Die Binomialverteilung und ihr mathematisches Umfeld

Stefan Götz
Universität Wien und Akademisches Gymnasium Wien

Zusammenfassung

In dieser Arbeit soll die Binomialverteilung (als wichtiges Beispiel einer diskreten Verteilung) mit Methoden der Analysis näher untersucht werden. Die Wahrscheinlichkeitsrechnung liefert dabei die Ausgangsprobleme, die Analysis wird hier ihrem Werkzeugcharakter bei der Lösung derselben gerecht. Die Interpretation der Lösungen übernimmt wieder die Wahrscheinlichkeitsrechnung.

Im Detail wird das Bernoullische Gesetz der großen Zahlen aus einer konkreten Fragestellung hergeleitet und mit der Tschebyscheff-Ungleichung verglichen, verschiedene Methoden zur Berechnung des Erwartungswertes und der Varianz werden vorgestellt. Beide Momente werden erst entsprechend motiviert. Weiters werden „verwandte“ Verteilungen wie die Hypergeometrische und die Poisson-Verteilung besprochen (ausgehend von einer wahrscheinlichkeitstheoretischen bzw. analytischen Problemstellung, welche die entsprechenden Herleitungen implizieren), schließlich wird die Verteilung der Summe von zwei binomialverteilten Zufallsvariablen berechnet.

Wird die Wahrscheinlichkeitsrechnung in dieser Art und Weise als ein Teilgebiet der Mathematik betrachtet, welches analytisch untersucht werden kann, so fügt sie sich harmonisch in das Gebäude der Schulmathematik ein. Zugleich gewinnt die Analysis ein schönes Anwendungsgebiet.

Grundlegende Definitionen

DEFINITION (nach *Kolmogoroff*): Sei P eine Funktion von $\Omega \to \mathbb{R}$, wobei Ω ein System von Ereignissen darstellt. P heißt *Wahrscheinlichkeit,* wenn gilt:

(I) $0 \leq P(A) \leq 1 \quad \forall A \subseteq \Omega$.

(II) $P(\Omega) = 1$.

(III) $P(A \cup B) = P(A) + P(B)$ für $A \cap B = \emptyset$ $(A, B \subseteq \Omega)$.

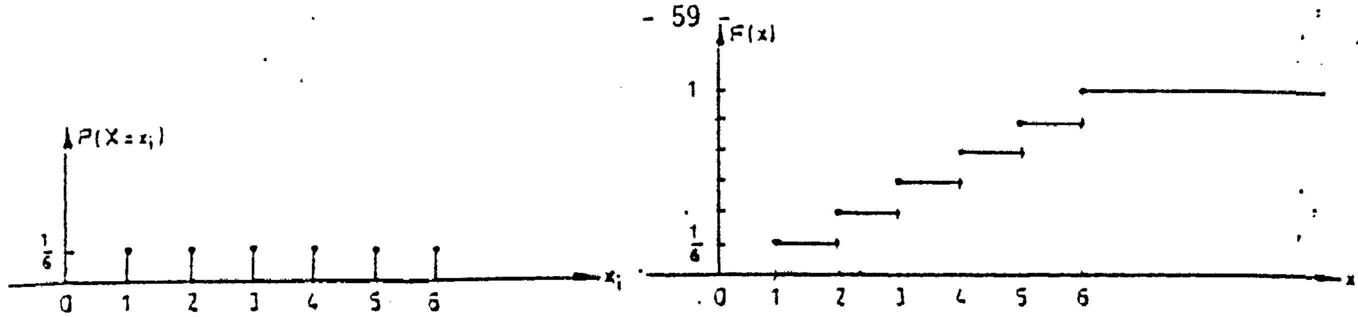

Abbildung 1: Verteilung der Wahrscheinlichkeiten — Verteilungsfunktion

DEFINITION: Eine auf Ω definierte reellwertige Funktion X heißt *Zufallsvariable (ZV)* oder *Zufallsgröße*, wenn sie folgende Eigenschaften besitzt:
Für jedes $x \in \mathbb{R}$ und jedes reelle Intervall $(a, b]$, $a < b$ ($a = -\infty$ ist auch möglich!), besitzen die Ereignisse

$$A_x = \{\omega \in \Omega | X(\omega) = x\} \text{ und } A_{(a,b]} = \{\omega \in \Omega | a < X(\omega) \leq b\}$$

Wahrscheinlichkeiten.
Statt $P(A_x)$ schreiben wir $P(X = x)$ und $P(a < X \leq b)$ statt $P(A_{(a,b]})$.
Unter ihrem *Wertebereich* verstehen wir die Menge

$$\{x \in \mathbb{R} | \exists \omega \in \Omega \text{ mit } X(\omega) = x\} \, .$$

DEFINITION: Eine ZV X heißt ***diskret***, wenn ihr Wertebereich abzählbar ist (das heißt insbesondere, wenn er endlich ist).

DEFINITION: Zwei diskrete ZV X (Wertebereich W_X) und Y (Wertebereich W_Y) heißen ***stochastisch unabhängig***, wenn

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

für alle Wertepaare (x_i, y_j) mit $x_i \in W_X$ und $y_j \in W_Y$ gilt.

Die folgenden beiden Definitionen dienen der *Charakterisierung* von (diskreten) ZV:

DEFINITION: Für eine diskrete ZV X mit dem Wertebereich $\{x_1, x_2, \ldots\}$ ist die *Verteilung der Wahrscheinlichkeiten* gegeben durch den Wahrscheinlichkeitsvektor

$$(p_1, p_2, \ldots), \text{ wobei } p_i = P(X = x_i) \quad (i = 1, 2, \ldots)$$

ist. Es ist $\sum_i p_i = 1$.

DEFINITION: Sei X eine beliebige ZV. Die Funktion $F : \mathbb{R} \to \mathbb{R}$ mit

$$F(x) = P(X \leq x).$$

ist die *Verteilungsfunktion* von X.

Beispiel 0: Es werde mit einem idealen Würfel gewürfelt, die ZV X beschreibe die geworfene Augenzahl. Wie sieht die Verteilung der Wahrscheinlichkeiten *und die* Verteilungsfunktion *aus?*

Lösung: Der Wertebereich ist $\{1, 2, 3, 4, 5, 6\}$. — Siehe Abbildung 1! ◇

Ausgangsproblem

Ein Zufallsexperiment habe nur zwei Ausgänge: Das Ereignis A tritt ein (mit der Wahrscheinlichkeit p) oder nicht (mit der Wahrscheinlichkeit $1-p$) *(dichotomes Merkmal)*. Dieses Zufallsexperiment werde n-mal wiederholt, wobei der Ausgang des i-ten Experiments nicht von den Ausgängen der vorangegangenen beeinflußt wird und auch die nachfolgenden nicht beeinflußt ($i = 1, \ldots, n$). (Die Ereignisse seien also *paarweise unabhängig.*) Insgesamt handelt es sich um ein sogenanntes *Bernoulli-Experiment.*

Man interessiert sich nun für die Wahrscheinlichkeit, daß A *k-mal* ($0 \leq k \leq n$) *eintritt.* Die *ZV X* zähle, wie oft A eintritt, dann gilt:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} .$$

Zu den einzelnen Faktoren:

- p^k: A tritt k-mal ein.
- $(1-p)^{n-k}$: $(n-k)$-mal tritt A *nicht* ein.
- $\binom{n}{k}$: Anzahl der gleichwahrscheinlichen Möglichkeiten, die k tatsächlichen Eintritte von A auf die n Durchführungen des Experiments aufzuteilen: *Kombination ohne Wiederholung.* (Suche aus n Plätzen $1, 2, \ldots, n$ k aus.)

Man sagt: *„X ist binomialverteilt mit den Parametern n und p."*

Beispiel 1: In einer Kugellageranfertigung *sind 90% der hergestellten Kugeln 1. Qualität, der Rest 2. Qualität. Wie viele Kugeln müssen mindestens hergestellt werden, damit die Wahrscheinlichkeit, daß darunter wenigstens eine Kugel von 2. Qualität ist, mindestens 0,5 beträgt?*

Lösung: Sei X die ZV, welche die Kugeln 2. Qualität unter den n hergestellten zählt. Dann ist X binomialverteilt mit den Parametern n und 0,1. Wir fordern

$$P(X \geq 1) \geq 0,5 \ , \quad \text{woraus wir}$$

$$1 - P(X = 0) \geq 0,5 \quad \text{und} \quad P(X = 0) \leq 0,5 \quad \text{erhalten.}$$

Einsetzen liefert schließlich das Gewünschte:

$$\binom{n}{0} \cdot 0,1^0 \cdot 0,9^n \leq 0,5 \Longleftrightarrow 0,9^n \leq 0,5$$

$$\begin{aligned} \ln 0,9^n &\leq \ln 0,5 \quad \text{(ln ist streng monoton wachsend!)} \\ n \cdot \ln 0,9 &\leq \ln 0,5 \\ n &\geq \frac{\ln 0,5}{\ln 0,9} = 6,5\ldots \quad (\ln 0,9 < 0\,!) \end{aligned}$$

Es müssen also wenigstens *sieben* Stück erzeugt werden. ◇

Beispiel 2: Zwei Kleinbusse *müssen insgesamt zehn Personen von A nach B transportieren. Jede Person besteige unabhängig von den anderen mit $p = \frac{1}{2}$ einen der beiden Busse. Wie viele Sitzplätze muß jeder der beiden Busse mindestens haben, wenn die Wahrscheinlichkeit dafür, daß alle Personen einen Sitzplatz erhalten, mindestens gleich 0,95 sein soll?*

Lösung: Die ZV X_1 beschreibe die Anzahl der Personen im ersten Bus, X_2 analog für den zweiten. Dann sind sowohl X_1 als auch X_2 binomialverteilt mit den Parametern $n = 10$ und $p = \frac{1}{2}$.
Es soll nun

$$P(X_1 \leq c, X_2 \leq c) \geq 0,95$$

gelten, wobei c die gesuchte Mindestanzahl ist.
Wir schließen

$$X_2 \leq c \Longleftrightarrow 10 - X_1 \leq c \Longleftrightarrow X_1 \geq 10 - c$$

und sehen, daß X_1 und X_2 hochgradig *abhängig* voneinander sind!
Wir fordern also

$$P(10 - c \leq X_1 \leq c) \geq 0,95 \ .$$

Einsetzen ergibt

$$\sum_{i=10-c}^{c} \binom{10}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2}\right)^{10-i} = \frac{1}{2^{10}} \cdot \sum_{i=10-c}^{c} \binom{10}{i} \geq 0,95 \text{ und schließlich}$$

$$\sum_{i=10-c}^{c} \binom{10}{i} \geq 972,8 \ .$$

Diese (Un)gleichung lösen wir durch *Probieren*: Für $c = 7$ beträgt die Summe $912 < 972,8$; für $c = 8$ dagegen wächst die Summe auf $1002 > 972,8$. — Jeder Bus muß wenigstens *acht* Plätze bieten. ◊

1. Frage: Ist dadurch wirklich eine Verteilung gegeben?

Dazu müßte vor allem gelten: $\sum_k P(X = k) = 1$.

Bei uns ergibt sich daraus

$$\sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1 \ ,$$

worin der *Binomische Lehrsatz* zur Anwendung gekommen ist, was auch den *Namen der Verteilung* klärt.

Die zugehörige *Verteilungsfunktion* lautet:

$$F(x) = P(X \le x) = \begin{cases} \sum_{k=0}^{[x]} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} & 0 \le x \le n\ , \\ 1 & x > n\ , \\ 0 & x < 0\ . \end{cases}$$

($[x]$ bezeichnet dabei die nächst kleinere ganze Zahl zu $x \in \mathbb{R} \setminus \mathbb{Z}$ bzw. x für $x \in \mathbb{Z}$ — Gauß-Klammer.)

Sie erfüllt

- Monotonie,
- $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$ und
- F ist rechtsseitig stetig: $\lim_{h\searrow 0} F(x+h) = F(x)$.

2. Frage: Woher bekomme ich das p?

Antwort: aus der *relativen Häufigkeit* $R_n = \frac{X}{n}$ (ebenfalls eine ZV!).

Rechtfertigung (innerhalb des Modells):
Es ist

$$P(X = k) = P\left(\frac{X}{n} = R_n = \frac{k}{n}\right) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad 0 \le k \le n\ .$$

Wir betrachten ($\varepsilon > 0$ beliebig, aber fest)

$$P(|R_n - p| > \varepsilon) = P(R_n < p - \varepsilon) + P(R_n > p + \varepsilon)\ .$$

Daraus folgt

$$\begin{aligned} P(|R_n - p| > \varepsilon) &= P[X < n\cdot(p-\varepsilon)] + P[X > n\cdot(p+\varepsilon)] = \\ &= \sum_{k<n\cdot(p-\varepsilon)} P(X=k) + \sum_{k>n\cdot(p+\varepsilon)} P(X=k)\ . \end{aligned}$$

Es ist

$$\begin{array}{lcl} k < n\cdot(p-\varepsilon) = np - n\varepsilon & \text{bzw.} & k > n\cdot(p+\varepsilon) = np + n\varepsilon \\ n\varepsilon < np - k & & k - np > n\varepsilon \\ n^2\varepsilon^2 < (np-k)^2 & & (k-np)^2 > n^2\varepsilon^2\ , \end{array}$$

woraus in beiden Fällen

$$\frac{(k-np)^2}{n^2\varepsilon^2} > 1$$

folgt.

Wir erhalten die Abschätzung $P(|R_n - p| > \varepsilon) <$

$$< \sum_{k<n\cdot(p-\varepsilon)} \frac{(k-np)^2}{n^2\varepsilon^2} \cdot P(X=k) + \sum_{k>n\cdot(p+\varepsilon)} \frac{(k-np)^2}{n^2\varepsilon^2} \cdot P(X=k) \ ,$$

woraus wir

$$n^2\varepsilon^2 \cdot P(|R_n - p| > \varepsilon) < \sum_{k=0}^{n} (k-np)^2 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

ableiten. Wir zerlegen den letzten Ausdruck in drei Summen:

$$S_1 = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ ,$$

$$S_2 = 2np \cdot \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \text{ und}$$

$$S_3 = n^2p^2 \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = n^2p^2 \ .$$

Für die Berechnung von S_2 benötigen wir

$$k \cdot \binom{n}{k} = k \cdot \frac{n!}{k! \cdot (n-k)!} = n \cdot \frac{(n-1)!}{(k-1)! \cdot (n-k)!} = n \cdot \binom{n-1}{k-1} \ .$$

Damit ist mit $q \doteq 1-p$

$$\begin{aligned} S_2 &= 2np \cdot \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} \cdot p^k \cdot q^{n-k} = \\ &= 2n^2p \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \cdot p^{j+1} \cdot q^{n-j-1} = \\ &= 2n^2p^2 \cdot \sum_{j=0}^{m} \binom{m}{j} \cdot p^j \cdot q^{m-j} = 2n^2p^2 \cdot [p + (1-p)]^m = 2n^2p^2 \ . \end{aligned}$$

$S_1 = ?$ — Dazu: $k^2 = k \cdot (k-1) + k$ (TRICK!). Eingesetzt ergibt sich für S_1

$$\sum_{k=1}^{n} \underbrace{k \cdot (k-1) \cdot \binom{n}{k}}_{=\frac{n!}{(k-2)! \cdot (n-k)!} = n \cdot (n-1) \cdot \binom{n-2}{k-2}} \cdot p^k \cdot (1-p)^{n-k} + \underbrace{\sum_{k=1}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}_{=np \text{ (siehe } S_2 \text{ !)}} \ .$$

Es folgt

$$\begin{aligned} S_1 &= n \cdot (n-1) \cdot p^2 \cdot \sum_{k=2}^{n} \binom{n-2}{k-2} \cdot p^{k-2} \cdot q^{n-k} + np = \\ &= n \cdot (n-1) \cdot p^2 \cdot \sum_{j=0}^{n-2} \binom{n-2}{j} \cdot p^j \cdot q^{n-2-j} + np = \\ &= n \cdot (n-1) \cdot p^2 \cdot \sum_{j=0}^{m} \binom{m}{j} \cdot p^j \cdot q^{m-j} + np = \\ &= n \cdot (n-1) \cdot p^2 \cdot [p + (1-p)]^m + np = n \cdot (n-1) \cdot p^2 + np \ . \end{aligned}$$

Insgesamt ergibt sich

$$\begin{aligned} n^2\varepsilon^2 \cdot P(|R_n - p| > \varepsilon) &< n \cdot (n-1) \cdot p^2 + np - 2n^2p^2 + n^2p^2 = \\ &= n^2p^2 - np^2 + np - n^2p^2 = np \cdot (1-p) \ , \end{aligned}$$

woraus sich

$$P(|R_n - p| > \varepsilon) < \frac{p \cdot (1-p)}{n\varepsilon^2}$$

ableiten läßt.
Wegen $p \cdot (1-p) \leq \frac{1}{4} \ \forall p \in [0,1]$ ist

$$P(|R_n - p| > \varepsilon) < \frac{1}{4n\varepsilon^2} \ .$$

[Mit $f(p) \doteq p \cdot (1-p)$ gewinnt man $f'(p) = 1 - p + p \cdot (-1) = 1 - 2p$ und Nullsetzen von $f'(p)$ liefert $p = \frac{1}{2}$. Wegen $f''(p) = -2 < 0$ liegt ein lokales Maximum an der Stelle $p = \frac{1}{2}$ vor, was $f\left(\frac{1}{2}\right) = \frac{1}{4}$ zur Folge hat.]
Im *Grenzwert* $n \to \infty$ erkennen wir wegen

$$\lim_{n\to\infty} P(|R_n - p| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

bzw.

$$\lim_{n\to\infty} P(|R_n - p| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$$

das *„Bernoullische Gesetz der großen Zahlen."*

Bemerkung: Das Ergebnis bedeutet *nicht* $\lim_{n\to\infty} r_n = p$, wobei r_n die Realisation der ZV R_n bezeichnet. Abweichungen von r_n um mehr als ε von p werden mit wachsendem n *immer unwahrscheinlicher*, sind aber für endliches n *nicht unmöglich*. Dagegen würde $\lim_{n\to\infty} r_n = p$ bedeuten, daß $\forall \varepsilon > 0 \ \exists n_0 \in \mathbb{N}$ mit $|r_n - p| < \varepsilon \ \forall n \geq n_0$. Diese Formulierung würde also Abweichungen um mindestens ε ab einem gewissen Index ausschließen, was aber nach dem eben Berechneten und Gesagten nicht getan werden darf.

Beispiel 3: Wie oft muß mit einer idealen Münze *mindestens geworfen werden, damit mit einer Wahrscheinlichkeit von wenigstens 0,95 die ZV der relativen Häufigkeit für Wappen* (R_n) *von* $p = \frac{1}{2}$ *um höchstens a) 0,01 oder b) 0,001 abweicht?*

Lösung: Wir setzen in die eben erhaltene Formel

$$P(|R_n - p| > \varepsilon) < \frac{p \cdot (1-p)}{n\varepsilon^2}$$

ein und bekommen

$$P\left(\left|R_n - \frac{1}{2}\right| > \varepsilon\right) < \frac{1}{4n\varepsilon^2} \quad \text{bzw.}$$
$$P\left(\left|R_n - \frac{1}{2}\right| \leq \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2} .$$

a) $\varepsilon = 0,01$: Es ergibt sich

$$1 - \frac{1}{4 \cdot n \cdot 0,01^2} \geq 0,95 \quad \text{und}$$
$$\frac{1}{4 \cdot n \cdot 0,01^2} \leq 0,05 , \quad \text{woraus}$$
$$n \geq \frac{1}{4 \cdot 0,01^2 \cdot 0,05} = 50000 \quad \text{folgt.}$$

b) $\varepsilon = 0,001$: Analog schließen wir $n \geq 5000000$. ◇

Beispiel 4: Ein Bernoulli-Experiment *werde n-mal durchgeführt, das Ereignis A mit* $P(A) = p =$ *const. trete dabei k-mal auf* $(0 \leq k \leq n)$. *Angenommen, p ist nicht bekannt, wie ist dann auf Grund des Versuchsausgangs (i.e. von n-mal ist A k-mal eingetreten) der Parameter p zu „schätzen"?*

Lösung: Es ist

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} ,$$

wobei n und k als bekannt anzunehmen sind und p der gesuchte Parameter ist. Wir legen nun die Philosophie zu Grunde, daß in der Welt das tatsächlich eintritt, was am wahrscheinlichsten ist. Also:

$$f(p) \doteq \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad p \in (0,1) .$$

Für welches p wird $f(p)$ maximal? — Dazu bilden wir die erste Ableitung $f'(p)$:

$$f'(p) = \binom{n}{k} \cdot \left[k \cdot p^{k-1} \cdot (1-p)^{n-k} + p^k \cdot (n-k) \cdot (1-p)^{n-k-1} \cdot (-1)\right] .$$

Nullsetzen derselben liefert

$$\begin{aligned}
k \cdot p^{k-1} \cdot (1-p)^{n-k} &= p^k \cdot (n-k) \cdot (1-p)^{n-k-1} \\
k \cdot (1-p) &= p \cdot (n-k) \\
k - kp &= np - kp \\
p &= \frac{k}{n} = r_n \ .
\end{aligned}$$

Um zu überprüfen, ob wir tatsächlich ein Maximum gefunden haben, könnten wir einfach $f''(r_n)$ ausrechnen, dies ist aber sehr mühsam! Stattdessen betrachten wir die Funktion $g \doteq \ln f$, welche an der selben Stelle wie f ihr Maximum (wenn vorhanden) annimmt, denn der Logarithmus ist eine stetige, streng monoton wachsende Funktion.
Einsetzen ergibt

$$g(p) = \ln \binom{n}{k} + k \cdot \ln p + (n-k) \cdot \ln(1-p) \ .$$

Wir leiten nun zweimal nach p ab und erhalten

$$\begin{aligned}
g'(p) &= k \cdot \frac{1}{p} + (n-k) \cdot \frac{1}{1-p} \cdot (-1) = \frac{k}{p} - \frac{n-k}{1-p} \quad \text{bzw.} \\
g''(p) &= -\frac{k}{p^2} - (n-k) \cdot \frac{1}{(1-p)^2} < 0 \quad (n \geq k) \ .
\end{aligned}$$

Ergo ist $p = \frac{k}{n} = r_n$ ein (lokales) Maximum.

Dies ist ein zweiter Hinweis dafür (innerhalb unseres Modells), daß der theoretische Parameter p und die ZV R_n in engem Zusammenhang stehen! ◇

3. Frage: Wie oft darf ich erwarten, daß A eintreten wird (bei gegebenem n und p)?

Idee: Suche jenes k aus $0, \ldots, n$, für welches $P(X = k)$ *maximal* wird.
Dazu zeigen wir die folgende *Rekursionsformel*

$$b(k+1, n, p) \doteq \binom{n}{k+1} \cdot p^{k+1} \cdot q^{n-k-1} =$$

$$= \frac{n! \cdot (n-k) \cdot p}{(k+1) \cdot k! \cdot (n-k)!} \cdot p^k \cdot \frac{q^{n-k}}{q} = \frac{(n-k) \cdot p}{(k+1) \cdot (1-p)} \cdot b(k, n, p) \ ,$$

welche

$$\begin{aligned}
b(k+1, n, p) \geq b(k, n, p) \Longleftrightarrow (n-k) \cdot p &\geq (k+1) \cdot (1-p) \\
np - kp &\geq k - kp + 1 - p \\
k &\leq np + p - 1 \qquad (*)
\end{aligned}$$

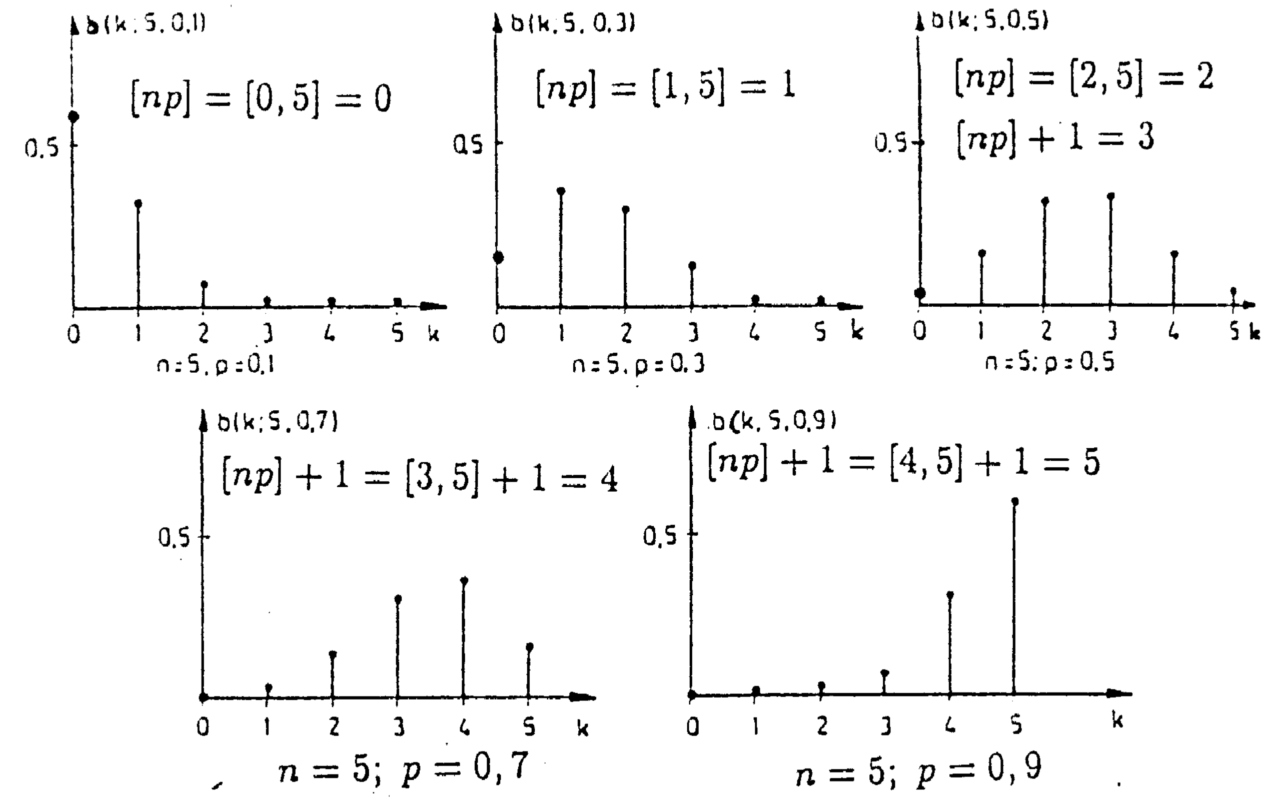

Abbildung 2: Verteilung der Wahrscheinlichkeiten verschiedener Binomialverteilungen

mit sich bringt. Sei nun $k^* \in \{0, \ldots, n\}$ das größte k, für das $(*)$ gilt. (Ein solches k^* mit $0 \leq k^* < n$ gibt es, da $np + p - 1 < n$ für $p < 1$ ist. Ist dagegen $np + p - 1 < 0$, so ist das *Maximum* bei $k = 0$ zu finden.)
Wir erhalten

$$b(k^* + 1, n, p) \geq b(k^*, n, p) \quad \text{und} \quad b(k^* + 2, n, p) < b(k^* + 1, n, p) .$$

Das *Maximum* liegt also bei $k^* + 1$, eventuell auch bei k^*, wegen

$$\left.\begin{array}{l} k^* + 1 \leq np + p \text{ und} \\ (k^* + 1) + 1 > np + p \end{array}\right\} \text{ sehen wir, daß } (k^* + 1) = [np] \text{ oder } [np] + 1 \text{ gilt.}$$

Im Falle von $k^* = np + p - 1$ ist das Maximum bei $np + p - 1$ und $np + p$ zu finden (siehe auch Abbildung 2).

Es gilt:

$$[np] \leq \underbrace{np}_{\text{„theoretischer Mittelwert"}} \leq [np] + 1 .$$

Betrachten wir nun m Versuchsserien (pro Versuchsserie werden n Zufallsexperimente durchgeführt). Die absolute Häufigkeit h des Eintretens von A ist dann

$$h = k_0 \cdot 0 + k_1 \cdot 1 + \ldots + k_n \cdot n ,$$

wenn k_i $(i = 0, \ldots, n)$ die Anzahl der Versuchsserien bezeichnet, bei denen A i-mal eingetreten ist: $\sum_{i=0}^{n} k_i = m$.

Es ist weiter

$$\overline{x} \doteq \frac{h}{m} = 0 \cdot \frac{k_0}{m} + 1 \cdot \frac{k_1}{m} + \ldots + n \cdot \frac{k_n}{m} \approx$$
$$\approx 0 \cdot P(X=0) + 1 \cdot P(X=1) + \ldots + n \cdot P(X=n) ,$$

wobei $\overline{x}$ also das *durchschnittliche Eintreten* von A pro Versuchsserie bedeutet. Der *zweite* Ausdruck stellt eine *theoretische Vorhersage* des *ersten* dar.
Wir *definieren* diese *theoretische Größe* für *diskrete ZV* X als *Erwartungswert von* X:

$$E(X)\,[= \mu(X)] = \sum_i x_i \cdot P(X = x_i) .$$

Speziell für die Binomialverteilung ergibt sich

$$E(X) = \sum_{i=0}^{n} i \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} .$$

- „Zu Fuß“ ausrechnen ergibt das Prophezeite:

$$E(X) = np \cdot \sum_{i=1}^{n} \binom{n-1}{i-1} \cdot p^{i-1} \cdot q^{n-i} = np \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \cdot p^j \cdot q^{n-1-j} =$$
$$= np \cdot [p + (1-p)]^{n-1} = np .$$

- Eine *zweite* Möglichkeit wäre: Sei die ZV

$$X_i \doteq \begin{cases} 1, & \text{wenn } A \text{ beim } i\text{-ten Durchgang eintritt, und} \\ 0 & \text{sonst } (i = 1, \ldots, n) . \end{cases}$$

Dann ist

$$P(X_i = 1) = p = 1 - P(X_i = 0) \quad \text{und} \quad X = \sum_{i=1}^{n} X_i ,$$

woraus

$$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$$

folgt und schließlich

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np .$$

Es muß also gelten:

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) .$$

BEWEIS: Für $n = 2$ ist $E(X+Y) =$

$$= \sum_j \sum_i (x_i + y_j) \cdot P(X = x_i, Y = y_j) =$$

$$= \sum_i x_i \cdot \sum_j P(X = x_i, Y = y_j) + \sum_j y_j \cdot \sum_i P(X = x_i, Y = y_j) =$$

$$= \sum_i x_i \cdot P(X = x_i) + \sum_j y_j \cdot P(Y = y_j) = E(X) + E(Y) \ .$$

Mit *vollständiger Induktion* folgt das Gewünschte. ■
Es ist weiters $E(aX) = a \cdot E(X) \quad \forall a \in \mathbb{R}$, was insgesamt zeigt, daß $E(X)$ *linear* ist.
BEWEIS:
$E(aX) = \sum_i ax_i \cdot P(aX = ax_i) = a \cdot \sum_i x_i \cdot P(X = x_i) = a \cdot E(X)$. ■
Natürlich ist $E(a) = a \ \forall a \in \mathbb{R}$ fest: $E(a) = a \cdot \underbrace{P(X = a)}_{=1} = a$. ■

- Eine *dritte* Möglichkeit wäre: Es ist

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k$$

wegen des *Binomischen Lehrsatzes;* differenzieren wir nun links und rechts:

$$n \cdot (1+x)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} \cdot k \cdot x^{k-1} = \sum_{k=1}^{n} \binom{n}{k} \cdot k \cdot x^{k-1}$$

und erhalten so

$$n \cdot x \cdot (1+x)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} \cdot k \cdot x^k \ .$$

Damit ist

$$E(X) = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot \underbrace{p^k \cdot (1-p)^{n-k}}_{= \underbrace{\left(\frac{p}{1-p}\right)^k}_{\doteq x} \cdot (1-p)^n} =$$

$$= n \cdot \frac{p}{1-p} \cdot \left(1 + \frac{p}{1-p}\right)^{n-1} \cdot (1-p)^n =$$

$$= np \cdot \left(\frac{1-p+p}{1-p}\right)^{n-1} \cdot (1-p)^{n-1} = np \ .$$

- Eine *vierte* Möglichkeit wäre mittels *erzeugender Funktionen.*
[Diese dienen zur *Charakterisierung* einer diskreten ZV (so wie die Verteilungsfunktion).]
X besitze die Verteilung $[i, P(X=i)]$, $i = 0, 1, 2, \ldots$
DEFINITION:
$$G_X(x) \doteq \sum_{i=0}^{\infty} x^i \cdot P(X=i) \ , \ x \in \mathbb{R}$$
heißt *erzeugende Funktion* G_X der ZV X.
Für $|x| \leq 1$ ist
$$|G_X(x)| \leq \sum_{i=0}^{\infty} P(X=i) = 1 \ ,$$
das heißt G_X ist für alle $|x| \leq 1$ jedenfalls erklärt. Es ist
$$G_X(0) = P(X=0) \quad \text{und weiters wegen}$$
$$G'_X(x) = \sum_{i=1}^{\infty} i \cdot x^{i-1} \cdot P(X=i) \quad \text{ist} \quad G'_X(0) = P(X=1) \ .$$
Nochmals differenzieren liefert
$$G''_X(x) = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot x^{i-2} \cdot P(X=i) \quad \text{und}$$
$$G''_X(0) = 2 \cdot (2-1) \cdot P(X=2) = 2 \cdot P(X=2) \ .$$
Allgemein ist
$$G_X^{(n)}(0) = n! \cdot P(X=n) \ , \quad \text{ja mehr noch:}$$
$$\underbrace{G'_X(1)}_{\text{linksseitige Ableitung}} = \sum_{i=1}^{\infty} i \cdot P(X=i) = E(X) \ .$$
Das *Problem* ist nun die konkrete Berechnung (also das Finden einer geschlossenen Darstellung) von G_X für eine spezielle ZV X, bei uns ist
$$\begin{aligned} G_X(x) &= \sum_{k=0}^{n} x^k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \\ &= \sum_{k=0}^{n} \binom{n}{k} \cdot (xp)^k \cdot (1-p)^{n-k} = (xp + 1 - p)^n = \\ &= [p \cdot (x-1) + 1]^n \ . \end{aligned}$$
Damit erhalten wir wieder
$$G'_X(x) = n \cdot [p \cdot (x-1) + 1]^{n-1} \cdot p$$
und
$$E(X) = G'_X(1) = np \ .$$

4. Frage: Wie stark „schwanken" die tatsächlichen Ergebnisse um den sie vorhersagenden theoretischen Wert $E(X)$?

Zur Beantwortung dieser Frage nehmen wir Anleihen aus der *beschreibenden* Statistik. [Wir wissen schon: $\overline{x} \approx E(X)$.]
Folgende *Streumaße* kennen wir dort:

- die *Spannweite* $r = x_{\max} - x_{\min}$,
- die *Halbweite* $R = q_3 - q_1$, wobei die q_i die i-ten Quartile darstellen,
- die *mittlere lineare Abweichung* $d_{\overline{x}} = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \overline{x}|$ und
- die *mittlere quadratische Abweichung* oder *empirische Varianz* $s_{\overline{x}}^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$.

Für die letzten beiden gilt:

$$d_{\mathcal{M}} \leq d_c \;\forall c \in \mathbb{R} \quad (\text{mit } \mathcal{M} \text{ Median}) \text{ und} \qquad s_{\overline{x}}^2 \leq s_c^2 \;\forall c \in \mathbb{R} \,.$$

In diesem Sinne paßt $\mathcal{M}$ ideal zu d und $\overline{x}$ zu s^2. Wir entscheiden uns daher für s^2, also für das arithmetische Mittel der quadratischen Abweichungen von $\overline{x}$!
Wir nehmen uns

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \approx E(X)$$

zum Vorbild und schließen analog

$$s_{\overline{x}}^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2 \approx E\left\{[X - E(X)]^2\right\} \,.$$

Diesen eben hingeschriebenen Erwartungswert definieren wir als die *Varianz einer ZV X*, sie ist eine theoretische Größe, ein Parameter. Ausgeschrieben lautet sie

$$\sum_i [x_i - E(X)]^2 \cdot P(X = x_i)$$

für eine diskrete ZV X und abgekürzt wird sie mit

$$Var(X) = D^2(X) = \sigma^2(X) \,.$$

1. *Es gilt:*

$$\begin{aligned} E\left\{[X - E(X)]^2\right\} &= E[X^2 - 2 \cdot X \cdot E(X) + E(X)^2] = \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + E(X)^2 = \\ &= E(X^2) - E(X)^2 = D^2(X) \,. \end{aligned}$$

Dies ist der *Verschiebungssatz,* welcher eine leichtere Berechnung von $D^2(X)$ erlaubt. (Wir werden ihn zur Berechnung der Varianz einer binomialverteilten ZV verwenden.)

2. *Folgerung:*
$$E(X^2) \geq E(X)^2 \text{ bzw. } \sqrt{E(X^2)} \geq E(X) .$$

3. *Abschätzung* für $D^2(X)$:
$$E[|X - E(X)|] \leq D(X) \doteq \sqrt{D^2(X)} ,$$
was eine Verbindung zu $d_{\overline{x}} = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \overline{x}|$ herstellt! (Siehe obigen Analogieschluß!)
Beweis: Sei $Y \doteq |X - E(X)|$ eine ZV. Wie immer ist $E(Y^2) \geq E(Y)^2$, was eingesetzt
$$E\left\{[X - E(X)]^2\right\} \geq E[|X - E(X)|]^2$$
zur Folge hat. Wurzelziehen links und rechts liefert die Behauptung. ■

4. $D^2(aX + b) = a^2 \cdot D^2(X)\ \forall a, b \in \mathbb{R}$, denn:
$$\begin{aligned} D^2(aX + b) &= E\left\{[aX + b - E(aX + b)]^2\right\} = \\ &= E\left\{[aX + b - a \cdot E(X) - b]^2\right\} = \\ &= a^2 \cdot E\left\{[X - E(X)]^2\right\} = a^2 \cdot D^2(X) . \end{aligned}$$

5. Die ZV
$$X^* \doteq \frac{X - E(X)}{D(X)} = \frac{X - \mu}{\sigma}$$
heißt „*Standardisierte ZV X^* zu X*" mit
$$E(X^*) \ = \ E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma} \cdot E(X - \mu) = \frac{1}{\sigma} \cdot [E(X) - \mu] = 0$$
und
$$D^2(X^*) \ = \ D^2\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \cdot D^2(X - \mu) = \frac{1}{\sigma^2} \cdot D^2(X) = 1 .$$

6. Für $D^2(X + Y)$ berechnen wir erst $[X + Y - E(X + Y)]^2 \overset{\text{TRICK!}}{\doteq}$
$$\begin{aligned} &= [X + Y - E(X) - E(Y)]^2 \overset{\text{TRICK!}}{\doteq} \\ &= \{[X - E(X)] + [Y - E(Y)]\}^2 = \\ &= [X - E(X)]^2 + [Y - E(Y)]^2 + 2 \cdot [X - E(X)] \cdot [Y - E(Y)] = \\ &= [X - E(X)]^2 + [Y - E(Y)]^2 + \\ &+ 2 \cdot [X \cdot Y - E(X) \cdot Y - E(Y) \cdot X + E(X) \cdot E(Y)] . \end{aligned}$$

Mittels Erwartungswertbildung folgt:

$$\begin{aligned} D^2(X+Y) &= D^2(X) + D^2(Y) + \\ &+ 2\cdot[E(X\cdot Y) - E(X)\cdot E(Y) - \\ &- E(Y)\cdot E(X) + E(X)\cdot E(Y)] = \\ &= D^2(X) + D^2(Y) + 2\cdot[E(X\cdot Y) - E(X)\cdot E(Y)] . \end{aligned}$$

Wenn X und Y *unabhängig* voneinander sind, folgt:

$$D^2(X) + D^2(Y) = D^2(X+Y) .$$

Wir sehen: Die Varianz ist dann *additiv!* Warum? — Nun:

$$\begin{aligned} E(X\cdot Y) &= \sum_i \sum_j x_i\, y_j \cdot P(X = x_i, Y = y_j) \overset{\text{Unabh.}}{\overset{!}{=}} \\ &= \sum_i \sum_j x_i\, y_j \cdot P(X = x_i)\cdot P(Y = y_j) = \\ &= \sum_i x_i \cdot P(X = x_i) \cdot \sum_j y_j \cdot P(Y = y_j) = \\ &= E(X)\cdot E(Y) . \end{aligned}$$

Mittels vollständiger Induktion folgt für *paarweise stochastisch unabhängige* ZV $X_1, \ldots, X_n$

$$E(X_1 \cdot \ldots \cdot X_n) = E(X_1)\cdot \ldots \cdot E(X_n) .$$

7. *Tschebyscheffsche Ungleichung:* Sei X eine beliebige (diskrete) ZV mit $E(X) = \mu \in \mathbb{R}$ und $D^2(X) = \sigma^2 \in \mathbb{R}_0^+$. Dann gilt:

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2} \qquad \forall a \in \mathbb{R}^+ .$$

BEWEIS: Es ist

$\sigma^2 \geq \sum_{|x_k-\mu|\geq a} (x_k-\mu)^2 \cdot P(X = x_k) \geq \sum_{|x_k-\mu|\geq a} a^2 \cdot P(X = x_k)$, was

$\frac{\sigma^2}{a^2} \geq \sum_{|x_k-\mu|\geq a} P(X = x_k) = P(|X-\mu| \geq a)$ zur Folge hat. ∎

8. Oft findet man in Büchern:

$$s^2 = \underbrace{\frac{1}{n-1}}_{?!} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2 .$$

[Dabei stellt $(x_1, \ldots, x_n)$ eine *einfache* Stichprobe dar, das heißt alle zu den x_i gehörenden ZV X_i sind paarweise voneinander unabhängig und haben dieselbe Verteilungsfunktion.]
Warum? — Dazu fassen wir s^2 als *Realisation* der *ZV*

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2$$

auf! Wir werden zeigen:

$$E\left(S^2\right) = \sigma^2 ,$$

man sagt: „S^2 ist ein *erwartungstreuer Schätzer* für σ^2."

- $\overline{X}$ ist ein erwartungstreuer Schätzer für μ:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} \cdot E\left(\sum_{i=1}^{n} X_i\right) =$$
$$= \frac{1}{n} \cdot \sum_{i=1}^{n} \underbrace{E(X_i)}_{=\mu \ \forall i=1,\ldots,n} = \frac{1}{n} \cdot n \cdot \mu = \mu .$$

- Zu zeigen ist also $E\left(S^2\right) = \sigma^2$: Wir beginnen dazu mit

$$\begin{aligned}
(n-1) \cdot S^2 &= \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i{}^2 - 2 \cdot \sum_{i=1}^{n} X_i \overline{X} + n\overline{X}^2 = \\
&= \sum_{i=1}^{n} X_i{}^2 - 2n\overline{X} \cdot \overline{X} + n\overline{X}^2 = \sum_{i=1}^{n} X_i{}^2 - n\overline{X}^2 = \\
&= \sum_{i=1}^{n} X_i{}^2 - n \cdot \frac{1}{n} \sum_{i=1}^{n} X_i \cdot \frac{1}{n} \sum_{j=1}^{n} X_j = \\
&= \sum_{i=1}^{n} X_i{}^2 - \frac{1}{n} \cdot \sum_{i=1}^{n} X_i{}^2 - \frac{1}{n} \cdot \sum_{i \neq j} X_i \cdot X_j = \\
&= \frac{n-1}{n} \cdot \sum_{i=1}^{n} X_i{}^2 - \frac{1}{n} \cdot \sum_{i \neq j} X_i \cdot X_j ,
\end{aligned}$$

Erwartungswertbildung zeigt uns schließlich

$$\begin{aligned}
E\left[(n-1) \cdot S^2\right] &= \frac{n-1}{n} \cdot \sum_{i=1}^{n} E\left(X_i{}^2\right) - \frac{1}{n} \cdot \sum_{i \neq j} E(X_i) \cdot E(X_j) = \\
&= \frac{n-1}{n} \cdot \sum_{i=1}^{n} E\left(X_i{}^2\right) - \frac{1}{n} \cdot n \cdot (n-1) \cdot \mu^2 \text{ und}
\end{aligned}$$

$$E\left(S^2\right) = \frac{1}{n}\cdot\sum_{i=1}^{n} E\left({X_i}^2\right) - \mu^2 =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n}\left[\underbrace{E\left({X_i}^2\right) - \mu^2}_{=D^2(X_i)=\sigma^2\ \forall i=1,\dots,n}\right] = \frac{1}{n}\cdot n\cdot\sigma^2 = \sigma^2 .$$

- Hingegen liefert ${S^*}^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

$$E\left({S^*}^2\right) = \frac{n-1}{n^2}\cdot\sum_{i=1}^{n} E\left({X_i}^2\right) - \frac{n-1}{n}\cdot\mu^2 =$$

$$= \frac{n-1}{n}\cdot\left[\frac{1}{n}\cdot\sum_{i=1}^{n} E\left({X_i}^2\right) - \mu^2\right] = \frac{n-1}{n}\cdot\sigma^2 ,$$

wir sagen: „${S^*}^2$ ist nur *asymptotisch erwartungstreu!*“
(Das heißt $\lim_{n\to\infty} E\left({S^*}^2\right) = \sigma^2$.)

- Kennt man μ und muß es daher nicht durch $\overline{X}$ schätzen, so ist

$$S_\mu^{*2} = \frac{1}{n}\cdot\sum_{i=1}^{n}(X_i-\mu)^2$$

erwartungstreu:

$$E\left(S_\mu^{*2}\right) = \frac{1}{n}\cdot E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right] =$$

$$= \frac{1}{n}\cdot E\left(\sum_{i=1}^{n} {X_i}^2 - 2\cdot X_i\cdot\mu + \mu^2\right) =$$

$$= \frac{1}{n}\cdot E\left(\sum_{i=1}^{n} {X_i}^2\right) - \frac{1}{n}\cdot 2\cdot\mu\cdot E\left(\sum_{i=1}^{n} X_i\right) + \frac{1}{n}\cdot n\cdot\mu^2 =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} E\left({X_i}^2\right) - \frac{1}{n}\cdot 2\cdot\mu\cdot n\cdot\mu + \mu^2 =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} E\left({X_i}^2\right) - \mu^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}\left[\underbrace{E\left({X_i}^2\right) - \mu^2}_{=\sigma^2\ \forall i=1,\dots,n}\right] =$$

$$= \frac{1}{n}\cdot n\cdot\sigma^2 = \sigma^2 .$$

- $S_\mu{}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \mu)^2$ hingegen liefert:

$$E\left(S_\mu{}^2\right) = \begin{array}{c}\text{wie}\\ \text{eben}\end{array} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left[\underbrace{E\left(X_i{}^2\right) - \mu^2}_{=\sigma^2\ \forall i=1,\ldots,n}\right] =$$

$$= \frac{1}{n-1} \cdot n \cdot \sigma^2 = \frac{n}{n-1} \cdot \sigma^2 .$$

$S_\mu{}^2$ ist in diesem Fall nur *asymptotisch erwartungstreu.*

Nun wollen wir konkret die *Varianz* einer *binomialverteilten ZV* unter Verwendung des *Verschiebungssatzes* berechnen:

$$D^2(X) = E\left(X^2\right) - E(X)^2 .$$

Wir wissen schon: $E(X)^2 = (np)^2 = n^2p^2$. Wir wenden uns also der *Frage* bzw. dem *Problem* der Berechnung von $E\left(X^2\right)$ zu.

- *„Zu Fuß“:*

$$E\left(X^2\right) = \sum_{i=0}^{n} i^2 \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} =$$

$$= \sum_{i=1}^{n} i \cdot (i-1) \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} + \sum_{i=1}^{n} i \cdot \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} .$$

Den zweiten Ausdruck kennen wir schon: np. — Es bleibt:

$$\sum_{i=2}^{n} \frac{n!}{(i-2)! \cdot (n-i)!} \cdot p^i \cdot (1-p)^{n-i} =$$

$$= n \cdot (n-1) \cdot p^2 \cdot \sum_{i=2}^{n} \binom{n-2}{i-2} \cdot p^{i-2} \cdot (1-p)^{n-i} =$$

$$= n \cdot (n-1) \cdot p^2 \cdot \sum_{j=0}^{n-2} \binom{n-2}{j} \cdot p^j \cdot (1-p)^{n-2-j} =$$

$$= n \cdot (n-1) \cdot p^2 \cdot [p + (1-p)]^{n-2} = n \cdot (n-1) \cdot p^2 .$$

Insgesamt erhalten wir

$$E\left(X^2\right) = n \cdot (n-1) \cdot p^2 + np .$$

Daraus folgt

$$E\left(X^2\right) - E(X)^2 = n^2p^2 - np^2 + np - n^2p^2 = np \cdot (1-p) = npq = D^2(X) .$$

- Die *zweite* Möglichkeit ist (siehe Frage 3 — Berechnung des Erwartungswertes einer binomialverteilten ZV):
  Sei X_i wie oben (ebendort) definiert. Wir wissen bereits:
  $E(X_i) = p \; \forall i = 1, \ldots, n$. — Weiters ist:

$$E\left({X_i}^2\right) = 1^2 \cdot P(X_i = 1) + 0^2 \cdot P(X_i = 0) = p \quad \text{und damit}$$
$$D^2(X_i) = E\left({X_i}^2\right) - E(X_i)^2 = p - p^2 = p \cdot (1-p) = pq \; .$$

Weil $X = \sum_{i=1}^n X_i$ ist und die X_i paarweise stochastisch unabhängig voneinander sind, gilt

$$D^2(X) = D^2\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n D^2(X_i) = \sum_{i=1}^n pq = npq \; .$$

- Die *dritte* Möglichkeit entspricht der dritten bei der Berechnung von $E(X)$: Wir hatten dort

$$n \cdot (1+x)^{n-1} = \sum_{k=1}^n \binom{n}{k} \cdot k \cdot x^{k-1} \; .$$

Nochmals differenzieren liefert:

$$n \cdot (n-1) \cdot (1+x)^{n-2} = \sum_{k=1}^n \binom{n}{k} \cdot k \cdot (k-1) \cdot x^{k-2} \quad \text{und weiters}$$

$$n \cdot (n-1) \cdot (1+x)^{n-2} \cdot x^2 \doteq \sum_{k=2}^n \binom{n}{k} \cdot k \cdot (k-1) \cdot x^k =$$
$$= \sum_{k=1}^n \binom{n}{k} \cdot k^2 \cdot x^k - \underbrace{\sum_{k=1}^n \binom{n}{k} \cdot k \cdot x^k}_{= n \cdot x \cdot (1+x)^{n-1} \quad \text{(siehe oben)}} \; .$$

Daraus folgt

$$\sum_{k=1}^n \binom{n}{k} \cdot k^2 \cdot x^k = n \cdot (n-1) \cdot (1+x)^{n-2} \cdot x^2 + n \cdot x \cdot (1+x)^{n-1} =$$
$$= n \cdot x \cdot (1+x)^{n-2} \cdot [(n-1) \cdot x + (1+x)] =$$
$$= n \cdot x \cdot (1+x)^{n-2} \cdot (n \cdot x - x + 1 + x) =$$
$$= n \cdot x \cdot (1+x)^{n-2} \cdot (n \cdot x + 1) \; .$$

Wir erhalten insgesamt

$$n \cdot x \cdot (1+x)^{n-2} \cdot (n \cdot x + 1) = \sum_{k=1}^n \binom{n}{k} \cdot k^2 \cdot x^k \; ,$$

woraus wir für die Varianz auf das altbekannte Ergebnis schließen:

$$\begin{aligned}
D^2(X) &\doteq \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} - (np)^2 = \\
&= \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} \cdot \underbrace{\left(\frac{p}{1-p}\right)}_{\doteq\, x}{}^{k} \cdot (1-p)^n - n^2p^2 = \\
&= (1-p)^n \cdot \sum_{k=0}^{n} \binom{n}{k} \cdot k^2 \cdot x^k - n^2p^2 = \\
&= (1-p)^n \cdot \frac{np}{1-p} \cdot \left(1 + \frac{p}{1-p}\right)^{n-2} \cdot \left(\frac{np}{1-p} + 1\right) - n^2p^2 = \\
&= (1-p)^n \cdot n \cdot \frac{p}{1-p} \cdot \frac{1^{n-2}}{(1-p)^{n-2}} \cdot \frac{np+1-p}{1-p} - n^2p^2 = \\
&= np \cdot (np - p + 1) - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 = \\
&= np \cdot (1-p) = npq \ .
\end{aligned}$$

- Die *vierte* Möglichkeit verwendet wieder als wesentliches Hilfsmittel *erzeugende Funktionen.* Wir hatten schon

$$P(X = n) \quad = \quad \frac{G_X{}^{(n)}(0)}{n!}$$

und

$$G'_X(1) \quad = \quad E(X) \ .$$

Ebenfalls sahen wir dort, daß

$$\begin{aligned}
G''_X(x) &= \sum_{i=2}^{\infty} i \cdot (i-1) \cdot x^{i-2} \cdot P(X = i) = \\
&= \sum_{i=2}^{\infty} i^2 \cdot x^{i-2} \cdot P(X = i) - \sum_{i=2}^{\infty} i \cdot x^{i-2} \cdot P(X = i)
\end{aligned}$$

ist.
Daraus erkennen wir

$$G''_X(1) = \sum_{i=1}^{\infty} i^2 \cdot P(X = i) - \sum_{i=1}^{\infty} i \cdot P(X = i) = E\left(X^2\right) - E(X) \ ,$$

und damit folgt:

$$D^2(X) = E\left(X^2\right) - E(X)^2 = G''_X(1) + G'_X(1) - \left[G'_X(1)\right]^2 \ .$$

Also:

$$\begin{aligned} G_X(x) &= [p \cdot (x-1)+1]^n \\ G'_X(x) &= n \cdot p \cdot [p \cdot (x-1)+1]^{n-1} \\ G''_X(x) &= n \cdot (n-1) \cdot p^2 \cdot [p \cdot (x-1)+1]^{n-2} . \end{aligned}$$

Einsetzen von $x = 1$ führt zu

$$\begin{aligned} G''_X(1) &= n \cdot (n-1) \cdot p^2 \text{ , was letztlich} \\ D^2(X) &= n \cdot (n-1) \cdot p^2 + np - n^2p^2 = np - np^2 = npq \end{aligned}$$

mit sich bringt.

Beispiel 5: Eine diskrete ZV X *habe den Wertebereich* $\{0, \ldots, 10\}$, *den Erwartungswert* $E(X) = 5$ *und die Varianz* $D^2(X) = 2{,}5$.
Schätzen Sie $P(|X - E(X)| \geq 3)$ *ab! Wie ändert sich das Ergebnis, wenn wir zusätzlich annehmen, daß* X *binomialverteilt (Parameter?) ist?*

Lösung: Die Tschebyscheff-Ungleichung führt uns zu

$$P(|X-5| \geq 3) \leq \frac{2{,}5}{9} = 0{,}2\dot{7} .$$

Unter der oben genannten zusätzlichen Annahme erhalten wir (mit den Parametern $n = 10$ und $p = \frac{1}{2}$)

$$\begin{aligned} & P(|X-5| \geq 3) = P(X \geq 8) + P(X \leq 2) = \\ &= \sum_{k=8}^{10} \binom{10}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{10-k} + \sum_{k=0}^{2} \binom{10}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{10-k} = \\ &= \left(\frac{1}{2}\right)^{10} \cdot \left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] = \\ &= \left(\frac{1}{2}\right)^{10} \cdot (2 + 20 + 90) = \frac{1}{2^{10}} \cdot 112 \approx 0{,}11 . \end{aligned}$$

Wir sehen: Die *zusätzliche Information* über die *Verteilung* von X verbessert das Ergebnis beträchtlich (hier: Faktor 0,4), das heißt Abweichungen von $E(X)$ können wesentlich kleinere Wahrscheinlichkeiten zugeordnet werden. Die *Tschebyscheff-Ungleichung* geht vom *schlechtesten* Fall aus, daher genügen auch die schwachen Voraussetzungen ebendort. ◇

*Beispiel 6: Bei einem „*Multiple-Choice-Test*" sind 18 Fragen zu beantworten, wobei zu jeder Frage drei Antwortmöglichkeiten angeboten werden, wovon nur eine richtig ist. Wie viele richtige Antworten werden erwartungsgemäß erraten, und wie stark schwankt diese Zahl?*

Lösung: Sei X die ZV, welche die Anzahl der richtig geratenen Antworten beschreibt. Dann ist X binomialverteilt mit den Parametern $p = \frac{1}{3}$ und $n = 18$. Wir erhalten daraus

$$E(X) = np = 18 \cdot \frac{1}{3} = 6 \quad \text{und} \quad D^2(X) = npq = 6 \cdot \frac{2}{3} = 4 \, ,$$

woraus $D(X) = 2$ folgt.
Wenn wir uns

$$E[|X - E(X)|] \leq D(X)$$

vor Augen halten, ist $x > 8$ oder $x < 4$, das heißt $x \in \{0, \ldots, 3, 9, \ldots, 18\}$, nicht zu erwarten! (Dabei ist x die Realisation von X.) Was heißt das?

$$\text{Nun:} \quad P(X < 4) + P(X > 8) = 1 - P(4 \leq X \leq 8) =$$
$$= 1 - \sum_{k=4}^{8} \binom{18}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{18-k} \approx 0{,}2 \, .$$

Also wird in rund 20% aller Fälle $x < 4$ oder $x > 8$ sein, in rund 80% der Fälle ist $x \in [E(X) - D(X), E(X) + D(X)] = [4, 8]$. ◇

5. Frage: Haben die „*Tschebyscheff-Ungleichung*" und das „*Bernoullische Gesetz der großen Zahlen*" etwas miteinander zu tun?

Wir erinnern uns an

$$\lim_{n \to \infty} P(|R_n - p| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0 \text{ (Bernoulli)}$$

und an

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2} \quad \forall a \in \mathbb{R}^+ \text{ (Tschebyscheff)}.$$

Es ist

$$R_n = \frac{X}{n} \, .$$

Damit folgt

$$E(R_n) = E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n} \cdot np = p \quad \text{und}$$
$$D^2(R_n) = D^2\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot D^2(X) = \frac{1}{n^2} \cdot npq = \frac{pq}{n} \, .$$

Tschebyscheff ergibt nun speziell

$$P(|R_n - p| \geq a) \leq \underbrace{\frac{pq}{a^2 \cdot n}}_{\to 0 \text{ für } n \to \infty} ,$$

woraus wir

$$\lim_{n\to\infty} P(|R_n - p| \geq a) = 0 \quad \text{und}$$
$$\lim_{n\to\infty} P(|R_n - p| < a) = 1 \quad \text{schließen.}$$

Die letzte Zeile gilt erst recht für „$\leq$", wenn wir nun „a" ($\in \mathbb{R}^+$) durch „ε" (> 0) ersetzen, steht *Bernoulli* vor uns!

6. Frage

Die eingangs vorgestellte Situation, die uns auf die Binomialverteilung brachte, kann mit dem Ziehen aus einer Urne *mit* Zurücklegen verglichen werden: Es seien N Kugeln in der Urne, davon M weiße, so daß $\frac{M}{N} = p$ ist, und $N - M$ schwarze. Das Ziehen einer weißen Kugel entspricht dann dem Eintreten von A. Es werde n-mal gezogen. X beschreibe die Anzahl der gezogenen weißen Kugeln. Dann ist X binomialverteilt mit den Parametern n und $p = \frac{M}{N}$. — Nun die *Frage:* Wie stellt sich die Situation beim Ziehen *ohne* Zurücklegen dar?

Gesucht ist also $P(Y = k)$ mit $0 \leq k \leq \min\{n, M\}$. ($Y$ zähle — in dieser neuen Situation — die gezogenen weißen Kugeln.) Es ist

$$P(Y = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \qquad 0 \leq k \leq \min\{n, M\} .$$

Man sagt: „Y ist *‚hypergeometrisch'* verteilt."

Bemerkung: Die hypergeometrische Verteilung ist — in mathematischer Hinsicht — schwierig zu behandeln:

$$\left.\begin{array}{l} E(Y) = np \\ D^2(Y) = npq \cdot \frac{N-n}{N-1} \end{array}\right\} \text{ mit } p = \frac{M}{N} \ : \ \textit{ohne} \text{ Beweis.}$$
$$G_Y(x) = ?$$

Die hypergeometrische Verteilung an sich ist nicht unser Thema, jedoch ihre *Beziehung* zur *Binomialverteilung:*

	P(weiße Kugel gezogen — mit Zurücklegen)	P(weiße Kugel gezogen — ohne Zurücklegen)
1. Zug	$\frac{M}{N}$	$\frac{M}{N}$
2. Zug	$\frac{M}{N}$	$\frac{M}{N-1}$ oder $\frac{M-1}{N-1}$
3. Zug	$\frac{M}{N}$	$\frac{M}{N-2}$ oder $\frac{M-1}{N-2}$ oder $\frac{M-2}{N-2}$

Idee: Für großes N stimmen die Wahrscheinlichkeiten in der linken Spalte mit jenen in der rechten (fast) überein, wenn n nicht zu groß wird.

Tatsächlich gilt:

$$\lim_{N\to\infty} P(Y=k) = \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} \quad \text{mit } p=\frac{M}{N} \quad (n \text{ fest}) .$$

BEWEIS:

$$\lim_{N\to\infty}\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}} = \lim_{N\to\infty}\frac{\frac{M\cdot(M-1)\cdot\ldots\cdot(M-k+1)}{1\cdot 2\cdot\ldots\cdot k}\cdot\frac{(N-M)\cdot\ldots\cdot(N-M-n+k+1)}{1\cdot\ldots\cdot(n-k)}}{\frac{N\cdot(N-1)\cdot\ldots\cdot(N-n+1)}{1\cdot 2\cdot\ldots\cdot n}} =$$

$$= \lim_{N\to\infty}\binom{n}{k}\cdot$$

$$\cdot\frac{\overbrace{M\cdot(M-1)\cdot\ldots\cdot(M-k+1)}^{k\text{ Faktoren}}\cdot\overbrace{(N-M)\cdot\ldots\cdot(N-M-n+k+1)}^{(n-k)\text{ Faktoren}}}{\underbrace{N\cdot(N-1)\cdot\ldots\cdot(N-n+1)}_{n\text{ Faktoren}}} =$$

$$= \lim_{N\to\infty}\binom{n}{k}\cdot$$

$$\cdot\frac{\frac{M}{N}\cdot\left(\frac{M}{N}-\frac{1}{N}\right)\cdot\ldots\cdot\left(\frac{M}{N}-\frac{k-1}{N}\right)\cdot\left(1-\frac{M}{N}\right)\cdot\ldots\cdot\left(1-\frac{M}{N}-\frac{n-k-1}{N}\right)}{1\cdot\left(1-\frac{1}{N}\right)\cdot\ldots\cdot\left(1-\frac{n}{N}+\frac{1}{N}\right)} =$$

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

Für *große* N gilt also mit $\frac{M}{N}=p$ die *Näherungsformel*

$$\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}} \approx \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$$

(*Faustregel:* $N\geq 60$ und $N>10n$).

Bemerkungen:

1. Dies ist in der Wahrscheinlichkeitsrechnung allgemein so: Jeder *Grenzwertsatz* bringt eine *Approximationsaussage* mit sich, die von praktischer Bedeutung ist.

2. Auch die Parameter gehen ineinander über:

$$E(Y) = np = E(X) \text{ und}$$

$$D^2(Y) = npq \cdot \frac{N-n}{N-1} = npq \cdot \frac{1-\frac{n}{N}}{1-\frac{1}{N}} \underbrace{\longrightarrow}_{N\to\infty} npq = D^2(X) .$$

Beispiel 7: Eine Lieferung von 100 Dioden *enthalte genau vier fehlerhafte. Aus der Lieferung werden zufällig fünf Dioden entnommen. Wie groß ist die Wahrscheinlichkeit, 0, 1, 2, 3, 4 oder 5 fehlerhafte zu entdecken?*

Lösung:

a) Sei Y jene ZV, die die Anzahl der fehlerhaften Stücke beschreibt. Dann ist Y hypergeometrisch verteilt! — Wir erhalten:

$$P(Y=0) = \frac{\binom{4}{0}\cdot\binom{96}{5}}{\binom{100}{5}} = 0,811875116 ,$$

$$P(Y=1) = \frac{\binom{4}{1}\cdot\binom{96}{4}}{\binom{100}{5}} = 0,17649459 ,$$

$$P(Y=2) = \frac{\binom{4}{2}\cdot\binom{96}{3}}{\binom{100}{5}} = 0,011386748 ,$$

$$P(Y=3) = \frac{\binom{4}{3}\cdot\binom{96}{2}}{\binom{100}{5}} = 0,000242271 ,$$

$$P(Y=4) = \frac{\binom{4}{4}\cdot\binom{96}{1}}{\binom{100}{5}} = 1,275112\cdot 10^{-6} \text{ und}$$

$$P(Y=5) = 0 .$$

b) Wegen $N = 100 \geq 60$ und $N = 100 > 10n = 50$ können wir die *Approximation durch die Binomialverteilung* wagen! Die Parameter ergeben sich zu $n = 5$ und $p = \frac{4}{100} = \frac{1}{25}$. — Wir berechnen für $k = 0,\ldots,4$:

$$P(Y=0) \approx \binom{5}{0}\cdot\left(\frac{1}{25}\right)^0\cdot\left(\frac{24}{25}\right)^5 = 0,815372698 ,$$

$$P(Y=1) \approx \binom{5}{1}\cdot\left(\frac{1}{25}\right)^1\cdot\left(\frac{24}{25}\right)^4 = 0,169869312 ,$$

$$P(Y=2) \approx \binom{5}{2}\cdot\left(\frac{1}{25}\right)^2\cdot\left(\frac{24}{25}\right)^3 = 0,014155776 ,$$

$$P(Y=3) \approx \binom{5}{3}\cdot\left(\frac{1}{25}\right)^3\cdot\left(\frac{24}{25}\right)^2 = 0,000589824 \text{ und}$$

$$P(Y=4) \approx \binom{5}{4}\cdot\left(\frac{1}{25}\right)^4\cdot\left(\frac{24}{25}\right)^1 = 1,2288\cdot 10^{-5} .$$

Wir bemerken, daß für kleine k die Ergebnisse aus a) und b) gut übereinstimmen ($k = 0, 1, 2$), für große k hingegen nicht mehr ($k = 3, 4$).
Der *Grund* hiefür ist folgender: Die Wahrscheinlichkeit, eine schlechte Diode zu erwischen, wenn man weiß, daß schon zwei schlechte gezogen worden sind, ist $\frac{2}{98}$, $\frac{2}{97}$ oder $\frac{2}{96}$, in jedem Fall ungefähr 0,02. Vergleiche dagegen $p = \frac{1}{25} = 0,04$! ◇

7. Frage: Wie ist die Summe zweier binomialverteilter ZV X_1 und X_2 verteilt?

Die Frage ist nur dann sinnvoll zu beantworten, wenn wir die beiden „zweiten" Parameter p_1 und p_2 gleichsetzen: $p_1 = p_2 = p$.

SATZ: Sei X_1 binomialverteilt mit den Parametern n_1 und p, X_2 sei binomialverteilt mit den Parametern n_2 und p, und seien X_1 und X_2 stochastisch unabhängig voneinander, dann ist $X = X_1 + X_2$ wieder binomialverteilt mit den Parametern $n_1 + n_2$ und p.

BEWEIS:

- Die mühsame Variante konstatiert zuerst:
  X hat den Wertebereich $\{0, \ldots, n_1 + n_2\}$. Schließlich berechnen wir

$$
\begin{aligned}
P(X = k) &= \sum_{\substack{i+j=k \\ 0 \leq i(j) \leq n_1(n_2)}} P(X_1 = i, X_2 = j) = \\
&= \sum_{i+j=k} P(X_1 = i) \cdot P(X_2 = j) = \\
&= \sum_{i=0}^{k} \binom{n_1}{i} \cdot p^i \cdot (1-p)^{n_1 - i} \cdot \binom{n_2}{k-i} \cdot p^{k-i} \cdot (1-p)^{n_2 - k + i} = \\
&= p^k \cdot (1-p)^{n_1 + n_2 - k} \cdot \underbrace{\sum_{i=0}^{k} \binom{n_1}{i} \cdot \binom{n_2}{k-i}}_{= \binom{n_1 + n_2}{k}}
\end{aligned}
$$

nach dem folgenden

LEMMA:

$$\sum_{i=0}^{k} \binom{n_1}{i} \cdot \binom{n_2}{k-i} = \binom{n_1 + n_2}{k} .$$

BEWEIS:

$$(a+b)^{n_1+n_2} = \sum_{k=0}^{n_1+n_2} \binom{n_1+n_2}{k} \cdot a^k \cdot b^{n_1+n_2-k} =$$

$$= (a+b)^{n_1} \cdot (a+b)^{n_2} = \sum_{i=0}^{n_1} \binom{n_1}{i} \cdot a^i \cdot b^{n_1-i} \cdot \sum_{j=0}^{n_2} \binom{n_2}{j} \cdot a^j \cdot b^{n_2-j} =$$

$$= \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \binom{n_1}{i} \cdot \binom{n_2}{j} \cdot a^{i+j} \cdot b^{n_1+n_2-(i+j)} .$$

Koeffizientenvergleich: Sei $i+j=k$. Dann folgt die Behauptung. ∎

Bemerkung: Wir haben hier zwei Funktionen *„gefaltet"*.

- Die leichte Variante (aber viel Theorie!) arbeitet mit Hilfe von erzeugenden Funktionen: Es gilt (*ohne* Beweis) unter gewissen Voraussetzungen:

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t) .$$

Damit bekommen wir wegen

$$G_{X_1}(t) = [p \cdot (t-1)+1]^{n_1} \text{ und } G_{X_2}(t) = [p \cdot (t-1)+1]^{n_2}$$
$$G_{X_1+X_2}(t) = G_{X_1}(t) \cdot G_{X_2}(t) = [p \cdot (t-1)+1]^{n_1+n_2} .$$

Dies ist die erzeugende Funktion einer binomialverteilten ZV mit den Parametern n_1+n_2 und p. Da

$$X = Y \iff G_X = G_Y$$

gilt, ist alles gezeigt. ∎

Beispiel 8 (Fortsetzung von Beispiel 6): Nicht einer, sondern vier *solche Tests sollen bearbeitet werden. Dieselben Fragen und Voraussetzungen wie bei Beispiel 6!*

Lösung: Die ZV Y zähle die bei den vier Tests richtig erratenen Antworten. Es ist

$$Y = X_1 + X_2 + X_3 + X_4 ,$$

wobei die ZV X_i $(i=1,\ldots,4)$ die richtig erratenen Antworten beim i-ten Test zählen. Die X_i sind daher alle binomialverteilt mit den Parametern $n=18$ und $p=\frac{1}{3}$.

Nach dem soeben bewiesenen Satz über die Verteilung der Summe binomialverteilter ZV ist Y ebenfalls binomialverteilt mit den Parametern $n=4\cdot 18$ und $p=\frac{1}{3}$. Damit ist

$$E(Y) = 4 \cdot 18 \cdot \frac{1}{3} = 24 \text{ und } D^2(Y) = 24 \cdot \frac{2}{3} = 16 ,$$

woraus $D(Y)=4$ folgt.

Man könnte natürlich auch so argumentieren: Die Gesamtanzahl der zu beantwortenden Fragen beträgt $n = 4 \cdot 18 = 72$. Die Wahrscheinlichkeit, bei einer Frage die richtige Antwort zu erraten, ist $p = \frac{1}{3}$. Y (wie eben definiert) ist daher binomialverteilt mit den Parametern $n = 72$ und $p = \frac{1}{3}$. Dies liefert uns

$$E(Y) = np = 72 \cdot \frac{1}{3} = 24 \text{ und } D^2(Y) = npq = 24 \cdot \frac{2}{3} = 16 \ ,$$

was wieder $D(Y) = 4$ zur Folge hat.

Im Lichte dieses Beispiels erscheint uns nun der Satz über die Verteilung der Summe zweier binomialverteilter ZV ganz klar! ◊

Beispiel 9 (Fortsetzung von Beispiel 8): Ein zweiter Test kommt zu den vieren (jetzt als ein Gesamttest aufgefaßt) hinzu: 36 Fragen mit je zwei Antwortmöglichkeiten. Dieselben Fragen und Voraussetzungen wie bei Beispiel 6 für

a) den zweiten Test alleine und

b) für ersten (bestehend aus vier Teilen) und zweiten Test!

Lösung:

a) Die ZV Z zähle die richtig erratenen Antworten, folglich ist Z binomialverteilt mit den Parametern $n = 36$ und $p = \frac{1}{2}$. Wir erhalten so

$$E(Z) = 36 \cdot \frac{1}{2} = 18 \text{ und } D^2(Z) = 18 \cdot \frac{1}{2} = 9 \ ,$$

was $D(Z) = 3$ impliziert.

b) Nun stellt sich die Frage: Wie ist $Y + Z$ verteilt? Da für Y $p = \frac{1}{3}$ ist und für Z $p = \frac{1}{2}$, ist der eben zitierte Satz über die Verteilung der Summe zweier binomialverteilter ZV *nicht* anwendbar!
Aber: gesucht ist ja nur $E(Y + Z)$ und $D^2(Y + Z)$! Die können wir berechnen:

$$\begin{aligned} E(Y+Z) &= E(Y) + E(Z) = 24 + 18 = 42 \quad \text{und} \\ D^2(Y+Z) &= D^2(Y) + D^2(Z) = 16 + 9 = 25 \ , \end{aligned}$$

somit ist $D(Y + Z) = 5$.
Voraussetzung dabei ist nur, daß die ZV Y und Z (so wie auch schon X_1, X_2, X_3 und X_4 bei Beispiel 8) (paarweise) stochastisch *unabhängig* sind. Letztlich müssen die *einzelnen* Beantwortungen voneinander unabhängig sein, ansonsten ist ja von Anfang an keine Binomialverteilung gegeben. ◊

8. Frage

Wir haben die Binomialverteilung u.a. als bestimmten Grenzwert der hypergeometrischen Verteilung kennengelernt. Strebt auch die Binomialverteilung selbst — unter gewissen Bedingungen — gegen eine andere — diskrete — Verteilung?

Wir betrachten folgenden *Spezialfall: großes* n und *kleines* p so, daß

$$\lim_{n\to\infty} n\cdot p = \lambda \in \mathbb{R}^+$$

gilt (also $p \to 0$ für $n \to \infty$).

$$\begin{aligned}
&\lim_{\substack{n\to\infty\\ n\cdot p\to\lambda}} P(X=k) = \lim_{\substack{n\to\infty\\ n\cdot p\to\lambda}} \binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} = \\
&= \lim_{n\to\infty} \frac{n!}{k!\cdot(n-k)!}\cdot\left(\frac{\lambda}{n}\right)^k\cdot\left(1-\frac{\lambda}{n}\right)^{n-k} = \\
&= \frac{\lambda^k}{k!}\cdot\lim_{n\to\infty}\frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n\cdot n\cdot\ldots\cdot n}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k} = \\
&= \frac{\lambda^k}{k!}\cdot\lim_{n\to\infty} 1\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k}{n}+\frac{1}{n}\right)\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k} = \\
&= \frac{\lambda^k}{k!}\cdot e^{-\lambda}\,.
\end{aligned}$$

Dies gibt Anlaß zu folgender DEFINITION: Eine diskrete ZV Z heißt *Poisson-verteilt* mit dem *Parameter* $\lambda \in \mathbb{R}^+$, wenn

$$P(Z=k) = \frac{\lambda^k}{k!}\cdot e^{-\lambda} \quad (k=0,1,2,\ldots)$$

gilt.

Damit haben wir auch folgende *Approximationsaussage* in der Tasche:

$$\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k} \approx \frac{\lambda^k}{k!}\cdot e^{-\lambda} \quad \text{mit } \lambda = n\cdot p\,,$$

wenn n *groß* und p *klein* ist (*Faustregel:* $p \le 0,05$ und $n \ge 10$ oder $p \le 0,1$ und $n \ge 50$).

Die Poisson-Verteilung heißt daher auch *„Verteilung der seltenen Ereignisse."*

Zwei Bemerkungen:

- Der Erwartungswert und die Varianz von Z ergeben sich so:

$$E(Z) = \sum_{i=0}^{\infty} i\cdot\frac{\lambda^i}{i!}\cdot e^{-\lambda} = \lambda\cdot\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}\cdot e^{-\lambda} = \lambda\cdot e^{\lambda}\cdot e^{-\lambda} = \lambda \text{ und}$$

$$E(Z^2) = \sum_{i=0}^{\infty} i^2 \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda} =$$

$$= \sum_{i=2}^{\infty} i \cdot (i-1) \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda} + \underbrace{\sum_{i=1}^{\infty} i \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda}}_{=\lambda} =$$

$$= \lambda^2 \cdot \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \cdot e^{-\lambda} + \lambda = \lambda^2 + \lambda .$$

Daraus erhalten wir schließlich

$$D^2(Z) = E\left(Z^2\right) - E(Z)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda .$$

Diese Ergebnisse sind nicht überraschend, wenn wir

$$E(X) = np \longrightarrow_{n\to\infty} \lambda = E(Z) \quad \text{und}$$
$$D^2(X) = npq = np \cdot q \quad \underbrace{\longrightarrow_{n\to\infty}}_{p\to 0 \Rightarrow q=1-p\to 1} \quad \lambda \cdot 1 = \lambda = D^2(Z) \quad \text{beachten.}$$

Selbstverständlich ist durch obige Definition auch eine *Verteilung* gegeben:

$$\sum_{i=0}^{\infty} P(Z=i) = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} \cdot e^{\lambda} = 1 .$$

- Obige Grenzwertbetrachtung kann auch mittels *erzeugender Funktionen* durchgeführt werden:

$$G_Z(x) = \sum_{i=0}^{\infty} x^i \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{i=0}^{\infty} \frac{(\lambda x)^i}{i!} = e^{-\lambda} \cdot e^{\lambda x} = e^{\lambda \cdot (x-1)} .$$

$$\lim_{\substack{n\to\infty \\ n\cdot p\to\lambda}} G_X(x) = \lim_{\substack{n\to\infty \\ n\cdot p\to\lambda}} [p \cdot (x-1) + 1]^n = \lim_{n\to\infty} \left[\frac{\lambda}{n} \cdot (x-1) + 1\right]^n =$$
$$= \lim_{n\to\infty} \left[1 + \frac{\lambda \cdot (x-1)}{n}\right]^n = e^{\lambda \cdot (x-1)} = G_Z(x) .$$

Es gilt nämlich (*ohne* Beweis) folgender
SATZ: Eine Folge von diskreten Verteilungen $\langle P_n(k)\,, k = 0, 1, 2, \ldots \rangle$ konvergiert genau dann *„schwach"* gegen die diskrete Verteilung $P(k)$ [das heißt $\lim_{n\to\infty} P_n(k) = P(k)\ \forall k$], wenn die Folge der entsprechenden erzeugenden Funktionen $G_n(x)$ *gleichmäßig* in jedem Kreis (wir betrachten jetzt die erzeugenden Funktionen auf ganz $\mathbb{C}$) $\{x \text{ mit } |x| \leq r < 1\}$ gegen die erzeugende Funktion $G(x)$ von $P(k)$ konvergiert: $\lim_{n\to\infty} G_n(x) = G(x)$.

Beispiel 10: 2% der Bevölkerung seien Alkoholiker. *Man berechne die Wahrscheinlichkeit dafür, daß unter 100 zufällig ausgewählten Personen mindestens drei Alkoholiker sind!*

Lösung:

a) Die ZV X zähle die Alkoholiker in der Stichprobe. Dann ist X binomialverteilt mit den Parametern $n = 100$ und $p = 0,02$. Daraus folgt:

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - P(X = 0) -$$
$$- P(X = 1) - P(X = 2) = 1 - \binom{100}{0} \cdot 0,02^0 \cdot 0,98^{100} -$$
$$- \binom{100}{1} \cdot 0,02^1 \cdot 0,98^{99} - \binom{100}{2} \cdot 0,02^2 \cdot 0,98^{98} = 0,323314378 \; .$$

b) Wegen $n = 100 \geq 10$ und $p = 0,02 \leq 0,05$ können wir die *Poisson-Verteilung zur Approximation* heranziehen. Mit $\lambda = np = 2$ ist

$$P(X \geq 3) = 1 - P(X \leq 2) \approx 1 - e^{-2} \cdot \left(1 + \frac{2^1}{1!} + \frac{2^2}{2!}\right) =$$
$$= 1 - 5 \cdot e^{-2} = 0,323323584 \; ,$$

worin wir große Übereinstimmung mit a) erkennen. ◇

Literatur

- Bosch, K.: Elementare Einführung in die Wahrscheinlichkeitsrechnung; Vieweg (Reihe „Basiswissen"), Braunschweig/Wiesbaden 1986⁵.
- Bürger, H., Fischer, R., Malle, G. u.a.: Mathematik Oberstufe 3; hpt, Wien 1991.
- Reichel, H.-C., Hanisch, G. und Müller, R.: Wahrscheinlichkeitsrechnung und Statistik; hpt, Wien 1989^2.
- Reichel, H.-C., Müller, R., Hanisch, G. und Laub, J.: Lehrbuch der Mathematik 7; hpt, Wien 1991.
- Rosanow, J.A.: Wahrscheinlichkeitstheorie; Vieweg (Serie „Mathematik Grundwissen"), Braunschweig (Redaktion); veröffentlicht bei Rowohlt, Reinbeck bei Hamburg 1974.

Anschrift des Verfassers:

Mag. Stefan Götz
Institut für Mathematik der Universität Wien und AKG Wien
A-1090 Wien, Strudlhofgasse 4 bzw. A-1010 Wien, Beethovenplatz 1