Über die Normalverteilung

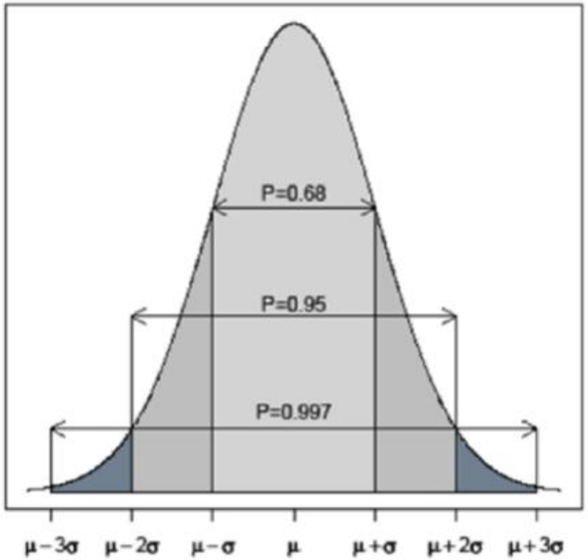
(Math. Statistik)

Zu Strukturintegritäts-Anw. 2 / Dickwandiger Behälter

Normalverteilung

Die Normalverteilung ist die wichtigste stetige Verteilung. Sie wurde als Modell für Messwerte bzw. Messfehler entwickelt und passt dort meist sehr gut. Doch das ist nicht der einzige Grund: wie wir später sehen werden, hat die Summe von stochastisch unabhängigen Zufallsvariablen mit beliebiger, identischer Verteilung approximativ eine Normalverteilung. Sie kann daher stets zum Einsatz kommen, falls sich ein Resultat aus einer additiven Überlagerung von vielen „kleinen Einflüssen" ergibt. Die Normalverteilung mit ihrer Glockenform ist also so etwas wie eine „Naturkonstante" in der Welt der Zufallsvariablen. Die Dichtefunktion einer Zufallsvariable X mit Normalverteilung $N(\mu,\sigma^2)$ ist gegeben durch:

$$f(x) := \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left\{-\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2}\right\} \text{ für } x\in(-\infty,+\infty)\,.$$



Die Normalverteilung hat offensichtlich zwei Parameter. Der Lageparameter $\mu \in (-\infty, +\infty)$ ist für Verschiebungen der Glockenkurve auf der x-Achse zuständig, ist Zentrum und Symmetriepunkt der Verteilung, und ist gleichzeitig auch der Erwartungswert $E[X]$. Der Skalenparameter σ^2 bestimmt die „Breite" der Verteilung, d.h. streckt oder staucht die Glockenkurve und ist gleichzeitig die Varianz $Var(X)$. Für die Interpretation ist die sogenannte Standardabweichung σ zugänglicher. Bei $\mu \pm \sigma$ befinden sich die Wendepunkte der Dichtefunktion und das Intervall $[\mu - \sigma, \mu + \sigma]$ enthält rund 68% der Wahrscheinlichkeit (d.h. der gesamten Fläche unter der Kurve). Im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ sind es rund 95%, und innerhalb von $\mu \pm 3\sigma$ liegen bereits ca. 99.7% der Wahrscheinlichkeit.

Um diese Wahrscheinlichkeiten zu berechnen, muss man bekanntlich die Dichtefunktion integrieren. Dies ist im Fall der Normalverteilung aber schwierig, gibt es doch keine Stammfunktion

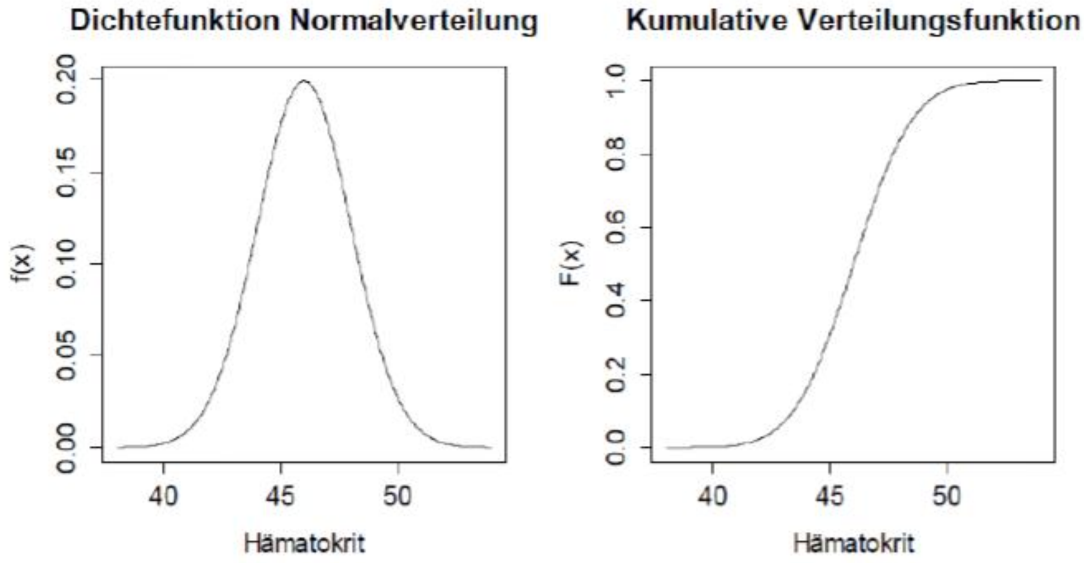
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2} \cdot \frac{(z-\mu)^2}{\sigma^2}\right) dz$$

welche man in geschlossener Form schreiben könnte. Die Integrale existieren aber, und man kann/muss zur Berechnung auf numerische Methoden zurückgreifen. Während früher häufig mit Tabellen gearbeitet wurde, verwendet man heute in der Regel Computerprogramme.

Ein nicht-technisches Beispiel

Als Beispiel betrachten wir den Hämatokrit. Er bezeichnet den Anteil der *Erythrozyten* am Volumen des Blutes. Diese stellen rund 99% des

Gesamtvolumens der Blutzellen dar, somit entspricht der Hämatokrit ungefähr dem Anteil des Zellvolumens im Blut und gibt Aufschluss über den Wasserhaushalt einer Person. Bei Menschen liegen die normalen Hämatokrit-Werte zwischen 42% und 50%. Wir können also den Hämatokrit mit einer Normalverteilung beschreiben, wo $\mu = 46$ und $\sigma = 2$. Dichte und kumulative Verteilungsfunktion sind dann wie folgt:



Im Nordischen Skisport wurde durch Doping-Kontrollbehörden ein Hämatokrit-Grenzwert von 51.5% festgelegt. Überschreitet ein Athlet diese Schwelle, so ist er nicht mehr zu den Wettkämpfen zugelassen. Wir können nun aufgrund der obigen Verteilung die Wahrscheinlichkeit $P(X > 51.5) = 0.3\%$ berechnen, mit welcher ein Athlet (ohne Manipulation) die Schwelle überschreitet. In der Praxis, d.h. wenn eine Stichprobe von Beobachtungen vorliegt, schätzt man die beiden Parameter μ und σ^2 oftmals aus den Daten durch Mittelwert und Stichproben-Varianz, d.h.: