

Grundlagen der Binomialverteilung
Approximation der Binomialverteilung
durch die
Normalverteilung

Auf wenigen Seiten eine Übersicht über
Fakten und Methoden

Datei Nr. 34512

Stand 14. März 2015

Friedrich W. Buckel

INTERNETBIBLIOTHEK FÜR SCHULMATHEMATIK

<https://mathe-cd.de>

Vorwort

Dieser Text fasst das Wichtigste über die Binomialverteilung und die Normalverteilung kompakt zusammen. Dabei spielen das Histogramm als Schaubild der Binomialverteilung eine ganz wichtige Rolle, zusammen mit Erwartungswert und Standardabweichung.

Dann wird gezeigt, dass mit dem Verfahren der Standardisierung eine Grundform der Binomial-Histogramme erzeugt werden kann, was zu den beiden standardisierten Gaußfunktionen ϕ und Φ führt. Mit ihrer Hilfe kann man die Binomialverteilung näherungsweise auch berechnen.

Es gibt aber auch Größen, die sich nicht mit der Binomialverteilung berechnen lassen, und die dann nur mit ϕ und Φ berechnet werden.

Wer diesen Text ausführlicher und mit Beispielen und Aufgaben haben möchte, findet ihn unter der Nummer 34510 Normalverteilung, und eine große Aufgabensammlung unter 34511.

Wer eine einfache Einführung zur Binomialverteilung benötigt, kann die Texte 34010 und 34011 lesen.

Inhalt

1	Grundwissen zur Binomialverteilung	3
2	Standardisierung der Binomialverteilung	7
3	Verteilungsfunktionen für „Höchstens-Ereignisse“	11
4	Normalverteilte Zufallsgrößen	16
5	Sigma-Umgebungen vom Erwartungswert	18

1. Grundwissen zur Binomialverteilung

Welche Experimente werden hier berechnet?

Experimente, bei denen die Binomialverteilung zum Einsatz kommt, bestehen aus n gleichen Stufen (Vorgängen), bei dem jede genau zwei mögliche Ergebnisse (Ausgänge, Merkmale) hat. Man nennt das Experiment dann eine **Bernoullikette**.

Die Wahrscheinlichkeiten:

Das typische Merkmal für eine Bernoullikette sind die in allen Stufen gleichbleibenden Wahrscheinlichkeiten.

Beispiele: Ziehen aus einer Urne mit Zurücklegen, Würfeln, Drehen eines Glücksrades, Produktion von fehlerhaften Bauteilen durch eine Maschine, usw.

Um welche Ereignisse geht es dabei?

Es sind Ereignisse, bei denen man zählt, wie oft ein Merkmal bei den insgesamt n Stufen (n heißt **Umfang der Stichprobe**) vorkommt. Diese Anzahl ordnet man einer Zufallsvariablen zu, der man meist Namen wie X , Y , Z gibt. Beispiele:

- X sei die Anzahl der gewürfelten Sechsen unter 100 Würfeln. Dann kann man das Ereignis A: „Man würfelt 7 mal die Sechsen“ schreiben als A: $X = 7$.
- Y sei die Anzahl der defekten Schrauben bei der Produktion von 1000 Schrauben. Das Ereignis B: „Man findet 25 defekte Schrauben“ kann man so darstellen: B: $Y = 25$.

Wie berechnet man die Wahrscheinlichkeiten dieser Ereignisse?

- $P(A) = P(X = 7) = \binom{100}{7} \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^{93} \approx 0,0025$ Dieser Term lässt sich so erklären:

Man denkt sich die 100 Würfelresultate auf einem „Pfad“ aufgeschrieben.

Die 7 gewürfelten Sechsen nehmen in diesem Pfad 7 bestimmte Plätze ein, die anderen 93 Würfelresultate sind keine Sechsen. Beispielpfad:



Jeder dieser Pfade hat, egal wo man nun die 7 Sechsen findet, die gleiche

Wahrscheinlichkeit: $\left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^{93}$. Nun muss man nur noch ermitteln, wie viele solche

Pfade es gibt. Die Kombinatorik sagt uns, dass man aus den 100 Plätzen eines solchen

Pfades die 7 Plätze für die Sechsen auf $\binom{100}{7} = 16.007.560.800$ Möglichkeiten

auswählen kann. So viele Pfade gibt es also.

- $P(B) = P(Y = 25) = \binom{1000}{25} \cdot 0,02^{25} \cdot 0,98^{975} \approx 0,0446$, wenn $p = 0,02$ die

Wahrscheinlichkeit dafür ist, dass eine defekte Schraube produziert worden ist.

FORMEL:

Es sei X die beobachtete Anzahl des Merkmals, das mit der Wahrscheinlichkeit p auftritt. n sei der Umfang der Stichprobe. Dann ist die Wahrscheinlichkeit dafür, dass dieses Merkmal genau k -mal gefunden wird:

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

Diese Funktion heißt **Binomialverteilung**.

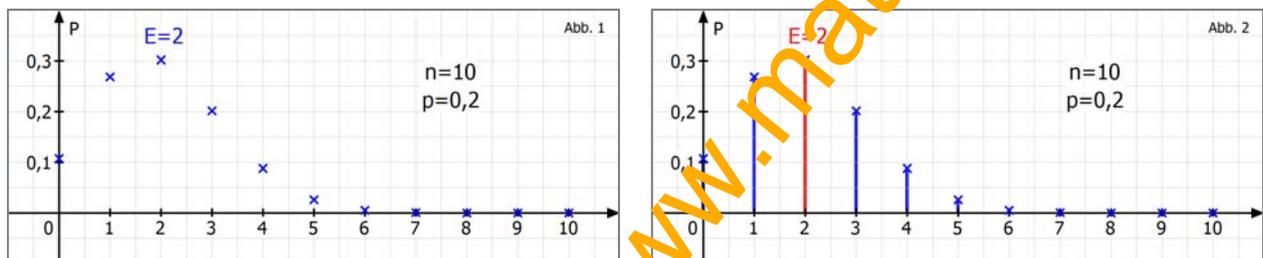
Erwartungswert einer Zufallsvariablen

Die Frage „Mit welchem X-Wert kann man rechnen“ oder „Welchen X-Wert kann man erwarten“ führt zum Erwartungswert der Zufallsvariablen, den man so berechnet: $E(X) = n \cdot p$. Dieser Erwartungswert hat von allen X-Werten die größte Wahrscheinlichkeit. Er ist der mit „großer“ Wahrscheinlichkeit vorhergesagte Mittelwert.

Schaubilder der Binomialverteilung

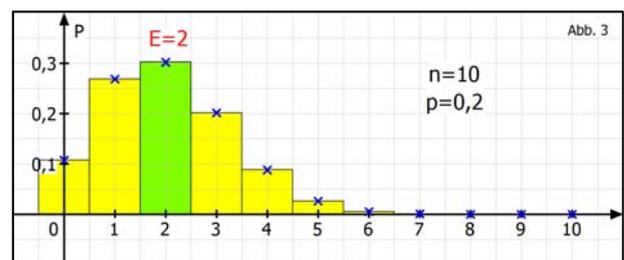
Man kann die aus k und den Wahrscheinlichkeitswerten $P(X=k)$ gebildeten Paare $(k | P(X=k))$ als Punkte abtragen und erhält dann das Schaubild (den Graph) der Binomialverteilung.

Die Binomialverteilung ist keine stetige Funktion, sondern hat einen Definitionsbereich aus genau $(n+1)$ Zahlen. Im Fall der Abbildung (also bei $n = 10$) besteht der Definitionsbereich aus diesen 11 Zahlen: $D = \{0; 1; 2; \dots; 9; 10\}$. Das Schaubild dieser Binomialverteilung besteht also aus genau 11 Punkten. (Abb. 1). Man darf also diese Punkte nicht zu einer Kurve verbinden, da es keine Zwischenwerte gibt. Oft erzeugt man ein **Stabdiagramm** (Abb. 2)



Psychologen haben festgestellt, dass Flächen sich besser zur Darstellung von Größen eignen. Daher hat man das sogenannte Histogramm erfunden. Es enthält Rechtecke der Breite 1. Da man die Wahrscheinlichkeit $P(X=k)$ als Höhe des Rechtecks verwendet, stellt auch der Flächeninhalt diese Wahrscheinlichkeit anschaulich dar.

Die in Abb. 3 eingetragenen Punkte (Kreuze) lässt man später weg. Man erkennt hier jedoch nur 7 Rechtecke, da die Höhe der zu $X=7$ bis 10 gehörenden Rechtecke zu klein für eine Darstellung ist.



Wichtig ist noch, dass man weiß, dass die **Summe aller Wahrscheinlichkeiten 1** ist, und dass aus diesem Grund auch die Gesamtfläche aller Histogramm-Rechtecke stets den Inhalt 1 hat.

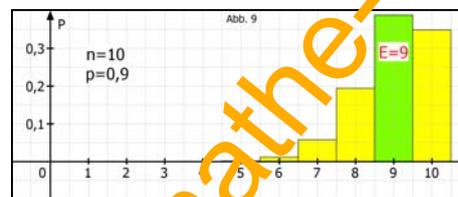
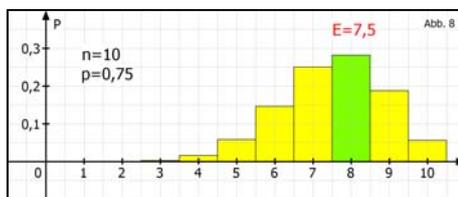
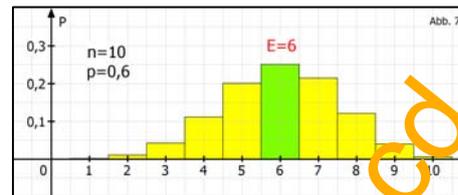
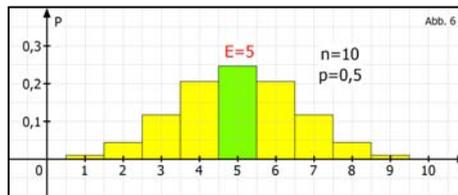
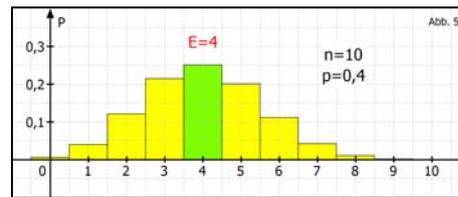
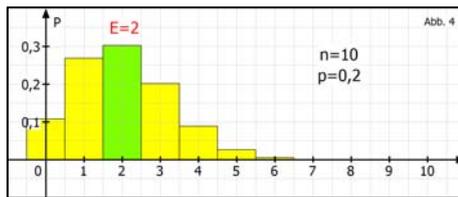
Unterschiedliche Formen der Histogramme

Zur Binomialverteilung eines bestimmten Umfangs n der Stichprobe gehört also stets ein Histogramm aus $(n+1)$ Rechtecken (auch wenn sie wegen zu geringer Höhe nicht mehr darstellbar sind. Mit zunehmendem n werden die Histogramme also immer breiter. Und weil die Gesamtfläche stets 1 ist, werden sie mit wachsendem n immer flacher.

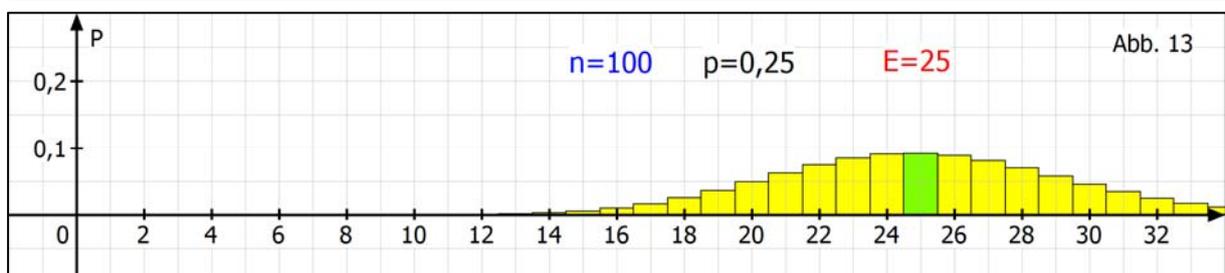
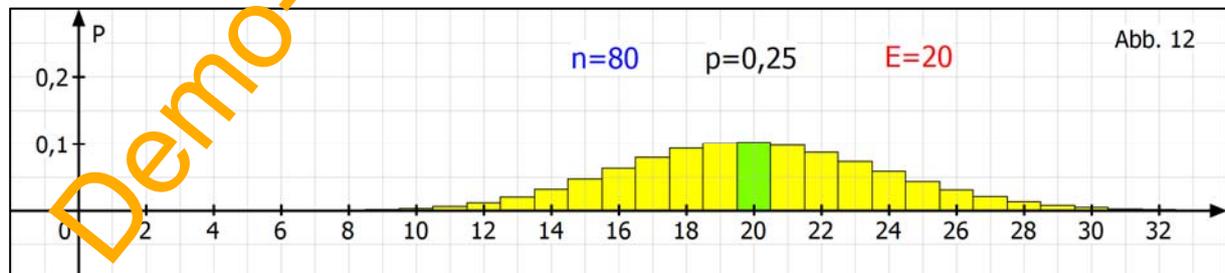
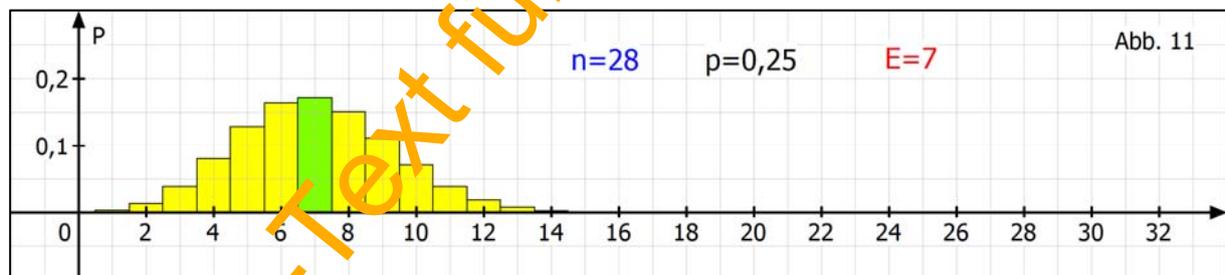
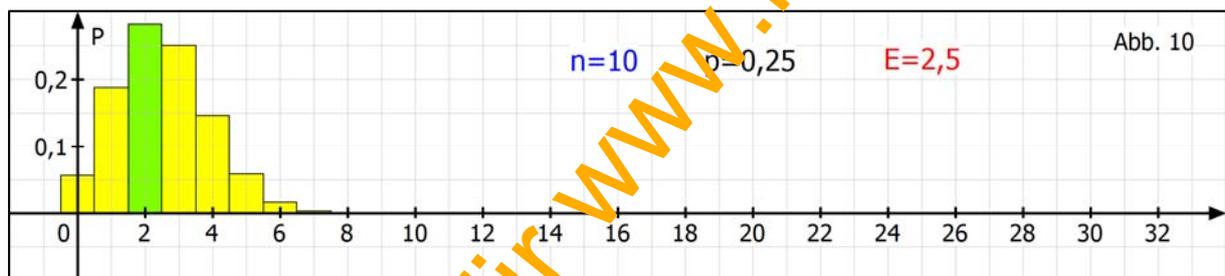
Die **Lage des Maximums** (beim Erwartungswert E) hängt von p ab. Bei $p = 0,5$ liegt der Erwartungswert bei $E(X) = \frac{1}{2}n$, also in der Mitte des Histogramms. Bei kleinerem p liegt das Maximum weiter links, bei größerem p weiter rechts.

Beispiele verschiedener Histogramme

$n = 10$, verschiedene p -Werte:

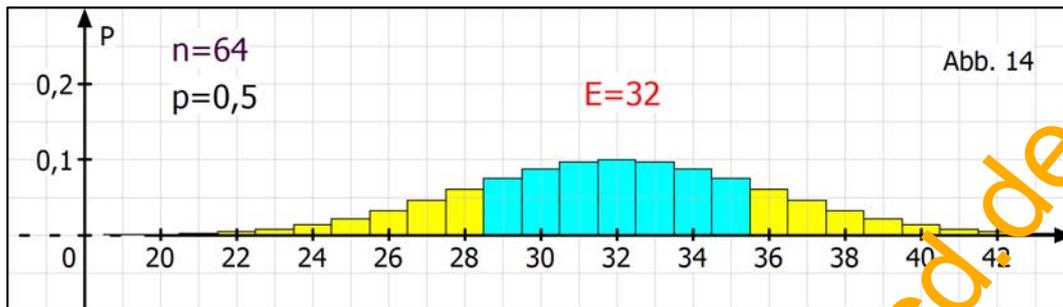


$p = 0,25$, verschiedene n -Werte:



Die Streuung der Werte um den Erwartungswert

Der Zufall sorgt bei jedem „Zufallsexperiment“ dafür, dass die X-Werte um den Erwartungswert herum schwanken (streu). Und das mit unterschiedlicher Wahrscheinlichkeit.



Hier wurde eine Umgebung von $E = 32$ mit dem Radius 3 gewählt. Man kann diese Umgebung durch eine Doppelungleichung darstellen: $32 - 3 \leq X \leq 32 + 3$ bzw. $29 \leq X \leq 35$.

Die Wahrscheinlichkeit dafür, dass ein Ergebnis für X in dieser Umgebung liegt, kann mit der **Verteilungsfunktion** (Summenfunktion) der Binomialverteilung berechnet werden.

In geeigneten Rechnern wird diese Funktion mit Binomial C.D oder BinomialCdf bezeichnet.

Hier wurde berechnet: $P(X = 29) + P(X = 30) + \dots + P(X = 35) = P(29 \leq X \leq 35) \approx 0.62$:

Als Maß für die Streuung der Werte hat

sich die sogenannte **Standardabweichung**

bewährt. Sie wird bei der Binomialverteilung

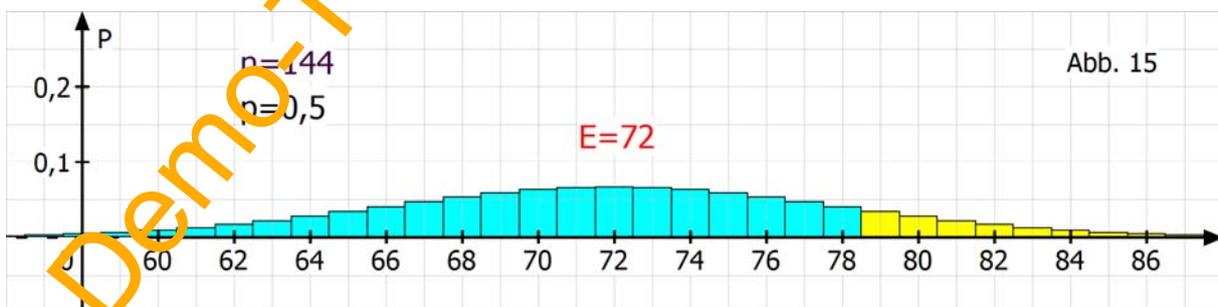
so berechnet: $\sigma = \sqrt{n \cdot p \cdot (1-p)}$

Binomial C.D
Data: Variable
Lower: 29
Upper: 35
Numtrial: 64
p: 0.5
Save Res: None
List Var

Binomial C.D
p=0.6182672

In unserem Beispiel ist sie $\sigma = \sqrt{36 \cdot 0.5 \cdot 0.5} = 3$. Nimmt man diesen Wert als Radius für eine Umgebung des Erwartungswerts, dann erhält man einen Bereich für X , in dem zu etwa **68%** die X -Werte liegen. Dieser Wert ist ein Näherungswert, der um so besser zutrifft, je größer die Standardabweichung ist. Auf jeden Fall sollte sie größer als 3 sein. Wir hatten hier $\sigma = 3$ und wir erkennen die leichte Abweichung.

Nun ein Beispiel mit größerer Standardabweichung:



$\sigma = \sqrt{144 \cdot 0.5 \cdot 0.5} = 6$. Die Umgebung von $E = 72$ mit Radius 6 ist: $66 \leq X \leq 78$.

$P(X = 66) + P(X = 67) + \dots + P(X = 78) = P(66 \leq X \leq 78) \approx 0.72$

Ein X -Wert liegt also im blau gefärbten Bereich mit etwa 72 % Wahrscheinlichkeit.

Binomial C.D
Data: Variable
Lower: 66
Upper: 78
Numtrial: 144
p: 0.5
Save Res: None
List Var

Binomial C.D
p=0.72138545

2. Standardisierung der Binomialverteilung

Was versteht man darunter?

Carl Friedrich Gauß hat entdeckt, dass man alle Histogramme auf eine Grundform zurückführen kann. Dazu muss man sie in x-Richtung verschieben und stauchen. Anschließend werden sie in y-Richtung gestreckt. Dann passen sie näherungsweise zur berühmten Gaußkurve.

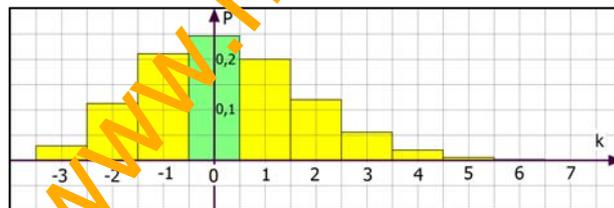
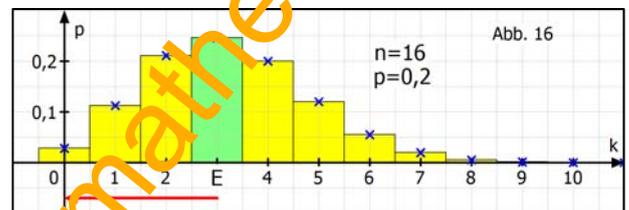
Das hat eine historische Bedeutung in dem Sinne, dass man aus diesem Grunde jede Binomialverteilung, sofern $\sigma > 3$ ist, näherungsweise durch diese Gaußfunktion berechnen kann. Schüler sollten dieses Verfahren nach Ansicht der Mathelehrer noch beherrschen.

Standardisierung an drei Beispielen gezeigt.

- (1) Die **Binomialverteilung zu $n = 16$, $p = 0,2$** hat den Erwartungswert $E = n \cdot p = 3,2$ und die Standardabweichung $\sigma = \sqrt{16 \cdot 0,2 \cdot 0,8} = 1,6$.

Das Histogramm der Binomialverteilung:

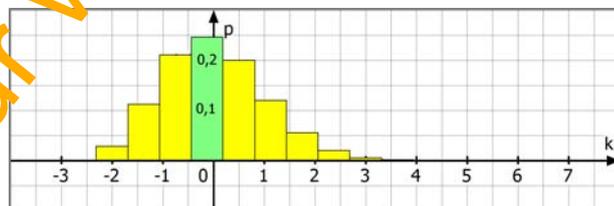
verschiebt man so, dass der Erwartungswert in den Ursprung kommt:



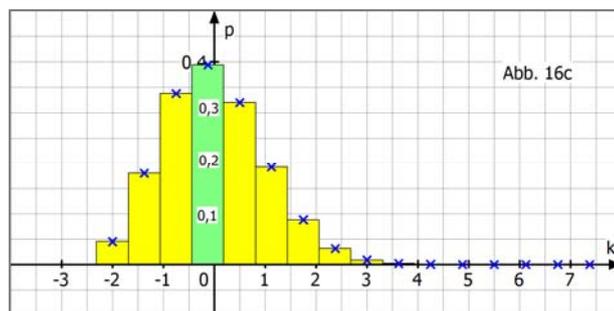
Dann **staucht** man es mit dem

$$\text{Faktor } \frac{1}{\sigma} = \frac{1}{1,6} = 0,625$$

Es wird also schmaler.



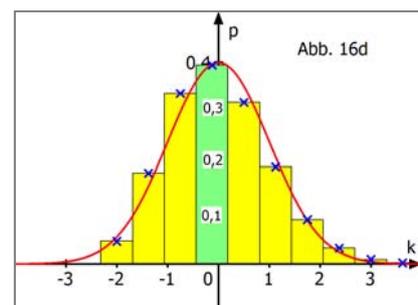
Weil damit die Gesamtfläche nicht mehr den Inhalt 1 hat, **streckt** man die Figur mit dem Faktor $\sigma = 1,6$ in y-Richtung. Jetzt stellen die Flächen der Rechtecke wieder die Wahrscheinlichkeitswerte dar.



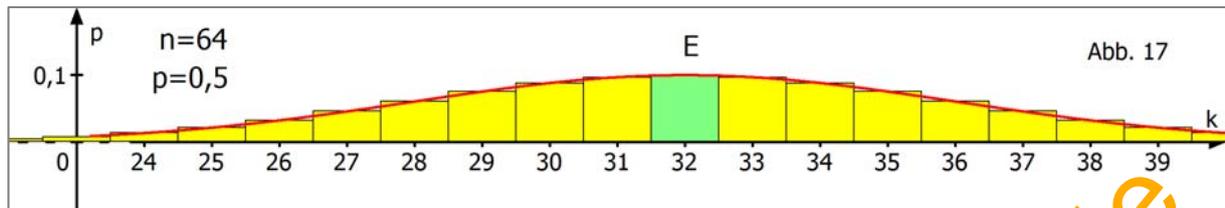
Die Kreuze geben die abgebildeten Punkte des Graphen der Binomialverteilung wieder.

Die blaue Kurve ist die Gaußsche Glockenkurve, die zur Funktion φ (phi) gehört.

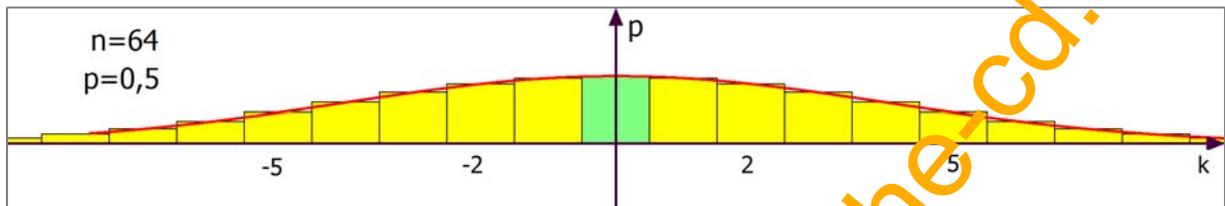
Man erkennt, dass sie (wegen $\sigma < 3$), deutlich an den Punkten vorbeigeht.



- (2) Die **Binomialverteilung zu $n = 64$, $p = 0,5$** hat den Erwartungswert $E = 32$ und $\sigma = 4$.
Ich habe über die Rechtecke eine passende Glockenkurve gelegt.

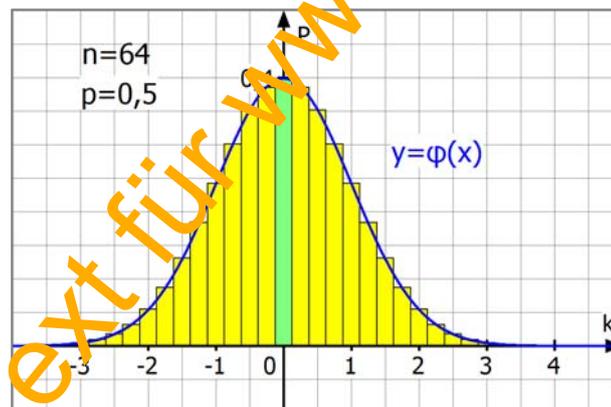


Nach der Verschiebung liegt E bei 0.



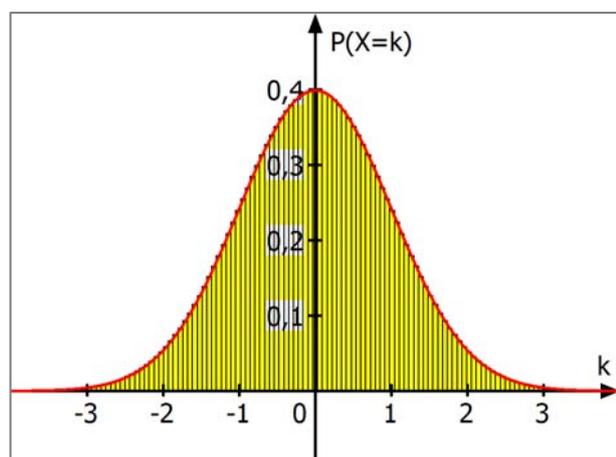
Die nächsten zwei Schritte fasse ich zusammen:

Zuerst wird die Anordnung in x-Richtung gestaucht mit dem Faktor $\frac{1}{4} = \frac{1}{\sigma}$, wodurch sich das Histogramm von der Breite 65 auf ein Viertel, also auf 16,25 zusammen. Und damit alle Flächeninhalte wieder die ursprünglichen Werte haben, wird die Anordnung mit dem Faktor $\sigma = 4$ in y-Richtung gestreckt.



Man erkennt, dass jetzt ($\sigma > 3$) die standardisierte Gaußkurve mit besserer Genauigkeit durch die Mittelpunkte der oberen Rechteckseiten geht.

- (3) Nun habe ich das Histogramm der Binomialverteilung für $n = 1000$, $p = 0,4$ gleich in die standardisierte Form gebracht. Durch die vielen vertikalen Rechteckslinien wird die Fläche dunkel. Man erkennt, dass die Ecken der Rechtecke kaum mehr sichtbar sind. Für $n \rightarrow \infty$ erhält man die glatte Gaußkurve.

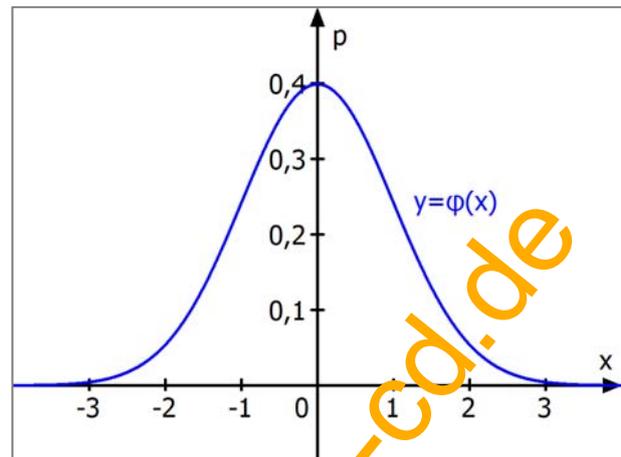


Welche Funktionsgleichung gehört zur standardisierten Gaußkurve?

Diese standardisierte Gaußfunktion hat die

Gleichung $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$.

Sie ist symmetrisch zur y-Achse und hat ihr Maximum bei $x = 0$ mit $\varphi(0) \approx 0,4$.



Um die Gleichungen der Glockenkurven zu erhalten, die zu den normalen, also nicht standardisierten Histogrammen gehören, muss man die drei besprochenen Abbildungen wieder rückgängig machen.

Zuerst staucht man die y-Koordinaten mit $\frac{1}{\sigma}$ zu:

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

Dann wird die Kurve in x-Richtung mit dem Faktor σ gestreckt:

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

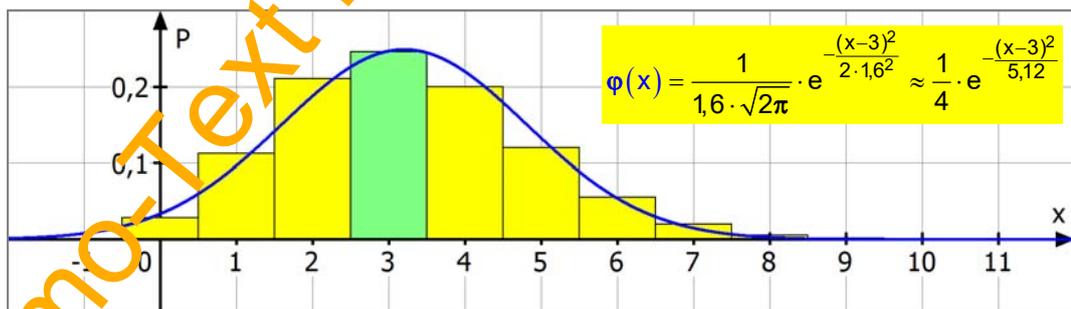
Schließlich wird die Kurve wieder nach rechts um die „Strecke E“ zurückverschoben.

Dann erhält man diese Glockenfunktion:

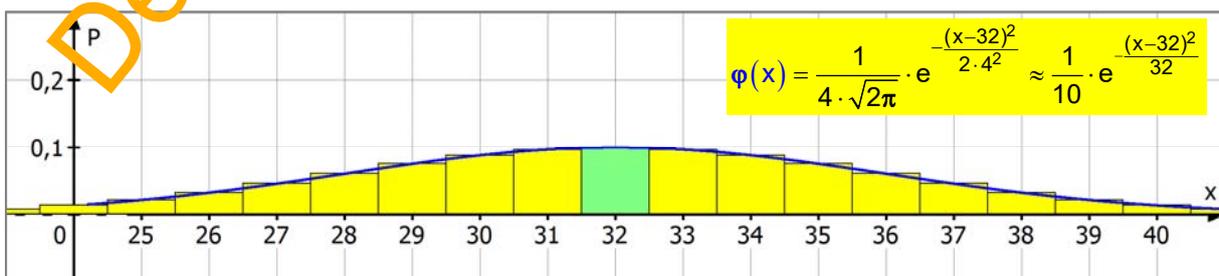
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-E)^2}{2\sigma^2}}$$

Sie passt dann näherungsweise zum Original-Histogramm der Binomialverteilung:

Zuerst $n = 16$, $p = 0,2$ (von Seite 7)



Dann $n = 64$, $p = 0,5$ (von Seite 8)



Berechnung von Wahrscheinlichkeiten mit φ :

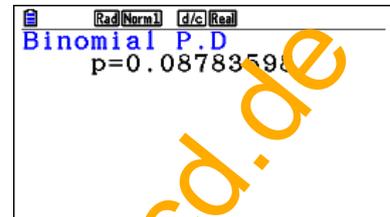
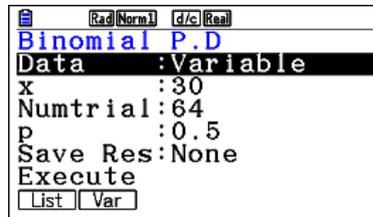
Im ersten Beispiel war $n = 64$, $p = 0,5$ mit $E = 32$ und $\sigma = 4$

Berechnung mit der **Binomialverteilung**:

$$P(X = 30) = \binom{64}{30} \cdot 0,5^{30} \cdot 0,5^{34} = \binom{64}{30} \cdot 0,5^{64} \approx 0,0878$$

$$64C30 \times 0.5^{64} \\ 0.08783598905$$

Grafikrechner:



Näherungsrechnung mit der **Normalverteilungsfunktion**:

$$P(X = 30) \approx \varphi(30) = \frac{1}{4\sqrt{2\pi}} \cdot e^{-\frac{(30-32)^2}{2 \cdot 4^2}} \approx 0,0880$$

$$e^{-((30-32)^2 \div 32)} \div (4\sqrt{2\pi}) \\ 0.08801633169$$

Näherungsrechnung mit der **standardisierten Normalverteilung** (Tabellenwert)

$$P(X = 30) \approx \frac{1}{4} \cdot \varphi\left(\frac{30-32}{4}\right) = \frac{1}{4} \varphi(-0,5) = \frac{1}{4} \varphi(0,5) \approx \frac{1}{4} \cdot 0,3521 \approx 0,0880$$

Und hier der passende Tabellenausschnitt für die **standardisierte φ -Funktion**:

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	,3989	,3989	,3988	,3986	,3984	,3982	,3980	,3977	,3973
0,1	0,3970	,3965	,3961	,3956	,3951	,3945	,3939	,3932	,3925	,3918
0,2	0,3910	,3902	,3894	,3885	,3877	,3867	,3857	,3847	,3836	,3825
0,3	0,3814	,3802	,3790	,3778	,3765	,3752	,3739	,3725	,3712	,3697
0,4	0,3683	,3668	,3653	,3637	,3621	,3605	,3589	,3572	,3555	,3538
0,5	0,3521	,3503	,3485	,3467	,3448	,3429	,3410	,3391	,3372	,3352
0,6	0,3332	,3312	,3292	,3271	,3251	,3230	,3209	,3187	,3166	,3144
0,7	0,3123	,3101	,3079	,3055	,3034	,3011	,2989	,2966	,2943	,2920
0,8	0,2897	,2874	,2850	,2827	,2803	,2780	,2756	,2732	,2709	,2685
0,9	0,2661	,2637	,2613	,2589	,2565	,2541	,2516	,2492	,2468	,2444
1,0	0,2420	,2396	,2371	,2344	,2323	,2299	,2275	,2251	,2227	,2203
1,1	0,2179	,2155	,2130	,2107	,2083	,2059	,2036	,2012	,1989	,1965
1,2	0,1942	,1915	,1890	,1872	,1849	,1826	,1804	,1781	,1758	,1736
1,3	0,1714	,1687	,1669	,1647	,1626	,1604	,1582	,1561	,1539	,1518
1,4	0,1497	,1476	,1456	,1435	,1415	,1394	,1374	,1354	,1334	,1315
1,5	0,1295	,1276	,1257	,1238	,1219	,1200	,1182	,1163	,1145	,1127
1,6	0,1109	,1092	,1074	,1057	,1040	,1023	,1006	,0989	,0973	,0957
1,7	0,0940	,0925	,0909	,0893	,0878	,0863	,0848	,0833	,0818	,0804
1,8	0,0790	,0775	,0761	,0748	,0734	,0721	,0707	,0694	,0681	,0669
1,9	0,0656	,0644	,0632	,0620	,0608	,0596	,0584	,0573	,0562	,0551
2,0	0,0540	,0529	,0519	,0508	,0498	,0488	,0478	,0468	,0459	,0449
2,1	0,0440	,0431	,0422	,0413	,0404	,0396	,0387	,0379	,0371	,0363
2,2	0,0355	,0347	,0339	,0332	,0325	,0317	,0310	,0303	,0297	,0290
2,3	0,0283	,0277	,0270	,0264	,0258	,0252	,0246	,0241	,0235	,0229
2,4	0,0224	,0219	,0213	,0208	,0203	,0198	,0194	,0189	,0184	,0180
2,5	0,0175	,0171	,0167	,0163	,0158	,0154	,0151	,0147	,0143	,0139
2,6	0,0136	,0132	,0129	,0126	,0122	,0119	,0116	,0113	,0110	,0107
2,7	0,0104	,0101	,0099	,0096	,0093	,0091	,0088	,0086	,0084	,0081
2,8	0,0079	,0077	,0075	,0073	,0071	,0069	,0067	,0065	,0063	,0061
2,9	0,0060	,0058	,0056	,0055	,0053	,0051	,0050	,0048	,0047	,0046
3,0	0,0044	,0043	,0042	,0040	,0039	,0038	,0037	,0036	,0035	,0034
3,1	0,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026	,0025	,0025
3,2	0,0024	,0023	,0022	,0022	,0021	,0020	,0020	,0019	,0018	,0018
3,3	0,0017	,0017	,0016	,0016	,0015	,0015	,0014	,0014	,0013	,0013
3,4	0,0012	,0012	,0012	,0011	,0011	,0010	,0010	,0010	,0009	,0009

3. Verteilungsfunktionen für „Höchstens-Ereignisse“

Demo-Text für www.mathe-cd.de