Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

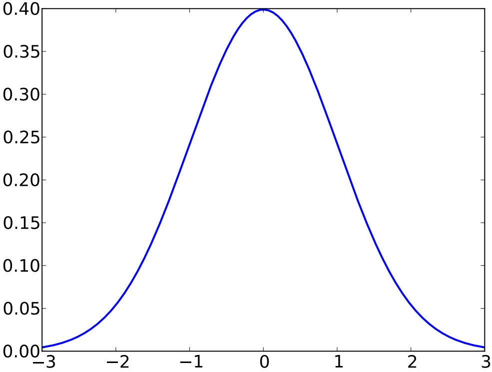
Heinrich-Heine-Universität Düsseldorf

18. November 2010

Gausche Glockenkurve

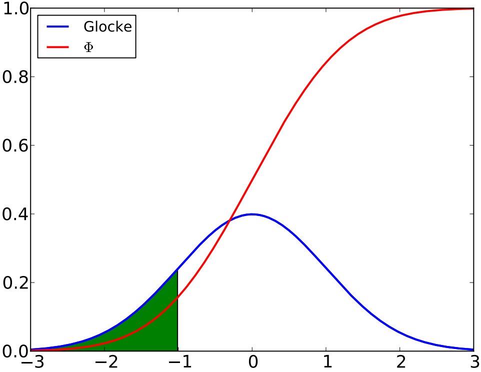
$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

heißt *Gaußsche Glockenkurve*



Standard-Normalverteilung als Fläche

$\Phi(x)$ ist die Fläche links von x unter der Gaußschen Glockenkurve



Normalverteilungen

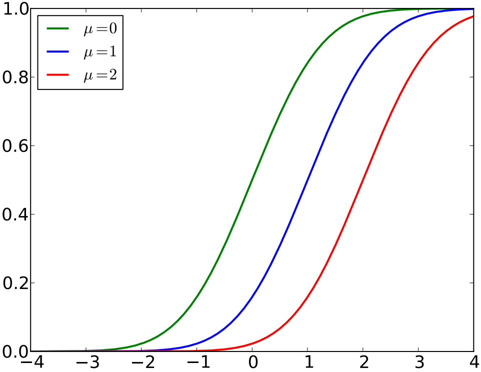
- Die Zufallsvariable X heißt *normalverteilt* zum Erwartungswert μ und der Varianz σ^2, wenn

$$Y = \frac{X - \mu}{\sigma}$$

  standard-normalverteilt ist. Man sagt dann, X sei $N(\mu, \sigma^2)$-verteilt
- $X = \mu + \sigma \cdot Y$
- Y ist die Standardisierung von X
- Vorerst ohne präzise mathematische Definition halten wir fest

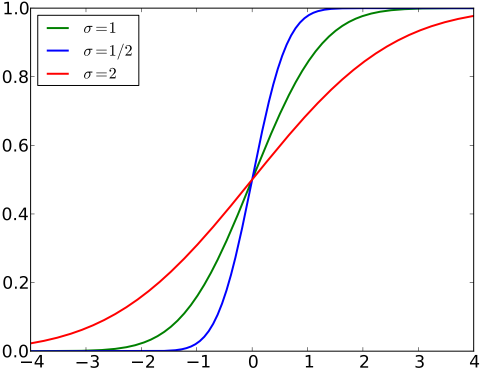
$$E(X) = \mu$$
$$Var(X) = \sigma^2$$

Normalverteilungen für verschiedene Erwartungswerte



Verteilungsfunktionen für $N(\mu, 1)$-verteilte Zufallsvariable, $\mu = 0, 1, 2$

Normalverteilungen für verschiedene Streuungen



Verteilungsfunktionen für $N(0, \sigma^2)$-verteilte Zufallsvariable,
$\sigma = 1, \frac{1}{2}, 2$

Umrechnung auf Standardnormalverteilung

Die Zufallsvariable X sei $N(\mu, \sigma^2)$-verteilt. Dann gelten für $a < b$

$$P(a < X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Beispiel: natürliche Variabilitäten

- Roggenpflanzen erreichen eine mittlere Höhe von $1m$. Dabei streut die Höhe um $20cm$. Welcher Prozentsatz aller Pflanzen erreicht mindestens $1.10m$ Höhe?
- $X =$ Höhe der Pflanze
- Wir rechnen in Metern. Dann $E(X) = 1$ und $Var(X) = 0.04$
- X ist nicht standard-normalverteilt
- $X = 1 + 0.2 \cdot Y$ für standard-normalverteiltes Y
- $X \geq 1.10$ bedeutet $Y \geq 0.5$
- $P(Y \geq 0.5) = 1 - P(Y \leq 0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085$
- ca 31% der Pflanzen sind mindestens $1.10m$ hoch

Kritische Betrachtung des Modells

- Das Modell erlaubt auch den unsinnigen Fall, dass Roggenpflanzen eine negative Höhe aufweisen
- Mit welcher Wahrscheinlichkeit geschieht das?
- $X \leq 0$ bedeutet $Y \leq -5$
- $P(Y \leq -5) = \Phi(-5) = 1 - \Phi(5) = 1 - (1 - 2.867 \cdot 10^{-7}) = 2.867 \cdot 10^{-7}$
- Das Modell sagt für weniger als eine unter 3 Millionen Pflanzen eine negative Höhe voraus
- Damit können wir leben

Beispiel: IQ-Tests

- IQ-Tests sind so skaliert, dass die Werte in der Population normalverteilt mit Erwartungswert $\mu = 100$ und Streuung $\sigma = 15$ sind
- Welcher Anteil der Bevölkerung hat einen IQ über 130?
- X messe den IQ
- X ist $N(100, 225)$-verteilt.
- Also

$$\begin{aligned} P(130 < X) &= 1 - \Phi\left(\frac{130 - 100}{15}\right) \\ &= 1 - \Phi(2) = 1 - 0.977250 = 0.02275 \end{aligned}$$

- Ungefähr 2.3% der Population weist einen IQ von mindestens 130 auf

Der zentrale Grenzwertsatz

- "Wenn man nur genügend große Stichproben hat, dann ist alles normalverteilt."
- Präziser: $X_1, \ldots, X_n$ unabhängig, alle mit derselben Verteilung
- $\mu = E(X_1) = \cdots = E(X_n)$ und
  $\sigma^2 = Var(X_1) = \cdots = Var(X_n)$
- Bilde die *standardisierte Summenvariable*

$$Z_n = \frac{\sum_{j=1}^{n}(X_j - \mu)}{\sigma \cdot \sqrt{n}}$$

- Wenn n ausreichend groß ist, gilt für jede reelle Zahl x

$$P(Z_n \leq x) \cong \Phi(x)$$

Die 3σ-Regel

- $X_1, \ldots, X_n$ unabhängig, alle mit derselben Verteilung
- $\mu = E(X_1) = \cdots = E(X_n)$ und
  $\sigma^2 = Var(X_1) = \cdots = Var(X_n)$
- Arithmetisches Mittel der X_j

$$Y = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

- n ausreichend groß, dann näherungsweise

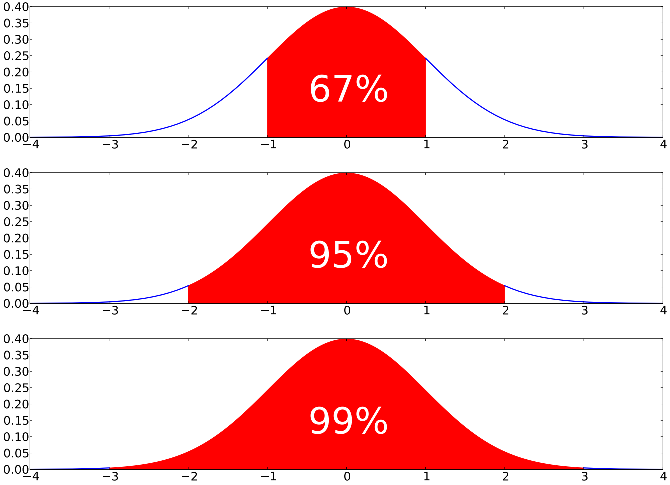
$$P\left(\left|Y - \mu\right| \leq \frac{\sigma}{\sqrt{n}}\right) \geq \frac{2}{3}$$

$$P\left(\left|Y - \mu\right| \leq \frac{2\sigma}{\sqrt{n}}\right) \geq 0.95$$

$$P\left(\left|Y - \mu\right| \leq \frac{3\sigma}{\sqrt{n}}\right) \geq 0.99$$

- Die 3σ-Regel ist nur eine Faustregel.

3σ-Regel, Zeichnung



3σ-Regel, Beispiel

- Der Wirkstoffgehalt von Düngetabletten wird 35-mal gemessen (in mg). Die Streuung betrage $8mg$
- Wie viele Messungen schreibt die 3σ-Regel vor, um mit 95% Wahrscheinlichkeit einen Messfehler von weniger als $4mg$ zu haben?
- Wir brauchen
$$4 = \frac{2\sigma}{\sqrt{n}}$$
- Also
$$4 = \frac{2 \cdot 8}{\sqrt{n}}$$
- Das bedeutet $n = 16$