Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

17. November 2010

1. Gesetze
   - Das Gesetz der seltenen Ereignisse
   - Das schwache Gesetz der großen Zahl

2. Verteilungsfunktionen
   - Definition
   - Binomialverteilung
   - Geometrische Verteilung
   - Poissonverteilung

3. Stetige Zufallsvariable
   - Standardisierte Verteilung
   - Standard-Normalverteilung

Poissonverteilung

Es sei $\lambda > 0$. Die Poissonverteilung zum Parameter λ ist definiert durch

$$P_\lambda(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

Unter den folgenden Voraussetzungen ist eine Zufallsvariable X poissonverteilt zum Parameter λ:

- X zählt das Auftreten eines Ereignisses pro Zähleinheit
- Im Mittel treten λ Ereignisse pro Zähleinheit auf
- Die Ereignisse beeinflussen sich nicht gegenseitig

Beispiel Tumor

- Ein Tumor aus 160 Zellen wird bestrahlt
- Im Mittel sterben 16 Tumorzellen pro Minute
- Mit welcher Wahrscheinlichkeit sterben 10 Zellen in der ersten Minute?
- Zwei Modelle sind angemessen
  - Binomialverteilung
  - Poissonverteilung

Beispiel Tumor: Rechnung mit Binomialverteilung

- Modell: 160 unabhängige ja/nein-Experimente

  Erfolg : Tod der Tumorzelle

- Erfolgswahrscheinlichkeit im Einzelfall $p = 0.1$
- Anzahl der Erfolge verteilt gemäß $B_{160,\,0.1}$
- Antwort:

$$B_{160,\,0.1}(10) = 0.03113$$

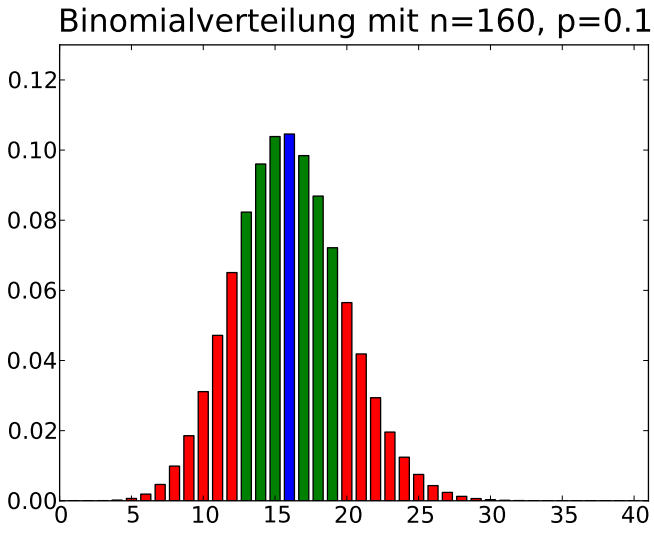
Beispiel Tumor: Rechnung mit Poissonverteilung

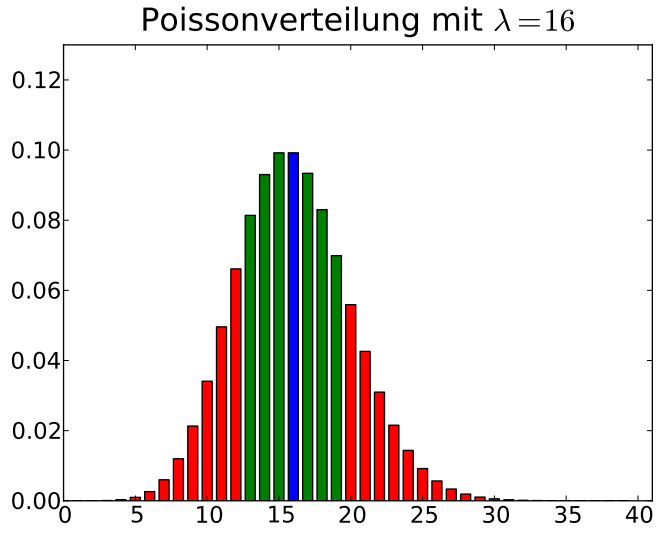
- Modell: seltenes Ereignis, das im Mittel 16 mal pro Minute auftritt
- Parameter der Poissonverteilung ist $\lambda = 16$
- Anzahl der Ereignisse pro Zähleinheit verteilt gemäß P_{16}
- Antwort

$$P_{16}(10) = \frac{16^{10}}{10!} e^{-16} = 0.03410$$

- Zum Vergleich $B_{160,\,0.1}(10) = 0.03113$

Vergleich Binomial- und Poissonverteilung

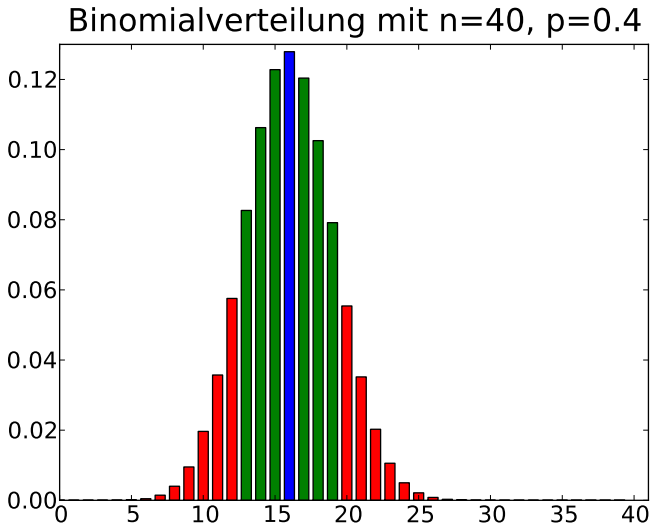


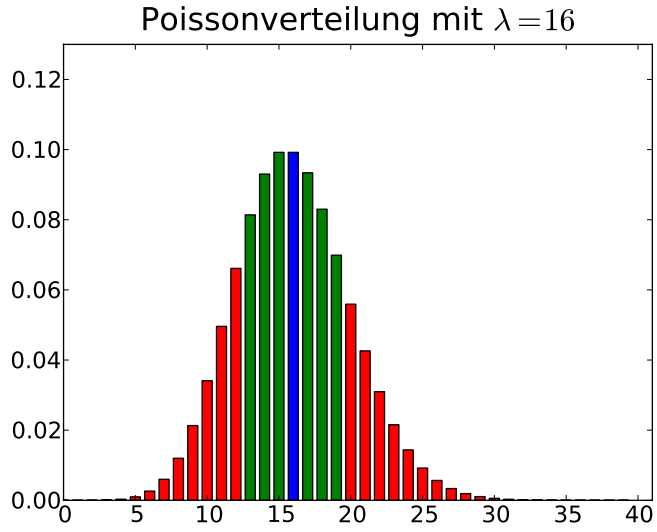


Beide beschreiben einen Prozess mit 16 Erfolgen im Mittel. Der Unterschied ist, dass beim Poissonprozess die Anzahl der Erfolge potenziell unbeschränkt ist.

Vergleich Binomial- und Poissonverteilung, Fortsetzung

$B_{40,0.4}$ und P_{16} besitzen ebenfalls beide den Erwartungswert 16





Gesetz der seltenen Ereignisse

Die Poisson-Verteilung P_λ mit $\lambda = n \cdot p$ ist eine sehr gute Annäherung an die Binomialverteilung $B_{n,p}$, falls $n \geq 100$ und $n \cdot p \leq 10$.

Im Beispiel waren

- $n = 160$
- $p = 0.1$

Die Annäherung ist daher nur gut

Messwiederholungen

- Warum erhöhen mehrere Messungen die Genauigkeit?
- Warum braucht man 100-mal so viele Messungen, um die Genauigkeit zu verzehnfachen?

Rechenregeln für den Erwartungswert

- Für jede Zahl c und jede Zufallsvariable X ist $E(c \cdot X) = c \cdot E(X)$
- Für Zufallsvariablen $X_1, \ldots, X_n$ ist $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$
- X und Y **unabhängige** Zufallsvariable. Dann

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Rechenregeln für die Varianz

- Für jede Zahl a und jede Zufallsvariable X gilt $Var(a + X) = Var(X)$
- Für Zahl c und jede Zufallsvariable X gilt $Var(c \cdot X) = c^2 \cdot Var(X)$
- X und Y **unabhängige** Zufallsvariable. Dann

$$Var(X + Y) = Var(X) + Var(Y)$$

Zwei unabhängige, identisch verteilte Zufallsvariable

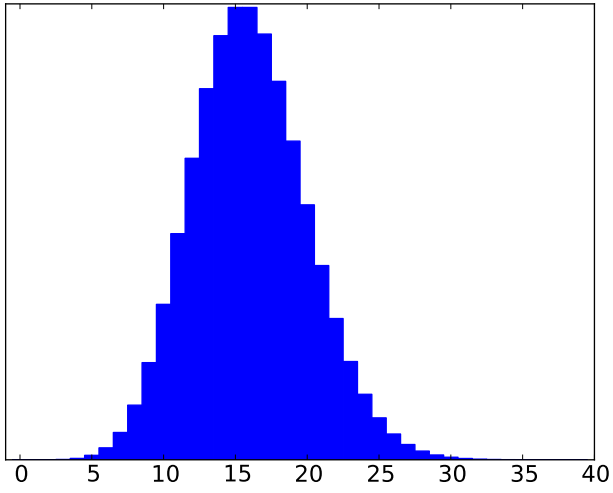
- X_1 und X_2 seien unabhängige Zufallsvariable, die derselben Verteilung gehorchen (also z. B. Messwiederholungen). Sei $Y = \frac{1}{2}(X_1 + X_2)$ der Durchschnittswert
- Der Erwartungswert von X_1 heiße μ, also $E(X_1) = E(X_2) = \mu$
- Die Streuung von X_1 heiße σ, also $Var(X_1) = Var(X_2) = \sigma^2$
- $E(Y) = \frac{1}{2}(E(X_1) + E(X_2)) = \mu$
- $Var(Y) = \left(\frac{1}{2}\right)^2 Var(X_1) + \left(\frac{1}{2}\right)^2 Var(X_2) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$
- Also ist $\frac{\sigma}{\sqrt{2}}$ die Streuung von Y

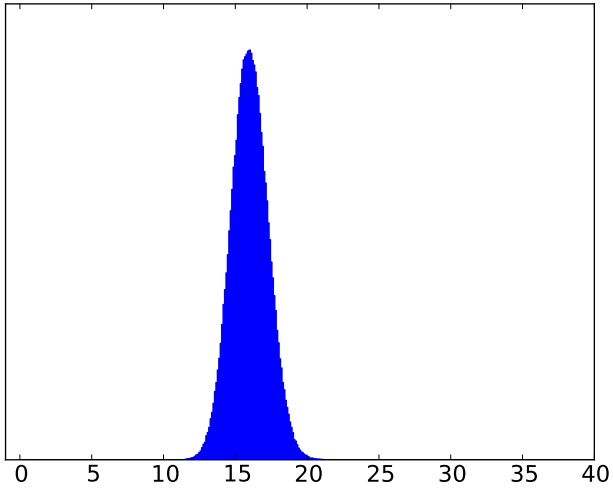
Das schwache Gesetz der großen Zahl

- "Mit ausreichend vielen Messwiederholungen lässt sich jede Genauigkeit erreichen"
- Präziser: $X_1, \ldots, X_n$ unabhängig, alle mit derselben Verteilung
- $\mu = E(X_1) = \cdots = E(X_n)$ und
  $\sigma^2 = Var(X_1) = \cdots = Var(X_n)$
- $Y = \frac{1}{n}(X_1 + \cdots + X_n)$
- Y ist das arithmetische Mittel der $X_1, X_2, \ldots, X_n$
- Dann $E(Y) = \mu$ und die Streuung von Y beträgt

$$\sigma_Y = \frac{\sigma}{\sqrt{n}}$$

Messwiederholungen: Beispiel





Links: Poissonverteilung P_{16}, Streuung ist 4

Rechts: Durchschnittswerte aus zehn P_{16}-verteilten Zufallsvariablen, Streuung ist

$$\frac{4}{\sqrt{10}} = 1.26$$

Beispiel: α-Strahler

- Ein Geigerzähler registriert pro Sekunde ungefähr 25 α-Teilchen
- Dieser Wert soll so genau bestimmt werden, dass die Streuung des Mittelwerts Y unter 2.0 liegt
- Wie viele Einzelversuche von einer Sekunde Dauer sind erforderlich?
- $Var(P_\lambda) = \lambda$. Also hat jeder Einzelversuch die Streuung $\sqrt{25} = 5$
- Gleichung

$$\frac{5}{\sqrt{n}} = 2.0$$

- Also $\sqrt{n} = 2.5$ und $n \cong 6$

Verteilungsfunktion

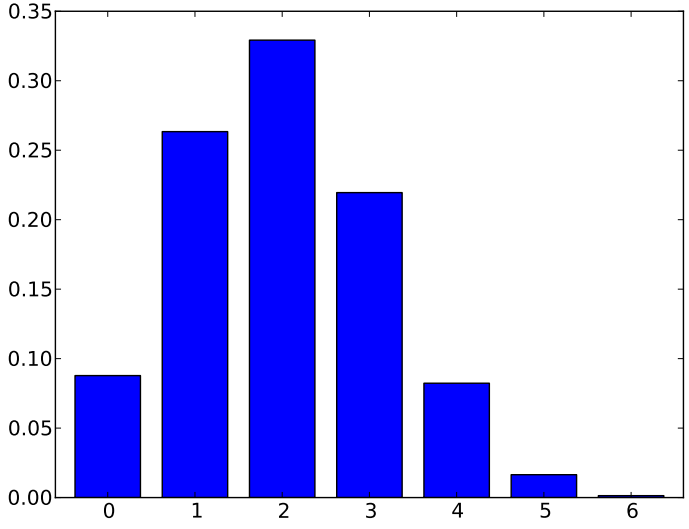
- X sei eine Zufallsvariable. Die Funktion

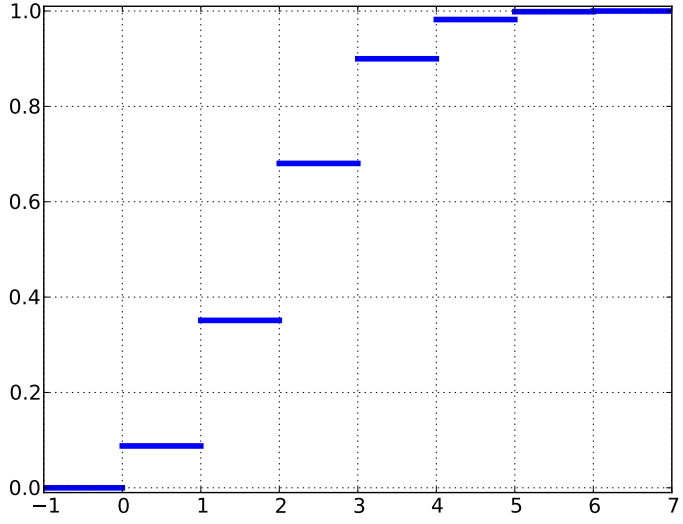
$$F(x) = P(X \leq x)$$

  ist die *Verteilungsfunktion* von X
- Die Verteilungsfunktion $F(x)$ gibt an, mit welcher Wahrscheinlichkeit ein **kleinerer** Wert als x angenommen wird (x eingeschlossen)

Datensatz	Zufallsvariable
empirische Häufigkeitsverteilung	Verteilung
empirische Verteilungsfunktion	Verteilungsfunktion

Beispiel: Verteilungsfunktion von $B_{6,1/3}$





Stabdiagramm und Graph der Verteilungsfunktion für $B_{6,1/3}$

Kumulierte Tabellen

- Kumulierte Tabellen der Binomialverteilung zeigen die Verteilungsfunktion
- Beispiel: 47 Versuche mit Erfolgswahrscheinlichkeit $p = 0.88$ im Einzelfall wurden gemacht. Mit welcher Wahrscheinlichkeit gelingen weniger als 44, aber mehr als 39, beide Grenzen eingeschlossen?

$$\begin{aligned} P(39 \leq X \leq 44) &= P(X \leq 44) - P(X \leq 38) \\ &= F(44) - F(38) \\ &= 0.93236 - 0.10408 \\ &= 0.82828 \end{aligned}$$

Tabelle der Werte $\sum_{k=0}^{r} B_{n,p}(k)$ für $n = 47$

r	p	0.85	0.86	0.87	0.88	0.89
27	0.	00001				
28		00002	00001			
29		00008	00003	00001		
30		00029	00012	00005	00002	00001
31		00093	00043	00018	00007	00002
32		00274	00137	00063	00026	00010
33		00742	00398	00199	00091	00038
34		01832	01060	00573	00286	00130
35		04128	02571	01503	00817	00408
36		08463	05663	03578	02115	01156
37		15768	11311	07707	04946	02957
38		26660	20441	14978	10408	06792
39		40904	33384	26208	19651	13952
40		57047	49285	41238	33208	25538
41		72665	65962	58411	50182	41543
42		85309	80597	74830	67964	60042
43		93639	91050	87606	83128	77447
44		97931	96887	95379	93236	90248
45		99552	99278	98847	98178	97153
46		99952	99917	99856	99754	99582

Geometrische Verteilung

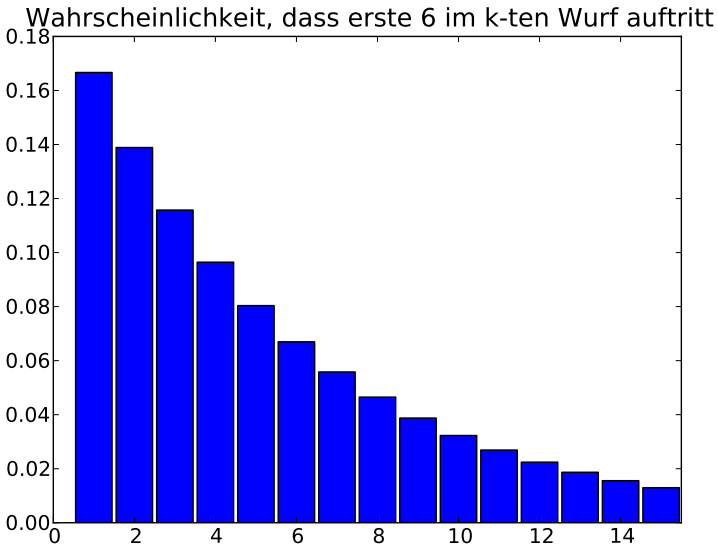
- Die Verteilungsfunktion der geometrischen Verteilung kann man ausrechnen
- Sie beträgt für $x \geq 0$

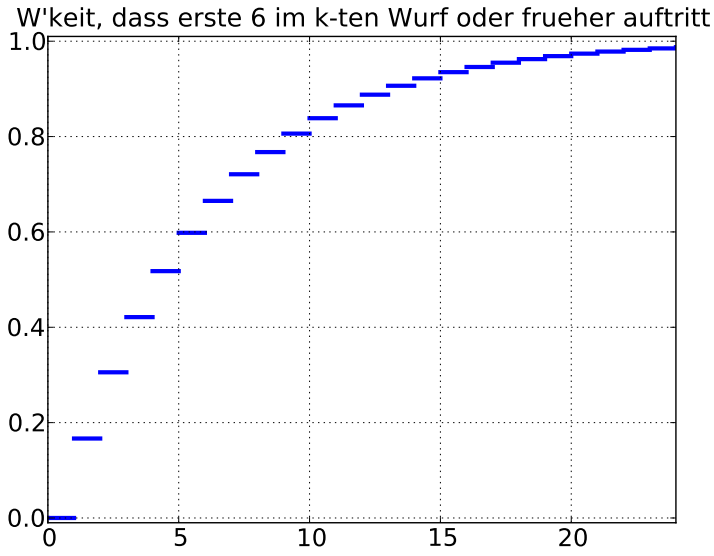
$$F(x) = 1 - (1-p)^j \qquad \text{falls } j \leq x < j+1$$

- Beispiel: Mit welcher Wahrscheinlichkeit fällt beim Wurf eines fairen Würfels die erste 6 spätestens im fünften Wurf?
- Antwort:

$$F(5) = 1 - \left(1 - \frac{1}{6}\right)^5 = 1 - \left(\frac{5}{6}\right)^5 = 0.5981$$

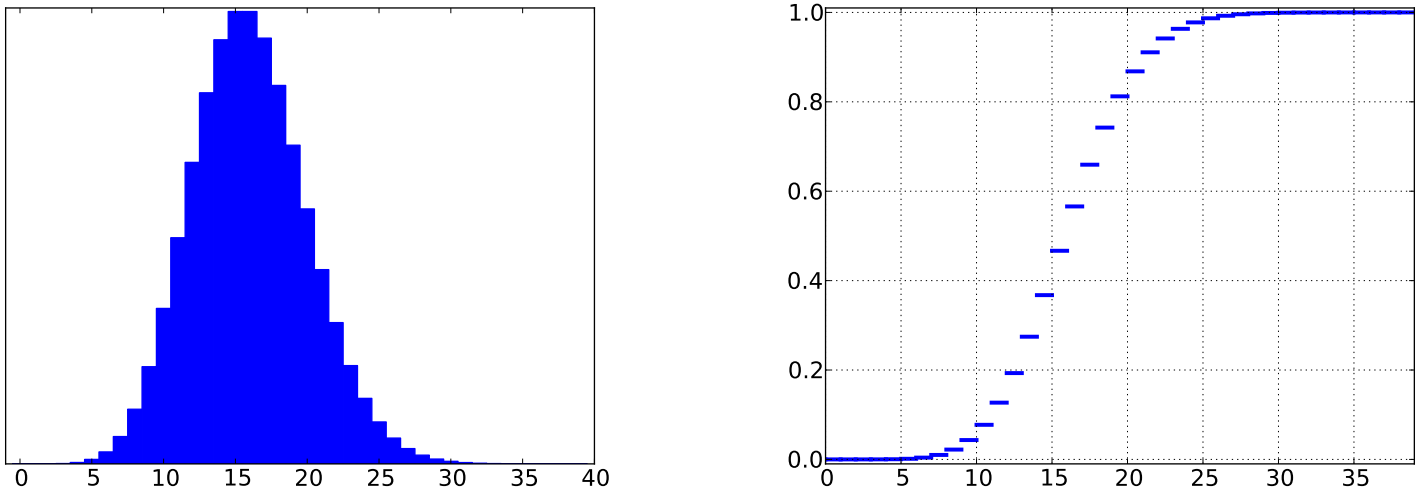
Verteilungsfunktion einer geometrischen Verteilung





Verteilungsfunktion einer Poissonverteilung

Die Verteilungsfunktion einer Poissonverteilung kann man nicht geschlossen ausdrücken

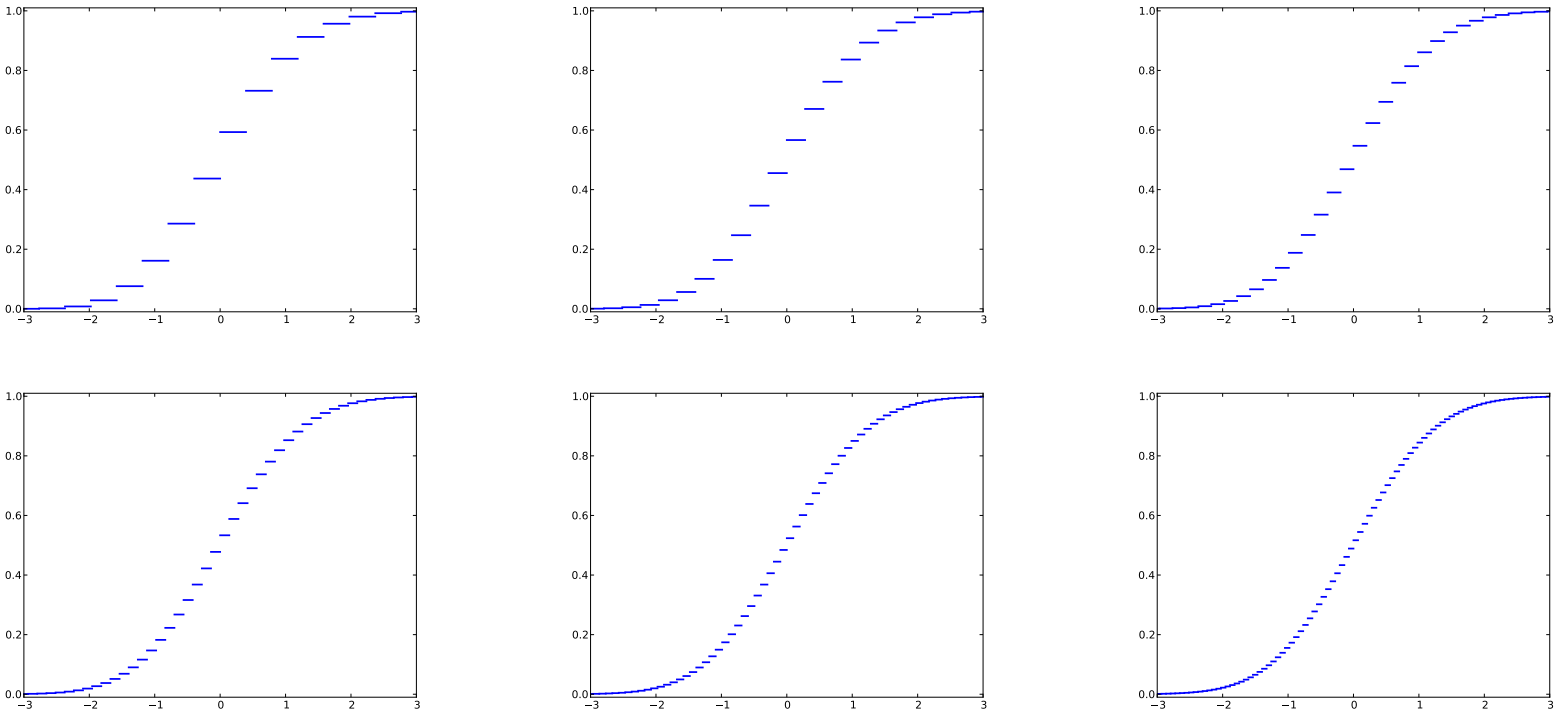


Verteilung und Verteilungsfunktion der Poissonverteilung P_{16}

Die standardisierte Verteilung

- Die Zufallsvariable X besitze den Erwartungswert $E(X) = \mu$ und die Varianz $Var(X) = \sigma^2$
- Setze
$$Y = \frac{X - \mu}{\sigma}$$
- Dann $E(Y) = 0$ und $Var(Y) = 1$
- Y ist die *standardisierte Zufallsvariable* zu X

Standardisierte Binomialverteilungen



Verteilungsfunktionen der Standardisierungen von $B_{n,p}$ für $p = 0.2$ und $n = 40, 80, 160, 320, 640$ und 1280

Zentraler Grenzwertsatz

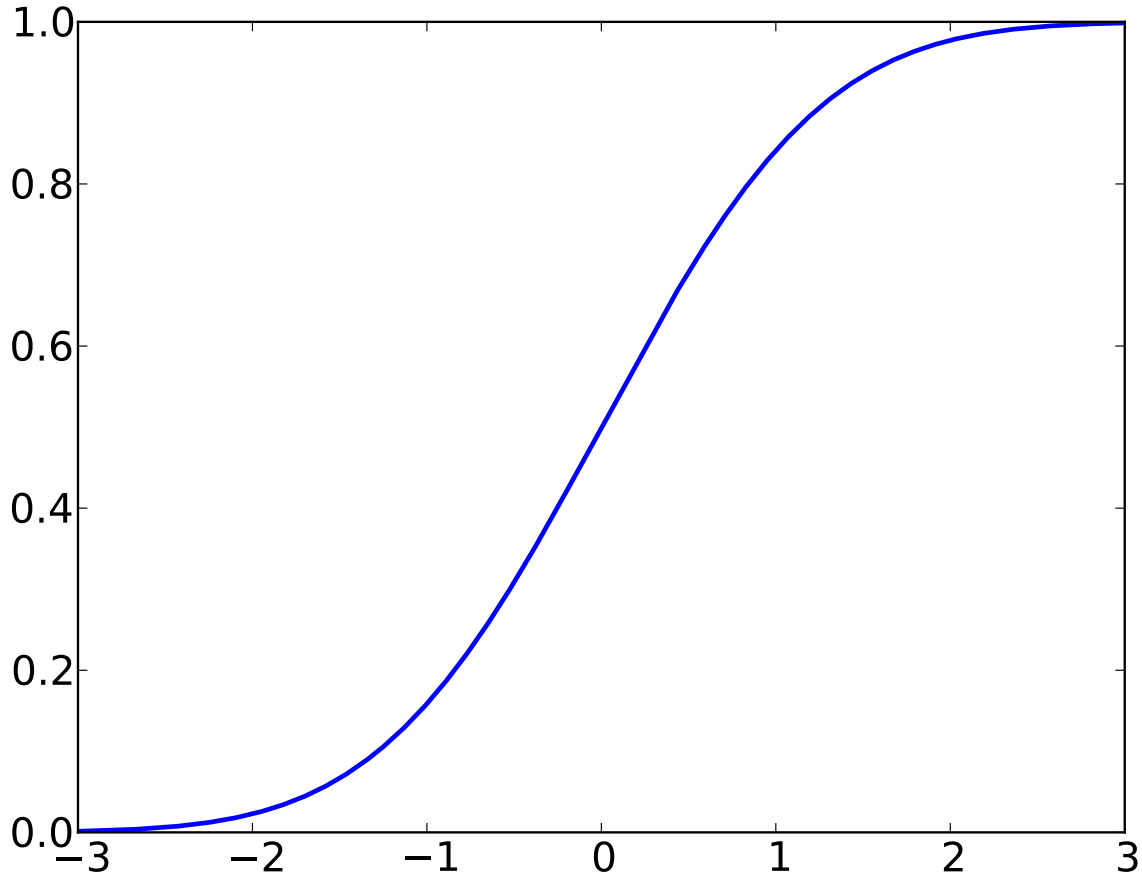
Für jeden Parameterwert p konvergiert der Grenzprozess auf der letzten Folie gegen dieselbe Funktion.

Diese Funktion heißt

Verteilungsfunktion der Standard-Normalverteilung

Sie wird mit Φ bezeichnet.

Verteilungsfunktion der Standard-Normalverteilung



Standard-normalverteilte Zufallsvariable

- X ist standard-normalverteilt, wenn

$$P(X \leq x) = \Phi(x) \qquad \text{für alle } x$$

- in diesem Fall

$$P(x < X \leq y) = \Phi(y) - \Phi(x) \qquad \text{für alle } x < y$$

Eigenschaften von Φ

- $0 < \Phi(x) < 1$ für alle Zahlen x
- Für $x \to \infty$ strebt $\Phi(x)$ gegen 1
- Für $x \to -\infty$ strebt $\Phi(x)$ gegen 0
- $\Phi(0) = \frac{1}{2}$
- Der Graph von Φ ist punktsymmetrisch am Punkt $(0, \frac{1}{2})$, d. h.

$$\Phi(-x) = 1 - \Phi(x)$$

- Die Verteilungsfunktion der Standard-Normalverteilung ist vertafelt

Tabelle der Standard-Normalverteilung, linke Seite

u	0.00	0.01	0.02	0.03	0.04
0.0	0,500000	0,503989	0,507978	0,511966	0,515953
0.1	,539828	,543795	,547758	,551717	,555670
0.2	,579260	,583166	,587064	,590954	,594835
0.3	,617911	,621720	,625516	,629300	,633072
0.4	,655422	,659097	,662757	,666402	,670031
0.5	,691462	,694974	,698468	,701944	,705401
0.6	,725747	,729069	,732371	,735653	,738914
0.7	,758036	,761148	,764238	,767305	,770350
0.8	,788145	,791030	,793892	,796731	,799546
0.9	,815940	,818589	,821214	,823814	,826391
1.0	,841345	,843752	,846136	,848495	,850830
1.1	,864334	,866500	,868643	,870762	,872857
1.2	,884930	,886861	,888768	,890651	,892512
1.3	,903200	,904902	,906582	,908241	,909877
1.4	0,919243	0,920730	0,922196	0,923641	0,925066

Tabelle der Standard-Normalverteilung, rechte Seite

u	0.05	0.06	0.07	0.08	0.09
0.0	0,519939	0,523922	0,527903	0,531881	0,535856
0.1	,559618	,563559	,567495	,571424	,575345
0.2	,598706	,602568	,606420	,610261	,614092
0.3	,636831	,640576	,644309	,648027	,651732
0.4	,673645	,677242	,680822	,684386	,687933
0.5	,708840	,712260	,715661	,719043	,722405
0.6	,742154	,745373	,748571	,751748	,754903
0.7	,773373	,776373	,779350	,782305	,785236
0.8	,802337	,805105	,807850	,810570	,813267
0.9	,828944	,831472	,833977	,836457	,838913
1.0	,853141	,855428	,857690	,859929	,862143
1.1	,874928	,876976	,879000	,881000	,882977
1.2	,894350	,896165	,897958	,899727	,901475
1.3	,911492	,913085	,914657	,916207	,917736
1.4	0,926471	0,927855	0,929219	0,930563	0,931888

Lesehinweise

- Auf dem Weblog gibt es die komplette Tabelle im Format A4 unter http://www.math.uni-duesseldorf.de/~braun/bio1011/tabNorm.pdf
- Beispiel $\Phi(0.31) = 0.621720$
- Wegen der Punktsymmetrie, also wegen

$$\Phi(-x) = 1 - \Phi(x)$$

kann man die Tabelle auch für negative Argumente verwenden. Also beispielsweise
$\Phi(-1.34) = 1 - \Phi(1.34) = 1 - 0.909877 = 0.090123$

- Durch Aufsuchen des Funktionswertes erhält man die Umkehrfunktion: Gesucht u mit $\Phi(u) = 0.79$. Man liest ab: $u = 0.81$

Tabelle der Standard-Normalverteilung, linke Seite

u	0.00	0.01	0.02	0.03	0.04
0.0	0,500000	0,503989	0,507978	0,511966	0,515953
0.1	,539828	,543795	,547758	,551717	,555670
0.2	,579260	,583166	,587064	,590954	,594835
0.3	,617911	,621720	,625516	,629300	,633072
0.4	,655422	,659097	,662757	,666402	,670031
0.5	,691462	,694974	,698468	,701944	,705401
0.6	,725747	,729069	,732371	,735653	,738914
0.7	,758036	,761148	,764238	,767305	,770350
0.8	,788145	,791030	,793892	,796731	,799546
0.9	,815940	,818589	,821214	,823814	,826391
1.0	,841345	,843752	,846136	,848495	,850830
1.1	,864334	,866500	,868643	,870762	,872857
1.2	,884930	,886861	,888768	,890651	,892512
1.3	,903200	,904902	,906582	,908241	,909877
1.4	0,919243	0,920730	0,922196	0,923641	0,925066

Beispiel

- Normalverteilungen werden beispielsweise zur Modellierung von Messfehlern benutzt
- Beispiel: Die Wirkstoffkonzentration in einem Heilmittel soll 4g/l betragen. Herstellungsabhängig beträgt die Streuung 100mg/l
- Mit welcher Wahrscheinlichkeit liegt die Konzentration über 4.12g/l?
- Y sei die Konzentration in g/l
- Dann $Y = 4 + 0.1 \cdot X$, wobei X standard-normalverteilt ist
- Gesucht $P(Y > 4.12)$
- Das ist $P(X > 1.2) = 1 - P(X \leq 1.2) = 1 - \Phi(1.2) = 1 - 0.884930 = 0.115070$