten, die Ausbreitung von Genen und Memen, Epidemien und sozial-virale Phänomene, Computer-Viren, Kettenbriefe, ebenso künftige Weltraummissionen oder Nanotechnologie mit selbstreplizierenden Maschinen. Kernspaltung, ob als Bombe oder als Reaktor, folgt demselben Mechanismus der Kettenreaktion: Ein freies Neutron erzeugt zufällig weitere. Im deterministischen Fall gilt $p_a = 1$ für ein $a \in \mathbb{N}$ und somit sicher $Z_n = a^n$ für alle $n \in \mathbb{N}$. Fortan schließen wir diesen trivialen Sonderfall aus.

Satz U4I: Aussterbewahrscheinlichkeit nach Galton–Watson

Die Aussterbewkt $q \in [0,1]$ ist die kleinste Lösung der Fixpunktgleichung

$$\boxed{G(q) = q.}$$

Aus der Erwartung $\mu := G'(1) > 1$ folgt $q < 1$: Überleben ist möglich.
Aus Erwartung $\mu \le 1$ (und $p_1 < 1$) folgt $q = 1$: fast sicheres Aussterben.

Aufgabe: Führen Sie das Modell und den Satz sorgfältig aus. **Lösung:**

Das Individuum $i \in \mathbb{N}$ in Generation $n \in \mathbb{N}$ erzeugt $X_{n,i}$ Nachkommen. Zur Vereinfachung nehmen wir hier an, alle $X_{n,i}$ sind unabhängig und identisch verteilt, mit derselben erzeugenden Funktion $G(z) = \mathbf{E}[z^{X_{n,i}}]$. Aus Z_n Individuuen in Generation n werden so $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$.

Wir beginnen mit einem Individuum, also $Z_0 = 1$ und $H_0(z) = z$.
Die Anzahl $Z_1 = X_{0,1}$ hat die erzeugende Funktion $H_1(z) = G(z)$.
Rekursiv ist $H_n = G \circ H_{n-1} = G^{\circ n}$ die n–fache Komposition dank U4H.

Dieser **Galton–Watson–Prozess** $(Z_n)_{n\in\mathbb{N}}$ ist eine Markov–Kette mit Zustandsraum $\mathbb{N}$ und Übergangswkten $\mathbf{P}[Z_n{=}b \mid Z_{n-1}{=}a] = p_b^{*a}$.

Aus $Z_n = 0$ folgt $Z_{n+1} = 0$: Aussterben ist ein absorbierender Zustand. Die gesuchte Wkt $q_n := \mathbf{P}[Z_n{=}0] = H_n(0)$ ist somit monoton wachsend.

Rekursiv gilt $q_n = G(q_{n-1})$. Wir suchen den Grenzwert $q_n \nearrow q \in [0,1]$. Dank Stetigkeit gilt $G(q) = G(\lim q_n) = \lim G(q_n) = \lim q_{n+1} = q$.

🙂 Mit dieser **Fixpunktgleichung** können wir q leicht berechnen.

Reproduktion und Aussterben: zufällige Zellteilung

Im Falle $p_0 + p_1 = 1$ entspricht der Prozess $(Z_n)_{n\in\mathbb{N}}$ dem Warten auf die erste Null wiederholter 0–1–Experimente: geometrische Verteilung T4A. Wir fordern daher $p_a > 0$ für mindestens ein $a \in \mathbb{N}_{\ge 2}$. Die erz. Funktion $G:[0,1] \to [0,1]$ ist dann strikt konvex. Die Fixpunktgleichung $G(q) = q$ hat neben $G(1) = 1$ somit höchstens eine weitere Lösung $q \in [0,1[$.

Zellteilung: Eine Zelle teilt sich ($p_2 > 0$), überlebt (p_1) oder stirbt (p_0). In unserem Modell gilt dann $p_0 + p_1 + p_2 = 1$ und $p_a = 0$ für alle $a \ge 3$.

Aufgabe: Berechnen Sie die Aussterbewkt q als Funktion von p_0, p_1, p_2.

Lösung: Die erz. Funktion $G(z) = p_0 + p_1 z + p_2 z^2$ ist quadratisch. Der Fixpunkt $q = G(q)$ ist dank Mitternachtsformel nach Vereinfachung

$$q = \frac{(p_0 + p_2) - |p_0 - p_2|}{2p_2} = \begin{cases} 1 & \text{falls } p_0 \ge p_2, \\ p_0/p_2 & \text{falls } p_0 \le p_2. \end{cases}$$

Aus $p_0 = p_2$ folgt $q = 1$. Das ist bemerkenswert und nicht intuitiv.

Allgemein gilt $\mathbf{E}[Z_n] = \mu^n$ dank Satz U4H. Im subkritischen Fall $\mu < 1$ ist die Aussterbezeit $T = \inf\{\, n \in \mathbb{N} \mid Z_n = 0 \,\}$ endlich mit Wkt $q = 1$. Genauer gilt $\mathbf{P}[T{>}n] = \mathbf{P}[Z_n{\ge}1] \le \mu^n$ dank Markov–Ungleichung T3G.

Reproduktion und Aussterben: Poisson–Verteilung

Aufgabe: Illustrieren Sie die Argumente für poisson–verteilte $X \sim P(\mu)$. Bestimmen Sie graphisch $q_n = \mathbf{P}[Z_n{=}0]$ und den Grenzwert $q_n \nearrow q$. Berechnen und zeichnen Sie die Aussterbewkt $q(\mu)$ für $\mu \in [1,4]$.

Lösung: Hier aktivieren Sie wunderbar Ihre numerischen Methoden: