Normalverteilung

Approximation der Binomialverteilung

Für großes **n** ist der rechnerische Aufwand zur Bestimmung von (Bernoulli-) Wahrscheinlichkeiten, dass eine binomialverteilte Zufallsfunktion die Funktionswerte aus dem Intervall [k_1 ; k_2] annimmt, kaum zu bewältigen. Es liegt der Gedanke nahe, deshalb Binomialverteilungen durch eine (stetige) Funktion zu approximieren, womit gleichzeitig das Problem eines kontinuierlichen (statt diskreten) Wahrscheinlichkeitsmaßes gelöst wäre.

Vom Säulendiagramm ausgehend ist die Höhe einer Säule, bzw der Flächeninhalt bei Säulenbreite 1, signifikant für eine konkrete Wahrscheinlichkeit. Im kontinuierlichen Fall einer stetigen Funktion kann man natürlich nicht Funktionswerte addieren sondern wird den Flächeninhalt unter dem Graphen eindeutig bestimmen können.

Verschiedene Binomialverteilungen sehen im Histogramm unterschiedlich aus. Die Symmetrie kann man bei „unsymmetrischer" Trefferwahrscheinlichkeit p und Nietenwahrscheinlichkeit q durch großes n „erzwingen" (Laplace-Bedingung: $\sigma^2 > 9$), der Erwartungswert μ wandert jedoch für wachsendes n „nach rechts", die Gesamtverteilung wird graphisch breiter und flacher.
Will man also alle Binomialverteilungen, die die Laplace-Bedingung erfüllen, durch eine Funktion $\boldsymbol{\varphi}$ approximieren, so sind die Verteilungen zunächst einmal auf ein einheitliches graphisches „Aussehen" zu transformieren!

1. Verschiebung des Erwartungswertes: womit für alle Verteilungen gilt: $\boldsymbol{\mu = 0}$. $\quad \mathbf{k \to k - \mu}$

2. Normierung (Stauchung) der Breite: $\quad \mathbf{k - \mu \to \frac{k - \mu}{\sigma}}$

   womit alle Verteilungen eine Einheitsbreite besitzen.

3. Normierung (Streckung) der Höhe: $\quad \mathbf{B(n;p;k) \to \sigma \cdot B(n;p;k)}$
   damit der Verlust an Flächeninhalt (von 2.) ausgeglichen wird und ein „einheitlicher" Graph entsteht.

Damit:

$$\boldsymbol{\sigma \cdot B(n;p;k) =: \varphi\left(\frac{k-\mu}{\sigma}\right)}$$

$$\boldsymbol{\Leftrightarrow \quad B(n;p;k) = \frac{1}{\sigma} \cdot \varphi\left(\frac{k-\mu}{\sigma}\right)}$$

(Kontinuierliche) Wahrscheinlichkeitsmaße lassen sich also nun über Flächeninhalte unter einem geeigneten Funktionsgraphen bestimmen:

$$P(X \le c) = \frac{1}{\sigma} \cdot \int_{-\infty}^{c} \varphi\left(\frac{t-\mu}{\sigma}\right) dt = \int_{-\infty}^{\frac{c-\mu}{\sigma}} \varphi(x)\, dx =: \Phi\left(\frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{c-\mu}{\sqrt{n \cdot p \cdot q}}\right)$$

Beachte: Substitution $x = \frac{t-\mu}{\sigma} \quad \Rightarrow \quad dx = \frac{1}{\sigma} \cdot dt$

Normalverteilung
Approximation der Binomialverteilung

Welches ist nun eine geeignete Funktion **φ** ?

Auf C.F. Gauß geht zurück: $\varphi(x) = b \cdot e^{-a \cdot x^2}$.

Doch wie sind die Parameter a und b zu wählen? - Nach der 2. Transformation ist das $1 \cdot \sigma$ - Intervall gerade [-1 ; 1] und der Graph der normierten Verteilung sieht so aus, als wäre an diesen Stellen (im Sinne der Analysis) eine Wendepunktstelle.

Bestätige, dass damit: $a = \frac{1}{2}$ ist !

Der Parameter b ist nun so zu wählen, dass der gesamte Flächeninhalt unter dem Graphen 1 ist, d.h.:

$$I = \frac{1}{b} := \int_{-\infty}^{\infty} e^{-\frac{1}{2} \cdot x^2} \, dx .$$

Nun kommen üble „Tricks“ (Oberflächenintegral), die man erst richtig auf der Universität lernt.

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \cdot x^2} \cdot e^{-\frac{1}{2} \cdot y^2} \, dx \, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \cdot (x^2 + y^2)} \, dx \, dy \\ &= \int_{0}^{2 \cdot \pi} \int_{0}^{\infty} e^{-\frac{1}{2} \cdot r^2} \cdot r \cdot d\varphi \, dr \\ &= 2 \cdot \pi \cdot \int_{0}^{\infty} r \cdot e^{-\frac{1}{2} \cdot r^2} \, dr \\ &= 2 \cdot \pi \cdot \left[-e^{-\frac{1}{2} \cdot r^2} \right]_0^{\infty} = 2 \cdot \pi \end{aligned}$$

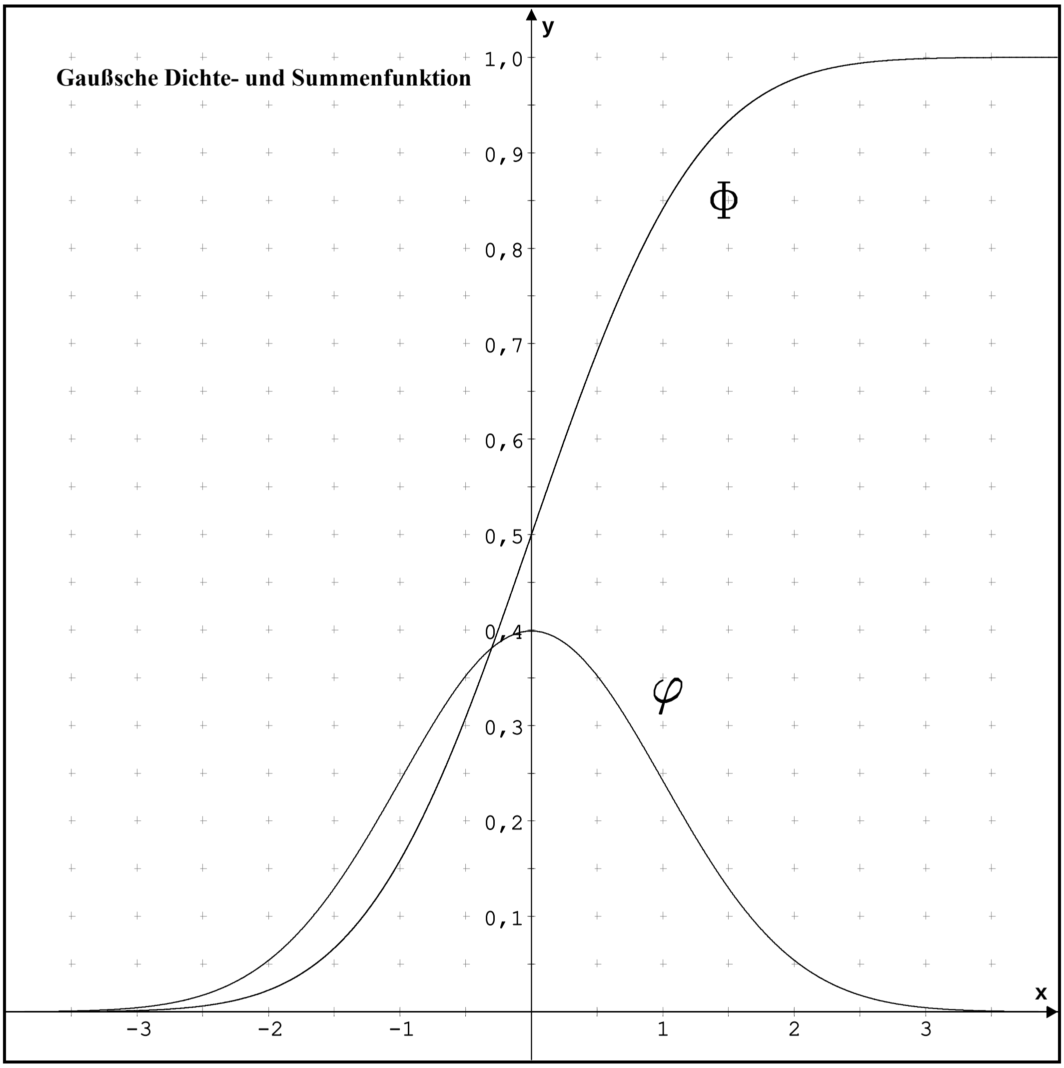
Beim Übergang zu Polarkoordinaten (3. Zeile) gilt für ein Flächendifferential: $dx \cdot dy = (r \cdot d\varphi) \cdot dr$.

Damit ist $b := \frac{1}{\sqrt{2 \cdot \pi}} \quad \Rightarrow \quad \int_{-\infty}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot x^2} \, dx = 1$

Die Funktion **φ** mit $\varphi(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot x^2}$ heißt Gaußsche Dichtefunktion, die zugehörige (uneigentliche) Integralfunktion **Φ** (mit unterer Grenze „- ∞“) heißt Gaußsche Summenfunktion.

Normalverteilung

Approximation der Binomialverteilung



Aufgaben:

(1) Begründe, dass gilt: $P(X \in \,]\, \mu - n\cdot\sigma \,;\, \mu + n\cdot\sigma\,[\,) = \int_{-n}^{n} \varphi(x)\cdot dx$

(2) Bestimmt werden soll ein Maß für die Sicherheit, mit der sich ein Ergebnis einer binomialverteilten Zufallsfunktion im $1\cdot\sigma$ - , $2\cdot\sigma$ - oder $3\cdot\sigma$ - Intervall befindet. Dies soll durch numerische Integration mit Hilfe des Simpson-Verfahrens geschehen. Aus dem Graphen der Dichtefunktion **φ** ist zu schließen, dass die Approximation durch 3 Parabeln (Keplersche Faßregel; d.h. Einteilung des Intervalls in 6 Teilintervalle) schon sehr gut sein müsste. - Ergänze die Tabelle durch die fehlenden Werte.

n	P(X ∈] μ-σ ; μ+σ [ ) ≈	P(X ∈] μ-2σ ; μ+2σ [ ) ≈	P(X ∈] μ-3σ ; μ+3σ [ ) ≈
2	0,69323686	1,13583405	1,60463283
4	0,68305811	0,94721077	0,92144452
6			
8	0,68271098	0,95440214	0,99695884
16	0,68269082	0,95449473	0,99728308

Normalverteilung

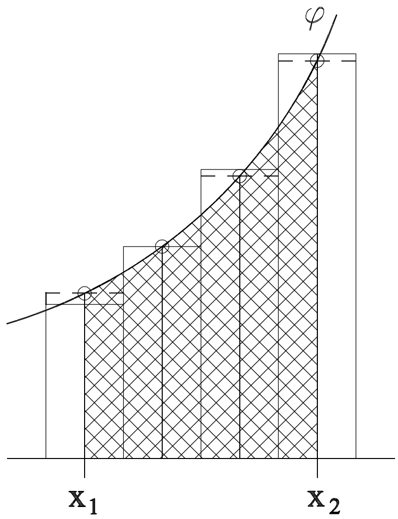
Approximation der Binomialverteilung

Nun zur praktischen Anwendung der Normalverteilung. Wir erinnern uns an die wesentlichen Beziehungen:

$$\boxed{x = \frac{k-\mu}{\sigma} \; ; \quad y = \sigma \cdot B(n;p;k)} \quad \text{und} \quad \boxed{P(k_1 \le X \le k_2) \approx \int_{x_1}^{x_2} \varphi(t)\cdot dt}$$

wobei φ die Gaußsche Dichtefunktion ist. Nun taucht jedoch bei der Approximation einer diskreten Binomialverteilung durch eine stetige Funktion noch ein kleines Problem auf.

Die Stellen x_1 und x_2 sind die jeweiligen Mitten der Histogrammsäulen, mit der Breite $\frac{1}{\sigma}$, für die Trefferwahrscheinlichkeiten k_1 und k_2.



Wir müssen also einen Randausgleich durchführen, d.h. x_1 nach links und x_2 nach rechts verschieben, und zwar um $\frac{1}{2}\cdot\frac{1}{\sigma}$. Damit ergibt sich:

$$x_1 := \frac{k_1 - 0{,}5 - n\cdot p}{\sqrt{n\cdot p\cdot q}}$$

$$x_2 := \frac{k_2 + 0{,}5 - n\cdot p}{\sqrt{n\cdot p\cdot q}}$$

Je größer n ist, umso weniger spielt der Summand 0,5 im Zähler eine Rolle. <u>Faustregel:</u> Ab n= 1000 ist der Randausgleich von 0,5 im Zähler zu vernachlässigen.

Wir erinnern uns weiterhin, für praktische Berechnungen, an die Definition der Gaußschen Summenfunktion Φ:

$$\boxed{\Phi(x) := \int_{-\infty}^{x} \varphi(t)\cdot dt}$$

und eine sehr nützliche Folgerung aus der Achsensymmetrie von φ zur y-Achse bzw. der Punktsymmetrie von Φ zum Punkt $\left(0 \;\middle|\; \frac{1}{2}\right)$:

$$\boxed{\begin{aligned} \Phi(-x) &= \int_{-\infty}^{-x} \varphi(t)\cdot dt \\ &= 1 - \Phi(x) = 1 - \int_{-\infty}^{x} \varphi(t)\cdot dt \end{aligned}}$$

Normalverteilung

Approximation der Binomialverteilung

Nun 2 Aufgaben:

1) Ein Händler bietet Gurkensamen an, die erfahrungsgemäß zu 95 % keimfähig sind. Wie groß ist die Wahrscheinlichkeit, dass von 500 ausgesäten Körnern

a) höchstens 470
b) mindestens 470 und höchstens 485
c) mindestens 480 keimen?

Lösung:[1] $\mu = 500 \cdot \frac{95}{100} = 475 \;; \quad \sigma = \sqrt{475 \cdot \frac{5}{100}} = \sqrt{23{,}75}$

a) $P(X \le 470) = \Phi\left(\dfrac{470 - 475}{\sqrt{23{,}75}}\right) \approx \Phi(-1{,}03) \approx 0{,}1515$

$P(X \le 470) = \Phi\left(\dfrac{470 + 0{,}5 - 475}{\sqrt{23{,}75}}\right) \approx \Phi(-0{,}92) \approx 0{,}1788$

b) $P(470 \le X \le 485) \approx \Phi(2{,}05) - \Phi(-1{,}03) \approx 0{,}8283$

c) $P(X \ge 480) = 1 - P(X \le 479) \approx 1 - \Phi(0{,}82) \approx 0{,}2061$

---

2) Die Lebensdauer X (in km) eines Automotors einer bestimmten Marke sei angenähert normalverteilt mit dem Erwartungswert $\mu = 105000$ und der Standardabweichung $\sigma = 10000$.[2]

a) Bei wie viel Prozent der Motoren übersteigt die Lebensdauer 120000 km?

b) Bei wie viel Prozent der Motoren weicht die Lebensdauer um mehr als 12000 km vom Erwartungswert ab?

Lösung:

a) $P(X > 120000) = 1 - \Phi\left(\dfrac{120000 - 105000}{10000}\right) \approx 1 - \Phi(1{,}5) \approx 0{,}0668$

b) $P(|X - 105000| > 12000) = 2 \cdot \left[1 - \Phi\left(\dfrac{12000}{10000}\right)\right] \approx 0{,}2302$

---

1 Beachte: Bei Aufgabenteil a) einmal ohne, einmal mit Randausgleich. Die Faustregel ist nicht erfüllt!

2 Dies ist eine reine Aufgabe zur Normalverteilung und hat mit der Approximation der Binomialverteilung nichts mehr zu tun.

Quelle: Lambacher-Schweizer: Stochastik Leistungskurs, Klett Verlag